On Cumulative Information Measures : Properties, Inferences and Applications

Ph. D. Thesis

Siddhartha Chakraborty

STATISTICAL QUALITY CONTROL & OPERATIONS RESEARCH DIVISION

INDIAN STATISTICAL INSTITUTE
203, B. T. ROAD, KOLKATA 700 108
INDIA

October 2023

On Cumulative Information Measures : Properties, Inferences and Applications

Siddhartha Chakraborty

Thesis submitted to the Indian Statistical Institute
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in
Quality, Reliability and Operations Research (QROR)

September 2023

Thesis supervisor: Prof. Biswabrata Pradhan

Indian Statistical Institute
203, B. T. Road, Kolkata 700 108
India

To my Mother and Father

Acknowledgements

First and foremost, my earnest acknowledgement must go to my supervisor Prof. Biswabrata Pradhan for being such a great advisor and nice human being as well. I can not thank him enough for everything he has done for me. His endless support and encouragement made this work possible. I have learned a lot from the fruitful discussions that we have throughout the entire period. Whenever needed, he was available for providing invaluable advice, no matter if it was academic or non-academic issues. He has provided important ideas, suggestions and comments and insisted on the highest standards for the research.

I would like to express my sincere gratitude to my parents Mr. Shyamal Chakraborty and Mrs. Ratna Chakraborty for their relentless efforts, blessings and having faith in me as always. They are my very first teachers as well and words can not explain how much I have learned from them. They sacrificed a lot for me and because of them, my life has a purpose. Words will not be sufficient to express how grateful I am to them for everything.

I would like to thank Prof. Gopal Krishna Basak of Statistics and Mathematics Unit and Dr. Samarjit Das of Economic Research Unit. I did my MSc. project under their guidance in Indian Statistical Institute, Kolkata. Interactions with them made me highly motivated to pursue research. During my PhD, I also had useful discussions with them which improved my work.

I am lucky enough to be accompanied by Prof. Anup Dewanji and Prof. Debasis Sengupta from Applied Statistics Unit; Dr. Md. Zafar Anis, Dr. Arup Ranjan Mukherjee and Dr. Prasun Das from SQC & OR Unit; Prof. Asok Kumar Nanda, Department of Mathematics and Statistics, Indian Institute of Science Education and Research, Kolkata, during my life in ISI. Their presence and suggestions in various discussion meetings help me to improve as a researcher.

I would like to take this opportunity to express my thanks to my teachers at the University of Kalyani where I learned Statistics during my 5-year Integrated MSc. course. I feel lucky to have Prof. R. N. Panda, Prof. P. Das, Prof. S. S. Maiti, Prof. N. K. Chanda, Prof. C. Pal, Prof. D. Bandyopadhyay, Dr. S. K. Samanta, Dr. M. Deekshithulu of University of Kalyani and Dr. M. N. Maji of Hooghly Mohsin College, as my teachers.

I must thank my collaborator Dr. Ritwik Bhattacharya of University of Texas at El Paso, Texas, USA, for his support and encouragement. I have learned a lot from him through the various discussions we have in the last few years. I highly appreciate his advice and suggestions.

Another person I must thank is my collaborator and long time friend Miss. Oindrali Das of Ranaghat College, Nadia, India. She encouraged and motivated me which help me pass through my troubled times.

I thank my relatives and friends for their goodwill, support and encouragement. My life would have been completely boring and monotonous without them.

I am grateful to the Statistical Quality Control & Operations Research Unit of Indian Statistical Institute, Kolkata for providing me with all the facilities to carry out my research. It was my dream to pursue my PhD from the Indian Statistical Institute and I am grateful for the opportunity.

Finally, I thank the Almighty for giving me the great gift of life and for providing me eternal strength to never lose hope and to never give up.

Monday 18th September, 2023 — Siddhartha Chakraborty

Kolkata — Indian Statistical Institute

Table of contents

List of figures

List of tables

Chapter 1

Introduction

Abbreviations

rv	random variable
rs	random sample
id	identically distributed
iid	independent and identically distributed
d.i.d.	dependent and identically distributed
pdf	probability density function
cdf	cumulative distribution function
edf	empirical distribution function
sf	survival function
qf	quantile function
hr	hazard rate
rhr	reversed hazard rate
MRL	mean residual life
MPL	mean past life
WMRL	weighted mean residual life
WMPL	weighted mean past life
MLE	maximum likelihood estimate
MSE	mean square error
CRE	cumulative residual entropy
CE	cumulative entropy
CREx	cumulative residual extropy
WCRE	weighted cumulative residual entropy
WCE	weighted cumulative entropy

GWSE	generalized weighted survival entropy
GWFE	generalized weighted failure entropy
GDWSE	generalized dynamic weighted survival entropy
GDWFE	generalized dynamic weighted failure entropy
IGDWSE	increasing generalized dynamic weighted survival entropy
DGDWFE	decreasing generalized dynamic weighted failure entropy
WCTRE	weighted cumulative Tsallis residual entropy
WCTPE	weighted cumulative Tsallis past entropy
DWCTRE	dynamic weighted cumulative Tsallis residual entropy
DWCTPE	dynamic weighted cumulative Tsallis past entropy
IDWCTRE	increasing dynamic weighted cumulative Tsallis residual entropy
DDWCTPE	decreasing dynamic weighted cumulative Tsallis past entropy
CRKL	cumulative residual Kullback-Leibler information
WCRKL	weighted cumulative residual Kullback-Leibler information
CKL	cumulative Kullback-Leibler information
WRKL	weighted cumulative Kullback-Leibler information
WSEx	weighted survival extropy
DWSEx	dynamic weighted survival extropy
WESEx	weighted extended survival extropy
DWESEx	dynamic weighted extended survival extropy
NCEx	negative cumulative extropy
WNCEx	weighted negative cumulative extropy

Notations

X	underlying random variable
$f(x)$	pdf of X
$F(x)$	cdf of X
$S(x)$	sf of X
$F_n(x)$	edf of X
$E(X)$	expectation of X
$\lambda_F(\cdot)$	hr function of a rv having cdf F
$r_F(\cdot)$	rhr function of a rv having cdf F
$m_F(\cdot)$	MRL function of a rv having cdf F
$\mu_F(\cdot)$	MPL function of a rv having cdf F
$m_F^w(\cdot)$	WMRL function of a rv having cdf F
$\mu_F(\cdot)$	WMPL function of a rv having cdf F

$H(X)$	Shannon entropy for continuous case or differential entropy
$H(X;t)$	residual Shannon entropy
$\bar{H}(X;t)$	past Shannon entropy
$H^w(X)$	weighted Shannon entropy
$H_\theta(X)$	Renyi entropy of X of order θ
$T_\theta(X)$	Tsallis entropy of X of order θ
$CRE(X)$	cumulative residual entropy of X
$CE(X)$	cumulative entropy of X
$CRE^w(X)$	weighted cumulative residual entropy of X
$CE^w(X)$	weighted cumulative entropy of X
$CRE(X;t)$	dynamic cumulative residual entropy of X
$CE(X;t)$	dynamic cumulative entropy of X
$CRE^w(X;t)$	dynamic weighted cumulative residual entropy of X
$CE^w(X;t)$	dynamic weighted cumulative entropy of X
$\xi^w_{\theta_1,\theta_2}(X)$	GWSE of X of order (θ_1,θ_2)
$\xi^w_{\theta_1,\theta_2}(X;t)$	GDWSE of X of order (θ_1,θ_2)
$f\xi^w_{\theta_1,\theta_2}(X)$	GWFE of X of order (θ_1,θ_2)
$f\xi^w_{\theta_1,\theta_2}(X;t)$	GDWFE of X of order (θ_1,θ_2)
$\hat{\xi}^w_{\theta_1,\theta_2}(X)$	non-parametric estimator of GWSE of X
$f\hat{\xi}^w_{\theta_1,\theta_2}(X)$	non-parametric estimator of GWFE of X
$\xi_\theta(X)$	CTRE of X
$\bar{\xi}_\theta(X)$	CTPE of X
$\xi^w_\theta(X)$	WCTRE of X
$\xi^w_\theta(X;t)$	DWCTRE of X
$\bar{\xi}^w_\theta(X)$	WCTPE of X
$\bar{\xi}^w_\theta(X;t)$	DWCTPE of X
$\hat{\xi}^w_\theta(X)$	empirical WCTRE of X
$\hat{\bar{\xi}}^w_\theta(X)$	empirical WCTPE of X
$CRKL^w(X,Y)$	WCRKL between X and Y
$CRKL^w(X,Y;t)$	dynamic WCRKL between X and Y
$CKL^w(X,Y)$	WCKL between X and Y
$CKL^w(X,Y;t)$	dynamic WCKL between X and Y
$J(X)$	extropy of X
$\xi J(X)$	CREx of X
$J^w(X)$	weighted extropy of X
$\xi^w J(X)$	WCREx of X
$\mathbf{s}$	signature vector
$JCREx\,(T : X_{1:n},\cdots,X_{n:n})$	Jensen-cumulative residual extropy divergence between T and $X_{1:n},\cdots,X_{n:n}$.

$\xi J^1_{F_n}(X)$	edf based non-parametric estimator of CREx
$\xi J^2_{F_n}(X)$	L- Statistics estimator of CREx
$\xi J^3_{F_n}(X)$	Kernel based estimator of CREx
$\xi J^w(X)$	WSEx of X
$\xi J^w(X;t)$	DWSEx of X
$\mathscr{C}(X)$	NCEx of X
$\mathscr{C}^w(X)$	WNCEx of X
$\xi J^\varepsilon(X)$	WESEx of X
$\xi J^\varepsilon(X;t)$	DWESEx of X
$\xi J^\varepsilon(F_n)$	edf based non-parametric estimator of WESEx
$\xi J^\varepsilon(F_n;t)$	non-parametric estimator of DWESEx
$\mathscr{C}^w(\hat{F}_n)$	non-parametric estimator of WNCEx
$\xi J^w(\hat{S}_n)$	recursive kernel based estimator of WSEx for d.i.d. observations
$H_{1\cdots r:n}$	Shannon entropy of first r order statistics
$H_{1\cdots m:m:n}$	Shannon entropy of progressive type-II censored order statistics
$CRE_{1\cdots r:n}$	CRE of first r order statistics
$CRE_{1\cdots m:m:n}$	CRE of progressive type-II censored order statistics
$CE_{s\cdots m:n}$	CE of last $(n-s+1)$ order statistics
PCII	Progressive type-II
PCOS	Progressively type-II censored order statistics
COD	Compound optimal design

1.1 Literature review

INFORMATION theory is a branch of applied probability and statistics that studies various aspects of information such as processing, transmission, compression etc. Entropy is by far regarded as the most important concept in information theory which was first introduced by physicists in the context of equilibrium thermodynamics. Boltzmann provided a formal definition of entropy that represents the disorder between particles of a microscopic system. Shannon (1948) proposed entropy measure for a discrete random variable (rv) X that takes values $\{x_1, x_2, \cdots, x_n\}$ with probabilities $\mathbf{P} = \{p_1, p_2, \cdots, p_n\}$ as

$$H(\mathbf{P}) = -\sum_{i=1}^{n} p_i \log p_i. \tag{1.1}$$

For a continuous rv X, entropy is defined as

$$H(X) = -\int_A f(x) \log f(x) dx, \tag{1.2}$$

where log is the natural logarithm, $0 \log 0 = 0$ for computational convenience, f is the probability density function (pdf) of X and A is the support of X. It is also known as the differential entropy. Shannon (1948) introduced entropy as a measure of uncertainty associated with probability distribution of the underlying rv. Higher values of entropy of a rv X means the probability distribution of X will represent more uncertainty. Also entropy has been widely regarded as a measure of information conveyed by the underlying distribution. According to Rényi (1961), "the amount of information which we get when we observe the result of an experiment (depending on chance) can be taken numerically equal to the amount of uncertainty concerning the outcome of the experiment before carrying it out." Jaynes (1968) stated that, "the probability distribution which maximizes the entropy is numerically identical with the frequency distribution which can be realized in the greatest number of ways". When inference has to be made on the basis of prior knowledge, the maximum entropy distribution subject to the given constraint will be the best model available, see Jaynes (1982). This is also supported by the fact that many popular distributions possess maximum entropy property given appropriate constraint. For example, normal distribution has the maximum entropy among all continuous distributions having the same 2nd raw moment ($E(X^2)$). Among all non-negative continuous distributions having the same mean, exponential distribution has the maximum entropy. For detailed discussions on maximum entropy inference see Jaynes (1968, 1982), Burnham and Anderson (2004) and the references therein where both intuitive and mathematical arguments are provided for the interpretation of entropy as an information measure. In probability and statistics, distributions with higher entropy are considered better than the ones having lower entropy. Throughout the course of this thesis all the random variables are assumed to be non-negative and absolutely continuous and by increasing (decreasing) we mean non-decreasing (non-increasing) unless otherwise specified.

Shannon's work on entropy gave rise to a new branch of applied probability with useful applications in a variety of fields such as economics, thermodynamics, statistical mechanics, mathematical biology, signal processing, statistics and reliability. An enormous amount of research has been carried out over the years by many scholars on generalizations of Shannon entropy and their applications. Some important ones are due to Rényi (1961), Varma (1966), Tsallis (1988) and Mathai and Haubold (2007). For two independent rvs X and Y, the joint entropy $H(XY) = H(X) + H(Y)$. This means that entropy is additive in nature. Renyi entropy of X is defined as

$$H_\theta(X) = \frac{1}{1-\theta}\log\int_0^{+\infty} f^\theta(x)dx, \quad 0<\theta\neq 1. \tag{1.3}$$

It is easy to see that as $\theta \to 1$, $H_\theta(X) \to H(X)$. Generalization parameter θ makes Renyi entropy more flexible and it preserves the additive property of $H(X)$. Another important generalization of $H(X)$ is Tsallis entropy which is given by

$$T_\theta(X) = \frac{1}{1-\theta}\left(\int_0^{+\infty} f^\theta(x)dx - 1\right), \quad 0<\theta\neq 1. \tag{1.4}$$

Tsallis entropy reduces to Shannon entropy when $\theta \to 1$. Renyi and Tsallis entropy are related through

$$H_\theta(X) = \frac{1}{1-\theta}\log\left(1-(\theta-1)T_\theta(X)\right).$$

The difference between the two measures is that one additive but the other is non-additive. It follows from Eq. (1.4) that,

$$T_\theta(X\cdot Y) = T_\theta(X)+T_\theta(Y)+(1-\theta)T_\theta(X)T_\theta(Y).$$

The advantages of using Tsallis entropy over Renyi entropy in generalized statistical mechanics are explained in detail in Beck (2009). They noted that Tsallis entropy satisfies concavity and Lesche stability properties that Renyi entropy does not possess. Tsallis entropy is more useful than Shannon entropy in studying information of complex correlated systems where Tsallis entropy becomes additive but Shannon entropy does not (Tsallis et al., 2005). Wilk and Włodarczyk (2008) studied situations where information can be computed only by Tsallis entropy as Shannon entropy fails to do so.

Entropy defined in Eq. (1.1) is always positive but differential entropy (entropy for continuous cases) may be negative. Differential entropy has some drawbacks such as it can not be defined for distributions that do not have densities. For discrete rvs X and Y, conditional entropy of X given Y is zero if and only if (iff) X is a function of Y. However, differential conditional entropy of X given Y is zero does not necessarily imply that X is a function of Y. For practical usage, approximation based on empirical distribution function (edf) has great significance among researchers due to its simplicity and convergence to the real value. But for differential entropy, one can not estimate it by edf. Rao et al. (2004) extended differential entropy by replacing the densities with the survival function (sf) of the rv. This measure is called cumulative residual entropy (CRE) and is defined as

$$CRE(X) = -\int_0^{+\infty} S(x)\log S(x)dx, \tag{1.5}$$

where S is the sf of X. The CRE overcomes the above mentioned challenges that differential entropy faces and it possesses fundamental properties that Shannon entropy has. Like Shannon entropy, CRE is always non-negative and it increases by adding independent components (observations) and decreases by conditioning. Apart from that, CRE also possesses some useful properties such as it can be defined for both discrete and continuous rvs and can be easily estimated by the edf. Also CRE$(X|Y) = 0$ iff X is a function of Y where CRE$(X|Y)$ is the conditional CRE of X given Y. Analogous to CRE, Di Crescenzo and Longobardi (2009) proposed cumulative entropy (CE) measure as

$$CE(X) = -\int_0^{+\infty} F(x)\log F(x)dx, \tag{1.6}$$

where F is the cumulative distribution function (cdf) of X. While CRE is used to measure information related to future lifetime of systems, CE is more suitable to measure information regarding past system lifetime.

Entropy, CRE and CE measures are defined using the pdf, sf and cdf of the rvs under consideration. They only considered the quantitative (i.e. probabilistic) information but in many applied fields it is often required to consider qualitative characteristics or the utility of the random events as well. For example, in a two person game, it is necessary to take into account the various random strategies (quantitative) of the players involved as well as the gain or loss (qualitative) corresponding to the chosen strategy. Noticing the importance of qualitative characteristics of the random events in various applied fields, Belis and Guiasu (1968) introduced the concept of weighted entropy by assigning non-negative weights to each event based on their utility. For further details and applications on weighted entropy measure, see Guiasu (1971, 1986). For the continuous case, weighted differential entropy can be defined as

$$H^w(X) = -\int_0^{+\infty} xf(x)\log f(x)dx, \tag{1.7}$$

where the factor x is the linear weight function that gives more importance to the larger values of the rv X. Weighted entropy can be generalized by using different weight functions depending on the utility of the events. Non-weighted information measures are position free in the sense that they do not depend on the change of location of the rv. In other words, they are the same for X and $X+\mu$, where μ is a constant. But $H^w(X)$ is not a position free measure. It is a shift-dependent measure of information. Misagh et al. (2011) proposed weighted CRE and CE measures and studied various properties.

In reliability study, a component or a system is better than the other if it survives more than the other counterparts. An experimenter has the information about the current age of a system. Therefore, it is necessary to have knowledge about the residual life of the system. Usually, hazard rate (hr) and mean residual life (MRL) function are used for this purpose. The differential entropy is not a suitable measure to analyze residual lifetime. Ebrahimi (1996) proposed an extension of differential entropy to the residual lifetime of the system which is called dynamic residual entropy measure. Suppose a rv X has survived upto time t, then the residual lifetime at t is defined as $X_t = [X - t|X > t]$. The residual Shannon entropy of X is the differential entropy of the residual rv X_t and it is defined as

$$H(X;t) = -\int_0^{+\infty} \frac{f(x)}{S(t)} \log \frac{f(x)}{S(t)} dx. \tag{1.8}$$

Note that $\frac{f(x)}{S(t)}$ is the pdf of X_t, $H(X;t)$ is a function of time (age) and $H(X;0) = H(X)$. Like hr and MRL, it also uniquely determines the underlying distribution. Now suppose at an inspection time t, a system is not working. Then one needs to study the past lifetime of the system to determine the reason for failure. So analysis of past lifetime is also an important topic in reliability. The past lifetime at time t is defined as ${}_tX = [t - X|X < t]$. The past Shannon entropy measure was suggested by Di Crescenzo and Longobardi (2002) which is defined as

$$\bar{H}(X;t) = -\int_0^{+\infty} \frac{f(x)}{F(t)} \log \frac{f(x)}{F(t)} dx, \tag{1.9}$$

where $\frac{f(x)}{F(t)}$ is the pdf of the past lifetime ${}_tX$ and $\bar{H}(X;0) = H(X)$. Dynamic entropy measures are mainly used in reliability for developing various ageing classes and characterizations of life distributions such as exponential, Rayleigh, power and generalized Pareto distributions. Dynamic versions of cumulative entropies are also studied extensively in the literature. The dynamic cumulative residual entropy (DCRE) was introduced by Asadi and Zohrevand (2007) and dynamic cumulative entropy (DCE) was proposed by Di Crescenzo and Longobardi (2009). Over the years, many generalizations of cumulative entropy and their dynamic version are considered along with their potential applications in the field of image processing, actuarial science, statistics and reliability. One may refer to Abbasnejad et al. (2010), Abbasnejad (2011), Kumar and Taneja (2011), Sunoj and Sankaran (2012), Psarrakos and Navarro (2013), Kayal (2015), Sati and Gupta (2015), Xiong et al. (2019) and the references therein. Extensive works on weighted entropy measures are also available in the literature. See, for example, Mirali et al. (2017), Mirali and Baratpour (2017), Kayal (2018), Khammar and Jahanshahi (2018) and Tahmasebi (2020).

Recently, Lad et al. (2015) introduced an alternative measure of information called extropy. Extropy is the complementary dual of entropy measure and for the discrete case it is defined as

$$J(\mathbf{P}) = -\sum_{i=1}^{n}(1-p_i)\log(1-p_i), \tag{1.10}$$

where p_i's are the same as defined earlier. They found that extropy possesses some properties similar to entropy such as maximum extropy distribution is the uniform distribution and it is invariant under monotone transformations and permutations of its probability mass function (pmf). Entropy and extropy of binary distributions are the same and the duality between them is expressed by the relation

$$H(\mathbf{P}) + J(\mathbf{P}) = \sum_{i=1}^{n} H(p_i, 1-p_i) = \sum_{i=1}^{n} J(p_i, 1-p_i). \tag{1.11}$$

One major difference between entropy and extropy is that the scale of maximum entropy $(\log n)$ is unbounded but the scale of maximum extropy $\big((n-1)\log(\frac{n}{n-1})\big)$ is 1 for $n \to +\infty$. Extropy is also defined for continuous rvs by Lad et al. (2015). If p_i's are small, then $J(\mathbf{P}) \approx 1 - \frac{1}{2}\sum_{i=1}^{n} p_i^2$ and based on this result the extropy for continuous rvs is defined as

$$J(X) = -\frac{1}{2}\int_0^{+\infty} f^2(x)dx. \tag{1.12}$$

Lad et al. (2015) used extropy in combination with entropy to develop a new logarithmic scoring rule called *total logarithmic scoring rule* for forecasting alternative distributions. For more details on extropy, see Lad et al. (2015). Motivated from the work of Rao et al. (2004), cumulative residual extropy (CREx) measure has been proposed in the literature by Jahanshahi et al. (2020) as

$$\xi J(X) = -\frac{1}{2}\int_0^{+\infty} S^2(x)dx. \tag{1.13}$$

Recently, Balakrishnan et al. (2022) proposed the shift-depended (weighted) extropy measure as

$$J^w(X) = -\frac{1}{2}\int_0^{+\infty} x f^2(x)dx. \tag{1.14}$$

Ever since its introduction, extropy has become very popular among statisticians and extensive research has been performed on extropy, dynamic extropy and their various generalized measures. Interested readers may refer to Qiu (2017), Qiu and Jia (2018b), Jose and Sathar (2019), Kamari and Buono (2021), Sathar and Nair R (2021), Nair and Sathar (2020), and Sathar and Nair (2021a,b).

1.2 Applications of information measures in reliability and statistics

Information measures have applications in many fields from theoretical physics (see Dong (2016), Nishioka (2014), Saridakis et al. (2018)) to portfolio optimizations in actuarial science (see Mercurio et al. (2020), Li and Zhang (2021)). Statistics and reliability are no exceptions and here we will focus our attention towards the applications of different information measures in reliability and statistics. First consider the relative entropy also known as the Kullback-Leibler (KL) divergence measure (Kullback and Leibler, 1951) that is widely used as a measure of closeness between two densities. The KL information measure between two rvs X and Y with densities f and g is defined as

$$K(X,Y) = \int_0^{+\infty} f(x) \log \frac{f(x)}{g(x)} dx. \tag{1.15}$$

It is important to note that $K(X,Y) \geq 0$ and equality holds iff $f(x) = g(x)$, $\forall x > 0$. Entropy and KL information measures are used quite extensively in statistical inference. Normality tests and other general purpose goodness-of-fit tests are derived using sample entropy and KL information measures. Vasicek (1976) first proposed a consistent estimator of entropy based on sample observations and performed normality tests. Other entropy based normality tests can be found in Prescott (1976), Esteban et al. (2001), Alizadeh Noughabi (2010) and Zamanzade and Arghami (2012). Dudewicz and Van Der Meulen (1981) studied entropy based uniformity test and a general purpose goodness-of-fit test statistic is developed using entropy by Gokhale (1983). Entropy based exponentiality tests are considered by Ebrahimi et al. (1992), Crzcgorzewski and Wirczorkowski (1999) and Taufer (2002). Like KL information measures, many closeness measures are introduced based on Renyi, Tsallis and cumulative entropy measures. Baratpour and Rad (2012) proposed cumulative residual Kullback-Leibler (CRKL) information measure and construct exponentiality test. Recently, Mehrali and Asadi (2021) developed a new method of estimation by minimizing CRKL. The extropy measure defined in Eq. (1.12) also has applications in various testing problems. Using maximum extropy principle, Qiu and Jia (2018a) developed uniformity tests based on extropy of order statistics and record values, Noughabi and Jarrahiferiz (2022) and Xiong et al. (2021) developed test of symmetry. Recently, an exponentiality test based on extropy of record statistics was studied by Xiong et al. (2022). Jahanshahi et al. (2020) used CREx as a measure of independence and Hashempour et al. (2022) performed testing equality between two cdfs using weighted CREx measure.

In reliability and life-testing, information measures gain quite popularity in recent years due to their huge potential of applicability in numerous problems in the said field. In life-testing, the concept of censoring plays an important role. In conducting a life-testing experiment, it is not feasible to continue the experiment until all the items fail because the experiment may run for a very long time and also the cost of the experiment will increase significantly. This is the reason various censored experiments are performed. Some commonly used censoring are Type-I, Type-II, hybrid censoring and progressive censoring. We will briefly discuss various censoring experiments later. An important problem a reliability practitioner often faces is goodness-of-fit tests for various distributions based on censored data. Extropy, extropy and related information measures are used in reliability to develop goodness-of-fit tests under various censoring experiments. Using KL and CRKL measures, Park (2005), Park and Lim (2015) proposed goodness-of-fit tests for exponential distribution for Type-II censored data. Exponentiality tests for Type-I and hybrid censored data are suggested by Pakgohar et al. (2020) and Noughabi and Chahkandi (2018), respectively, using different information measures. For exponentiality tests under progressively Type-II censored data based on various entropy measures, see Balakrishnan et al. (2007), Park and Pakyari (2015), Baratpour and Rad (2016) and Noughabi (2017). Recently, entropy and extropy are used to perform uniformity tests for Progressively Type-II censored samples, see Hazeb et al. (2021a) and Hazeb et al. (2021b). Based on the dynamic information measures, many aging classes are proposed in the literature and characterization results for some life distributions such as uniform, exponential, Rayleigh, Power and generalized Pareto are obtained.

In reliability engineering, information measures are used for measuring complexity of systems i.e. how far away a system is, in terms of complexity, than a k-out-of-n system having the same number of components. Asadi et al. (2016) proposed Jensen-Shannon information measure for comparing coherent systems. They also studied the complexity of coherent systems. Similar studies are performed in terms of extropy (Jensen-Extropy) measure by Qiu et al. (2019). Toomaj et al. (2017) studied CRE of coherent and mixed systems and they proposed a new ordering to compare two systems when usual stochastic order comparisons can not be made. Applications of information measures in "used but still working" systems can be found in Toomaj et al. (2021).

1.3 Preliminary concepts

Here we provide basic definitions of some preliminary concepts that are used in reliability analysis. We will use them throughout the course of this thesis. Also we mention some important inequalities which we use throughout the course of the thesis.

1.3.1 Hazard rate

The conditional probability that an item will fail in $(t, t+\Delta t)$, given that the item has survived upto time t, is given by

$$P(t < X < t+\Delta t | X > t) = \frac{F(t+\Delta t) - F(t)}{S(t)}.$$

Then, the hr function can be defined as

$$\begin{aligned} \lambda_F(t) &= \lim_{\Delta t \to 0+} \frac{P(t < X < t+\Delta t | X > t)}{\Delta t} \\ &= \frac{f(t)}{S(t)}, \quad \text{[for continuous rvs]}. \end{aligned} \tag{1.16}$$

It is also called instantaneous failure rate. The hr function uniquely determines the underlying distribution function by the relation

$$S(t) = \exp\left(-\int_0^t \lambda_F(u) du\right).$$

It has many applications in reliability analysis in characterizing distributions, developing aging classes and many testing problems, see Barlow and Proschan (1975), Nanda and Shaked (2001) and Noughabi et al. (2013).

1.3.2 Reversed hazard rate

The reversed hazard rate (rhr) function of a rv X is defined as

$$\begin{aligned} r_F(t) &= \lim_{\Delta t \to 0+} \frac{P(t < X < t+\Delta t | X > t)}{\Delta t} \\ &= \frac{f(t)}{F(t)}, \quad \text{[for continuous rvs]}. \end{aligned} \tag{1.17}$$

Like $\lambda_F(t)$, rhr function also uniquely determines the distribution through the relationship

$$F(t) = \exp\left(-\int_t^{+\infty} r_F(u) du\right).$$

For detailed discussions and applications of $r_F(t)$, one may refer to Gupta and Nanda (2001), Nanda and Shaked (2001) and Kundu and Ghosh (2017).

1.3.3 Mean residual and mean past life

Mean residual life function (MRL) of a rv variable X is the expected value of the residual rv X_t. It can be defined as

$$m_F(t) = E[X-t|X>t] = \int_t^{+\infty} \frac{S(x)}{S(t)}dx. \tag{1.18}$$

The mean past life function (MPL), also known as mean inactivity time function (MIT), of X is the expectation of the inactivity time ${}_tX$. It is given by

$$\mu_F(t) = E[t-X|X<t] = \int_0^t \frac{F(x)}{F(t)}dx. \tag{1.19}$$

1.3.4 Weighted mean residual and mean past life

Misagh et al. (2011) first introduced the weighted MRL and MPL functions for obtaining bounds of weighted CRE and CE measures. These functions are often used along with dynamic cumulative information measures for characterizations of Rayleigh and power distributions. The weighted MRL (WMRL) of a rv X is defined as

$$m_F^w(t) = \int_t^{+\infty} x\frac{S(x)}{S(t)}dx. \tag{1.20}$$

Note that, $m_F^*(0) = \int_0^{+\infty} x\bar{F}(x)dx = \frac{1}{2}E(X^2)$. The weighted MPL (WMPL) is given by

$$\mu_F^w(t) = \int_0^t x\frac{F(x)}{F(t)}dx. \tag{1.21}$$

Recently, analysis of WMRL and WMPL functions became a problem of interest among researchers. The relationships between WMRL and WMPL with other functions like variance and some generalized informations measures can be found in Toomaj and Di Crescenzo (2020) and Di Crescenzo and Toomaj (2022). They also proposed extensions of WMRL and WMPL measures and provide various applications.

1.3.5 Inequalities

Bernoulli inequality

Bernoulli inequality has many different variants and the one we will use states that, for $\theta \geq 1$ and $0 \leq u \leq 1$,

$$(1-u)^\theta \geq 1 - u\theta.$$

Markov inequality

For a non-negative continuous rv X Markov inequality states that

$$P(X \geq C) \leq \frac{E(X)}{C},$$

where $C > 0$ is a constant.

Log-sum inequality

Suppose a non-negative continuous rv X has support A and f and g are positive functions then

$$\int_A f(x) \log\left(\frac{f(x)}{g(x)}\right) dx \geq \left(\int_A f(x) dx\right) \log \frac{\int_A f(x) dx}{\int_A g(x) dx}.$$

1.3.6 Stochastic orders

Here we provide definitions of some basic stochastic orderings which will be utilized later. Consider the following definitions. For details one may refer to Shaked and Shanthikumar (2007).

Definition 1.3.1. *Let X_1 and X_2 be two rvs with sfs S_1 and S_2, respectively. Then X_1 is smaller than X_2 in stochastic ordering, denoted by $X_1 \overset{st}{\leq} X_2$, if $S_1(x) \leq S_2(x)$, for all x.*

Definition 1.3.2. *X_1 is smaller than X_2 in hazard rate ordering, denoted by $X_1 \overset{hr}{\leq} X_2$, if $\lambda_F(t) \geq \lambda_G(t)$, $\forall t \geq 0$ or equivalently $\frac{\bar{G}(t)}{\bar{F}(t)}$ is increasing in t.*

Definition 1.3.3. *X_1 is smaller than X_2 in rhr ordering, denoted by $X_1 \overset{rh}{\leq} X_2$, if $r_F(t) \leq r_G(t)$, $\forall t \geq 0$ or equivalently $\frac{G(t)}{F(t)}$ is increasing in t.*

Definition 1.3.4. *Let X_1 and X_2 be absolutely continuous rvs having pdfs f and g and sfs S_1 and S_2, respectively. Then X_1 is said to be smaller than X_2 in dispersive order, denoted by $X_1 \overset{disp}{\leq} X_2$, if $f(S_1^{-1}(v)) \geq g(S_2^{-1}(v))$, $0 < v < 1$.*

1.4 Salient features of the thesis

Salient features of the thesis are as follows:

1. We propose various weighted generalized cumulative information measures along with their dynamic versions and study numerous properties and bounds. Characterization results for Rayleigh and power distributions are obtained using the dynamic

measures. Characterizations for distributions in terms of information of the smallest and largest order statistics are also developed. New aging classes based on the dynamic information measures are proposed.

2. We investigate different methods of estimation of the proposed measures as well as some existing cumulative measures of information. Various non-parametric estimators of the measures are proposed throughout the course of this thesis, their asymptotic properties are studied and their performances are investigated by simulation.

3. One important feature of the thesis is the variety of applications, of the proposed and existing information measures, that are considered throughout the course of this thesis. The majority of the said applications are concerned in the context of statistics and reliability. Goodness-of-fit tests for exponential and uniform distributions and testing equality between two cdfs are developed. Some potential applications in the field of actuarial science and model discrimination are also discussed.

4. Applications in system reliability are investigated using information measures. New method of analyzing system complexity is proposed and comparisons between two systems are studied when usual stochastic ordering can not be implemented. Applications involving redundancy are also considered.

5. Optimal life testing plans for progressive Type-II censored experiments are developed by maximizing new cumulative entropy based design criteria. Maximum cumulative residual entropy design subject to cost constraint is studied. Also compound optimal design strategy is implemented where cumulative entropy and cost is optimized simultaneously.

1.5 The summary of the thesis

The chapter-wise summary of the thesis is presented in the following.

[Chapter 2] Generalized weighted survival and failure entropies and their dynamic versions

In this Chapter, two new information measures called generalized weighted survival entropy (GWSE) and generalized weighted failure entropy (GWFE) are proposed. The dynamic versions of these proposed measures are also introduced. For a rv X, GWSE of order (θ_1, θ_2) is defined as

$$\xi^w_{\theta_1,\theta_2}(X) = \frac{1}{\theta_2 - \theta_1} \log \int_0^{+\infty} x S^{\theta_1+\theta_2-1}(x)dx,\ \theta_2 \geq 1,\ \theta_2 - 1\ < \theta_1 <\ \theta_2.$$

We define the GWFE measure for a rv X with finite support $[0,l]$, as

$$f\xi^w_{\theta_1,\theta_2}(X) = \frac{1}{\theta_2 - \theta_1} \log \int_0^{l} x F^{\theta_1+\theta_2-1}(x)dx,\ \theta_2 \geq 1,\ \theta_2 - 1\ < \theta_1 <\ \theta_2.$$

The advantages of these weighted measures are shown and their properties are studied. The dynamic GWSE is defined as

$$\xi^w_{\theta_1,\theta_2}(X;t) = \frac{1}{\theta_2 - \theta_1} \log \int_t^{+\infty} x \frac{S^{\theta_1+\theta_2-1}(x)}{S^{\theta_1+\theta_2-1}(t)}dx,\ \theta_2 \geq 1,\ \theta_2 - 1\ < \theta_1 <\ \theta_2.$$

We propose new stochastic ordering based on these measures and also propose new aging classes using the dynamic information measures. The dynamic GWSE and GWFE measures uniquely determine the underlying distribution. Characterization results for the Rayleigh distribution are obtained using $\xi^w_{\theta_1,\theta_2}(X;t)$ and it is shown that $\xi^w_{\theta_1,\theta_2}(X;t)$ is constant iff X follows the Rayleigh distribution. We also study characterization for the power distribution based on dynamic GWFE measure. More results on characterization using information of extreme order statistics, some generalized inequalities and stochastic ordering results are discussed in detail. Non-parametric estimators of GWSE and GWFE measures are proposed based on edf of the underlying rv. A test of exponentiality is considered using empirical GWSE as an application.

[Chapter 3] Weighted cumulative Tsallis residual and past entropy measures

In this Chapter, we propose weighted cumulative Tsallis residual entropy (WCTRE) and its dynamic version and study various properties and bounds. For a rv X, WCTRE is defined as

$$\xi^w_\theta(X) = \frac{1}{\theta - 1} \int_0^{+\infty} x \left(S(x) - S^\theta(x)\right) dx,\ 0 < \theta \neq 1.$$

When $\theta \to 1$, WCTRE reduces to weighted cumulative residual entropy (WCRE) proposed by Misagh et al. (2011) as $CRE^w(X) = -\int_0^{+\infty} xS(x) \log S(x)dx$. So WCRE is a special case of WCTRE measure. It is shown that Dynamic WCTRE uniquely determines the underlying distribution. Also we propose weighted cumulative Tsallis past entropy (WCTPE) as

$$\bar{\xi}^w_\theta(X) = \frac{1}{\theta - 1} \int_0^{+\infty} x \left(F(x) - F^\theta(x)\right) dx,\ 0 < \theta \neq 1.$$

We propose non-parametric estimators of these measures based on edf and discuss their asymptotic properties. We obtain empirical WCTPE as

$$\hat{\hat{\xi}}_{\theta}^{w}(X) = \frac{1}{2(\theta-1)} \sum_{i=1}^{n-1} (X_{(i+1):n}^{2} - X_{i:n}^{2}) \left[\frac{i}{n} - \left(\frac{i}{n} \right)^{\theta} \right].$$

For rs comes from Rayleigh distribution, it is shown that for every $0 < \theta \neq 1$,

$$\frac{\hat{\hat{\xi}}_{\theta}^{w}(X) - E[\hat{\hat{\xi}}_{\theta}^{w}(X)]}{\sqrt{Var[\hat{\hat{\xi}}_{\theta}^{w}(X)]}} \rightarrow N(0,1)$$

in distribution as $n \rightarrow +\infty$. This result can be used to test whether data comes from the Rayleigh distribution.

[Chapter 4] On weighted cumulative residual Kullback-Leibler information with application in testing exponentiality

This Chapter considers the study of weighted cumulative Kullback-Leibler type information measures. We introduce weighted cumulative residual Kullback-Leibler (WCRKL) information measure which is based on the WCRE measure. The WCRKL between two rvs X and Y is defined as

$$CRKL^{w}(X,Y) = \int_{0}^{+\infty} xS_1(x) \log \frac{S_1(x)}{S_2(x)} dx + \frac{1}{2}(E(Y^2) - E(X^2)).$$

It is shown that WCRKL is non-negative and it is zero when X and Y have the same distribution. Next we introduce weighted cumulative past Kullback-Leibler information measure using the cdfs. Their dynamic versions are also considered and various properties are studied.

Using WCRKL measure, we construct a goodness-of-fit test for the exponential distribution under complete and Type-I and Type-II censored samples. The performance of the test is compared with KL based test and cumulative residual KL information based test. Proposed test performs better than the other two tests when the alternative distribution has a decreasing hazard rate.

[Chapter 5] Applications of cumulative residual extropy in system reliability and hypothesis testing problems

This Chapter focuses on the applications of CREx measure defined in Eq. 1.13. First we discuss CREx for coherent and mixed reliability systems. We represent the CREx measure for mixed systems and compare two systems consisting of the same structure but different components. We define a new Jensen type divergence measure called Jensen cumulative residual extropy divergence (JCREx) and study the complexity of systems using this divergence measure. Consider a system with lifetime T having n iid components X_1, X_2, $\cdots, X_n$ and suppose $X_{1:n}, X_{2:n}, \cdots, X_{n:n}$ are the ordered components lifetime. Then the JCREx divergence between system lifetime T and $X_{1:n}, X_{2:n}, \cdots, X_{n:n}$ can be defined as

$$JCREx(T : X_{1:n}, \cdots, X_{n:n}) = \xi J(T) - \sum_{i=1}^{n} s_i \xi J(X_{i:n}),$$

where $\mathbf{s} = (s_1, s_2, \cdots, s_n)$ is the signature vector (Samaniego, 2007). Signature vector is a probability vector whose j-th component represents the probability that the system fails due to the failure of the j-th component. Note that, $JCREx(T : X_{1:n}, \cdots, X_{n:n}) = 0$ for k-out-of-n systems and higher values of $JCREx(T : X_{1:n}, \cdots, X_{n:n})$ implies the system is more complex than the k-out-of-n systems having same number of components. Also we define a new discrimination measure which can be used to compare between systems when usual stochastic orders can not be used (Navarro et al., 2008).

So far, we have proposed estimators of the information measures using edf of the rv under consideration. In this Chapter we propose edf based estimator for CREx and estimator based on L-Statistics and compare their performance with kernel based smooth estimator of CREx measure. The L-Statistics estimator is defined as

$$\xi J^2_{F_n}(X) = -\frac{1}{n} \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) X_{i:n},$$

which is asymptotically normally distributed and has better performance than the other two estimators. Using this estimator, we develop a consistent test of equality between two distribution functions. The asymptotic distribution of the test statistic follows normal distribution under the null hypothesis.

[Chapter 6] On some weighted generalized extropy measures with applications

In this Chapter, we propose a generalized weighted cumulative measure of information called weighted extended survival extropy (WESEx) measure. Instead of x, we use a non-negative continuous function of x as the weight function. For a rv X, WESEx is defined as

$$\xi J^{\varepsilon}(X) = -\frac{1}{2} \int_0^{+\infty} \varepsilon(x) S^2(x) dx,$$

where $\varepsilon(x)$ is a non-negative continuous function of x. We also propose the dynamic WESEx measure as

$$\xi J^{\varepsilon}(X;t) = -\frac{1}{2S^2(t)} \int_t^{+\infty} \varepsilon(x) S^2(x) dx.$$

Depending on the choice of $\varepsilon(x)$, one can obtain a variety of extropy measures from $\xi J^{\varepsilon}(X)$. Various interesting properties of these measures are studied. We consider estimations of these measures for iid samples and also when observations are identically distributed but not independent. For the dependent setup, a recursive kernel based estimator is used. Asymptotic properties of these estimators are studied in detail and various examples are provided for illustrative purposes. Applications in model discrimination and quantitative risk analysis are considered. We also propose weighted negative cumulative extropy measure and study its properties. For a non-negative continuous rv X it is defined as

$$\mathscr{C}^w(X) = \frac{1}{2} \int_0^{+\infty} x \left[1 - F^2(x)\right] dx.$$

Using non-parametric estimator of $\mathscr{C}^w(X)$, a uniformity test is developed.

[Chapter 7] Application of cumulative entropy measures in life testing

This Chapter discusses the applications of cumulative entropy measures in the context of life-testing. We consider designing optimal progressive Type-II (PCII) censored experiments by optimizing CRE (CE) based criterion. Abo-Eleneen et al. (2018) defined joint CRE of progressively type-II censored ordered statistics (PCOS) as

$$CRE_{1\cdots m:m:n} = -\int_0^{+\infty} \frac{1}{h(x)} \log S(x) \sum_{i=1}^{m} f_{X_{i:m:n}}(x) dx,$$

where $f_{X_{i:m:n}}$ is the pdf of $X_{1:m:n}$ and $h(x)$ is the hr of X. This can be treated as a measure of information of the PCII censored experiment and it is natural to design experiment by maximizing information. But $CRE_{1\cdots m:m:n}$ is not scale invariant so we obtain optimal design by maximizing a scale invariant criterion

$$\phi_A(\mathscr{R}) = \frac{CRE_{1\cdots m:m:n}}{E[X_{1:m:n}]}.$$

Maximizing information of a PCII censored experiment will increase the duration of the experiment and thus the cost of running that experiment will also increase. So we study a constraint design problem in which we maximize $\phi_A(\mathscr{R})$ subject to a bound on the cost associated with the experiment. Finally, we study optimal design problems by simultaneously maximizing $\phi_A(\mathscr{R})$ and minimizing the cost of the experiment. For this purpose, we implement a simultaneous optimal design strategy for life-testing called compound optimal design (Bhattacharya, 2020). Similar design problems are considered using CE based criterion as well.

[Chapter 8] Conclusions and future work

This chapter indicates areas of future research. We discuss applications in reliability acceptance sampling plan (RASP). Using the information measures studied in this thesis, a RASP can be developed for both classical and bayesian frameworks. Some remarks are made in this regard.

We propose weighted extended survival extropy measure by taking a continuous function of x as the weight function. We introduce various weighted generalized entropy measures and their extended version can be defined similarly. We discuss some problems regarding weighted extended entropy measures.

Some problems regarding double truncated (interval) information measures are discussed. New methods of estimation of the proposed measures are also discussed.

1.6 Publications from the thesis

1. Chakraborty, S., & Pradhan, B. (2023). Generalized weighted survival and failure entropies and their dynamic versions. *Communications in Statistics-Theory and Methods*, **52(3)**: 730-750 [Out of Chapter 2].

2. Chakraborty, S., & Pradhan, B. (2023). On weighted cumulative Tsallis residual and past entropy measures. *Communications in Statistics-Simulation and Computation*, **52(5)**: 2058-2072 [Out of Chapter 3].

3. Chakraborty, S., & Pradhan, B. (2022). Weighted cumulative residual Kullback-Leibler divergence: properties and applications. *Communications in Statistics- Simulation and Computation*, DOI: 10.1080/03610918.2022.2108053, Published online. [Out of Chapter 4].

4. Chakraborty, S., & Pradhan, B. Cumulative residual extropy of coherent and mixed systems. Under revision. [Out of Chapter 5].

5. Chakraborty, S., & Pradhan, B. (2023). On estimation of cumulative residual extropy and its quantile version. *Ricerche di Matematica*, DOI: 10.1007/s11587-022-00757-7, Published online. [Out of Chapter 5].

6. Chakraborty, S., & Pradhan, B. (2022). Some properties of weighted survival extropy and its extended measures. *Communications in Statistics-Theory and Methods*, DOI: 10.1080/03610926.2022.2076118, Published online. [Out of Chapter 6].

7. Chakraborty, S., Das, O. & Pradhan, B. On negative cumulative extropy with application in testing uniformity. Physica A: Statistical Mechanics and its Applications, DOI: 10.1016/j.physa.2023.128957. [Out of Chapter 6].

8. Chakraborty, S., Bhattacharya , R. & Pradhan, B. (2023). Cumulative entropy of progressively type-II censored order statistics and associated optimal life testing-plans. *Statistics*, **57(1)**: 161-174 [Out of Chapter 7].

9. Chakraborty, S., Bhattacharya , R. & Pradhan, B. On the application of compound optimal design strategy in progressively type-II censored life-testing experiments. Communicated. [Out of Chapter 7].

Chapter 2

Generalized Weighted Survival and Failure Entropies and their Dynamic Versions

GENERALIZATION of cumulative entropy measures became a problem of interest among researchers in the last two decades. Ever since the introduction of CRE and CE measures, many generalizations of them are proposed in the literature. The dynamic versions of CRE and CE are introduced in the literature by Asadi and Zohrevand (2007) and Di Crescenzo and Longobardi (2009), respectively. Zografos and Nadarajah (2005) introduced survival entropy of order θ_1 as $\xi_{\theta_1}(X) = \frac{1}{1-\theta_1}\log\int_0^{+\infty} S^{\theta_1}(x)dx,\ \theta_1(\neq 1) > 0$ and Abbasnejad (2011) proposed its dynamic version. For a rv X having bounded support $[0,l]$, Abbasnejad (2011) introduced the failure entropy of order θ_1 as $f\xi_{\theta_1}(X) = \frac{1}{1-\theta_1}\log\int_0^l F^{\theta_1}(x)dx$ and also obtained its dynamic version. Further generalizations of survival and failure entropy and their dynamic measures are considered by Kayal (2015). Generalized survival entropy (GSE) of order (θ_1, θ_2) is defined as

$$\xi_{\theta_1,\theta_2}(X) = \frac{1}{\theta_2-\theta_1}\log\int_0^{+\infty} S^{\theta_1+\theta_2-1}(x)dx,\ \theta_2 \geq 1,\ \theta_2 - 1 < \theta_1 < \theta_2, \tag{2.1}$$

where the integral in Eq. (2.1) is extended to the support of the rv. Note that, the survival entropy of order θ_1 ($\xi_{\theta_1}(X)$) is the cumulative residual Renyi entropy which defined by replacing the pdf in Renyi entropy, defined in Eq. (1.3), with the sf of the underlying rv. Whereas the GSE of order (θ_1, θ_2) is the cumulative residual Varma entropy.

Varma entropy is a generalization of Renyi entropy and a two parameter generalization of the Shannon entropy. For a non-negative continuous rv, Varma entropy of order θ_1 and type θ_2 is defined as

$$H_{\theta_1,\theta_2}(X) = \frac{1}{\theta_2 - \theta_1} \log \int_0^{+\infty} f^{\theta_1-\theta_2-1}(x)dx, \;\; \theta_2 \geq 1, \; \theta_2 - 1 \; < \theta_1 < \; \theta_2.$$

Note that, when $\theta_2 = 1$, it reduces to Renyi entropy and when $\theta_2 = 1$ and $\theta_1 \to 1$, $H_{\theta_1,\theta_2}(X)$ tends to Shannon entropy. A more popular and useful version of Varma entropy is

$$\bar{H}_{\theta_1,\theta_2}(X) = \frac{1}{\theta_2 - \theta_1} \log \int_0^{+\infty} f^{\theta_1+\theta_2-1}(x)dx, \;\; \theta_2 \geq 1, \; \theta_2 - 1 \; < \theta_1 < \; \theta_2.$$

This measure is also called generalized entropy of order (θ_1, θ_2). Varma's work opened a new area regarding two parameters generalized entropy measures and various works in this area followed ever since, see Sharma and Taneja (1975), Mittal (1975), Kapur (1967) and Kattumannil et al. (2022). Varma entropy is an important complexity measure in physics where two parameters are useful for determining uncertainties for chaotic systems because of the increased flexibility of entropy measure due to various choices of the two generalizing parameters. Also two-parameter generalized entropy measures are more sensitive to the shape of the underlying distribution and have a large range which is often useful in many applied problems (Pharwaha and Singh, 2009). Ullah (1996) observed that optimization of Shannon entropy often required complicated moment conditions to determine the distribution. They pointed out that in many situations, it is useful to have a complicated information measure but simpler and fewer moment conditions. For detailed discussions on the usefulness on two parameter entropy measures, see Kundu and Singh (2020). Over the years, these measures became popular because of their diverse applications and their cumulative versions are explored recently by Kayal (2015).

For a bounded rv with support $[0,l]$, Kayal (2015) defined generalized failure entropy (GFE) as

$$f\xi_{\theta_1,\theta_2}(X) = \frac{1}{\theta_2 - \theta_1} \log \int_0^{l} F^{\theta_1+\theta_2-1}(x)dx, \; \theta_2 \geq 1, \; \theta_2 - 1 \; < \theta_1 < \; \theta_2. \tag{2.2}$$

In this chapter, we propose weighted GSE and GFE measures along with their dynamic versions and study various properties and bounds. We introduce new aging classes based on the dynamic measures and obtain various characterization results. Also we propose nonparametric estimators for the newly defined weighted information measures and a test of exponentiality is considered as an application.

The organization of this chapter is as follows. We define weighted versions of GSE and its dynamic measure and study their properties in Section 2.1. Weighted GFE is introduced along with its dynamic form in Section 2.2. Two new aging classes based on the proposed dynamic weighted information measures are considered in Section 2.3. Characterization results based on the proposed dynamic information measures are studied in Section 2.4. Some general class of bounds of the proposed information measures are provided in Section 2.5. Non-parametric estimators are considered and a goodness-of-fit test for exponential distribution is proposed in Section 2.6. Finally, an overall discussion is made in Section 2.7.

2.1 Generalized weighted survival entropy of order (θ_1, θ_2) and its dynamic version

Here we introduce generalized weighted survival entropy (GWSE) and its dynamic version, and obtain various properties.

Definition 2.1.1. *For a non-negative continuous rv X, GWSE of order (θ_1, θ_2) is defined as*

$$\xi^w_{\theta_1,\theta_2}(X) = \frac{1}{\theta_2-\theta_1}\log\int_0^{+\infty} xS^{\theta_1+\theta_2-1}(x)dx,\ \theta_2 \geq 1,\ \theta_2-1 < \theta_1 < \theta_2. \tag{2.3}$$

Usefulness of this weighted measure over the non-weighted version can be explained by the following example.

Example 2.1.1. *Suppose X has uniform distribution, denoted by $U(a,b)$, with pdf*

$$f(x) = \frac{1}{b-a},\ a < x < b$$

and Y has $U(a+h, b+h)$ distribution with pdf

$$g(x) = \frac{1}{b-a},\ a+h < x < b+h,\ h > 0.$$

Then from Eq. (2.1), we have

$$\xi_{\theta_1,\theta_2}(X) = \xi_{\theta_1,\theta_2}(Y) = \frac{1}{\theta_2-\theta_1}\log\frac{b-a}{\theta_1+\theta_2}.$$

From Eq. (2.3) we get,

$$\xi^w_{\theta_1,\theta_2}(X) = \frac{1}{\theta_2-\theta_1}\log\left[\frac{(b-a)(a(\theta_1+\theta_2)+b)}{(\theta_1+\theta_2)(\theta_1+\theta_2+1)}\right],$$
$$\xi^w_{\theta_1,\theta_2}(Y) = \frac{1}{\theta_2-\theta_1}\log\left[\frac{(b-a)(a(\theta_1+\theta_2)+b+h(\theta_1+\theta_2+1))}{(\theta_1+\theta_2)(\theta_1+\theta_2+1)}\right].$$

It is easy to see that $\xi^w_{\theta_1,\theta_2}(X) < \xi^w_{\theta_1,\theta_2}(Y)$. So it is observed that $\xi_{\theta_1,\theta_2}(X) = \xi_{\theta_1,\theta_2}(Y)$, but GWSE of X is smaller than GWSE of Y. This is due to the fact that GSE measure is position free i.e. shift independent measure. The GSE for a rv X and $X+h$, where h is a constant, are the same. However, GWSE measures for X and $X+h$ are different. So the information contained in X and Y are the same in terms of GSE measure but they are different in terms of GWSE measure. This is one scenario where use of GWSE will be beneficial instead of GSE measure since GWSE is a shift-dependent measure of information. The shift-dependency property of $\xi^w_{\theta_1,\theta_2}(X)$ is expressed in the following lemma.

Lemma 2.1.1. *Consider the linear transformation $Z = cX + d$, where $c > 0$ and $d \geq 0$, then*

$$\exp[(\theta_2-\theta_1)\xi^w_{\theta_1,\theta_2}(Z)] = c^2 \exp[(\theta_2-\theta_1)\xi^w_{\theta_1,\theta_2}(X)] + cd\exp[(\theta_2-\theta_1)\xi_{\theta_1,\theta_2}(X)] \quad (2.4)$$

Proof. The results follows using $S_{cX+d}(x) = S_X(\frac{x-d}{c})$, $x \in R$. □

Let S_{X_η} and S denote the sfs of the rvs X_η and X, respectively. Then X_η and X satisfy proportional hazard rate model if $S_{X_\eta}(x) = [S(x)]^\eta$, $\eta > 0$, see Cox (1972). The following lemma compares the GWSE of X, X_η and ηX.

Lemma 2.1.2. *The following results hold:*

$$(a)\ \xi^w_{\theta_1,\theta_2}(X_\eta) = \left(\frac{\eta\theta_2 - \eta\theta_1 - \eta + 1}{\theta_2 - \theta_1}\right)\xi^w_{\eta\theta_1,\eta\theta_2-\eta+1}(X);$$
$$(b)\ \xi^w_{\theta_1,\theta_2}(X_\eta) \leq \xi^w_{\theta_1,\theta_2}(X) \leq \xi^w_{\theta_1,\theta_2}(\eta X),\ if\ \eta > 1;$$
$$(c)\ \xi^w_{\theta_1,\theta_2}(X_\eta) \geq \xi^w_{\theta_1,\theta_2}(X) \geq \xi^w_{\theta_1,\theta_2}(\eta X),\ if\ 0 < \eta < 1.$$

Proof. Proof is straight forward hence omitted. □

We provide GWSE for exponential and Pareto I distributions in Table 2.1 to support Lemma 2.1.2. The exponential distribution has the cdf

$$F(x) = 1 - e^{-\lambda x};\ x > 0, \lambda > 0,$$

and Pareto I distribution has the cdf

$$F(x) = 1 - \left(\frac{b}{x}\right)^a;\ x \geq b > 0,\ a > 0.$$

It is important to note that for Pareto I distribution when $a\gamma$, $a\eta\gamma$ and $\eta\gamma$ are less than 2, GWSE won't be finite.

Table 2.1: *GWSE for exponential and Pareto I distributions where* $\gamma = \theta_1 + \theta_2 - 1$.

Distribution	$(\theta_2 - \theta_1)\xi^w_{\theta_1,\theta_2}(X)$	$(\theta_2 - \theta_1)\xi^w_{\theta_1,\theta_2}(X_\eta)$	$(\theta_2 - \theta_1)\xi^w_{\theta_1,\theta_2}(\eta X)$
Exponential	$-2\log(\lambda\gamma)$	$-2\log(\lambda\eta\gamma)$	$2\log\eta - 2\log(\lambda\gamma)$
Pareto I	$\log\frac{b^2}{a\gamma-2}$; $a\gamma > 2$	$\log\frac{b^2}{a\eta\gamma-2}$; $a\eta\gamma > 2$	$\log\frac{b^2\eta^2}{\eta\gamma-2}$; $a\gamma > 2$

Now we define the dynamic version of GWSE to study the uncertainty in the residual life $X_t = [X - t | X \geq t]$ of a unit. It is the GWSE of the residual life X_t.

Definition 2.1.2. *Generalized dynamic weighted survival entropy (GDWSE) of order* (θ_1, θ_2) *of a continuous rv X is defined as*

$$\xi^w_{\theta_1,\theta_2}(X;t) = \frac{1}{\theta_2 - \theta_1}\log\int_t^{+\infty} x\frac{S^{\theta_1+\theta_2-1}(x)}{S^{\theta_1+\theta_2-1}(t)}dx,\ \theta_2 \geq 1,\ \theta_2 - 1 < \theta_1 < \theta_2. \tag{2.5}$$

Note that, $\xi^w_{\theta_1,\theta_2}(X;0) = \xi^w_{\theta_1,\theta_2}(X)$. In the following theorem we provide bounds for GWSE and GDWSE measures in terms of WMRL of the rv.

Theorem 2.1.1. *For a rv X with WMRL* $m^w_F(t)$*, the following inequalities hold for* $\theta_1 + \theta_2 - 1 \geq (\leq) 1$*:*

$$(i)\quad \xi^w_{\theta_1,\theta_2}(X) \leq (\geq)\frac{1}{\theta_2 - \theta_1}\log m^w_F(0);$$

$$(ii)\quad \xi^w_{\theta_1,\theta_2}(X;t) \leq (\geq)\frac{1}{\theta_2 - \theta_1}\log m^w_F(t).$$

Proof. (i) For $\theta_1 + \theta_2 - 1 \geq (\leq) 1$ we have $xS^{\theta_1+\theta_2-1}(x) \leq (\geq) xS(x)$. Taking integral on both sides and then taking logarithm, and dividing by $(\theta_2 - \theta_1)$ we get the result.

(ii) Since $\frac{S(x)}{S(t)} < 1$ for $x > t$, we have $\left(\frac{S(x)}{S(t)}\right)^{\theta_1+\theta_2-1} \leq (\geq)\frac{S(x)}{S(t)}$ for $\theta_1 + \theta_2 - 1 \geq (\leq) 1$. The result follows by taking integrals. □

Table 2.2: *GDWSE and WMRL for exponential and Pareto I distributions where* $\gamma = \theta_1 + \theta_2 - 1$. *For Pareto I distribution we assume* $t > b$.

Distribution	$\xi^w_{\theta_1,\theta_2}(X)$	$m^w_F(0)$	$\xi^w_{\theta_1,\theta_2}(X;t)$	$m^w_F(t)$
Exponential	$\frac{2}{\theta_2-\theta_1}\log(\frac{1}{\lambda\gamma})$	$\frac{1}{\lambda^2}$	$\frac{1}{\theta_2-\theta_1}\log\left(\frac{1+t\lambda\gamma}{\lambda^2\gamma^2}\right)$	$\frac{1+t\lambda}{\lambda^2}$
Pareto I	$\frac{1}{\theta_2-\theta_1}\log\frac{b^2}{a\gamma-2}$; $a\gamma > 2$	$\frac{ab^2}{2(a-2)}$; $a > 2$	$\log\frac{t^2}{a\gamma-2}$; $a\gamma > 2$	$\frac{t^2}{a-2}$; $a > 2$

We provide GWSE and its dynamic version along with WMRL for exponential and Pareto I distributions in Table 2.2 for illustration and verification of Theorem 2.1.1.
Next, the effect of linear transformation on GDWSE measure is considered. This result will be useful for stochastic comparison purposes later. Consider the following lemma.

Lemma 2.1.3. *Suppose $Z = cX + d$, where $c > 0$ and $d \geq 0$, then*

$$\exp[(\theta_2 - \theta_1)\xi^w_{\theta_1,\theta_2}(Z;t)] = c^2 \exp\left[(\theta_2 - \theta_1)\xi^w_{\theta_1,\theta_2}\left(X; \frac{t-d}{c}\right)\right] + cd\exp\left[(\theta_2 - \theta_1)\xi_{\theta_1,\theta_2}\left(X; \frac{t-d}{c}\right)\right].$$

Proof. The proof is similar to Lemma 2.1.1. □

Remark 2.1.1. *If $d = 0$, then from Lemma 2.1.3 we have*

$$\xi^w_{\theta_1,\theta_2}(Z;t) = \frac{2\log c}{\theta_2 - \theta_1} + \xi^w_{\theta_1,\theta_2}\left(X; \frac{t}{c}\right). \tag{2.6}$$

Now we define two new stochastic orderings in terms of GWSE and GDWSE measures and study their relationships with some popular stochastic orderings.

Definition 2.1.3. *X_1 is smaller than X_2 in GWSE ordering, denoted by $X_1 \overset{GWSE}{\leq} X_2$, if*

$$\xi^w_{\theta_1,\theta_2}(X_1) \leq \xi^w_{\theta_1,\theta_2}(X_2).$$

Definition 2.1.4. *X_1 is smaller than X_2 in GDWSE ordering, denoted by $X_1 \overset{GDWSE}{\leq} X_2$, if*

$$\xi^w_{\theta_1,\theta_2}(X_1;t) \leq \xi^w_{\theta_1,\theta_2}(X_2;t), \quad \forall\, t \geq 0.$$

In the following theorem, we provide the relationship of GWSE and GDWSE ordering with stochastic and hr orderings.

Theorem 2.1.2. *Let X_1 and X_2 be two non-negative continuous rvs with sfs S_1 and S_2, respectively. Then,*

(i) $X_1 \overset{st}{\leq} X_2 \Rightarrow X_1 \overset{GWSE}{\leq} X_2$,

(ii) $X_1 \overset{hr}{\leq} X_2 \Rightarrow X_1 \overset{GDWSE}{\leq} X_2$.

Proof. Proof of (i) follows from the fact that, $X_1 \overset{st}{\leq} X_2$, if $S_1(x) \leq S_2(x)$, for all $x > 0$.

Proof of (ii) follows using the fact that, if $X_1 \overset{hr}{\leq} X_2$ then $\frac{S_2(t)}{S_1(t)}$ is increasing in t. This implies $\frac{S_1(x)}{S_1(t)} \leq \frac{S_2(x)}{S_2(t)}$ $\forall x \geq t$. Hence the proof. □

2.2 Generalized weighted failure and dynamic failure entropies of order (θ_1, θ_2)

In this section, we define the weighted failure entropy measures for rvs having bounded support. Failure entropy measures are used to obtain information associated to the past lifetime of the rv. Consider the following definition.

Definition 2.2.1. *For a rv with finite support $[0, l]$, generalized weighted failure entropy (GWFE) of order (θ_1, θ_2) is defined as*

$$f\xi^w_{\theta_1,\theta_2}(X) = \frac{1}{\theta_2 - \theta_1} \log \int_0^l x F^{\theta_1+\theta_2-1}(x)dx, \ \theta_2 \geq 1, \ \theta_2 - 1 \ < \theta_1 < \ \theta_2. \tag{2.7}$$

Example 2.2.1. *Suppose X and Y have $U(0, a)$ and $U(h, a+h)$ distributions with pdfs*

$$f(x) = \frac{1}{a}, \ 0 < x < a$$

and

$$g(y) = \frac{1}{a}, \ h < y < a+h, \ h > 0,$$

respectively. From Eq. (2.2), we have

$$f\xi_{\theta_1,\theta_2}(X) = f\xi_{\theta_1,\theta_2}(Y) = \frac{1}{\theta_2 - \theta_1} \log \frac{a}{\theta_1 + \theta_2}.$$

Now from (2.7), we get

$$f\xi^w_{\theta_1,\theta_2}(X) = \frac{1}{\theta_2 - \theta_1} \log \left[\frac{a^2}{\theta_1 + \theta_2 + 1} \right]$$

and

$$f\xi^w_{\theta_1,\theta_2}(Y) = \frac{1}{\theta_2 - \theta_1} \log \left[\frac{a(a(\theta_1 + \theta_2) + h(\theta_1 + \theta_2 + 1))}{(\theta_1 + \theta_2)(\theta_1 + \theta_2 + 1)} \right].$$

So it is observed that, although $f\xi_{\theta_1,\theta_2}(X) = f\xi_{\theta_1,\theta_2}(Y)$ but $f\xi^w_{\theta_1,\theta_2}(X) \neq f\xi^w_{\theta_1,\theta_2}(Y)$.

In the subsequent lemmas, we study the shift-dependency properties of GWFE measure and various relations related to proportional reversed hazard model (PRHM).

Lemma 2.2.1. *Suppose X has finite support $[0,l]$ and $Z = cX + d$, where $c > 0$ and $d \geq 0$, then*

$$\exp[(\theta_2-\theta_1) f\xi^w_{\theta_1,\theta_2}(Z)] = c^2 \exp[(\theta_2-\theta_1) f\xi^w_{\theta_1,\theta_2}(X)] + cd \exp[(\theta_2-\theta_1) f\xi_{\theta_1,\theta_2}(X)]. \tag{2.8}$$

Let F_{X_η} and F denote the cdfs of the rvs X_η and X having finite support, then PRHM (see, Gupta et al. (1998)) is described by the relation $F_{X_\eta}(x) = [F(x)]^\eta$, where $\eta > 0$. The following Lemma compares the GWFE of X, X_η and ηX. Proofs are omitted.

Lemma 2.2.2. *The following relations hold:*

$$(a)\ f\xi^w_{\theta_1,\theta_2}(X_\eta) = \left(\frac{\eta\theta_2 - \eta\theta_1 - \eta + 1}{\theta_2 - \theta_1}\right) f\xi^w_{\eta\theta_1,\eta\theta_2-\eta+1}(X);$$
$$(b)\ f\xi^w_{\theta_1,\theta_2}(X_\eta) \leq f\xi^w_{\theta_1,\theta_2}(X) \leq f\xi^w_{\theta_1,\theta_2}(\eta X),\ if\ \eta > 1;$$
$$(c)\ f\xi^w_{\theta_1,\theta_2}(X_\eta) \geq f\xi^w_{\theta_1,\theta_2}(X) \geq f\xi^w_{\theta_1,\theta_2}(\eta X),\ if\ 0 < \eta < 1.$$

We illustrate Lemma 2.2.2 in Table 2.3 for $U(0,a)$ distribution and power distribution with cdf

$$F_P(x) = x^\alpha;\ 0 < x < 1,\ \alpha > 0.$$

Table 2.3: *GWFE for uniform and Power distributions where $\gamma = \theta_1 + \theta_2 - 1$.*

Distribution	$(\theta_2-\theta_1) f\xi^w_{\theta_1,\theta_2}(X)$	$(\theta_2-\theta_1) f\xi^w_{\theta_1,\theta_2}(X_\eta)$	$(\theta_2-\theta_1) f\xi^w_{\theta_1,\theta_2}(\eta X)$
Uniform	$\log \frac{a^2}{2+\gamma}$	$\log \frac{a^2}{2+\gamma\eta}$	$\log \frac{a^2\eta^2}{2+\gamma}$
Power	$\log \frac{1}{2+\gamma\alpha}$	$\log \frac{1}{2+\gamma\eta\alpha}$	$\log \frac{\eta^2}{2+\gamma\alpha}$

Now we define the dynamic version of GWFE measure. Generalized dynamic weighted failure entropy (GDWFE) of order (θ_1, θ_2) of a rv X is the GWFE of the past life ${}_tX = [t - X | X < t]$.

Definition 2.2.2. *The GDWFE of a rv X having finite support $[0,l]$ is defined as*

$$f\xi^w_{\theta_1,\theta_2}(X;t) = \frac{1}{\theta_2-\theta_1} \log \int_0^t x \frac{F^{\theta_1+\theta_2-1}(x)}{F^{\theta_1+\theta_2-1}(t)} dx,\ \theta_2 \geq 1,\ \theta_2 - 1 < \theta_1 < \theta_2. \tag{2.9}$$

Remark 2.2.1. *Let X be a rv with bounded support $\mathscr{A}$ and $l = \sup \mathscr{A} < +\infty$. Then, $f\xi^w_{\theta_1,\theta_2}(X;l) = \xi^w_{\theta_1,\theta_2}(X)$. Note that, GDWFE is defined even when the support is unbounded.*

Next lemma considers the effect of linear transformation on the GDWFE measure.

Lemma 2.2.3. *Suppose X has finite support $[0,l]$ and $Z = cX + d$, where $c > 0$ and $d \geq 0$, then*

$$\exp[(\theta_2 - \theta_1) f\xi^w_{\theta_1,\theta_2}(Y;t)] = c^2 \exp\left[(\theta_2 - \theta_1) f\xi^w_{\theta_1,\theta_2}\left(X;\frac{t-d}{c}\right)\right] + cd\exp\left[(\theta_2 - \theta_1) f\xi_{\theta_1,\theta_2}\left(X;\frac{t-d}{c}\right)\right].$$

Remark 2.2.2. *If $d = 0$, then from Lemma 2.2.3 we have*

$$f\xi^w_{\theta_1,\theta_2}(Y;t) = \frac{2\log c}{\theta_2 - \theta_1} + f\xi^w_{\theta_1,\theta_2}\left(X;\frac{t}{c}\right). \tag{2.10}$$

Now we provide some bounds and stochastic ordering results related to these failure entropy measures. Proofs are similar to that of Theorem 2.1.1, hence omitted.

Theorem 2.2.1. *Let X be a non-negative continuous rv having finite support $[0,l]$ with WMPL $\mu^w_F(t)$, GWFE $f\xi^w_{\theta_1,\theta_2}(X)$ and GDWFE $f\xi^w_{\theta_1,\theta_2}(X;t)$. Then for $\theta_1 + \theta_2 - 1 \geq (\leq) 1$ we have,*

(i) $f\xi^w_{\theta_1,\theta_2}(X) \leq (\geq) \frac{1}{\theta_2 - \theta_1} \log[\mu^w_F(l)]$;

(ii) $f\xi^w_{\theta_1,\theta_2}(X;t) \leq (\geq) \frac{1}{\theta_2 - \theta_1} \log[\mu^w_F(t)]$.

Definition 2.2.3. *Let X_1 and X_2 be two rvs having finite support $[0,l]$. Then, X_1 is smaller than X_2 in GWFE ordering, denoted by $X_1 \overset{GWFE}{\leq} X_2$, if*

$$f\xi^w_{\theta_1,\theta_2}(X_1) \leq f\xi^w_{\theta_1,\theta_2}(X_2).$$

Definition 2.2.4. *Let X_1 and X_2 be two rvs having finite support $[0,l]$. Then, X_1 is smaller than X_2 in GDWFE ordering, denoted by $X_1 \overset{GDWFE}{\leq} X_2$, if*

$$f\xi^w_{\theta_1,\theta_2}(X_1;t) \leq f\xi^w_{\theta_1,\theta_2}(X_2;t), \quad \forall\ 0 < t < l.$$

Theorem 2.2.2. *Let X_1 and X_2 be two rvs having finite support $[0,l]$ with cdfs F and G and rhr functions $r_F(t)$ and $r_G(t)$, respectively. Then,*

(i) $X_1 \overset{st}{\leq} X_2 \Rightarrow X_1 \overset{GWFE}{\leq} X_2$.

(ii) $X_1 \overset{rh}{\leq} X_2 \Rightarrow X_1 \overset{GDWFE}{\leq} X_2$.

Proof. Proof of (i) follows from the definition of GWFE measure. Proof of (ii) follows using the fact that if $X \overset{rh}{\leq} Y$ then $\frac{F(x)}{F(t)} \geq \frac{G(x)}{G(t)}$. □

2.3 Aging classes

In this section we define two new aging classes based on the dynamic weighted information measures and study various properties.

Definition 2.3.1. *A non-negative continuous rv X is said to be increasing (decreasing) generalized dynamic weighted survival entropy (IGDWSE (DGDWSE)), if $\xi^w_{\theta_1,\theta_2}(X;t)$ is increasing (decreasing) in t (≥ 0).*

Theorem 2.3.1. *A non-negative continuous rv X is IGDWSE (DGDWSE) iff*

$$\lambda_F(t) \geq (\leq) \frac{t}{\theta_1+\theta_2-1} \exp[-(\theta_2-\theta_1)\xi^w_{\theta_1,\theta_2}(X;t)], \quad \forall\, t \geq 0,$$

where $\lambda_F(t) = \frac{f(t)}{S(t)}$, is the hr function of X.

Proof. We have

$$(\theta_2-\theta_1)\xi^w_{\theta_1,\theta_2}(X;t) = \log\left[\int_t^{+\infty} xS^{\theta_1+\theta_2-1}(x)dx\right] - (\theta_1+\theta_2-1)\log S(t). \tag{2.11}$$

Differentiating (2.11) wrt t we get,

$$(\theta_2-\theta_1)\frac{d}{dt}\xi^w_{\theta_1,\theta_2}(X;t) = (\theta_1+\theta_2-1)\lambda_F(t) - t\frac{S^{(\theta_1+\theta_2-1)}(t)}{\int_t^{+\infty} xS^{(\theta_1+\theta_2-1)}(x)dx}.$$

Using Eq. (2.5) we get,

$$(\theta_2-\theta_1)\frac{d}{dt}\xi^w_{\theta_1,\theta_2}(X;t) = (\theta_1+\theta_2-1)\lambda_F(t) - t\exp[-(\theta_2-\theta_1)\xi^w_{\theta_1,\theta_2}(X;t)]. \tag{2.12}$$

The result follows from Eq. (2.12). □

In the following theorem, we show that GDWSE measure uniquely determines the underlying distribution.

Theorem 2.3.2. *Let X be a rv with finite $\xi^w_{\theta_1,\theta_2}(X;t)$. Then, $\xi^w_{\theta_1,\theta_2}(X;t)$ uniquely determines the sf of X.*

Proof. From Eq. (2.12) we have

$$\lambda_F(t) = \frac{1}{\theta_1+\theta_2-1}\left((\theta_2-\theta_1)\frac{d}{dt}\xi^w_{\theta_1,\theta_2}(X;t)+t\exp[-(\theta_2-\theta_1)\xi^w_{\theta_1,\theta_2}(X;t)]\right). \quad (2.13)$$

Now, let X_1 and X_2 be two rvs with sfs $S_1(t)$ and $S_2(t)$, GDWSEs $\xi^w_{\theta_1,\theta_2}(X_1;t)$ and $\xi^w_{\theta_1,\theta_2}(X_2;t)$ and hr functions $\lambda_{F_1}(t)$ and $\lambda_{F_2}(t)$, respectively.
Suppose $\forall\, t \geq 0$,

$$\xi^w_{\theta_1,\theta_2}(X_1;t) = \xi^w_{\theta_1,\theta_2}(X_2;t),$$

then from Eq. (2.13) we get $\lambda_{F_1}(t) = \lambda_{F_2}(t)$. Since hr function uniquely determines the sf of the underlying distribution, we conclude that,

$$S_1(t) = S_2(t),\ \forall t \geq 0.$$

□

Next theorem shows the preservation of dynamic weighted survival entropy order under scale transformation.

Theorem 2.3.3. *Let X_1 and X_2 be two rvs and $X_1 \overset{GDWSE}{\leq} (\geq) X_2$. Let $Z_1 = a_1X_1$ and $Z_2 = a_2X_2$, where $a_1, a_2 > 0$. Then $Z_1 \overset{GDWSE}{\leq} (\geq) Z_2$, if $\xi^w_{\theta_1,\theta_2}(X_1;t)$ is decreasing in $t > 0$ and $a_1 \leq (\geq) a_2$.*

Proof. Suppose $a_1 \leq a_2$. Since $\xi^w_{\theta_1,\theta_2}(X_1;t)$ is decresasing in t, we have,

$$\xi^w_{\theta_1,\theta_2}\left(X_1;\frac{t}{a_1}\right) \leq \xi^w_{\theta_1,\theta_2}\left(X_1;\frac{t}{a_2}\right).$$

Again, $\xi^w_{\theta_1,\theta_2}\left(X_1;\frac{t}{a_2}\right) \leq \xi^w_{\theta_1,\theta_2}\left(X_2;\frac{t}{a_2}\right)$ since $X_1 \overset{GDWSE}{\leq} X_2$. Combining these two inequalities and using Eq. (2.6), we have

$$\xi^w_{\theta_1,\theta_2}(Z_1;t) = \frac{2\log a_1}{\theta_2-\theta_1} + \xi^w_{\theta_1,\theta_2}\left(X_1;\frac{t}{a_1}\right) \leq \frac{2\log a_2}{\theta_2-\theta_1} + \xi^w_{\theta_1,\theta_2}\left(X_2;\frac{t}{a_2}\right) = \xi^w_{\theta_1,\theta_2}(Z_2;t).$$

Hence the result. Similarly, when $a_1 \geq a_2$ and $X_1 \overset{GDWSE}{\geq} X_2$, it can be easily shown that $Z_1 \overset{GDWSE}{\geq} Z_2$. □

Definition 2.3.2. *A non-negative rv X having finite support $[0,l]$ is said to be increasing (decreasing) generalized dynamic weighted failure entropy (IGDWFE (DGDWFE)), if $f\xi^w_{\theta_1,\theta_2}(X;t)$ is increasing (decreasing) in t (≥ 0).*

Theorem 2.3.4. *A non-negative continuous rv X having finite support* $[0,l]$ *is IGDWFE (DGDWFE) iff*

$$r_F(t) \leq (\geq) \frac{t}{\theta_1+\theta_2-1} \exp[-(\theta_2-\theta_1) f\xi^w_{\theta_1,\theta_2}(X;t)], \ \forall t \geq 0,$$

where $r_F(t) = \frac{f(t)}{F(t)}$ *is the rhr function.*

Proof. Differentiating Eq. (2.9) we get

$$(\theta_2-\theta_1)\frac{d}{dt} f\xi^w_{\theta_1,\theta_2}(X;t) = t \exp[-(\theta_2-\theta_1) f\xi^w_{\theta_1,\theta_2}(X;t)] - (\theta_1+\theta_2-1) r_F(t). \quad (2.14)$$

The result follows from Eq. (2.14). □

The GDWFE measure also uniquely determines the underlying distribution.

Theorem 2.3.5. *Let X be rv with* $f\xi^w_{\theta_1,\theta_2}(X;t) < +\infty;\ \forall t \geq 0,\ \theta_2 - 1 < \theta_1 < \theta_2,\ \theta_2 \geq 1.$ *Then for each* θ_1 *and* θ_2*,* $f\xi^w_{\theta_1,\theta_2}(X;t)$ *uniquely determines the cdf of X.*

Proof. Proceeding along the same line as Theorem 2.3.2, it can be easily shown that for two rvs X_1 and X_2,

$$f\xi^w_{\theta_1,\theta_2}(X_1;t) = f\xi^w_{\theta_1,\theta_2}(X_2;t) \Rightarrow r_{F_1}(t) = r_{F_2}(t) \ \ \forall t > 0,$$

where $r_{F_1}(t)$ and $r_{F_2}(t)$ are the rhr functions of X_1 and X_2, respectively. □

Theorem 2.3.6. *Let* X_1 *and* X_2 *be two rvs having finite support* $[0,l]$ *and* $X_1 \overset{GDWFE}{\leq} (\geq) X_2$*. Let* $Z_1 = a_1 X_1$ *and* $Z_2 = a_2 X_2$*, where* $a_1, a_2 > 0$*. Then* $Z_1 \overset{GDWFE}{\leq} (\geq) Z_2$*, if* $f\xi^w_{\theta_1,\theta_2}(X_1;t)$ *is decreasing in* $t > 0$ *and* $a_1 \leq (\geq) a_2$*.*

Proof. Proof follows from that of Theorem 2.3.3. □

2.4 Characterization results

In this section we study some important characterization results using the proposed measures. First we provide characterization results for Rayleigh distribution in terms of GDWSE and WMRL function.

Theorem 2.4.1. *The rv X has constant GDWSE iff it has a Rayleigh distribution with sf* $S(x) = e^{-\lambda x^2};\ x \geq 0,\ \lambda > 0.$

Proof. If $S(x) = e^{-\lambda x^2}$ then $\xi^w_{\theta_1,\theta_2}(X;t) = \frac{1}{\theta_2-\theta_1}\log\frac{1}{2\lambda(\theta_1+\theta_2-1)}$ which is a constant. Next assume that

$$\xi^w_{\theta_1,\theta_2}(X;t) = c.$$

Then,

$$\begin{aligned} \frac{d}{dt}\xi^w_{\theta_1,\theta_2}(X;t) &= 0 \\ \Rightarrow (\theta_1+\theta_2-1)\lambda_F(t) &= t\exp[-(\theta_2-\theta_1)\xi^w_{\theta_1,\theta_2}(X;t)]. \end{aligned}$$

This implies $\lambda_F(t) = \frac{e^{(\theta_1-\theta_2)c}}{\theta_1+\theta_2-1}t$, which is the hazard function of a Rayleigh distribution with sf $S(t) = e^{-\lambda t^2}$; $t \geq 0$, where $\lambda = \frac{e^{(\theta_1-\theta_2)c}}{2(\theta_1+\theta_2-1)} > 0$ as $\theta_1+\theta_2 > 1$. □

So it is observed that Rayleigh distribution provides a bridge between IGDWSE and DGDWSE classes of distributions.

Theorem 2.4.2. *For a rv X, the relation* $(\theta_2-\theta_1)\xi^w_{\theta_1,\theta_2}(X;t) = \log m^w_F(t) - \log(\theta_1+\theta_2-1)$ *holds iff X has a Rayleigh distribution.*

Proof. If X has Rayleigh distribution then it follows that

$$(\theta_2-\theta_1)\xi^w_{\theta_1,\theta_2}(X;t) = \log m^w_F(t) - \log(\theta_1+\theta_2-1).$$

Now suppose the relation $(\theta_2-\theta_1)\xi^w_{\theta_1,\theta_2}(X;t) = \log m^w_F(t) - \log(\theta_1+\theta_2-1)$ holds. Differentiating wrt t we get

$$(\theta_2-\theta_1)\frac{d}{dt}\xi^w_{\theta_1,\theta_2}(X;t) = \frac{\frac{d}{dt}m^w_F(t)}{m^w_F(t)}.$$

Using Eq. (2.12) we have

$$(\theta_1+\theta_2-1)\lambda_F(t) - t\ \exp[-(\theta_2-\theta_1)\xi^w_{\theta_1,\theta_2}(X;t)] = \frac{\frac{d}{dt}m^w_F(t)}{m^w_F(t)}.$$

Now substituting $\frac{d}{dt}m^w_F(t) = \lambda_F(t)m^w_F(t) - t$ and after simplification, we get

$$\lambda_F(t)m^w_F(t) = t \Rightarrow \frac{d}{dt}m^w_F(t) = 0.$$

Which means $m^w_F(t) = c$, where c is a constant and thus we obtain $\lambda_F(t) = \frac{t}{c}$, which is the hr function of a Rayleigh distribution with sf $S(t) = e^{-\frac{t^2}{2c}}$. □

Now we study characterization of power distribution in terms of GDWFE and WMPL function.

Theorem 2.4.3. *Let X be a rv having support $(0,b)$, then X has a power distribution with cdf $F(x) = \left(\frac{x}{b}\right)^c$, $0 < x < b$, $c > 0$ iff*

$$(\theta_2 - \theta_1) f\xi^w_{\theta_1,\theta_2}(X;t) = \log k + \log \mu^w_F(t),$$

where $\mu^w_F(t)$ is the WMPL function of X, $k(>0)$ is a constant and $\theta_1 + \theta_2 \neq 2$.

Proof. Suppose the relation

$$(\theta_2 - \theta_1) f\xi^w_{\theta_1,\theta_2}(X;t) = \log k + \log \mu^w_F(t)$$

holds. Now differentiating this wrt t and then using (2.14) we get

$$t \exp[-(\theta_2 - \theta_1) f\xi^w_{\theta_1,\theta_2}(X;t)] - (\theta_1 + \theta_2 - 1) r_F(t) = \frac{\frac{d}{dt}\mu^w_F(t)}{\mu^w_F(t)}.$$

Substituting the value of $(\theta_2 - \theta_1) f\xi^w_{\theta_1,\theta_2}(X;t)$ and $\frac{d}{dt}\mu^w_F(t) = t - r_F(t)\mu^w_F(t)$ and after some calculations, we obtain

$$r_F(t)\mu^w_F(t) = \frac{1-k}{k(\theta_1 + \theta_2 - 2)} t. \tag{2.15}$$

This implies

$$\frac{d}{dt}\mu^w_F(t) = \frac{k(\theta_1 + \theta_2 - 1) - 1}{k(\theta_1 + \theta_2 - 2)} t.$$

Integrating wrt t and taking $\mu^w_F(0) = 0$ we get,

$$\mu^w_F(t) = \frac{k(\theta_1 + \theta_2 - 1) - 1}{k(\theta_1 + \theta_2 - 2)} \frac{t^2}{2}.$$

From Eq. (2.15) we obtain

$$r_F(t) = \frac{2(1-k)}{k(\theta_1 + \theta_2 - 1) - 1} \frac{1}{t} = \frac{c}{t},$$

where $c = \frac{2(1-k)}{k(\theta_1+\theta_2-1)-1} > 0$. Now to determine the appropriate ranges of k, first suppose that $\theta_1 + \theta_2 - 1 > 1$. Then, for $c > 0$ we have $1 > k > \frac{1}{\theta_1+\theta_2-1}$. Again if $\theta_1 + \theta_2 - 1 < 1$ then we have $1 < k < \frac{1}{\theta_1+\theta_2-1}$ for $c > 0$. So we see that $r_F(t)$ is the rhr function of the power distribution with cdf $F(x) = \left(\frac{x}{b}\right)^c$, $0 < x < b$, $c > 0$.
Conversely, if X has power distribution with cdf $F(x) = \left(\frac{x}{b}\right)^c$, $0 < x < b$, $c > 0$ then

$$(\theta_2 - \theta_1) f\xi^w_{\theta_1,\theta_2}(X;t) = c(\theta_1 + \theta_2 - 1)\log t - \log(c(\theta_1 + \theta_2 - 1) + 2).$$

By taking $c = \frac{2}{\theta_1+\theta_2-1}$ and $k = \frac{c+2}{4}$, we get the result. □

Remark 2.4.1. *Note that for $\theta_1 + \theta_2 = 2$, the above relation in Theorem 2.4.3 becomes an identity. Since $\theta_1 + \theta_2 = 2 \Rightarrow \theta_1 + \theta_2 - 1 = 1$ and from Eq. (2.9) we get,*

$$(\theta_2 - \theta_1) f\xi^w_{\theta_1,\theta_2}(X;t) = \log \mu^w_F(t).$$

Next we provide two important rseults regarding the characterization of identically distributed rvs using GWSE (GWFE) of smallest (largest) order statistic. Consider the following definitions of GWSE (GWFE) of the smallest (largest) order statistic.

Let $X_1, X_2, \cdots, X_n$ be a rs of size n from F. Denote the corresponding order statistics as $X_{1:n}, X_{2:n}, \cdots, X_{n:n}$, where $X_{i:n}$ $(1 \le i \le n)$ is the i-th order statistic. The sf of $X_{1:n}$ is given by $S_{1:n}(x) = S^n(x)$ and GWSE of $X_{1:n}$ is obtained as

$$\begin{aligned} \xi^w_{\theta_1,\theta_2}(X_{1:n}) &= \frac{1}{\theta_2-\theta_1} \log \int_0^{+\infty} x S^{n(\theta_1+\theta_2-1)}(x)dx \\ &= \frac{1}{\theta_2-\theta_1} \log \int_0^1 \frac{v^{n(\theta_1+\theta_2-1)} F^{-1}(1-v)}{f(F^{-1}(1-v))} dv. \end{aligned} \tag{2.16}$$

Suppose X has finite support $[0,l]$ then the GDWFE of $X_{n:n}$ is obtained as

$$\begin{aligned} f\xi^w_{\theta_1,\theta_2}(X_{n:n};t) &= \frac{1}{\theta_2-\theta_1} \log \int_0^l x F^{n(\theta_1+\theta_2-1)}(x)dx \\ &= \frac{1}{\theta_2-\theta_1} \log \int_0^1 \frac{v^{n(\theta_1+\theta_2-1)} F^{-1}(v)}{f(F^{-1}(v))} dv, \end{aligned} \tag{2.17}$$

where $F_{n:n}(x) = F^n(x)$ is the cdf of $X_{n:n}$.
The following lemma of will be useful to prove the next theorems (Psarrakos and Toomaj, 2017) .

Lemma 2.4.1. *If ϕ is a continuous function on $[0,1]$, such that $\int_0^1 x^n \phi(x)dx = 0$, for $n \ge 0$, then $\phi(x) = 0$, $\forall x \in [0,1]$.*

Theorem 2.4.4. *Let X and Y be two non-negative continuous rvs having common support $[0,+\infty)$ with cdfs F and G, respectively. Then $F(x) = G(x)$ $\forall x > 0$, iff*

$$\xi^w_{\theta_1,\theta_2}(X_{1:n}) = \xi^w_{\theta_1,\theta_2}(Y_{1:n}), \forall n \ge 1.$$

Proof. If $\xi^w_{\theta_1,\theta_2}(X_{1:n}) = \xi^w_{\theta_1,\theta_2}(Y_{1:n})$, then from Eq. (2.16) we have

$$\int_0^1 v^{n(\theta_1+\theta_2-1)} \left[\frac{F^{-1}(1-v)}{f(F^{-1}(1-v))} - \frac{G^{-1}(1-v)}{g(G^{-1}(1-v))} \right] dv = 0.$$

Then from Lemma 2.4.1 we get $\frac{F^{-1}(1-v)}{f(F^{-1}(1-v))} = \frac{G^{-1}(1-v)}{g(G^{-1}(1-v))}$ for all $v \in (0,1)$. This reduces to $F^{-1}(w)\frac{d}{dw}F^{-1}(w) = G^{-1}(w)\frac{d}{dw}G^{-1}(w)$, where $w = 1-v$ and $\frac{d}{dw}F^{-1}(w) = \frac{1}{f(F^{-1}(w))}$. So we have $F^{-1}(w) = G^{-1}(w), 0 \le w \le 1$. If $F(x) = G(x)$ holds then it is obvious that $\xi^w_{\theta_1,\theta_2}(X_{1:n}) = \xi^w_{\theta_1,\theta_2}(Y_{1:n})$. Hence the proof. □

Theorem 2.4.5. *Let X and Y be two non-negative continuous rvs having common finite support $(0,l)$ with cdfs F and G, respectively. Then $F(x) = G(x)$ iff $f\xi^w_{\theta_1,\theta_2}(X_{n:n};t) = f\xi^w_{\theta_1,\theta_2}(Y_{n:n};t)$, $\forall n \ge 1$.*

Proof. Suppose $f\xi^w_{\theta_1,\theta_2}(X_{n:n};t) = f\xi^w_{\theta_1,\theta_2}(Y_{n:n};t)$ holds. Now from Eq. (2.17) we have,

$$\int_0^1 v^{n(\theta_1+\theta_2-1)}\left[\frac{F^{-1}(v)}{f(F^{-1}(v))} - \frac{G^{-1}(v)}{g(G^{-1}(v))}\right]dv = 0.$$

□

Then from Lemma 2.4.1 we get $\frac{F^{-1}(v)}{f(F^{-1}(v))} = \frac{G^{-1}(v)}{g(G^{-1}(v))}$ for all $v \in (0,1)$. The rest of the proof is similar to the proof of Theorem 2.4.4.

Remark 2.4.2. *Proof of Theorem 2.4.4 is immediate from Theorem 2.3.2 since the distributions of $X_{1:n}$ and $Y_{1:n}$ are $1-(1-F)^n$ and $1-(1-G)^n$, respectively. Similarly, the proof of Theorem 2.4.5 can be obtained from Theorem 2.3.5.*

2.5 Some inequalities and bounds

In this section we provide some upper and lower bounds for the proposed measures.

Theorem 2.5.1. *Let X be a non-negative continuous rv with pdf f, cdf F and sf S. The following inequalities hold:*

$$(i)\ (\theta_2-\theta_1)\xi^w_{\theta_1,\theta_2}(X) + (\theta_1+\theta_2-1) \ge H(X) + E(\log X);$$
$$(ii)\ (\theta_2-\theta_1)f\xi^w_{\theta_1,\theta_2}(X) + (\theta_1+\theta_2-1) \ge H(X) + E(\log X).$$

Inequality (ii) is related to rvs having bounded support.

Proof. Using log-sum inequality, we have

$$\begin{aligned}\int_0^{+\infty} f(x)\log\frac{f(x)}{xS^{\theta_1+\theta_2-1}(x)}dx &\ge \left(\int_0^{+\infty} f(x)dx\right)\log\frac{\int_0^{+\infty} f(x)dx}{\int_0^{+\infty} xS^{\theta_1+\theta_2-1}(x)dx}\\ &= -\log\int_0^{+\infty} xS^{\theta_1+\theta_2-1}(x)dx\\ &= -(\theta_2-\theta_1)\xi^w_{\theta_1,\theta_2}(X).\end{aligned} \tag{2.18}$$

Now the L.H.S. of Eq. (2.18) equals

$$\int_0^{+\infty}(\log f(x))f(x)dx-\int_0^{+\infty}(\log x)f(x)dx-(\theta_1+\theta_2-1)\int_0^{+\infty}\log S(x)f(x)dx,$$

which reduces to $-H(X)-E(\log X)+(\theta_1+\theta_2-1)$. The result follows from Eq. (2.18). The proof of part (ii) is similar to that of part (i). □

In the next theorem we provide a lower bound for GDWSE and GDWFE measures in terms of the residual and past entropy.

Theorem 2.5.2. *For a non-negative continuous rv X, the following inequalities hold:*

$$(i)\ (\theta_2-\theta_1)\xi^w_{\theta_1,\theta_2}(X;t)+(\theta_1+\theta_2-1)\geq H(X;t)+\int_t^{+\infty}\frac{f(x)}{S(t)}\log(x)\,dx;$$

$$(ii)\ (\theta_2-\theta_1)f\xi^w_{\theta_1,\theta_2}(X;t)+(\theta_1+\theta_2-1)\geq \bar{H}(X;t)+\int_0^{t}\frac{f(x)}{F(t)}\log(x)\,dx.$$

Proof. (i). From log-sum inequality, we get

$$\begin{aligned}\int_t^{+\infty}f(x)\log\frac{f(x)}{x\left(\frac{S(x)}{S(t)}\right)^{\theta_1+\theta_2-1}}dx\ &\geq \log\frac{\int_t^{+\infty}f(x)dx}{\int_t^{+\infty}x\left(\frac{S(x)}{S(t)}\right)^{\theta_1+\theta_2-1}dx}\int_t^{+\infty}f(x)dx\\ &=S(t)[\log S(t)-(\theta_2-\theta_1)\xi^w_{\theta_1,\theta_2}(X;t)].\end{aligned}\tag{2.19}$$

After some simplifications, Eq. (2.19) reduces to

$$\begin{aligned}\int_t^{+\infty}f(x)\log\frac{f(x)}{x\left(\frac{S(x)}{S(t)}\right)^{\theta_1+\theta_2-1}}dx\ \geq\ &\int_t^{+\infty}(\log f(x))f(x)dx-\int_t^{+\infty}(\log x)f(x)dx\\ &+(\theta_1+\theta_2-1)S(t).\end{aligned}$$

Using the definition of $H(X;t)$ and after some simplifications, the results follows from Eq. (2.19). Proof of part (ii) follows similarly. □

Now we provide an upper bound for GDWSE and GDWFE measures for rvs having bounded support.

Theorem 2.5.3. *For a non-negative continuous rv X having support $[0,b]$, the following inequality holds:*

$$\xi^w_{\theta_1,\theta_2}(X;t)\leq\frac{\int_t^b x\left(\frac{S(x)}{S(t)}\right)^{(\theta_1+\theta_2-1)}\log\left[x\left(\frac{S(x)}{S(t)}\right)^{(\theta_1+\theta_2-1)}\right]dx}{(\theta_2-\theta_1)\int_t^b x\left(\frac{S(x)}{S(t)}\right)^{(\theta_1+\theta_2-1)}dx}+\frac{\log(b-t)}{\theta_2-\theta_1},\ t<b.$$

Proof. Using log-sum inequality, we have

$$\begin{aligned}
&\int_t^b x\left(\frac{S(x)}{S(t)}\right)^{(\theta_1+\theta_2-1)}\log\left[x\left(\frac{S(x)}{S(t)}\right)^{(\theta_1+\theta_2-1)}\right]dx\\
&\geq \log\frac{\int_t^b x\left(\frac{S(x)}{S(t)}\right)^{(\theta_1+\theta_2-1)}dx}{b-t}\int_t^b x\left(\frac{S(x)}{S(t)}\right)^{(\theta_1+\theta_2-1)}dx\\
&=\left[(\theta_2-\theta_1)\xi^w_{\theta_1,\theta_2}(X;t)-\log(b-t)\right]\int_t^b x\left(\frac{S(x)}{S(t)}\right)^{(\theta_1+\theta_2-1)}dx. \qquad (2.20)
\end{aligned}$$

The proof follows from Eq. (2.20). □

Proposition 2.5.1. *Let X be a non-negative continuous rv. Then,*

$$f\xi^w_{\theta_1,\theta_2}(X;t)\leq\frac{\int_0^t x\left(\frac{F(x)}{F(t)}\right)^{(\theta_1+\theta_2-1)}\log\left[x\left(\frac{F(x)}{F(t)}\right)^{(\theta_1+\theta_2-1)}\right]dx}{(\theta_2-\theta_1)\int_0^t x\left(\frac{F(x)}{F(t)}\right)^{(\theta_1+\theta_2-1)}dx}+\frac{\log(t)}{\theta_2-\theta_1}.$$

Proof. Proof is similar to that of Theorem 2.5.1. □

In the following section, we provide non-parametric estimators for GWSE and GWFE measures and develope a goodness-of-fit test for exponential distribution using the estimator of GWSE measure.

2.6 Estimation and application

Let $X_1,X_2,\cdots,X_n$ be a random sample of size n drawn from a distribution with cdf F and $X_{1:n}\leq X_{2:n}\leq\cdots\leq X_{n:n}$ be the corresponding order statistics. Let F_n be the edf of X which is defined as

$$F_n(x)=\begin{cases}0, \text{ if } x<X_{1:n},\\ \frac{i}{n}, \text{ if } X_{i:n}\leq x<X_{(i+1):n},\ i=1,2,\cdots,n-1\\ 1, \text{ if } x\geq X_{n:n}.\end{cases}$$

The non-parametric estimator of GWSE measure is defined as

$$\hat{\xi}^w_{\theta_1,\theta_2}(X)=\frac{1}{\theta_2-\theta_1}\log\int_0^{+\infty}x\,S_n^{\theta_1+\theta_2-1}(x)dx,\ \theta_2\geq 1,\ \theta_2-1<\theta_1<\theta_2. \qquad (2.21)$$

Substituting $S_n(x) = 1 - \frac{i}{n}$, $i = 1, 2, \cdots, n-1$ in Eq. (2.21), we get

$$\begin{aligned}
\hat{\xi}^w_{\theta_1,\theta_2}(X) &= \frac{1}{\theta_2-\theta_1}\log\left[\sum_{i=0}^{n-1}\int_{X_{i:n}}^{X_{(i+1):n}} x\left(1-\frac{i}{n}\right)^{\theta_1+\theta_2-1}(x)dx\right] \\
&= \frac{1}{\theta_2-\theta_1}\log\left[\sum_{i=0}^{n-1}\frac{X^2_{(i+1):n}-X^2_{i:n}}{2}\left(1-\frac{i}{n}\right)^{\theta_1+\theta_2-1}\right] \\
&= \frac{1}{\theta_2-\theta_1}\log\left[\sum_{i=0}^{n-1}U_{i+1}\left(1-\frac{i}{n}\right)^{\theta_1+\theta_2-1}\right], \qquad (2.22)
\end{aligned}$$

where $U_{i+1} = \dfrac{X^2_{(i+1):n}-X^2_{i:n}}{2}$ and $X_{0:n} = 0$.
Similarly, we can define the estimator for GWFE measure for a rv X with bounded support $\mathscr{A}$ and $sup\ \mathscr{A} = l$. Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the corresponding order statistics and $X_{(n+1):n} = l$. Then the empirical GWFE can be obtained as

$$\hat{f\xi}^w_{\theta_1,\theta_2}(X) = \frac{1}{\theta_2-\theta_1}\log\left[\sum_{i=1}^{n}U_{i+1}\left(\frac{i}{n}\right)^{\theta_1+\theta_2-1}\right]. \qquad (2.23)$$

Using the estimator of GWSE measure, we develop a test for exponentiality. Let $X_1, X_2, \cdots, X_n$ be iid rvs from a non-negative absolutely continuous cdf F. Suppose

$$F_0(x,\lambda) = 1 - e^{-\lambda x},\ x > 0,\ \lambda > 0,$$

denotes the cdf of an exponential distribution with parameter λ. We want to test the hypothesis

$$H_0 : F(x) = F_0(x,\lambda) \quad vs. \quad H_1 : F(x) \neq F_0(x,\lambda).$$

Now consider the absolute difference between $\xi^w_{\theta_1,\theta_2}(X)$ and $\hat{\xi}^w_{\theta_1,\theta_2}(X)$ as

$$D = \left|\xi^w_{\theta_1,\theta_2}(X) - \hat{\xi}^w_{\theta_1,\theta_2}(X)\right|.$$

If $X \sim F_0(x,\lambda)$ then $\xi^w_{\theta_1,\theta_2}(X) = \frac{2}{\theta_1-\theta_2}\log(\lambda(\theta_1+\theta_2-1))$ and D reduces to

$$D = \left|\hat{\xi}^w_{\theta_1,\theta_2}(X) - \frac{2}{\theta_1-\theta_2}\log(\lambda(\theta_1+\theta_2-1))\right|.$$

We estimate λ by its maximum likelihood estimate (mle) $\hat{\lambda} = 1/\bar{X}$ and obtain the test statistic as

$$\begin{aligned} D &= \left| \hat{\xi}^{w}_{\theta_1,\theta_2}(X) + \frac{2}{\theta_2 - \theta_1} \log((\theta_1+\theta_2-1)/\bar{X}) \right| \\ &= \left| \tfrac{1}{\theta_2-\theta_1} \log \left[\tfrac{(\theta_1+\theta_2-1)^2}{2} \sum_{i=0}^{n-1} \left(\left(\tfrac{X_{(i+1):n}}{\bar{X}} \right)^2 - \left(\tfrac{X_{i:n}}{\bar{X}} \right)^2 \right) \left(1 - \tfrac{i}{n}\right)^{\theta_1+\theta_2-1} \right] \right|. \end{aligned}$$

Since $\frac{X_{i:n}}{\bar{X}}$ is scale invariant $\forall\, i = 1(1)n$, hence D is scale invariant. It measures the distance between the parametric and the non-parametric estimate of GWSE measure. Large values of D indicate that the sample is from the exponential distribution. Note that the statistic D is scale invariant. Now consider the monotone transformation $T = \exp(-D)$, we have $0 < T < 1$. Under the null hypothesis, $D \xrightarrow{p} 0$ and hence $T \xrightarrow{p} 1$. So we reject H_0 at the significance level α if $T < T_{\alpha,n}$, where $T_{\alpha,n}$ is the lower α-quantile of the edf of T.
The sampling distribution of T under H_0 is intractable. Critical points of the distribution of T are obtained by simulations. To obtain the critical points $T_{\alpha,n}$ by simulations we generate 10000 samples of size n from a standard exponential distribution for $n = 4(1)30,\ 30(5)50$ and $50(10)100$. For each n the lower α-quantile of the edf of T is used to determine $T_{\alpha,n}$. The critical points vary for different choices of (θ_1, θ_2). The critical points of 90%, 95% and 99% are presented in Table 2.4 for $\theta_1 = 0.26$ and $\theta_2 = 1.25$.

Table 2.4: *Critical values of T.*

n	$T_{0.01,n}$	$T_{0.05,n}$	$T_{0.10,n}$	n	$T_{0.01,n}$	$T_{0.05,n}$	$T_{0.10,n}$
4	0.15452	0.18056	0.19797	22	0.30519	0.34955	0.38022
5	0.17084	0.19897	0.22030	23	0.30818	0.35743	0.39111
6	0.18293	0.21504	0.23581	24	0.31159	0.36074	0.39332
7	0.19529	0.23172	0.25352	25	0.32099	0.36921	0.40241
8	0.20218	0.24136	0.26678	26	0.32446	0.37260	0.40443
9	0.21664	0.25408	0.28284	27	0.33040	0.37802	0.41150
10	0.22490	0.26454	0.29226	28	0.33125	0.38068	0.41445
11	0.23571	0.27731	0.30546	29	0.33722	0.38670	0.41889
12	0.24354	0.28555	0.31272	30	0.33753	0.38741	0.42157
13	0.24885	0.29078	0.32062	35	0.35176	0.40709	0.44064
14	0.25726	0.30001	0.33029	40	0.37078	0.42298	0.45628
15	0.26489	0.31010	0.34048	45	0.38136	0.43473	0.47075
16	0.27563	0.31945	0.34892	50	0.39466	0.44945	0.48360
17	0.27848	0.32157	0.35377	60	0.41155	0.47032	0.50530
18	0.28509	0.33113	0.36187	70	0.43122	0.48464	0.52112
19	0.29347	0.33772	0.36784	80	0.45398	0.50520	0.54058
20	0.29454	0.34134	0.37401	90	0.46436	0.51746	0.55323
21	0.29890	0.34726	0.38058	100	0.47475	0.52805	0.56257

2.6.1 Power Comparison

The power of the proposed test is compared with two other exponentiality tests based on entropy. Ebrahimi et al. (1992) proposed a goodness-of-fit test for exponential distribution based on the Kullback-Leibler divergence measure and Baratpour and Rad (2012) proposed an exponentiality test based on cumulative residual entropy.
We compare the performance of T with the test statistic proposed by Baratpour and Rad (2012)

$$T^* = \frac{\sum_{i=1}^{n-1} \frac{n-i}{n} \log\left(\frac{n-i}{n}\right)(X_{(i+1):n} - X_{i:n}) + \frac{\sum_{i=1}^{n} X_i^2}{2\sum_{i=1}^{n} X_i}}{\frac{\sum_{i=1}^{n} X_i^2}{2\sum_{i=1}^{n} X_i}}$$

and with the test statistic provided by Ebrahimi et al. (1992)

$$KL_{mn} = exp(H_{mn} - \log \bar{X} - 1),$$

where $H_{mn} = \frac{1}{n}\sum_{i=1}^{n} \log[\frac{n}{2m}(X_{(i+m):n} - X_{(i-m):n})]$ is the Vasicek's entropy estimate. The window size m is a positive integer less that $\frac{n}{2}$, $X_{i:n} = X_{1:n}$ if $i < 1$ and $X_{i:n} = X_{n:n}$ if $i > n$. We reject the null hypothesis for large values of T^* and for small values of KL_{mn}.

Power of the tests are computed based on 10000 samples of size $n = 10(5)25$ for significance levels $\alpha = 0.01$ and 0.05. For power computation, we consider two alternative distributions

- Weibull $(p, 1)$ with pdf $f_W(x) = px^{p-1}e^{-x^p}$; $x, p > 0$.
- Gamma $(q, 1)$ with pdf $f_{GA}(x) = \frac{e^{-x}x^{q-1}}{\Gamma(q)}$; $x, q > 0$.

The power for Weibull and gamma alternatives are reported in Tables 2.5 and 2.6, respectively. It is observed that the power of the test T is similar to that of T^* but lower than that of KL_{mn} for small sample size $n = 10$. However, for moderate to large sample sizes the proposed test T behaves similar to KL_{mn} and T^*.

2.6.2 Data analysis

Consider the data set given in Grubbs (1971) that provides the mileages for 19 military personnel carriers that failed in service. The mileages are:
162, 200, 271, 320, 393, 508, 539, 629, 706, 778, 884, 1003, 1101, 1182, 1463, 1603, 1984, 2355, 2880.
The observed value of T is 0.41271. For $n = 19$ and $\alpha = 0.01$, from Table 2.4 we obtain the critical point as $T_{0.01,19} = 0.29347$. So we can not reject the null hypothesis that the failure time follows exponential distribution.

Table 2.5: *Power comparison for the tests T, T^* and KL_{mn} when the alternative is Weibull distribution.*

n	p	α	KL_{mn}	T^*	T
10	2	0.01	0.4267	0.3441	0.3237
		0.05	0.7170	0.6504	0.6627
	3	0.01	0.9013	0.8554	0.8404
		0.05	0.9858	0.9775	0.9758
	4	0.01	0.9947	0.9911	0.9880
		0.05	1	0.9997	0.9997
15	2	0.01	0.6526	0.5480	0.5577
		0.05	0.8628	0.8334	0.8308
	3	0.01	0.9914	0.9816	0.9805
		0.05	0.9995	0.9989	0.9990
	4	0.01	1	1	1
		0.05	1	1	1
20	2	0.01	0.7355	0.7080	0.7299
		0.05	0.9390	0.9173	0.9276
	3	0.01	0.9986	0.9986	0.9990
		0.05	1	1	1
	4	0.01	1	1	1
		0.05	1	1	1
25	2	0.01	0.8930	0.8287	0.8530
		0.05	1	0.9650	0.9642
	3	0.01	1	1	1
		0.05	1	1	1
	4	0.01	1	1	1
		0.05	1	1	1

2.7 Discussions

In this chapter, we proposed two new weighted information measures and obtained various properties. We also introduced their dynamic version and showed that dynamic information measures uniquely determine the underlying distribution. We defined new stochastic orderings based on the proposed measure. It is shown that dynamic weighted survival and failure entropy order is preserved under scale transformation under some specific conditions. An important use of these measures is characterization of distributions. We obtained characterization of Rayleigh and power distributions and identically distributed rvs using the proposed measures. Also we suggested non-parametric estimators based on the edf function and developed a goodness-of-fit test for exponential distribution. The proposed test

Table 2.6: *Power comparison for the tests T, T^* and KL_{mn} when the alternative is gamma distribution.*

n	q	α	KL_{mn}	T^*	T
10	5	0.01	0.7418	0.5652	0.5334
		0.05	0.9393	0.8306	0.8160
	6	0.01	0.8500	0.6876	0.6567
		0.05	0.9749	0.9155	0.8962
	7	0.01	0.9180	0.7856	0.7430
		0.05	0.9898	0.9578	0.9369
15	5	0.01	0.9344	0.7858	0.7636
		0.05	1	0.9458	0.9316
	6	0.01	0.9786	0.8987	0.8687
		0.05	1	0.9823	0.9725
	7	0.01	0.9933	0.9498	0.9221
		0.05	1	0.9950	0.9861
20	5	0.01	1	0.9006	0.8752
		0.05	1	0.9818	0.9669
	6	0.01	1	0.9677	0.9433
		0.05	1	0.9970	0.9898
	7	0.01	1	0.9908	0.9773
		0.05	1	0.9993	0.9953
25	5	0.01	1	1	0.9403
		0.05	1	1	0.9873
	6	0.01	1	1	0.9793
		0.05	1	1	0.9966
	7	0.01	1	1	0.9923
		0.05	1	1	0.9992

performed reasonably well for monotone increasing hazard alternatives.

In the testing problem, we took $\theta_1 = 0.26$ and $\theta_2 = 1.25$. However, there is no particular rule of choosing the generalizing parameters involved with the corresponding information measure. In our case, one may choose the parameters in such a way that the asymptotic variance of the estimator is minimum. If one is performing a testing problem, the parameters can be chosen such that the power of the test will be maximum. The choice of the parameters will depend on the underlying problem of interest.

We proposed non-parametric estimators based on the edf function. Other estimators like Kernel based estimators, estimators based on L-statistics may be used as well. It will be interesting to see under which circumstances these estimators outperform each other.

Chapter 3

Weighted cumulative Tsallis residual and past entropy measures

TSALLIS entropy is perhaps the most important non-additive generalization of Shannon entropy measure. It revolutionises statistical mechanics, thermodynamics and related fields. For detailed review and applications of Tsallis entropy see Cartwright (2014). Recently, cumulative Tsallis residual entropy (CTRE) measure has been introduced in the literature by Sati and Gupta (2015) for studying cumulative information based on non-additive entropy measure. For a rv X, CTRE is defined as

$$CT_\theta(X) = \frac{1}{\theta - 1}\left(1 - \int_0^{+\infty} S^\theta(x)dx\right), \quad 0 < \theta < 1. \tag{3.1}$$

Note that, if $\theta \to 1$, $CT_\theta(X) \to CRE(X)$. They also proposed the dynamic version of CTRE measure and provide application in characterizing some well known lifetime distributions. After the introduction of CTRE, various generalized Tsallis information measures have been proposed. Tsallis entropy defined in Eq. (1.4) can also be represented as

$$T_\theta(X) = \frac{1}{\theta - 1}\int_0^{+\infty}\left(f(x) - f^\theta(x)\right)dx, \quad 0 < \theta < 1. \tag{3.2}$$

An alternative version of CTRE was proposed by Rajesh and Sunoj (2019) by replacing the pdf in Eq. (3.2) with the sf. The alternative form of CTRE is defined as

$$\xi_\theta(X) = \frac{1}{\theta - 1}\int_0^{+\infty}\left(S(x) - S^\theta(x)\right)dx, \quad 0 < \theta \neq 1. \tag{3.3}$$

As $\theta \to 1$, $\xi_\theta(X)$ reduces to $CRE(X)$ and for $\theta = 2$, it becomes Gini's mean difference (GMD). This alternative measure is more flexible than $CT_\theta(X)$ because it has more relationships with other measures related to reliability and information theory, see Rajesh and Sunoj (2019) and Toomaj and Atabay (2022). A cdf based version of this measure is introduced by Calì et al. (2017), which is called cumulative Tsallis past entropy (CTPE) and is defined as

$$\bar{\xi}_\theta(X) = \frac{1}{\theta - 1}\int_0^{+\infty}\left(F(x) - F^\theta(x)\right)dx, \quad 0 < \theta \neq 1. \tag{3.4}$$

Recently, a weighted measure based on CTRE has been studied by Khammar and Jahanshahi (2018) which is defined as

$$CT_\theta^w(X) = \frac{1}{\theta - 1}\left(1 - \int_0^{+\infty} xS^\theta(x)\right)dx, \quad 0 < \theta \neq 1. \tag{3.5}$$

In the present chapter, we consider the weighted forms of (3.3) and (3.4) and their dynamic versions. We study numerous properties, define new aging classes based on the dynamic measures and provide characterization theorems for Rayleigh and power distribution. We propose non-parametric estimators of the proposed measures and study their asymptotic properties.

The rest of the chapter is organised as follows. The weighted cumulative Tsallis residual entropy (WCTRE) and its dynamic version are proposed and their properties are studied in Section 3.1. The weighted cumulative Tsallis past entropy (WCTPE) and its dynamic version are proposed in Section 3.2. Aging classes and characterization results based on dynamic entropy measures are studied in Section 3.3. Non-parametric estimators are developed in Section 3.4. Some concluding remarks are made in Section 3.5.

3.1 Weighted cumulative Tsallis residual entropy and its dynamic version

In this section, we propose WCTRE of order θ and its dynamic version and study some interesting properties of these measures.

Definition 3.1.1. *For a non-negative continuous rv X the WCTRE is defined as*

$$\xi_\theta^w(X) = \frac{1}{\theta - 1}\int_0^{+\infty} x\left(S(x) - S^\theta(x)\right)dx, \; 0 < \theta \neq 1. \tag{3.6}$$

Note that WCTRE is a generalization of weighted cumulative residual entropy (WCRE) proposed by Misagh et al. (2011). The WCRE is defined as

$$CRE^{w}(X) = -\int_{0}^{+\infty} xS(x)\log S(x)dx \tag{3.7}$$

and WCTRE reduces to WCRE when $\theta \to 1$. The usefulness of the proposed entropy measure is illustrated through the following example.

Example 3.1.1. *Suppose X and Y are continuous rvs having $U(0,a)$ and $U(h,a+h)$, $a,h > 0$, distributions. Then from Eq. (3.3), it follows that $\xi_\theta(X) = \xi_\theta(Y) = \frac{a}{2(\theta-1)}$. Now from Eq. (3.6) we get, $\xi^{w}_\theta(X) = \frac{a^2(\theta+4)}{6(\theta+1)(\theta+2)}$ and $\xi^{w}_\theta(Y) = \frac{ah}{2(\theta+1)} + \frac{a^2(\theta-1)(\theta+4)}{6(\theta+1)(\theta+2)}$. So it is seen that although $\xi_\theta(X) = \xi_\theta(Y)$ but $\xi^{w}_\theta(X) \neq \xi^{w}_\theta(Y)$.*

It is always interesting to express information measures in terms of expectations of a function of rv. Misagh et al. (2011) proved that WCRE and WCE measures of a rv X are the expectation of WMRL and WMPL of X. In the following lemma, a relationship between WCTRE and WMRL function is provided.

Lemma 3.1.1. *For a non-negative continuous rv X with sf S,*

$$\xi^{w}_\theta(X) = E[m^{w}_F(X)S^{\theta-1}(X)].$$

Proof. Note that $\frac{d}{dx}(m^{w}_F(x)S(x)) = -xS(x)$. Using this fact in Eq. (3.6) we get

$$\xi^{w}_\theta(X) = \frac{1}{\theta-1}\left[-\int_{0}^{+\infty}\frac{d}{dx}(m^{w}_F(x)S(x))(1-S^{\theta-1}(x))dx\right].$$

Using integration by parts we get

$$\xi^{w}_\theta(X) = \frac{1}{\theta-1}\left[0+\int_{0}^{+\infty}(\theta-1)S^{\theta-2}(x)f(x)m^{w}_F(x)S(x)dx\right].$$

Hence the proof. □

Using this result Kattumannil et al. (2022) express WCTRE measure as a special case of generalized entropy measure. The following examples illustrate Lemma 3.1.1.

Example 3.1.2. *Suppose X have an exponential distribution with cdf $F(x) = 1-e^{-\lambda x}$, $x > 0$, $\lambda > 0$, then $\xi^{w}_\theta(X) = \frac{\theta+1}{\lambda^2\theta^2}$, $m^{w}_F(x) = \frac{x}{\lambda} + \frac{1}{\lambda^2}$ and $E[m^{w}_F(X)S^{\theta-1}(X)] = \frac{\theta+1}{\lambda^2\theta^2}$.*

Example 3.1.3. *Suppose X have Pareto distribution with cdf $F(x) = 1-(\frac{b}{x})^a$, $x \geq b$, $b > 0$, $a > 0$. Then $\xi^{w}_\theta(X) = \frac{ab^2}{(a-2)(a\theta-2)}$, $m^{w}_F(x) = \frac{x^2}{a-2}$ and $E[m^{w}_F(X)S^{\theta-1}(X)] = \frac{ab^2}{(a-2)(a\theta-2)}$.*

Next we obtain a bound for WCTRE of a rv X in terms of WCRE of X.

Theorem 3.1.1. *For a rv X, $\xi_\theta^w(X) \leq (\geq)\, CRE^w(X)$ if $\theta > 1 (0 < \theta < 1)$.*

Proof. Suppose $\theta > 1$, then

$$\begin{aligned}
\xi_\theta^w(X) &= \frac{1}{\theta-1}\int_0^{+\infty} x(S(x)-S^\theta(x))dx \\
&= \frac{1}{\theta-1}\int_0^{+\infty} xS(x)(1-S^{\theta-1}(x))dx \\
&\leq -\frac{1}{\theta-1}\int_0^{+\infty} xS(x)\log S^{\theta-1}(x)dx \\
&= -\int_0^{+\infty} xS(x)\log S(x)dx \\
&= CRE^w(X),
\end{aligned}$$

where the inequality follows from the fact that for $u > 0$, $1-u < -\log u$. The inequality will reverse for $(0 < \theta < 1)$. □

Now we obtain upper bounds for WCTRE of smallest and largest order statistics.

Proposition 3.1.1. *Let $X_1, X_2, \cdots, X_n$ be a rs from a cdf F having finite support $[0, l]$, then for $\theta > 1$*

$$\begin{aligned}
&(i) \quad \xi_\theta^w(X_{1:n}) \leq n\int_0^l xF(x)dx; \\
&(ii) \quad \xi_\theta^w(X_{1:n}) \leq n\xi_\theta^w(X); \\
&(iii) \quad \xi_\theta^w(X_{n:n}) \leq \int_0^l xF(x)dx.
\end{aligned}$$

Proof.

$$\begin{aligned}
\xi_\theta^w(X_{1:n}) &= \frac{1}{\theta-1}\int_0^l x(S^n(x)-S^{n\theta}(x))dx \\
&= \tfrac{1}{\theta-1}\int_0^l x(S(x)-S^\theta(x))(S^{n-1}(x)+S^{n-2}(x)S^\theta(x)+\cdots+S^{(n-1)\theta}(x))dx \\
&\leq \frac{n}{\theta-1}\int_0^l x(S(x)-S^\theta(x))dx \\
&= \frac{n}{\theta-1}\int_0^l x[S(x)-(1-F(x))^\theta]dx \\
&\leq \frac{n}{\theta-1}\int_0^l x(S(x)+\theta F(x)-1)dx \\
&= n\int_0^l xF(x)dx,
\end{aligned}$$

where second inequality is obtained by using Bernoulli inequality. First inequality yields $\xi_\theta^w(X_{1:n}) \le n\xi_\theta^w(X)$. Similarly using Bernoulli's inequality in

$$\xi_\theta^w(X_{n:n}) = \frac{1}{\theta-1}\int_0^l x[(1-F^n(x))-(1-F^n(x))^\theta]dx,$$

it follows that $\xi_\theta^w(X_{n:n}) = \frac{1}{\theta-1}\int_0^l xF^n(x)dx \le \int_0^l xF(x)dx.$ □

Note that $X_{1:n}$ and $X_{n:n}$ represent the lifetimes of series and parallel systems, respectively, if the random variables $X_1,\cdots,X_n$ represent the lifetimes of components. Then using Proposition 3.1.1, bounds for the WCTRE of lifetimes of series and parallel systems can be obtained. Now we define dynamic weighted cumulative Tsallis residual entropy (DWCTRE) measure which is the WCTRE of $X_t = [X-t|X\ge t]$.

Definition 3.1.2. *Let X be an absolutely continuous non-negative rv then DWCTRE of order θ of X is given by*

$$\begin{aligned}
\xi_\theta^w(X,t) &= \frac{1}{\theta-1}\int_t^{+\infty} x\left(S_{X_t}(x)-(S_{X_t}(x))^\theta\right)dx \\
&= \frac{1}{\theta-1}\int_t^{+\infty} x\left(\frac{S(x)}{S(t)}-\left(\frac{S(x)}{S(t)}\right)^\theta\right)dx \\
&= \frac{1}{\theta-1}\left(m_F^w(t)-\int_t^{+\infty} x\left(\frac{S(x)}{S(t)}\right)^\theta dx\right). \qquad (3.8)
\end{aligned}$$

The relationship of DWCTRE and WMRL is provided in the next theorem.

Theorem 3.1.2. *Let X be an absolutely continuous non-negative rv with WMRL function $m_F^w(t)$ then $\xi_\theta^w(X,t) = \dfrac{E[m_F^w(X)S^{\theta-1}(X)|X>t]}{S^{\theta-1}(t)}$.*

Proof. From Eq. (3.8) we have,

$$\begin{aligned}
\xi_\theta^w(X,t) &= \frac{1}{\theta-1}\left[m_F^w(t)+\frac{1}{S^\theta(t)}\int_t^{+\infty}\left(\frac{d}{dx}(m_F^w(x)S(x))S^{\theta-1}(x)\right)dx\right] \\
&= \tfrac{1}{\theta-1}\left[m_F^w(t)+\tfrac{1}{S^\theta(t)}\left(-m_F^w(t)S^\theta(t)+(\theta-1)\textstyle\int_t^{+\infty} m_F^w(x)S^{\theta-1}(x)f(x)dx\right)\right] \\
&= \frac{1}{S^{\theta-1}(t)}\int_t^{+\infty} m_F^w(x)S^{\theta-1}(x)\frac{f(x)}{S(t)}dx. \qquad (3.9)
\end{aligned}$$

Hence the proof. □

Corollary 3.1.2.1. *If X has decreasing (increasing) weighted mean residual life (DWMRL (IWMRL)) then $\xi_\theta^w(X,t) \le (\ge)\frac{m_F^w(t)}{\theta}$.*

Proof. If X is DWMRL (IWMRL) then for $x \geq t$ we have $m_F^w(x) \leq (\geq)\, m_F^w(t)$. Using this fact in Eq. (3.9) and after some simplifications, we get the result. □

Next, the effect of linear transformation on DWCTRE measure is studied. Proofs are analogous to the results studied in Chapter 2 regarding linear transformation.

Proposition 3.1.2. *If $Y = cX + d$ with $c > 0$ and $d \geq 0$ then,*

(i) $\xi_\theta^w(Y) = c^2 \xi_\theta^w(X) + cd\xi_\theta(X)$;

(ii) $\xi_\theta^w(Y,t) = c^2 \xi_\theta^w\left(X, \frac{t-d}{c}\right) + cd\xi_\theta\left(X, \frac{t-d}{c}\right)$.

We express the relationship of DWCTRE measure between the rvs X_η and X, where X_η and X satisfies the PHRM i.e. $S_{X_\eta}(x) = [S(x)]^\eta$, $\eta > 0$.

Proposition 3.1.3. *The DWCTRE between the rvs X, X_η and ηX can be expressed as follows:*

(i) $(\theta - 1)\, \xi_\theta^w(X_\eta) - (\theta\eta - 1)\, \xi_{\theta\eta}^w(X) = (1-\eta)\, \xi_\theta^w(X)$;

(ii) $(\theta - 1)\, \xi_\theta^w(X_\eta, t) - (\theta\eta - 1)\, \xi_{\theta\eta}^w(X, t) = (1-\eta)\, \xi_\theta^w(X, t)$.

Proof. Proof follows from the definition of WCTRE (DWCTRE). □

3.2 Weighted cumulative Tsallis past entropy and its dynamic version

In this section, we define WCTPE of order θ along with its dynamic version.

Definition 3.2.1. *For a rv X the WCTPE of order θ is defined as*

$$\bar{\xi}_\theta^w(X) = \frac{1}{\theta - 1} \int_0^{+\infty} x \left(F(x) - F^\theta(x)\right) dx,\ 0 < \theta \neq 1. \tag{3.10}$$

Note that WCTPE is a generalization of the weighted cumulative entropy (WCE) measure, which was introduced by Misagh et al. (2011) as

$$CE^w(X) = -\int_0^{+\infty} xF(x) \log F(x) dx. \tag{3.11}$$

As $\theta \to 1$, WCTPE reduces to $CE^w(X)$. It is related to the WMPL measure which is demonstrated through the following lemma and example.

Lemma 3.2.1. *For a rv X with cdf F,* $\bar{\xi}^w_\theta(X) = E[\mu^w_F(X)F^{\theta-1}(X)]$.

Proof. Note that $\frac{d}{dx}(\mu^w_F(x)F(x)) = xF(x)$. Then using this fact in Eq. (3.10), we get

$$\bar{\xi}^w_\theta(X) = \frac{1}{\theta-1}\int_0^{+\infty}\frac{d}{dx}(\mu^w_F(x)F(x))(1-F^{\theta-1}(x))dx.$$

Now the result follows using integration by parts. □

Example 3.2.1. *Suppose X have U(0,1) distribution with cdf* $F(x) = x,\ 0 < x < 1$ *then* $\bar{\xi}^w_\theta(X) = \frac{1}{3(\theta+2)}$, $\mu^w_F(x) = \frac{x^2}{3}$. *It can be easily shown that,*

$$E[\mu^w_F(X)F^{\theta-1}(X)] = E\left(\frac{X^{\theta+1}}{3}\right) = \frac{1}{3(\theta+2)}.$$

Next the relationship between WCTPE and WCE measures is provided. Also some bounds for WCTPE of extreme order statistics are obtained.

Theorem 3.2.1. *Let X be a rv with cdf F, then* $\bar{\xi}^w_\theta(X) \le (\ge)\ CE^w(X)$ *if* $\theta > 1 (0 < \theta < 1)$.

Proof. Proof is similar to that of Theorem 3.1.1. □

Proposition 3.2.1. *Let* $X_1, X_2, \cdots, X_n$ *be i.i.d. rvs then for* $\theta > 1$

$$(i)\quad \bar{\xi}^w_\theta(X_{n:n}) \le \frac{n}{2}E(X^2);$$
$$(ii)\quad \bar{\xi}^w_\theta(X_{1:n}) \le n\,\bar{\xi}^w_\theta(X);$$
$$(iii)\quad \bar{\xi}^w_\theta(X_{1:n}) \le \frac{n}{2}E(X^2).$$

Proof. Proofs are analogous to that of Proposition 3.1.1. □

Next we propose dynamic weighted cumulative Tsallis past entropy (DWCTPE) measure and study some properties.

Definition 3.2.2. *For a rv X the DWCTPE measure is given by*

$$\begin{aligned}
\bar{\xi}^w_\theta(X,t) &= \frac{1}{\theta-1}\int_0^t x\left(F_{X_{t-}}(x) - (F_{{}_tX}(x))^\theta\right)dx \\
&= \frac{1}{\theta-1}\int_0^t x\left(\frac{F(x)}{F(t)} - \left(\frac{F(x)}{F(t)}\right)^\theta\right)dx \\
&= \frac{1}{\theta-1}\left(\mu^w_F(t) - \int_0^t x\left(\frac{F(x)}{F(t)}\right)^\theta dx\right). \qquad (3.12)
\end{aligned}$$

Theorem 3.2.2. *Let X be an absolutely continuous non-negative rv with WMPL function $\mu_F^w(t)$ then,*

$$\bar{\xi}_\theta^w(X,t) = \frac{E[\mu_F^w(X)F^{\theta-1}(X)|X<t]}{F^{\theta-1}(t)}. \tag{3.13}$$

Proof. We have,

$$\bar{\xi}_\theta^w(X,t) = \frac{1}{\theta-1}\left[\mu_F^w(t) - \frac{1}{F^\theta(t)}\int_0^t \frac{d}{dx}(\mu_F^w(x)F(x))F^{\theta-1}(x)dx\right].$$

Applying integration by parts and after some simplification, we get

$$\begin{aligned}\bar{\xi}_\theta^w(X,t) &= \tfrac{1}{\theta-1}\left[\mu_F^w(t) - \tfrac{1}{F^\theta(t)}\left(\mu_F^w(t)F^\theta(t) - (\theta-1)\textstyle\int_0^t \mu_F^w(x)F^{\theta-1}(x)f(x)dx\right)\right]\\ &= \frac{1}{F^{\theta-1}(t)}\int_0^t \mu_F^w(x)F^{\theta-1}(x)\frac{f(x)}{F(t)}dx.\end{aligned}$$

Hence the result. □

Corollary 3.2.2.1. *If X has decreasing (increasing) WMPL then $\bar{\xi}_\theta^w(X,t) \geq (\leq) \frac{\mu_F^w(t)}{\theta}$.*

Proof. If X has decreasing (increasing) WMPL then $\mu_F^w(x) \geq \mu_F^w(t)$ for $x \leq t$. Using this fact in Eq. (3.13) we get the result. □

Next, the effect of linear transformation of X on DWCTPE is considered.

Proposition 3.2.2. *If $Y = cX + d$ with $c > 0$ and $d \geq 0$ then,*

(i) $\bar{\xi}_\theta^w(Y) = c^2\bar{\xi}_\theta^w(X) + cd\ \bar{\xi}_\theta(X)$,

(ii) $\bar{\xi}_\theta^w(Y,t) = c^2\bar{\xi}_\theta^w\left(X, \frac{t-d}{c}\right) + cd\ \bar{\xi}_\theta\left(X, \frac{t-d}{c}\right)$.

We obtain the expression of WCTPE (DWCTPE) under PRHRM. Proof follows from their definitions, hence omitted.

Proposition 3.2.3. *Let X_η and X satisfies the PRHRM i.e. $F_{X_\eta}(x) = [F(x)]^\eta$, $\eta > 0$. Then,*

(i) $(\theta-1)\ \bar{\xi}_\theta^w(X_\eta) - (\theta\eta-1)\ \bar{\xi}_{\theta\eta}^w(X) = (1-\eta)\ \bar{\xi}_\theta^w(X)$;

(ii) $(\theta-1)\ \bar{\xi}_\theta^w(X_\eta,t) - (\theta\eta-1)\ \bar{\xi}_{\theta\eta}^w(X,t) = (1-\eta)\ \bar{\xi}_\theta^w(X,t)$.

3.3 Aging classes and charaterizations

In this section, two new aging classes are introduced based on the proposed dynamic information measures and some characterization results for Rayleigh and power distributions are obtained. First we show that DWCTRE uniquely determines the distribution. The following lemma will be needed to prove the uniqueness result.

Lemma 3.3.1. *Let X_1 and X_2 be two non-negative continuous rvs with cdfs F_1 and F_2, respectively. If $X_1 \overset{HR}{\leq} X_2$, then $m^w_{F_1}(t) \leq m^w_{F_2}(t)$, where $m^w_{F_1}(t)$ and $m^w_{F_2}(t)$ are the WMRL of X_1 and X_2, respectively.*

Proof. If $X_1 \overset{hr}{\leq} X_2$ then for $x \geq t$, $\frac{S_1(x)}{S_1(t)} \leq \frac{S_2(x)}{S_2(t)}$. Now by multiplying both sides with x and taking integration from t to $+\infty$, the required result follows. □

Theorem 3.3.1. *Let X be a non-negative continuous rv having pdf f and sf S. Assume that $\xi^w_\theta(X,t) < +\infty$; $t \geq 0$, $0 < \theta \neq 1$. Then for each θ, $\xi^w_\theta(X,t)$ uniquely determines the sf of X.*

Proof. Suppose X_1 and X_2 be two rvs with cdfs F_1 and F_2, respectively. Assume that

$$\xi^w_\theta(X_1,t) = \xi^w_\theta(X_2,t).$$

Differentiating both sides wrt t we get,

$$\lambda_{F_1}(t)[\theta\xi^w_\theta(X_1,t) - m^w_{F_1}(t)] = \lambda_{F_2}(t)[\theta\xi^w_\theta(X_2,t) - m^w_{F_2}(t)]. \tag{3.14}$$

Now if $\forall\, t \geq 0$, $\lambda_{F_1}(t) = \lambda_{F_2}(t)$ then $F_1 = F_2$ and the proof is complete. But suppose for some $t = t_0$, $\lambda_{F_1}(t_0) \neq \lambda_{F_2}(t_0)$ and without loss of generality, assume that $\lambda_{F_1}(t_0) > \lambda_{F_2}(t_0)$. Then (3.14) implies

$$\theta\xi^w_\theta(X_1,t_0) - m^w_{F_1}(t_0) < \theta\xi^w_\theta(X_2,t_0) - m^w_{F_2}(t_0)$$

and hence $m^w_{F_1}(t_0) > m^w_{F_2}(t_0)$. This is a contradiction since from Lemma 3.3.1 we have $m^w_{F_1}(t_0) < m^w_{F_2}(t_0)$ when $\lambda_{F_1}(t_0) > \lambda_{F_2}(t_0)$. Hence the proof. □

Remark 3.3.1. *Although DWCTRE uniquely determines the underlying distribution, this can not be said for DWCTPE measure. The DWCTPE can not determine the underlying cdf uniquely. i.e $\bar{\xi}^w_\theta(X,t)$=$\bar{\xi}^w_\theta(Y,t)$ does not necessarily imply that X and Y have the same distribution. Suppose X has uniform distribution with cdf $F(x) = \frac{x}{a}$, $0 < x < a$ and Y has power distribution with cdf $G(x) = \left(\frac{x}{a}\right)^c$; $0 < x < a$, $c > 0$. Then $\bar{\xi}^w_\theta(X,t) = \frac{t^2}{3(\theta+2)}$ and $\bar{\xi}^w_\theta(Y,t) = \frac{c\,t^2}{(c\theta+2)(c+2)}$ and for $c = \frac{4}{\theta}$, $\bar{\xi}^w_\theta(X,t) = \bar{\xi}^w_\theta(Y,t)$.*

Next we define some aging classes based on the proposed dynamic measures and study some properties.

Definition 3.3.1. *X is said to be increasing (decreasing) dynamic weighted cumulative Tsallis residual entropy (IDWCTRE) (DDWCTRE) of order θ iff $\xi^w_\theta(X,t)$ is increasing (decreasing) in t.*

Definition 3.3.2. *X is said to be increasing (decreasing) dynamic weighted cumulative Tsallis past entropy (IDWCTPE) (DDWCTPE) of order θ iff $\bar{\xi}_\theta^w(X,t)$ is increasing (decreasing) in t.*

The following theorem provide conditions for which the aging classes hold.

Theorem 3.3.2. *The measure $\xi_\theta^w(X,t)$ is increasing (decreasing) in t iff*

$$\xi_\theta^w(X,t) \geq (\leq) \frac{m_F^w(t)}{\theta}.$$

Proof. Differentiating Eq. (3.8) we get

$$\frac{d}{dt}\xi_\theta^w(X,t) = \frac{1}{\theta-1}\left(\lambda_F(t)m_F^w(t) - \theta\lambda_F(t)\int_t^{+\infty} x\frac{S^\theta(x)}{S^\theta(t)}\right)dx,$$

which reduces to

$$\frac{d}{dt}\xi_\theta^w(X,t) = \lambda_F(t)\left(\theta\xi_\theta^w(X,t) - m_F^w(t)\right). \tag{3.15}$$

The result follows from (3.15). □

Theorem 3.3.3. *The DWCTPE of X, $\bar{\xi}_\theta^w(X,t)$ is increasing (decreasing) in t iff*

$$\bar{\xi}_\theta^w(X,t) \leq (\geq) \frac{\mu_F^w(t)}{\theta}.$$

Proof. Differentiating Eq. (3.12) and after some simplifications, we get

$$\frac{d}{dt}\bar{\xi}_\theta^w(X,t) = r_F(t)(\mu_F^w(t) - \theta\bar{\xi}_\theta^w(X,t)), \tag{3.16}$$

where $r_F(t) = \frac{f(t)}{F(t)}$ is the rhr of X. The result follows from (3.16). □

The following theorems address characterization results for Rayleigh and power distributions.

Theorem 3.3.4. *For a non-negative, continuous rv X, DWCTRE is constant iff X has a Rayleigh distribution.*

Proof. If X has Rayleigh distribution then it immediately follows that $\xi_\theta^w(X,t)$ is constant. Now suppose $\xi_\theta^w(X,t) = c$, where c is a constant. This gives

$$\frac{d}{dt}\xi_\theta^w(X,t) = 0$$

which implies

$$m_F^w(t) = c\theta$$

and hence

$$\lambda_F(t) = \frac{t}{c\theta},$$

which is the hr of Rayleigh distribution with cdf $F(t) = 1 - e^{-\frac{t^2}{2c\theta}}$, $t > 0,\ c > 0,\ \theta \neq 1$. Hence the result. □

Theorem 3.3.5. *Let X be a non-negative rv having support* $(0,b)$*, with absolutely continuous cdf F and rhr function* r_F*. Then X has a power distribution with* $F(x) = \left(\frac{x}{b}\right)^c$, $0 < x < b,\ c > 0$ *iff*

$$\bar{\xi}_\theta^w(X,t) = c\,\mu_F^w(t),$$

where $\mu_F^w(t)$ *is the WMPL function of X.*

Proof. If part is straight forward. Now suppose $\bar{\xi}_\theta^w(X,t) = c\mu_F^w(t)$ holds. Then differentiating with respect to t and after some simplification, we get

$$r_F(t)\mu_F^w(t) = \frac{ct}{1-(\theta-1)c}.$$

This implies

$$\frac{d}{dt}\mu_F^w(t) = \frac{1-\theta c}{1-(\theta-1)c}t.$$

Integrating wrt t and taking $\mu_F^w(0) = 0$ we get

$$\mu_F^w(t) = \frac{1-\theta c}{1-(\theta-1)c}\frac{t^2}{2},$$

which implies

$$r_F(t) = \frac{2c}{(1-\theta c)}\frac{1}{t},\ c\theta < 1.$$

Clearly $r_F(t)$ is the rhr of power distribution with $F(t) = \left(\frac{t}{b}\right)^{\frac{2c}{1-c\theta}}$, $0 < t < b$, $c\theta < 1$. Hence the proof. □

In the following theorem we generalize Theorem 3.3.5 in the sense that instead of a constant c we take a function of t, $c(t)$, say, and obtain a generalized relation.

Theorem 3.3.6. *Let X be a non-negative continuous rv having support* $(0,b)$ *such that* $\bar{\xi}_\theta^w(X,t) = c(t)\mu_F^w(t)$*, then*

$$\mu_F^w(t) = (P_\theta(t))^{\frac{\theta}{\theta-1}} \int_0^t \frac{x(1-\theta c(x))}{(P_\theta(x))^{\frac{\theta}{\theta-1}}} dx, \tag{3.17}$$

where $P_\theta(x) = 1-(\theta-1)c(x)$.

Proof. Suppose the relation $\bar{\xi}_\theta^w(X,t) = c(t)\mu_F^w(t)$ holds. Then differentiating this and after simplification we have

$$\mu_F^{\prime *}(t) + \frac{c'(t)}{P_\theta(t)}\mu_F^w(t) = \frac{1-\theta c(t)}{(P_\theta(t))}t.$$

This is a first order differential equation in $\mu_F^w(t)$, with solution of the form (3.17). □

3.4 Estimation

In this section, we propose edf based non-parametric estimators of WCTRE and WCTPE measures. Asymptotic normality of these estimators are established when the random sample comes from the Rayleigh distribution. Let $X_1, X_2, \cdots, X_n$ be a rs from a continuous distribution and $X_{1:n}, X_{2:n}, \cdots, X_{n:n}$ are the corresponding order statistics. We propose non-parametric estimators of WCTRE and WCTPE measures using the edf function of X. The empirical WCTRE is defined as

$$\hat{\xi}_\theta^w(X) = \frac{1}{\theta-1}\int_0^{+\infty} x(S_n(x) - S_n^\theta(x))dx, \;\; \theta \neq 1, \; \theta > 0. \tag{3.18}$$

Substituting $S_n(x) = 1-\frac{i}{n}$, $i = 1,2,\cdots,n-1$, in (3.18) we get,

$$\hat{\xi}_\theta^w(X) = \frac{1}{2(\theta-1)}\sum_{i=1}^{n-1} U_{(i+1)}\left[\left(1-\frac{i}{n}\right) - \left(1-\frac{i}{n}\right)^\theta\right], \tag{3.19}$$

where $U_{i+1} = X^2_{(i+1):n} - X^2_{i:n}$ and $U_1 = X^2_{1:n}$. Similarly, empirical WCTPE can be expressed as

$$\hat{\bar{\xi}}_\theta^w(X) = \frac{1}{2(\theta-1)}\sum_{i=1}^{n-1} U_{(i+1)}\left[\frac{i}{n} - \left(\frac{i}{n}\right)^\theta\right]. \tag{3.20}$$

Now we provide two central limit theorems (CLT) for the estimators when data comes from the Rayleigh distribution.

Theorem 3.4.1. *Let* $X_1, X_2, \cdots, X_n$ *be a random sample from the Rayleigh distribution with pdf* $f(x) = 2\lambda x e^{-\lambda x^2}$; $x > 0$, $\lambda > 0$, *then* $\frac{\hat{\xi}_\theta^w(X) - E[\hat{\xi}_\theta^w(X)]}{\sqrt{Var[\hat{\xi}_\theta^w(X)]}} \to N(0,1)$ *in distribution as* $n \to +\infty$ *and for* $\theta > 2/3 (\neq 1)$.

Proof. The empirical WCTRE can be expressed as $\hat{\xi}^w_\theta(X) = \sum_{i=1}^{n-1} Z_i$, where

$$Z_i = \frac{1}{2(\theta-1)} U_{(i+1)} \left[\left(1-\frac{i}{n}\right) - \left(1-\frac{i}{n}\right)^\theta \right],\ i = 1,2,\cdots,n-1.$$

Since X has Rayleigh distribution with pdf $f(x) = 2\lambda x e^{-\lambda x^2}$; $x > 0$, $\lambda > 0$, then X^2 has exponential distribution with mean $1/\lambda$ and $U_{i+1} = X^2_{(i+1):n} - X^2_{i:n}$ also has exponential distribution with mean $1/\lambda(n-i)$ (Pyke, 1965). So,

$$\begin{aligned} E(Z_i) &= \frac{1}{2(\theta-1)\lambda(n-i)} \left[\left(1-\frac{i}{n}\right) - \left(1-\frac{i}{n}\right)^\theta \right], \\ Var[Z_i] &= \frac{1}{4(\theta-1)^2\lambda^2(n-i)^2} \left[\left(1-\frac{i}{n}\right) - \left(1-\frac{i}{n}\right)^\theta \right]^2. \end{aligned}$$

For any exponentially distributed rv Z_i, Di Crescenzo and Longobardi (2009) showed that $E[|Z_i - E(Z_i)|^3] = 2e^{-1}(6-e)[E(Z_i)]^3$. Denote

$$A^n_{i,\delta} = E[|Z_i - E(Z_i)|^\delta],$$

then for large n we have,

$$\begin{aligned} \sum_{i=1}^n A^n_{i,2} &= \sum_{i=1}^n E[|Z_i - E(Z_i)|^2] \\ &= \frac{1}{4\lambda^2(\theta-1)^2} \sum_{i=1}^n \frac{1}{(n-i)^2} \left[\left(1-\frac{i}{n}\right) - \left(1-\frac{i}{n}\right)^\theta \right]^2 \\ &= \frac{1}{4\lambda^2(\theta-1)^2 n^2} \sum_{i=1}^n \left[1 - \left(1-\frac{i}{n}\right)^{\theta-1} \right]^2 \\ &\approx \frac{C_2}{4\lambda^2(\theta-1)^2 n}, \end{aligned}$$

where $C_2 = \int_0^1 [1-(1-x)^{\theta-1}]^2 dx$ and C_2 converges for $\theta > 1/2$. Again,

$$\begin{aligned} \sum_{i=1}^n A^n_{i,3} &= \sum_{i=1}^n E[|Z_i - E(Z_i)|^3] \\ &= 2e^{-1}(6-e) \sum_{i=1}^n \frac{1}{8(\theta-1)^3\lambda^3(n-i)^3} \left[\left(1-\frac{i}{n}\right) - \left(1-\frac{i}{n}\right)^\theta \right]^3 \\ &= \frac{(6-e)}{4e(\theta-1)^3\lambda^3 n^3} \sum_{i=1}^n \left[1 - \left(1-\frac{i}{n}\right)^{\theta-1} \right]^3 \approx \frac{(6-e)C_3}{4e(\theta-1)^3\lambda^3 n^2}, \end{aligned}$$

where $C_3 = \int_0^1 [1-(1-x)^{\theta-1}]^3 dx$ which converges for $\theta > 2/3$. So both integral converges for $\theta > 2/3$.
Now

$$\frac{(\sum_{i=1}^n \theta_{i,3}^n)^{1/3}}{(\sum_{i=1}^n \theta_{i,2}^n)^{1/2}} \approx \frac{\sqrt[3]{2e^{-1}(6-e)C_3}}{\sqrt{C_2}} n^{-\frac{1}{6}},$$

which goes to 0 as $n \to +\infty$ and for $\theta > 2/3$. Therefore, Lyapunov's condition for CLT is satisfied. Hence the proof. □

Theorem 3.4.2. *Let $X_1, X_2, \cdots, X_n$ be a random sample from the Rayleigh distribution with pdf $f(x) = 2\lambda x e^{-\lambda x^2}$; $x > 0$, $\lambda > 0$, then for every $0 < \theta \neq 1$, $\frac{\hat{\hat{\xi}}_\theta^w(X) - E[\hat{\hat{\xi}}_\theta^w(X)]}{\sqrt{Var[\hat{\hat{\xi}}_\theta^w(X)]}} \to N(0,1)$ in distribution as $n \to +\infty$.*

Proof. Note that $\hat{\hat{\xi}}_\theta^w(X)$ can be represented as $\hat{\hat{\xi}}_\theta^w(X) = \sum_{i=1}^{n-1} Z_i^1$, where

$$Z_i^1 = \frac{1}{2(\theta-1)} U_{(i+1)} \left[\frac{i}{n} - \left(\frac{i}{n}\right)^\theta\right], \; i = 1, 2, \cdots, n-1.$$

Along the same line as Theorem 3.4.1 the mean and variance of Z_i^1 can be obtained as

$$E(Z_i^1) = \frac{1}{2(\theta-1)\lambda(n-i)} \left[\frac{i}{n} - \left(\frac{i}{n}\right)^\theta\right],$$

$$Var(Z_i^1) = \frac{1}{4(\theta-1)^2\lambda^2(n-i)^2} \left[\frac{i}{n} - \left(\frac{i}{n}\right)^\theta\right]^2.$$

Denote $B_{i,\delta}^n = E[|Z_i^1 - E(Z_i^1)|^\delta]$. For large n,

$$\sum_{i=1}^n B_{i,2}^n = \sum_{i=1}^n E[|Z_i^1 - E(Z_i^1)|^2] = \frac{1}{4(\theta-1)^2\lambda^2 n^2} \sum_{i=1}^n \frac{\left[\frac{i}{n} - \left(\frac{i}{n}\right)^\theta\right]^2}{(1-\frac{i}{n})^2} \approx \frac{C_2^1}{4(\theta-1)^2\lambda^2 n}$$

and

$$\sum_{i=1}^n B_{i,3}^n = \sum_{i=1}^n E[|Z_i^1 - E(Z_i^1)|^3] = \frac{(6-e)}{4e(\theta-1)^3\lambda^3 n^3} \sum_{i=1}^n \frac{\left[\frac{i}{n} - \left(\frac{i}{n}\right)^\theta\right]^3}{(1-\frac{i}{n})^3} \approx \frac{(6-e)C_3^1}{4e(\theta-1)^3\lambda^3 n^2},$$

where $C_2^1 = \int_0^1 \frac{(x-x^\theta)^2}{(1-x)^2} dx < +\infty$ and $C_3^1 = \int_0^1 \frac{(x-x^\theta)^3}{(1-x)^3} dx < +\infty \; \forall \theta > 0 (\neq 1)$.

Since,

$$\frac{(\sum_{i=1}^{n} B_{i,3}^{n})^{1/3}}{(\sum_{i=1}^{n} B_{i,2}^{n})^{1/2}} \approx \frac{\sqrt[3]{2e^{-1}(6-e)C_3^1}}{\sqrt{C_2^1}} n^{-\frac{1}{6}}$$

goes to 0 as $n \to +\infty$, Lyapunov's condition for CLT is satisfied. Hence the result. □

3.5 Discussions

In this chapter, we proposed weighted cumulative Tsallis residual and past entropy measures and their dynamic versions and studied various properties, developed aging classes and obtained characterization results for Rayleigh and power distributions. It is shown that dynamic weighted cumulative Tsallis residual entropy uniquely determines the underlying distribution; however, the same can not be said for dynamic weighted cumulative Tsallis past entropy measure. Also we proposed non-parametric estimators and it is observed that CLT holds for the estimators when a random sample comes from Rayleigh distribution.

The choice of the parameter θ remains a problem of interest. One can choose θ in such a way that the asymptotic variance of the estimators will be minimum. More work is needed in this direction. Also we have found that weighted cumulative Tsallis residual entropy is a generalization of weighted cumulative residual entropy measure and it has similar properties that weighted cumulative residual entropy possesses. Being a generalized information measure, it is much more useful than weighted cumulative residual entropy (which is a special case) because of the generalization parameter θ which give this measure a wide variety of possibilities in computing information.

Chapter 4

On weighted cumulative Kullback-Leibler information with application in testing exponentiality

THE Kullback-Leibler (KL) divergence defined in Eq. (1.15) is widely used as a measure of closeness between two models. It is extensively used in goodness-of-fit tests and model discrimination problems. An enormous amount of applications of KL divergence makes it a very popular measure and motivates researchers in further study regarding extensions and applications of this measure. Over the years, various KL type measures have been proposed in the literature based on different entropy measures like Renyi and Tsallis entropy. In recent years, KL type information measures are introduced based on cumulative entropies. Baratpour and Rad (2012) first proposed cumulative residual Kullback-Leibler information (CRKL) between two rvs X and Y as

$$CRKL(X,Y) = \int_0^{+\infty} S_1(x) \log \frac{S_1(x)}{S_2(x)} dx + (E(Y) - E(X)),$$

where S_1 and S_2 are the sfs of X and Y, respectively. Similarly, based on cumulative entropy measure, Di Crescenzo and Longobardi (2015) proposed the cumulative Kullback-Leibler information (CKL) between two rvs X and Y as

$$CKL(X,Y) = \int_0^{+\infty} F(x) \log \frac{F(x)}{G(x)} dx + (E(X) - E(Y)),$$

where F and G are the cdfs of X and Y, respectively. It may be noted that, CRKL and CKL are non-negative and equal to zero when $F(x) = G(x)$ for all x. For further details on CRKL, see Chamany and Baratpour (2014). Both CRKL and CKL measures do not consider the realization of the rvs. Recently, weighted information measures gained popularity among authors but not much attention is given towards weighted KL type divergence measures. Moharana and Kayal (2019) studied weighted Kullback-Leibler divergence measure for double truncated data. In this chapter, we propose weighted cumulative Kullback-Leibler information measures and study their properties. Using weighted cumulative residual Kullback-Leibler information we develop goodness-of-fit tests for exponential distribution for complete and censored data. We assess the performance of the proposed tests by means of power. Real data sets are analysed for illustrations. The organization of the Chapter is as follows.

Weighted cumulative residual Kullback-Leibler information (WCRKL) measure and its dynamic version, are introduced and their properties are studied in Section 4.2. The weighted cumulative Kullback-Leibler information (WCKL) measure is studied along with its dynamic version in Section 4.3. Goodness-of-fit tests for exponential distribution using WCRKL measure for complete and censored data are developed and their performance with other entropy based exponentiality tests are compared in Section 4.4. Finally, we conclude the chapter in Section 4.5.

4.1 Weighted cumulative residual Kullback-Leibler information measure

In this section, based on WCRE measure defined in Eq. (3.7), we define WCRKL measure. The WCRKL between two rvs X and Y is defined as

$$\begin{aligned} CRKL^w(X,Y) &= \int_0^{+\infty} xS_1(x)\left(\frac{S_2(x)}{S_1(x)} - \log\frac{S_2(x)}{S_1(x)} - 1\right)dx \\ &= \int_0^{+\infty} xS_1(x)\log\frac{S_1(x)}{S_2(x)}dx + \int_0^{+\infty} xS_2(x)dx - \int_0^{+\infty} xS_1(x)dx \\ &= \int_0^{+\infty} xS_1(x)\log\frac{S_1(x)}{S_2(x)}dx + \frac{1}{2}(E(Y^2) - E(X^2)). \end{aligned} \tag{4.1}$$

Note that $CRKL^w(X,Y) \geq 0$ using the fact that $x - \log x - 1 \geq 0$, $\forall x$ and $CRKL^w(X,Y) = 0$ iff $S_1(x) = S_2(x)$, $\forall x > 0$. Next, some basic results regarding WCRKL measure are provided.

Remark 4.1.1. *For two continuous non-negative rvs X and Y*

$$CRKL^w(X,Y) = K^w(X,Y) - CRE^w(X) + \frac{1}{2}(E(Y^2) - E(X^2)),$$

where $K^w(X,Y) = -\int_0^{+\infty} xS_1(x)\log S_2(x)dx$ *is the weighted cumulative residual inaccuracy measure proposed by Daneshi et al. (2019) and* $CRE^w(X)$ *is the WCRE of* X.

Example 4.1.1. *Suppose X and Y follow exponential distributions with respective means* $\frac{1}{\lambda}$ *and* $\frac{1}{\theta}$, *then WCRKL between X and Y is* $\frac{2\theta}{\lambda^3} - \frac{3}{\lambda^2} + \frac{1}{\theta^2}$, *which is zero when* $\lambda = \theta$.

Proposition 4.1.1. *If* $X \overset{st}{\geq} Y$ *then* $CRE^w(X) + \frac{1}{2}E(X^2) \leq CRE^w(Y) + \frac{1}{2}E(Y^2)$.

Proof. Since $CRKL^w(X,Y) \geq 0$, from Remark 4.1.1 we have,

$$CRE^w(X) + \frac{1}{2}E(X^2) \leq K^w(X,Y) + \frac{1}{2}E(Y^2).$$

Now $X \overset{st}{\geq} Y$ implies $K^w(X,Y) \leq CRE^w(Y)$. Hence the result. □

Next we provide lower bound for WCRKL measure in terms of the second raw moments of X and Y.

Proposition 4.1.2. *For two rvs X and Y*

$$CRKL^w(X,Y) \geq \frac{1}{2}\left(E(X^2)\log\frac{E(X^2)}{E(Y^2)} + E(Y^2) - E(X^2)\right). \tag{4.2}$$

Proof. Proof follows using log-sum inequality. Note that the RHS of (4.2) is non-negative as $x\log\frac{x}{y} \geq x - y$. □

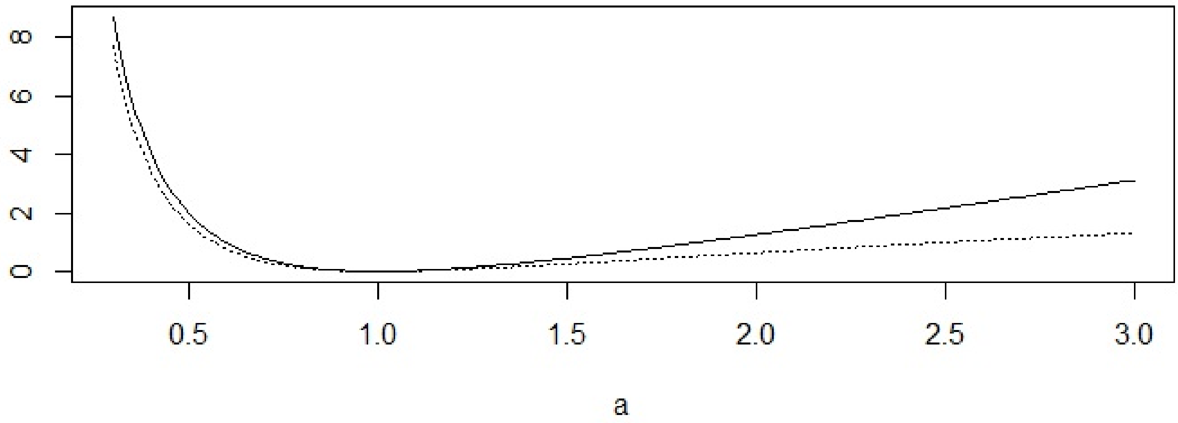
Example 4.1.2. *Suppose X and Y are exponentially distributed rvs with mean* 1 *and* $\frac{1}{a}$, *respectively. Then WCRKL between X and Y is* $2a + \frac{1}{a^2} - 3$ *and the lower bound in Eq. (4.2) is* $2\log a + \frac{1}{a^2} - 1$. *We plot these values for* $0.3 \leq a \leq 3$ *in Figure 4.1.*

From Figure 4.1 it is observed that, WCRKL and its lower bound is zero when $a = 1$, which is obvious. The difference between WCRKL and its lower bound increases as a moves away from 1.

Next we define the dynamic version of WCRKL measure. Let X and Y be two non-negative continuous rvs with sfs S_1 and S_2, respectively. Let $X_t = [X - t|X > t]$ and $Y_t = [Y - t|Y > t]$ be the respective residual lifetimes. Then, dynamic WCRKL between X and Y is given by

$$CRKL^w(X,Y;t) = CRKL^w(X_t,Y_t) = \int_t^{+\infty} x\frac{S_1(x)}{S_1(t)}\log\frac{S_1(x)/S_1(t)}{S_2(x)/S_2(t)}dx + m_G^W(t) - m_F^W(t).$$

Fig. 4.1: *WCRKL and its lower bound.*



Remark 4.1.2. *For two rvs X and Y, the following relation holds.*

$$CRKL^w(X,Y;t) = K^w(X,Y;t) - CRE^w(X;t) + m_G^w(t) - m_F^w(t),$$

where $K^w(X,Y;t) = \int_t^{+\infty} x\frac{S_1(x)}{S_1(t)} \log \frac{S_2(x)}{S_2(t)} dx$ is the dynamic weighted cumulative residual inaccuracy measure (Daneshi et al., 2019) and $CRE^w(X;t) = -\int_t^{+\infty} \frac{S(x)}{S(t)} \log \frac{S(x)}{S(t)} dx$ is the dynamic WCRE of X, see Mirali and Baratpour (2017).

Now we provide lower bound for $CRKL^w(X,Y;t)$ in terms of WMRL of X and Y.

Proposition 4.1.3. *The following inequality holds for dynamic WCRKL measure.*

$$CRKL^w(X,Y;t) \geq m_F^w(t) \log \frac{m_F^w(t)}{m_G^w(t)} + m_G^w(t) - m_F^w(t).$$

Proof. The result follows by substituting $U = \frac{S_1(x)}{S_1(t)}$ and $V = \frac{S_2(x)}{S_2(t)}$ in the log-sum inequality $U \log \frac{U}{V} \geq U - V$ and integrating from t to $+\infty$ wrt x. □

Following theorem addresses the monotonicity of $CRKL^w(X,Y;t)$.

Theorem 4.1.1. *The $CRKL^w(X,Y;t)$ is increasing (decreasing) in t, iff*

$$CRKL^w(X,Y;t) \geq (\leq) \psi(t),\ t \geq 0,$$

where $\psi(t) = \left(1 - \frac{\lambda_G(t)}{\lambda_F(t)}\right)(m_G^w(t) - m_F^w(t))$ and, $\lambda_F(t)$ and $\lambda_G(t)$ are the hrs of X and Y, respectively.

Proof. Differentiating $CRKL^w(X,Y;t)$ wrt t we have

$$\frac{d}{dt}CRKL^w(X,Y;t) = \lambda_F(t)CRKL^w(X,Y;t) + (\lambda_G(t) - \lambda_F(t))(m_G^w(t) - m_F^w(t)),\ t \geq 0,$$

which gives the result. □

Proposition 4.1.4. *If $X \overset{hr}{\leq} Y$ then $\psi(t)$ is increasing in t.*

Proof. If $X \overset{hr}{\leq} Y$ the $\lambda_F(t) \geq \lambda_G(t)$ so $\left(1 - \frac{\lambda_G(t)}{\lambda_F(t)}\right)$ is increasing in t. Again from Lemma 3.3.1, $(m_G^w(t) - m_F^w(t))$ is increasing in t. Hence the result. □

4.2 Weighted cumulative Kullback-Leibler information and its dynamic version

Now we propose WCKL and its dynamic version and study their properties. Let X and Y be two non-negative continuous rvs having cdfs F and G, respectively. Then, WCKL between X and Y can be defined as

$$\begin{aligned} CKL^w(X,Y) &= \int_0^{+\infty} xF(x)\left(\frac{G(x)}{F(x)} - \log\frac{G(x)}{F(x)} - 1\right)dx \\ &= \int_0^{+\infty} xF(x)\log\frac{F(x)}{G(x)}dx + \int_0^{+\infty} xG(x)dx - \int_0^{+\infty} xF(x)dx \\ &= \int_0^{+\infty} xF(x)\log\frac{F(x)}{G(x)}dx + \frac{1}{2}(E(X^2) - E(Y^2)). \end{aligned} \tag{4.3}$$

Note that, $CKL^w(X,Y) \geq 0$ and equality holds iff $F(x) = G(x)$, $\forall x$.

Remark 4.2.1. *For two continuous non-negative rvs X and Y*

$$CKL^w(X,Y) = CK^w(X,Y) - CE^w(X) + \frac{1}{2}(E(X^2) - E(Y^2)),$$

where $CK^w(X,Y) = -\int_0^{+\infty} xF(x)\log G(x)dx$ is the weighted cumulative inaccuracy measure (Daneshi et al., 2019) and $CE^w(X)$ is the WCE of X.

Example 4.2.1. *Suppose $X \sim U(0,1)$ and Y follows the power distribution with cdf $G(x) = x^c$, $0 < x < 1$, $c > 0$. Then WCKL between X and Y is $\frac{c-1}{9} + \frac{1}{c+2} - \frac{1}{3}$.*

Proposition 4.2.1. *If $X \overset{st}{\leq} Y$ then $CE^w(X) - \frac{1}{2}E(X^2) \leq CE^w(Y) - \frac{1}{2}E(Y^2)$.*

Proof. Proof follows proceeding similarly as in the proof of Proposition 4.1.1 and using the fact that $X \overset{st}{\leq} Y$ implies $CK^w(X,Y) \leq CE^w(Y)$. □

Proposition 4.2.2. *For two non-negative continuous rvs X and Y having cdfs F and G with common support* $[0,u]$,

$$CKL^w(X,Y) \geq \frac{1}{2}\left[(u^2-E(X^2))\log\frac{u^2-E(X^2)}{u^2-E(Y^2)}+E(X^2)-E(Y^2)\right].$$

Proof. Proof follows using log-sum inequality. □

Now we define the dynamic version of $CKL^w(X,Y)$ measure. Dynamic weighted cumulative Kullback-Leibler information between X and Y is the WCKL between past lifetimes ${}_tX=[t-X|X<t]$, $t>0$ and ${}_tY=[t-Y|Y<t]$, $t>0$. Dynamic $CKL^w(X,Y)$ is defined as

$$CKL^w(X,Y;t)=\int_0^t x\frac{F(x)}{F(t)}\log\frac{F(x)/F(t)}{G(x)/G(t)}dx+\mu_G^w(t)-\mu_F^w(t),$$

where $\mu_F^w(t)=\int_0^t x\frac{F(x)}{F(t)}dx$ and $\mu_G^w(t)=\int_0^t x\frac{G(x)}{G(t)}dx$ are the weighted mean past lifetimes (see, Misagh et al. (2011)) of X and Y, respectively.

Remark 4.2.2. *The following relation holds for* $CKL^w(X,Y;t)$*, analogous to dynamic WCRKL measure.*

$$CKL^w(X,Y;t)=CK^w(X,Y;t)-CE^w(X;t)+\mu_G^w(t)-\mu_F^w(t),$$

where $CK^w(X,Y;t)=\int_0^t x\frac{F(x)}{F(t)}\log\frac{G(x)}{G(t)}dx$*, is the dynamic weighted cumulative inaccuracy measure (Daneshi et al., 2019) and* $CE^w(X;t)=-\int_0^t\frac{F(x)}{F(t)}\log\frac{F(x)}{F(t)}dx$ *is the dynamic WCE measure.*

Proposition 4.2.3. *For two non-negative, continuous rvs X and Y,*

$$CKL^w(X,Y;t)\geq\mu_F^w(t)\log\frac{\mu_F^w(t)}{\mu_G^w(t)}+\mu_G^w(t)-\mu_F^w(t).$$

Proof. Proof follows along the same line as Proposition 4.1.3. □

In the following theorem we study the monotonicity property of $CKL^w(X,Y;t)$ measure.

Theorem 4.2.1. $CKL^w(X,Y;t)$ *is increasing (decreasing) in t, iff*

$$CKL^w(X,Y;t)\leq(\geq)\left[\frac{r_G(t)}{r_F(t)}-1\right]\left[\mu_F^w(t)-\mu_G^w(t)\right],\ t\geq 0. \tag{4.4}$$

$r_F(t)$ *and* $r_G(t)$ *are the rhr functions of X and Y, respectively.*

Proof. Differentiating $CKL^w(X,Y;t)$ with respect to t we have

$$\frac{d}{dt}CKL^w(X,Y;t)=(r_G(t)-r_F(t))(\mu_F^w(t)-\mu_G^w(t))-r_F(t)CKL^w(X,Y;t),\ t\geq 0.$$

Thus if $CKL^w(X,Y;t)$ is increasing (decreasing) in t iff (4.4) holds. □

In the following section we develope goodness-of-fit test for exponential distribution using WCRKL.

4.3 Exponentiality test for complete sample

Exponential distribution is the most popular distribution used in reliability. In life-testing, one of the important problem is to check whether a random sample comes from an exponential distribution or not. Here we develop a test statistic based on the WCRKL measure for testing exponentiality. Let F_θ be the cdf of X under consideration, and F_n, the edf of X. Then WCRKL between $F_n(x)$ and $F_\theta(x)$ can be obtained as

$$CRKL^w(F_n,F_\theta) = \int_0^{+\infty} xS_n(x)\log\frac{S_n(x)}{S_\theta(x)}dx + \int_0^{+\infty} xS_\theta(x)dx - \int_0^{+\infty} xS_n(x)dx,$$

where S_θ and S_n are the sf and empirical sf of X. Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from F_θ and $X_{1:n}, X_{2:n}, \cdots, X_{n:n}$ be the corresponding order statistics. Then WCRKL can be expressed as

$$\begin{aligned}
CRKL^w(F_n,F_\theta) &= \sum_{i=0}^{n-1}\int_{X_{i:n}}^{X_{i+1:n}} xS_n(x)\log\frac{S_n(x)}{S_\theta(x)}dx + \int_0^{X_{n:n}} xS_\theta(x)dx - \int_0^{X_{n:n}} xS_n(x)dx \\
&= \sum_{i=0}^{n-1}\frac{X_{i+1:n}^2 - X_{i:n}^2}{2}\left(\frac{n-i}{n}\right)\log\left(\frac{n-i}{n}\right) - \sum_{i=0}^{n-1}\frac{n-i}{n}\int_{X_{i:n}}^{X_{i+1:n}} x\log S_\theta(x)dx \\
&\quad + \int_0^{X_{n:n}} xS_\theta(x)dx - \frac{1}{2n}\sum_{i=1}^{n}X_i^2.
\end{aligned}$$

Now suppose $X_1, X_2, \cdots, X_n$ come from a non-negative continuous cdf F. Let $F_\theta(x) = 1 - e^{-\frac{x}{\theta}}$, $x > 0$, $\theta > 0$, be the cdf of an exponential distribution with parameter θ. We want to test the hypothesis

$$H_0 : F(x) = F_\theta(x) \quad vs. \quad H_1 : F(x) \neq F_\theta(x).$$

The WCRKL between F_θ and its edf F_n can be written as

$$\begin{aligned}CRKL^{w}(F_n,F_\theta) &= \sum_{i=0}^{n-1}\frac{X_{i+1:n}^2-X_{i:n}^2}{2}\left(\frac{n-i}{n}\right)\log\left(\frac{n-i}{n}\right)\\ &+\frac{1}{3\theta}\sum_{i=0}^{n-1}\frac{n-i}{n}(X_{i+1:n}^3-X_{i:n}^3)+\theta^2-\frac{1}{2n}\sum_{i=0}^{n-1}X_i^2. \qquad (4.5)\end{aligned}$$

We estimate θ by its mle $\hat{\theta}=\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$. The expression in (4.5) is not scale invariant. To obtain a scale invariant test statistic we divide equation (4.5) by $\int_0^{X_{n:n}}xS_n(x)dx$. So the scale invariant test statistic T is defined as

$$\begin{aligned}T &= \frac{1}{\int_0^{X_{n:n}}xS_n(x)dx}CRKL^{w}\left(F_n,F_{\hat{\theta}}\right)\\ &= \frac{n}{\sum_{i=1}^{n-1}X_i^2}\sum_{i=1}^{n-1}(X_{i+1:n}^2-X_{i:n}^2)\left(\frac{n-i}{n}\right)\log\left(\frac{n-i}{n}\right)\\ &+ \frac{2n}{3\bar{X}\sum_{i=1}^{n-1}X_i^2}\sum_{i=0}^{n-1}\frac{n-i}{n}(X_{i+1:n}^3-X_{i:n}^3)+\frac{2n\bar{X}^2}{\sum_{i=1}^{n}X_i^2}-1. \qquad (4.6)\end{aligned}$$

The null hypothesis is rejected for large values of T. Reject H_0 in favour of H_1 at significance level α, if $T>T_{\alpha,n}$, where $T_{\alpha,n}$ is the 100(1-α)% quantile point of the distribution of T. The sampling distribution of T is intractable, so we use Monte Carlo simulation to compute the critical points $T_{\alpha,n}$. We generate 10000 samples from standard exponential distribution for $n=8(1)20$, $25(5)50$ and $50(10)100$. The critical points of 1%, 5% and 10% are presented in Table 4.1.

The performance of T is compared with two entropy based test statistics proposed by Ebrahimi et al. (1992) and Baratpour and Rad (2012). Baratpour and Rad test statistic is given by

$$T^{*}=\frac{\sum_{i=1}^{n-1}\frac{n-i}{n}\log\left(\frac{n-i}{n}\right)\left(X_{(i+1):n}-X_{i:n}\right)+\frac{\sum_{i=1}^{n}X_i^2}{2\sum_{i=1}^{n}X_i}}{\frac{\sum_{i=1}^{n}X_i^2}{2\sum_{i=1}^{n}X_i}}$$

and Ebrahimi et al. (1992) test statistic is defined as

$$KL_{mn}=\exp(H_{mn}-\log\bar{X}-1),$$

where $H_{mn}=\frac{1}{n}\sum_{i=1}^{n}\log[\frac{n}{2m}(X_{(i+m):n}-X_{(i-m):n})]$ is the Vasicek's entropy estimate (Vasicek, 1976). The window size m is a positive integer less that $\frac{n}{2}$, $X_{i:n}=X_{1:n}$ if $i<1$ and $X_{i:n}=X_{n:n}$ if $i>n$. We reject the null hypothesis for large values of T^{*} and for small values of KL_{mn}.

Table 4.1: *Critical values of T.*

n	$T_{0.01,n}$	$T_{0.05,n}$	$T_{0.10,n}$	n	$T_{0.01,n}$	$T_{0.05,n}$	$T_{0.10,n}$
8	0.9985	0.8138	0.7056	20	0.7592	0.5160	0.4300
9	0.9765	0.7613	0.6718	25	0.6991	0.4577	0.3807
10	0.9270	0.7208	0.6244	30	0.6716	0.4129	0.3411
11	0.9110	0.7030	0.6035	35	0.6424	0.3869	0.3196
12	0.8941	0.6713	0.5715	40	0.6341	0.3554	0.2903
13	0.8497	0.6389	0.5502	45	0.6138	0.3361	0.2722
14	0.8442	0.6199	0.5255	50	0.5951	0.3113	0.2539
15	0.8312	0.5899	0.5024	60	0.5433	0.2840	0.2303
16	0.8175	0.5681	0.4827	70	0.4987	0.2634	0.2066
17	0.7979	0.5573	0.4702	80	0.4846	0.2401	0.1920
18	0.7919	0.5456	0.4564	90	0.4365	0.2261	0.1766
19	0.7806	0.5328	0.4486	100	0.4301	0.2089	0.1648

We calculate power of the tests based on 10000 samples of size n = 10, 15, 20 and 25 and with significance level $\alpha = 0.05$ and provide them in Tables 4.2 and 4.3. We consider monotone decreasing and increasing, and non-monotone hazard alternatives. The scale parameters are taken to be 1. The alternative distributions are taken as follows:

- Weibull distribution, denoted by WE(β), with pdf $f_{WE}(x) = \beta x^{\beta-1} e^{-x^{\beta}}$, $x, \beta > 0$.
- Gamma distribution, denoted by GA(β), with pdf $f_{GA}(x) = \frac{e^{-x} x^{\beta-1}}{\Gamma(q)}$, $x, \beta > 0$.
- Lognormal distribution. denoted by LN(β), with pdf $f_{LN}(x) = \frac{1}{\beta x \sqrt{2\pi}} e^{-\frac{\log^2 x}{2\beta^2}}$, $x, \beta > 0$.

For monotone decreasing hazard alternatives we take WE(0.5), GA(0.4) and LN(2). For monotone increasing hazard alternatives we consider WE(2), GA(2) and GA(3) and for non-monotone hazard alternatives we consider LN(0.6) and LN(1.2).

From Tables 4.2 and 4.3, it is observed that neither test dominates the others for all the alternatives. Test based on WCRKL performs better than the other two tests when alternative distribution has monotone decreasing hazard rate. All the tests attain the nominal significance level.

Table 4.2: *Power (%) of the test for* $\alpha = 0.05$.

n	Alternatives	T	$KL_{m,n}$	T^*	n	Alternatives	T	$KL_{m,n}$	T^*
10	WE(0.5)	32.59	11.26	17.53	15	WE(0.5)	50.06	24.52	37.80
	WE(2)	62.52	71.58	65.75		WE(2)	76.64	86.95	82.73
	GA(0.4)	16.56	5.30	4.81		GA(0.4)	26.91	11.60	16.07
	GA(2)	23.63	34.92	26.76		GA(2)	27.08	44.71	33.15
	GA(3)	47.15	65.29	50.81		GA(3)	57.79	80.27	65.99
	LN(0.6)	38.71	67.21	43.22		LN(0.6)	44.65	81.07	50.82
	LN(1.2)	14.21	5.20	7.68		LN(1.2)	22.89	6.01	15.39
	LN(2)	45.10	20.44	30.35		LN(2)	62.68	38.56	54.87
	Exp(1)	5.35	5.38	4.81		Exp(1)	5.04	5.09	4.82

Table 4.3: *Power (%) of the test for* $\alpha = 0.05$.

n	Alternatives	T	$KL_{m,n}$	T^*	n	Alternatives	T	$KL_{m,n}$	T^*
20	WE(0.5)	61.64	54.84	54.22	25	WE(0.5)	72.49	63.10	66.42
	WE(2)	85.18	93.68	91.66		WE(2)	91.63	97.25	96.79
	GA(0.4)	34.88	33.89	25.02		GA(0.4)	40.73	39.11	33.03
	GA(2)	30.81	52.41	38.93		GA(2)	34.42	58.82	44.56
	GA(3)	65.17	89.28	73.59		GA(3)	72.71	94.92	83.14
	LN(0.6)	47.09	90.86	57.56		LN(0.6)	50.47	95.37	64.56
	LN(1.2)	29.80	12.17	21.48		LN(1.2)	37.16	12.32	29.09
	LN(2)	77.58	66.02	71.98		LN(2)	86.10	74.28	82.53
	Exp(1)	4.93	5.37	4.72		Exp(1)	5.15	4.99	4.82

4.3.1 Data analysis

In this section three real data sets are analysed for illustrative purposes. It is shown how the proposed test works for these data sets. Exponential distribution provides good fit for the first two data sets. However, Chen distribution (Yousaf et al., 2019) provides better fit than exponential distribution for the third data set.

Data set 1

The data set is given in Grubbs (1971) that provides the mileages for 19 military personnel carriers that failed in service. The mileages are:
162, 200, 271, 320, 393, 508, 539, 629, 706, 778, 884, 1003, 1101, 1182, 1463, 1603, 1984, 2355, 2880.

Ebrahimi et al. (1992) showed that this data can be fitted well with exponential distribution. Here sample size is $n = 19$ and $T = 0.2916$. From Table 4.1 it is observed that the critical value at 5% significance level is 0.5328. Since $T < T_{0.05,19}$, we can not reject the null hypothesis at 5% significance level. Also the corresponding p-value is 0.3346, which is very high.

Data set 2

Here we consider a data set from Lawless (2011) that consists of failure times of 36 appliances. Baratpour and Rad (2012) studied this data set and found that exponential distribution fits the data reasonably well. The failure times are:
11, 35, 49, 170, 329, 381, 708, 958, 1062, 1167, 1594, 1925, 1990, 2223, 2327, 2400, 2451, 2471, 2551, 2565, 2568, 2694, 2702, 2761, 2831, 3034, 3059, 3112, 3214, 3478, 3504, 4329, 6367, 6976, 7846, 13403.
Here $n = 36$, $T = 0.1283$ and $T_{0.05,36} = 0.3741$. The corresponding p-value is 0.7041. So we can not reject the null hypothesis for this data as well.

Data set 3

Now we consider a data set represent quantity of 1000s of cycles to failure for electrical appliances. This data is from Lawless (2011). The observations are:
0.014, 0.034, 0.059, 0.061, 0.069, 0.080, 0.123, 0.142, 0.165, 0.210, 0.381, 0.464, 0.479, 0.556, 0.574, 0.839, 0.917, 0.969, 0.991, 1.064, 1.088, 1.091, 1.174, 1.270, 1.275, 1.355, 1.397, 1.477, 1.578, 1.649, 1.702, 1.893, 1.932, 2.001, 2.161, 2.292, 2.326, 2.337, 2.628, 2.785, 2.811, 2.886, 2.993, 3.122, 3.248, 3.715, 3.790, 3.857, 3.912, 4.100.
Yousaf et al. (2019) showed that Chen distribution fits this data better than exponential distribution. Xiong et al. (2022) found that for this data many popular tests like Kolmogorov-Smirnov, Kuiper, Cramer-von Misess, Anderson-Darling etc. can not reject the null hypothesis that the data follows exponential distribution. Here $n = 50$, $T = 0.3537$ and $T_{0.05,50} = 0.3133$. Also the corresponding p-value is 0.0344. So we can reject the null hypothesis for a 5% level of significance. So our proposed test can detect the difference between exponential and Chen distribution for a 5% significance level for this data set.

4.4 Testing exponentiality for censored data

Type-I and Type-II censoring are the two most common and widely used censoring schemes in life-testing. The simplicity as well as effectiveness of these schemes make them very popular among practitioners. In Type-I censoring, the experiment is continued until a prefixed time t_c and the number of failures is random. The Type-II censoring is continued until a pre-specified number of failures r is obtained. Here the duration of the experiment is random. The problem of testing exponentiality under Type-I and Type-II censored data are often encountered in life-testing. We propose a test for exponentiality using WCRKL based on censored data.

4.4.1 Exponentiality test under type-II censoring

Suppose n units are put on a Type-II censored experiment and $X_{1:n}, X_{2:n}, \cdots, X_{r:n}$ be the Type-II censored sample. We want to test the hypothesis

$$H_0 : F(x) = 1 - e^{-\frac{x}{\theta}},\ x, \theta > 0 \quad vs. \quad H_1 : F(x) \neq 1 - e^{-\frac{x}{\theta}},\ x, \theta > 0.$$

The WCRKL between F_n and $F_{\hat{\theta}}$, where $\hat{\theta}$ is the mle of θ for Type-II censored sample, can be written as

$$\begin{aligned} CRKL_C^w(F_n, F_{\hat{\theta}}) &= \int_0^{X_{r:n}} xS_n(x) \log \frac{S_n(x)}{S_{\hat{\theta}}(x)} dx + \int_0^{X_{r:n}} xS_{\hat{\theta}}(x) dx - \int_0^{X_{r:n}} xS_n(x) dx \\ &= \sum_{i=1}^{r-1} \frac{X_{i+1:n}^2 - X_{i:n}^2}{2} \left(\frac{n-i}{n}\right) \log \frac{n-i}{n} + \frac{1}{\hat{\theta}} \sum_{i=1}^{r-1} (X_{i+1:n}^3 - X_{i:n}^3) \frac{n-i}{n} \\ &+ \hat{\theta}^2 \left(1 - e^{-\frac{X_{r:n}}{\hat{\theta}}}\right) - \hat{\theta} X_{r:n} e^{-\frac{X_{r:n}}{\hat{\theta}}} - \frac{\sum_{i=1}^r X_{i:n}^2 + (n-r) X_{r:n}^2}{2n}. \end{aligned} \tag{4.7}$$

We use the scale invariant test statistic $T_C = \frac{1}{\int_0^{X_{r:n}} xS_n(x)dx} CRKL_C^w(F_n, F_{\hat{\theta}})$ to perform goodness-of-fit tests. We generate 10000 Type-II censored samples from standard exponential distribution for various values of n and r and compute T_C. We reject the null hypothesis for large values of T_C. We calculate the power of the test when alternatives are Weibull, gamma and log-normal. It is observed from simulation that the proposed test is suitable for monotone decreasing hazard alternatives. We compare the power of the test with tests based on KL and CRKL. The test based on KL divergence was proposed by Park (2005) as

$$T_{KL} = \frac{r}{n} \log(\hat{\theta} + 1) - \frac{1}{n} \sum_{i=1}^{r} \log\left(\frac{n}{2m}(X_{i+m:n} - X_{i-m:n})\right) + \left(1 - \frac{r}{n}\right) \log\left(1 - \frac{r}{n}\right).$$

The CRKL based test was developed by Park and Lim (2015) as

$$\begin{aligned} T_{CRKL} &= \frac{1}{\hat{\theta}} \sum_{i=0}^{r-1} (X_{i+1:n} - X_{i:n}) \frac{n-i}{r} \log \frac{n-i}{n} + \frac{1}{2\hat{\theta}^2} \sum_{i=0}^{r-1} \frac{n-i}{r} (X_{i+1:n}^2 - X_{i:n}^2) \\ &+ \frac{n}{r}\left(1 - exp\left(-\frac{X_{r:n}}{\hat{\theta}}\right)\right) - 1. \end{aligned}$$

The results corresponding to sample sizes $n = 10$, 20 and 30 are provided in Tables 4.4, 4.5 and 4.6, respectively. From the tables it is observed that the proposed test performs well than the other two tests. As r increases (number of censored items reduces), performance of T_C and T_{CRKL} become similar.

Table 4.4: *Power of the test when n = 10.*

Alternatives	$r = 4$			$r = 6$			$r = 8$		
	T_C	CRKL	KL	T_C	CRKL	KL	T_C	CRKL	KL
WE(0.5)	30.26	8.37	3.18	40.06	27.18	3.28	49.04	49.57	3.20
WE(0.8)	10.10	4.12	3.43	12.22	7.21	2.47	14.71	13.17	1.70
GA(0.5)	23.75	6.62	3.05	28.04	17.74	2.04	30.11	29.55	1.15
LN(2)	13.22	3.64	3.21	25.67	15.11	2.19	44.20	42.36	2.92
EXP(1)	5.05	5.05	4.89	4.79	4.90	4.89	5.11	4.94	4.85

Table 4.5: *Power of the test when n = 20.*

Alternatives	$r = 8$			$r = 12$			$r = 16$		
	T_C	CRKL	KL	T_C	CRKL	KL	T_C	CRKL	KL
WE(0.5)	49.05	25.49	3.66	62.47	51.86	18.43	73.61	49.57	31.01
WE(0.8)	13.62	5.36	2.23	17.26	10.36	2.57	20.29	17.87	2.06
GA(0.5)	38.56	17.90	2.26	44.61	33.17	8.60	45.97	45.66	8.87
LN(2)	20.15	7.53	2.16	42.39	29.50	7.88	69.10	66.22	23.36
EXP(1)	4.89	5.08	4.92	4.85	5.25	5.09	5.02	4.94	5.02

Table 4.6: *Power of the test when n = 30.*

Alternatives	$r = 12$			$r = 18$			$r = 24$		
	T_C	CRKL	KL	T_C	CRKL	KL	T_C	CRKL	KL
WE(0.5)	65.43	41.53	17.40	79.20	69.27	37.67	87.78	87.62	66.36
WE(0.8)	18.80	7.46	2.70	22.44	12.90	3.30	27.49	22.02	5.07
GA(0.5)	51.21	30.31	9.76	58.22	46.96	18.17	60.20	59.50	30.55
LN(2)	27.84	11.11	4.44	56.18	42.47	16.27	83.95	80.24	49.98
EXP(1)	5.08	5.04	4.78	4.96	4.98	5.90	5.09	4.99	5.01

Data Analysis

Consider Data set 2 where 36 units are put into test and assume that only first 20 failures are observed and rest are censored. So this becomes a Type-II censored sample with $n = 36$ and $r = 20$. The test statistic $T_C = 0.96085$ and the p-value is 0.6904. So we can not reject the null hypothesis.

4.4.2 Exponentiality test under type-I censoring

Suppose n units are put on a life-testing experiment which is terminated after time t_c. Let $X_{1:n}, X_{2:n}, \cdots, X_{d:n}$ be the failure times from the Type-I censored experiment. We want to test

$$H_0 : F(x) = 1 - e^{-\frac{x}{\theta}},\ x, \theta > 0 \quad vs. \quad H_1 : F(x) \neq 1 - e^{-\frac{x}{\theta}},\ x, \theta > 0.$$

Using WCRKL, we can test the hypothesis under type-I censored data by following the procedure suggested by Pakyari and Balakrishnan (2013). They used the idea that conditional on $D = d$, $(X_{1:n}, \cdots, X_{d:n}) = (Z_{1:n}, \cdots, Z_{d:d})$, where $(Z_{1:n}, \cdots, Z_{d:d})$ can be treated as an iid sample of size d from a scaled exponential distribution truncated at time t_c, see Arnold et al. (2008) and David and Nagaraja (2004). The mle of θ under Type-I censoring is $\hat{\theta} = \frac{\sum_{i=1}^{d} X_{i:n} + (n-d)t_c}{d}$. Now according to Pakyari and Balakrishnan (2013), we can use the following transformation to uniformity as

$$U_{i:d} = \frac{1 - e^{-\frac{X_{1:n}}{\hat{\theta}}}}{1 - e^{-\frac{t_c}{\hat{\theta}}}}, \quad i = 1, 2, \cdots, d.$$

Note that, $U_{1:d}, \cdots, U_{d:d}$ can be treated as iid standard uniform observations and based on this sample one can perform edf based goodness-of-fit tests. By using this approach, the goodness-of-fit tests can be performed under Type-I censoring based on WCRKL. So the test of exponentiality under Type-I censoring has now been reduced to a test of uniformity under a complete sample of size d. The test statistic is

$$\begin{aligned} T_C^1 &= \sum_{i=1}^{d-1} \frac{U_{i+1:d}^2 - U_{i:d}^2}{2} \left(\frac{d-i}{d}\right) \log\left(\frac{d-i}{d}\right) + \frac{U_{d:d}^2}{2} - \frac{U_{d:d}^3}{3} \\ &\quad - \sum_{i=0}^{d-1} \frac{d-i}{d} \int_{U_{i:d}}^{U_{i+1:d}} u \log(1-u)\,du - \frac{1}{2d} \sum_{i=1}^{d} U_{i:d}^2, \end{aligned}$$

where $\int u \log(1-u)du = \frac{u^2-1}{2}\log(1-u) - \frac{u^2}{4} - \frac{u}{2}$. Reject the null hypothesis at significance level α if $T_C^1 > \tau_{\alpha,n}$ where $\tau_{\alpha,n}$ is the 100(1-α)% quantile point of the distribution of T_C^1. The quantile points are generated by simulation. We calculate the power of the test at 5% level of significance with Weibull, gamma and generalized exponential alternatives having decreasing hr function. Generalized exponential distribution has the cdf $F_{GE}(x) = \left(1 - e^{-(\lambda x)}\right)^{\beta}$, $x, \beta, \lambda > 0$. We will denote it by GE(β, λ). We generate 10000 samples from standard exponential distribution with sample size 20 and 30 for various proportions of failures and compute the power. The results are provided in Table 4.7. From the table it is observed that the proposed test attains the specified level of significance. Power increases when the proportions of failures and sample size increases.

Table 4.7: *Power of the test with various proportions of failures.*

n	Alternatives	Proportions of failures					
		0.5	0.6	0.7	0.8	0.9	0.95
20	WE(0.5)	21.72	28.63	35.60	44.35	54.54	63.04
	GA(0.3)	19.84	21.72	22.51	27.06	37.47	51.77
	GE(0.3)	23.51	25.82	26.79	30.15	37.56	50.86
	Exp(1)	5.24	5.26	5.06	5.02	4.92	5.05
30	WE(0.5)	37.77	47.66	50.84	59.82	71.00	78.06
	GA(0.3)	28.42	30.87	32.12	40.11	57.40	70.41
	GE(0.3)	34.32	37.25	37.58	42.81	55.16	68.97
	Exp(1)	4.92	5.05	4.91	5.02	5.15	5.01

Data Analysis

For illustrative purposes, we analyse a real data set consisting of breakdown times of an insulating fluid (in minutes) tested at 34KVs given in Nelson (2003). The breakdown times are: 0.19, 0.78, 0.96, 1.31, 2.78, 3.16, 4.15, 4.67, 4.85, 6.50, 7.35, 8.01, 8.27, 12.06, 31.75*, 32.52*, 33.91*, 36.71*, 72.89*.

In this experiment, test termination time is $t_c = 15$, number of units at risk is $n = 19$ and number of failures is $d = 14$. Number of censored observations are 5. The censored observations are marked by an asterisk. The mle of exponential parameter θ is $\hat{\theta} = 10.003$. The observed value of the test statistics is $T_C^1 = 0.009$ and the p-value is 0.3015. The p-value is very high so we can not reject the null hypothesis.

4.5 Discussions

In this chapter, we proposed weighted cumulative residual Kullback-Leibler information measure as an alternative measure of closeness between two lifetime models. We introduced the dynamic version as well and studied its bound and monotonicity properties. Also we introduced weighted cumulative Kullback-Leibler information and its dynamic version. Based on weighted cumulative residual Kullback-Leibler information measure, goodness-of-fit tests are developed for exponential distribution under complete, Type-I and Type-II censored samples. The proposed tests perform well. The tests perform better for censored data when the alternative distributions have decreasing hr function.

These tests can be modified to perform goodness-of-fit tests for other models like Weibull, log-logistic, inverse Rayleigh etc. The proposed test for Type-II censored data can be extended to progressive Type-II censoring as well. Using the dynamic version, a goodness-of-fit test for exponential distribution in residual lifetime can be developed. More work is needed in this direction.

Chapter 5

Analysis and applications of cumulative residual extropy in system reliability and hypothesis testing problem

CUMULATIVE residual extropy (CREx) measure has been introduced recently by Jahanshahi et al. (2020). For a non-negative continuous rv X, CREx is defined as

$$\xi J(X) = -\frac{1}{2}\int_0^{+\infty} S^2(x)dx.$$

For two rvs X and Y, $\xi J(X) < \xi J(Y)$ means that X is more uncertain than Y. Note that for non-negative iid rvs, CREx can be expressed in terms of the expectation of $X_{1:2} = \min\{X_1, X_2\}$ which has sf $S^2(x)$ i.e.

$$E(X_{1:2}) = \int_0^{+\infty} S^2(x)dx = -2CREx.$$

Jahanshahi et al. (2020) studied various properties of CREx and they provided applications of CREx measure as a risk measure and also developed a test of independence between two rvs using conditional CREx measure. They showed that the CREx measure has a lot of potential in actuarial science due to its relationship with Gini's coefficient. Gini's coefficient is widely used in economics and actuarial science. It is effectively used as a risk measure in finance. The Gini's coefficient between two iid rvs X and Y is defined as

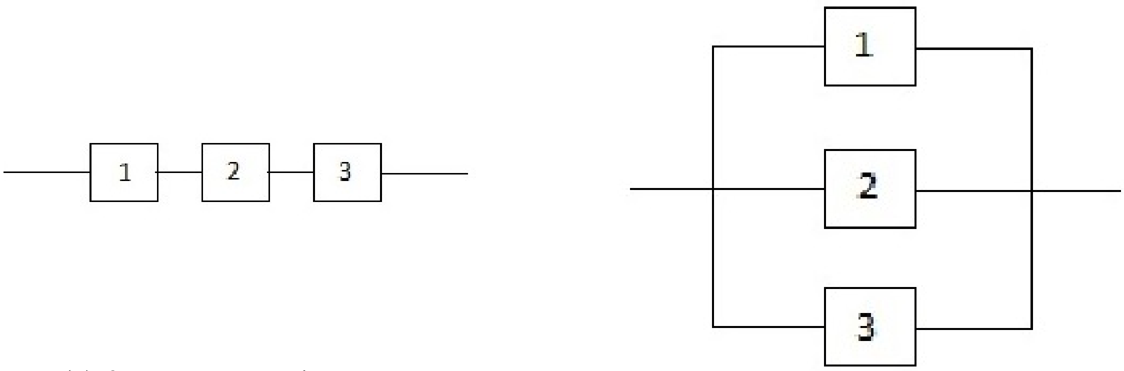
$$G(X) = \frac{E(|X-Y|)}{E(X+Y)} = 1 - \frac{\int_0^{+\infty} S^2(x)dx}{E(X)} = 1 + \frac{2\xi J(X)}{E(X)}.$$

So one can estimate Gini's coefficient from the estimate of CREx measure and vice-versa. In this Chapter we discuss some applications of CREx in reliability engineering and statistics.

Information measures have many applications in system reliability for comparing between systems, measuring system complexity and reliability optimization problems. The problem of comparing different coherent and mixed systems has emerged in recent years after the introduction of signature representation of systems, following the works of Samaniego (2007). A system is said to be coherent if all the components are relevant to the system and the structure function of the system is monotone (Barlow and Proschan, 1975). A k-out-of-n system is a coherent system which fails when the k component fails. This system is also referred to as the k-out-of-n:F system. Note that there is another system denoted as the k-out-of-n:G system, which functions as long as k components are working (good). In this study, by k-out-of-n system we mean k-out-of-n:F system. A mixed system is a stochastic mixture of coherent systems i.e. a mixed system is constructed by selecting a number of coherent systems according to a known probability distribution. So any coherent system is a special case of a mixed system. The practical implementation of a mixed system is randomization of a coherent system in which an experimenter selects a coherent system according to a mixture probability distribution and uses that coherent system thereafter. Boland and Samaniego (2004) showed that the mixture of coherent systems can always be represented as mixture of k-out-of-n systems.

The concept of system signature opens the way of analysing information properties of coherent and mixed systems consisting of iid components and various comparisons of systems are studied using system information measures. Entropy, extropy, CRE and other generalized measures are used for this purpose. Toomaj and Doostparast (2014, 2016) first analyzed entropy and KL divergence for the lifetime of a mixed system when the lifetimes of components are iid. Toomaj et al. (2017) studied CRE of coherent and mixed systems and Qiu et al. (2019) analyzed extropy for mixed reliability systems. For related works on analysis of different information measures of mixed systems, see for example, Kayal (2019), Rahimi et al. (2020), Toomaj (2017), Toomaj et al. (2021, 2018) and the references therein. These works mainly considered the study of various density based entropy measures in the context of system reliability. The component lifetimes are assumed to be iid. Comparison results between systems are developed in terms of the respective information measures and numerous bounds are obtained. However, the authors did not consider dependency among components. Toomaj (2017) first considered properties of information measure for series and parallel systems with dependent components. The series and parallel systems also belong to the class of k-out-of-n systems. We provide the structure of a 3-component series and parallel system in Figures 5.1a and 5.1b, respectively. An n-component series system fails when any one component fails and an n-component parallel system fails when all the components fail. A series system is the least reliable system and a parallel system is the most reliable system. Entropy of coherent and mixed systems consisting of dependent and identically distributed (d.i.d.) components are analyzed by Toomaj et al. (2017). Recently, Qiu et al. (2019) obtained properties of extropy for mixed systems with iid components.

Fig. 5.1: *Structure function of series and parallel systems.*



(a) *3-component series system*

(b) *3-component parallel system*

To the best of our knowledge, there is no work on CREx for mixed systems. In this work, we express the CREx measure for a coherent and mixed system consisting of iid as well as d.i.d. components using system signature and distortion functions. We obtain some comparison results between systems having the same structure but different components. We provide some bounds for CREx of coherent and mixed systems which is useful for highly complex systems when exact CREx can not be computed due to the complicated structure of the systems. Our study makes two contributions in the literature of reliability engineering. We introduced a new divergence measure, called Jensen-Cumulative residual extropy divergence, to measure the complexity of mixed systems having iid components. Complexity of a system is measured with respect to the k-out-of-n systems (least complex systems). Using the proposed divergence measure we compute how much more complex an n-component system is than the k-out-of-n systems consisting of the same number of iid components. Another major finding of the work is a new discrimination measure for comparing two systems. For this purpose, first we introduce a relative CREx measure between two rvs X and Y by replacing the pdfs in the relative extropy measure of Lad et al. (2015) with the sfs of the rvs. Using the relative CREx measure, we develop a discrimination measure that calculates how close (far) a system is towards a series or parallel system having the same components.

A system is good (bad) if it is close to a parallel (series) system. Also we provide some potential applications of these proposed measures in systems involving redundancy.

We used the CREx measure for testing equality between two distribution functions. We propose three non-parametric estimators of CREx measure, study their asymptotic properties and compare their performances by evaluating their MSE. Using asymptotic normality of one of the proposed estimators, we develop a test for equality between two distributions.

The rest of the chapter is organized as follows. We obtain CREx of mixed systems having iid components and obtain stochastic ordering results and some bounds in Section 5.1. The CREx for d.i.d. components are discussed in Section 5.2. We discuss the applications in system reliability in Section 5.3. We consider estimations of CREx measure in Section 5.4. We construct an equality test between two distributions in Section 5.5. Finally, we conclude the Chapter in Section 5.6.

5.1 CREx of mixed systems consisting of iid components

In this section, we obtain expression of CREx measure for coherent and mixed reliability systems and study several properties. The CREx defined in Eq. (1.13) can be expressed as

$$\begin{aligned} \xi J(X) &= -\frac{1}{2}\int_0^{+\infty} S^2(x)dx \\ &= -\frac{1}{2}\int_0^1 \frac{u^2}{f(S^{-1}(u))}du. \end{aligned} \tag{5.1}$$

Note that $S^{-1}(u) = \sup\{x : S(x) \geq u\}$ is called the quantile function (qf) of $S = 1-F$. Now, consider a mixed system constructed by n iid components having lifetimes $X_1, X_2, \cdots, X_n$ with common cdf F and sf S. Suppose T represents the lifetime of the system. Then the sf of T is given by (Samaniego, 2007)

$$S_T(t) = \sum_{i=1}^{n} s_i S_{i:n}(t), \tag{5.2}$$

where $S_{i:n}(t) = \sum_{j=0}^{i-1}\binom{n}{j}[F(t)]^j[S(t)]^{n-j}$, $\forall i = 1(1)n$ are the sfs of the ordered component lifetimes $X_{1:n}, X_{2:n}, \cdots, X_{n:n}$, and $s_i = P(T = X_{i:n})$ is the probability that the system fails due to the failure of i-th component, $i = 1, 2, \cdots, n$, $\sum_{i=1}^{n} s_i = 1$ and $\mathbf{s} = (s_1, s_2, \cdots, s_n)$ is known as system signature. Note that, the k-th order statistics $X_{k:n}$ denotes the lifetime of a $(n-k+1)$-out-of-n system.

Now consider the transformation $U = S(X)$. Then for $i = 1(1)n$ we have,

$U_i = S(X_i) \sim U(0,1)$ and $W_{i:n} = S(X_{i:n}) \sim$Beta $(n-i+1, i)$ with the cdf

$$G_{i:n}(w) = \sum_{j=0}^{i-1} \binom{n}{j} (1-w)^j w^{n-j}; \quad 0 \le w \le 1.$$

To obtain the CREx of T, we consider the transformation $V = S(T)$. Then the cdf of V is given by

$$G_V(v) = \sum_{i=1}^{n} s_i G_{i:n}(v); \quad 0 \le v \le 1.$$

Using $S_T(t) = G_V(S(t))$ in Eq. (5.1), the CREx of a mixed system is expressed as

$$\begin{aligned} \xi J(T) &= -\frac{1}{2}\int_0^{+\infty} S_T^2(t)dt \\ &= -\frac{1}{2}\int_0^1 \frac{G_V^2(v)}{f(S^{-1}(v))}dv. \end{aligned} \quad (5.3)$$

We consider some examples for illustration.

Example 5.1.1. *Suppose the coherent system with lifetime* $T = \max\{\min\{X_1, X_2\}, \min\{X_3, X_4\}\}$ *has iid components with common sf* $S(t) = \exp(-\frac{t}{\lambda})$*;* $\lambda > 0$*,* $t \ge 0$*. This is a parallel-series system where two 2-component series systems are put in parallel. For exponential distribution,* $f(S^{-1}(v)) = \frac{v}{\lambda}$*. The signature vector for* T *is* $\mathbf{s} = \left(0, \frac{2}{3}, \frac{1}{3}, 0\right)$*. The CREx of the system is obtained as*

$$\xi J(T) = -\frac{1}{2}\lambda \int_0^1 \frac{(2v^2 - v^4)^2}{v} dv = -0.2292\lambda.$$

Example 5.1.2. *Suppose the mixed system with signature* $\mathbf{s} = \left(\frac{1}{2}, 0, \cdots, 0, \frac{1}{2}\right)$ *has n standard exponentially distributed components. This is the uniform mixture of n-component series and parallel systems. This system is constructed by selecting n-component series and parallel systems with probability* $\frac{1}{2}$*. The system cdf is*

$$G_V(v) = \frac{1}{2}[1 + v^n - (1-v)^n]; \; 0 \le v \le 1,$$

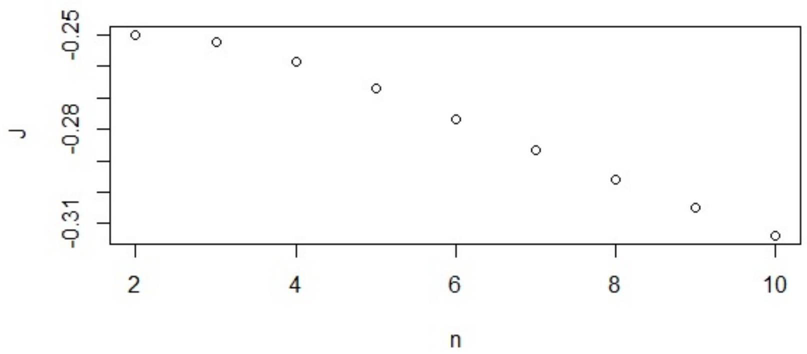
and the CREx of the system is given by

$$\xi J(T) = -\frac{1}{8}\int_0^1 \frac{[1 + v^n - (1-v)^n]^2}{v} dv.$$

We plot the CREx of T for different values of n in Figure 5.2 and find that the CREx of the system decreases as the number of components increases. This is natural, since the value of the information measure increases if the number of elements increases.

Now suppose T^X and T^Y are the lifetimes of two mixed systems having the same structure (signature) and n iid components with cdfs F and G, respectively. The component

Fig. 5.2: *Values of CREx of the system in Example 5.1.2 for* $n = 2(1)10$.



lifetimes of T^X and T^Y are $(X_1, X_2, \cdots, X_n)$ and $(Y_1, Y_2, \cdots, Y_n)$, respectively. Since the components of each systems are identically distributed, we use the notation $X \overset{disp}{\leq} Y$ which implies that the components lifetimes of T^X are smaller than that of T^Y in dispersive ordering. In the following proposition, we will compare CREx of two systems having the same structure but different components when the component lifetimes maintain dispersive ordering.

Proposition 5.1.1. *Let T^X and T^Y be the lifetimes of two mixed systems having same signature and n iid components with cdfs F and G, respectively. If $X \overset{disp}{\leq} Y$ then,*

$$\xi J(T^X) \geq \xi J(T^Y).$$

Proof. Proof follows applying the definition of dispersive order in Eq. (5.1). □

In the next proposition, we compare two systems but relax the dispersive ordering assumption among component lifetimes. Here we only assume that $\xi J(X) \geq \xi J(Y)$ i.e. the CREx of the components of the first system is higher than that of the second system.

Proposition 5.1.2. *Let $D(v) = \frac{v^2}{g(S_2^{-1}(v))} - \frac{v^2}{f(S_1^{-1}(v))}$, $A_1 = \left\{v \in [0,1] : f(S_1^{-1}(v)) > g(S_2^{-1}(v))\right\}$ and $A_2 = \left\{v \in [0,1] : f(S_1^{-1}(v)) \leq g(S_2^{-1}(v))\right\}$. If $\inf_{v \in A_1} \frac{G_V^2(v)}{v^2} \geq \sup_{v \in A_2} \frac{G_V^2(v)}{v^2}$ and $\xi J(X) \geq \xi J(Y)$ then, $\xi J(T^X) \geq \xi J(T^Y)$.*

Proof. From Eq. (5.1) we have,

$$\xi J(X) \geq \xi J(Y) \quad \Rightarrow \quad \int_0^1 D(v)dv \geq 0.$$

Now using Eq. (5.3) we can write

$$\begin{aligned}
2[\xi J(T^X) - \xi J(T^Y)] &= \int_0^1 \frac{G_V^2(v)}{v^2} D(v)dv \\
&= \int_{A_1} \frac{G_V^2(v)}{v^2} D(v)dv + \int_{A_2} \frac{G_V^2(v)}{v^2} D(v)dv. \\
&\geq \inf_{v\in A_1} \frac{G_V^2(v)}{v^2} \int_{A_1} D(v)dv + \sup_{v\in A_2} \frac{G_V^2(v)}{v^2} \int_{A_2} D(v)dv \\
&\geq \sup_{v\in A_2} \frac{G_V^2(v)}{v^2} \int_{A_1} D(v)dv + \sup_{v\in A_2} \frac{G_V^2(v)}{v^2} \int_{A_2} D(v)dv \\
&= \sup_{v\in A_2} \frac{G_V^2(v)}{v^2} \int_0^1 D(v)dv \geq 0.
\end{aligned}$$

The inequalities follows since

$$\int_{A_1} \frac{G_V^2(v)}{v^2} D(v)dv \geq \inf_{v\in A_1} \frac{G_V^2(v)}{v^2} \int_{A_1} D(v)dv$$

and as $D(v) < 0$ in A_2

$$\int_{A_1} \frac{G_V^2(v)}{v^2} D(v)dv \geq \sup_{v\in A_2} \frac{G_V^2(v)}{v^2} \int_{A_2} D(v)dv.$$

Hence the proof. □

Next we provide some bounds of CREx of a mixed system. These will be useful to approximate system CREx in situations when it can not be calculated due to the highly complex nature of the structure function.

Proposition 5.1.3. *Let T be the lifetime of a mixed system with signature* **s** *having n iid components. Let $X_1, X_2, \cdots, X_n$ be the component lifetimes with common sf F. Then*

$$B_1 \xi J(X) \leq \xi J(T) \leq B_2 \xi J(X),$$

where $B_1 = \sup_{v\in[0,1]} \frac{G_V^2(v)}{v^2}$ and $B_2 = \inf_{v\in[0,1]} \frac{G_V^2(v)}{v^2}$.

Proof. We have,

$$\begin{aligned}
\xi J(T) &= -\frac{1}{2} \int_0^1 \frac{G_V^2(v)}{v^2} \frac{v^2}{f(S^{-1}(v))} dv \geq - \sup_{v\in[0,1]} \frac{G_V^2(v)}{v^2} \int_0^1 \frac{v^2}{f(S^{-1}(v))} dv \\
&= B_1 \xi J(X).
\end{aligned}$$

Upper bound can be obtained similarly. □

Proposition 5.1.4. *Suppose T denote the lifetime of a mixed system of n iid components having lifetimes with common pdf f. Then*

$$-\frac{1}{2m}\int_0^1 G_V^2(v)dv \le \xi J(T) \le -\frac{1}{2M}\int_0^1 G_V^2(v)dv,$$

where $m = \inf_{x\in A} f(x)$, $M = \sup_{x\in A} f(x)$ and A is the support of f.

Proof. Note that $m \le f(S^{-1}(v)) \le M$. Now

$$\xi J(T) = -\frac{1}{2}\int_0^1 \frac{G_V^2(v)}{f(S^{-1}(v))}dv \le -\frac{1}{2M}\int_0^1 G_V^2(v)dv.$$

Lower bound can be obtained similarly. □

Example 5.1.3. *For the system considered in Example 5.1.1, $m = 0$ and $M = \frac{1}{\lambda}$. From Proposition 5.1.4 we have, $\xi J(T) \le -0.1698\lambda$. Also here $B_1 = 1.1851$ and $B_2 = 0$. So from Proposition 5.1.3 we get, $-0.2963\lambda \le J_S(T) \le 0$. The exact value of $J_S(T)$ is -0.2292λ and $J_S(X) = -0.25\lambda$.*

We compute the CREx and its upper bound using Proposition 5.1.2 for a set of systems with four iid components having standard exponentially distributed lifetimes. This set of systems was studied by Shaked and Suarez-Llorens (2003). We present the CREx and its upper bound in Table 5.1. It is observed that the CREx of the parallel system is minimum but maximum for the series system. It is natural since the parallel (series) system is the most (least) reliable system so the CREx of parallel (series) system is minimum (maximum). So if CREx of a system with lifetime T_1 is less than that of a system with lifetime T_2 then T_1 is more close towards a parallel system than T_2. From the table we find that the CREx of the 3-out-of-4 system is smaller than that of the 2-out-of-4 system so the 3-out-of-4 system is better than the 2-out-of-4 system. Since, lower value of CREx implies better performance of systems so we can determine which systems are more reliable by comparing their CREx measure.

Next we provide an important lower bound for $\xi J(T)$ in terms of CREx of k-out-of-n systems. Later we will use this result to study the system complexity.

Proposition 5.1.5. *Consider a mixed system of n iid components having lifetimes $X_1, X_2, \cdots, X_n$. Then the CREx of the system lifetime T satisfies $\xi J(T) \ge \sum_{j=1}^n s_j \xi J(X_{j:n})$, where $X_{1:n}, \cdots, X_{n:n}$ are the ordered lifetimes of the components. Equality holds for k-out-of-n systems.*

Proof. Applying Jensen's inequality in Eq. (5.3) we get,

$$\xi J(T) \ \ge\ -\frac{1}{2}\int_0^1 \frac{\sum_{j=1}^n s_j G_{j:n}^2(v)}{f(S^{-1}(v))}dv = \sum_{j=1}^n s_j \xi J(X_{j:n}).$$

Equality holds for a k-out-of-n systems since $s_j = 0$ for all $j \neq n-k+1$ and $s_j = 1$ for $j = n-k+1$. □

Table 5.1: *CREx and its upper bound for coherent systems.*

Systems	**s**	$J_S(T)$	Upper bound
Group 1			
1) $X_{1:4}$ (Series)	(1,0,0,0)	-0.0625	-0.0556
2) $X_{4:4}$ (Parallel)	(0,0,0,1)	-0.7244	-0.3556
3) $X_{2:4}$ (3-out-of-4)	(0,0,1,0)	-0.3673	-0.2429
4) $X_{3:4}$ (2-out-of-4)	(0,1,0,0)	-0.1815	-0.1429
Group 2			
5) $\max\{X_{1:2}, \min\{X_1,X_3,X_4\}, \min\{X_2,X_3,X_4\}\}$	$(0,\frac{5}{6},\frac{1}{6},0)$	-0.2036	-0.1556
6) $\min\{X_{2:2}, \max\{X_1,X_3,X_4\}, \max\{X_2,X_3,X_4\}\}$	$(0,\frac{1}{6},\frac{5}{6},0)$	-0.3274	-0.2223
Group 3			
7) $\max\{X_{1:2}, \min\{X_3,X_4\}\}$	$(0,\frac{2}{3},\frac{1}{3},0)$	-0.2292	-0.1698
8) $\max\{X_{1:2}, \min\{X_1,X_3\}, \min\{X_2,X_3,X_4\}\}$	$(0,\frac{2}{3},\frac{1}{3},0)$	-0.2292	-0.1698
9) $\max\{X_{2:2}, \max\{X_3,X_4\}\}$	$(0,\frac{1}{3},\frac{2}{3},0)$	-0.2911	-0.2032
10) $\max\{\min\{X_1, \max\{X_2,X_3,X_4\}\}, \min\{X_2,X_3\}\}$	$(0,\frac{1}{3},\frac{2}{3},0)$	-0.2911	-0.2032
Group 4			
11) $\max\{X_{1:2}, \min\{X_2,X_3\}, \min\{X_3,X_4\}\}$	$(0,\frac{1}{2},\frac{1}{2},0)$	-0.2583	-0.1857
12) $\max\{\min\{X_1, \max\{X_2,X_3,X_4\}\}, \min\{X_2,X_3,X_4\}\}$	$(0,\frac{1}{3},\frac{2}{3},0)$	-0.2911	-0.2032
Group 5			
13) $\min\{X_{1:3}, X_4\}$	$(\frac{1}{4},\frac{3}{4},0,0)$	-0.1429	-0.1151
14) $\max\{X_{1:3}, X_4\}$	$(0,0,\frac{1}{4},\frac{3}{4})$	-0.6083	-0.3214
Group 6			
15) $\max\{X_{1:3}, \min\{X_2,X_3,X_4\}\}$	$(\frac{1}{2},\frac{1}{2},0,0)$	-0.1101	-0.0913
16) $\max\{X_1,X_2, \min\{X_3,X_4\}\}$	$(0,0,\frac{1}{2},\frac{1}{2})$	-0.5101	-0.2913
Group 7			
17) $\min\{X_{2:2}, \max\{X_1,X_3\}, X_4\}$	$(\frac{3}{12},\frac{7}{12},\frac{2}{12},0)$	-0.1613	-0.1258
18) $\max\{X_1, \min\{X_2,X_4\}, \min\{X_3,X_4\}\}$	$(0,\frac{2}{12},\frac{7}{12},\frac{3}{12})$	-0.3839	-0.2425
Group 8			
19) $\min\{X_{3:3}, X_4\}$	$(\frac{1}{4},\frac{1}{4},\frac{2}{4},0)$	-0.2089	-0.1520
20) $\max\{X_1, \min\{X_2,X_3,X_4\}\}$	$(0,\frac{2}{4},\frac{1}{4},\frac{1}{4})$	-0.3030	-0.2020

We provide bounds for system CREx in terms of CREx of k-out-of-n systems for the same set of systems discussed in Table 5.1. The set of systems have four iid components with standard exponential lifetimes. From Tables 5.1 and 5.2, it is observed that bounds in terms of the k-out-of-n system perform better than the bounds discussed in Proposition 5.1.4.

In the following section we express CREx of a mixed system consisting of d.i.d. components using distortion function and copula.

5.2 CREx of mixed systems consisting of d.i.d. components

Here we obtain CREx of mixed systems having n d.i.d. components. Let $X_1, X_2, \cdots, X_n$ be the lifetimes of the components and T be the system lifetime. It is assumed that the component lifetimes are d.i.d. with common marginal cdf F and pdf f. We consider the representation of sf of the system using a continuous and increasing distortion function $h : [0,1] \rightarrow [0,1]$ as (Navarro et al., 2013).

$$S_T(t) = h(S(t)). \tag{5.4}$$

Now if the components are exchangeable, then we have

$$h(v) = \sum_{i=1}^{n} c_i K(\underbrace{v, \cdots, v}_{\text{i times}}, \underbrace{1, \cdots, 1}_{\text{n-i times}}), \tag{5.5}$$

where $c = (c_1, \cdots, c_n)$ is the maximal signature of the mixed system and K is the diagonal section of the survival copula of the random vector $(X_1, X_2, \cdots, X_n)$. For iid components

$$h(v) = \sum_{i=1}^{n} c_i v^i. \tag{5.6}$$

The pdf of T can be expressed as

$$f_T(t) = -\frac{d}{dt} h(S(t)) = f(t) h'(S(t)), \tag{5.7}$$

where $h'(t) = \frac{d}{dt} h(t)$. Now using Eq. (5.1) we can express the CREx of a mixed system T consisting of d.i.d. components with common cdf F as

$$\xi J(T) = -\frac{1}{2} \int_0^1 \frac{h^2(v)}{f(S^{-1}(v))} dv. \tag{5.8}$$

Example 5.2.1. *Suppose the system given in Example 5.1.1 consisting of d.i.d. and exchangeable components having FGM copula:*

$$K(v_1, v_2, v_3, v_4) = v_1 v_2 v_3 v_4 (1 + a(1 - v_1)(1 - v_2)(1 - v_3)(1 - v_4)), \ a \in [-1, 1].$$

Then

$$S_T(t) = 2K(S(t),S(t),1,1) - K(S(t),S(t),S(t),S(t)).$$

So we have, $h(v) = 2v^2 - v^4(1+a(1-v)^4)$*. If the common sf of the components is* $S(t) = e^{-t}$ *and* $a = 0.5$*, then*

$$\begin{aligned}\xi J(T) &= -\frac{1}{2}\int_0^1 \frac{(2v^2 - v^4(1+\frac{1}{2}(1-v)^4))^2}{v}dv \\ &= -0.2285.\end{aligned}$$

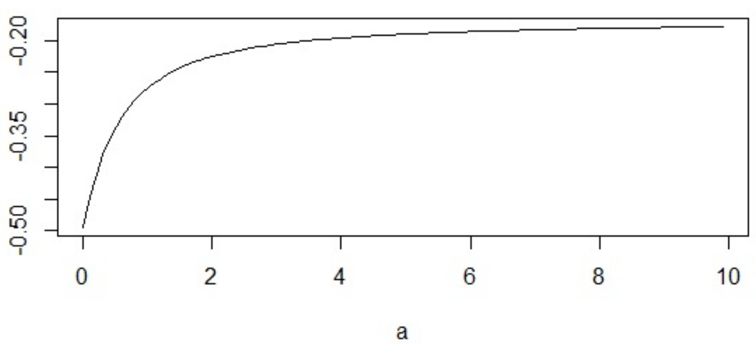
It is observed that CREx of T *for iid components having standard exponential distribution is* -0.2292*, which is less than the CREx of* T *consisting of dependent components. So in this case, the uncertainty is lower for d.i.d. components.*

Example 5.2.2. *We consider a series system having three components with d.i.d. components having standard exponential distribution. The lifetime of the system is* $T = X_{1:3}$*. Suppose the d.i.d. components have Clayton survival copula:* $K(v_1,v_2,v_3,) = \left(\sum_{i=1}^{3} v_i^{-\frac{1}{a}} - 2\right)^{-a}$*, where* $a > 0$*. So the diagonal section of the copula is* $h(v) = k(v,v,v) = \left(3v^{-\frac{1}{a}} - 2\right)^{-a}$*. Hence the CREx of* T *is*

$$\xi J(T) = -\frac{1}{2}\int_0^1 \frac{(3v^{-\frac{1}{a}} - 2)^{-a}}{v}dv.$$

We plotted the CREx with respect to the dependency parameter a *in Figure 5.3. From Figure 5.3 we see that, as the dependency parameter* a *increases the CREx of* T *increases hence the uncertainty associated with* T *decreases.*

Fig. 5.3: *Values of CREx in Example 5.2.2 for different values of* a*.*



Now we consider some comparison results and bounds for CREx of mixed systems consisting of d.i.d. components. The results are analogous to the iid case so we omit the proofs.

Proposition 5.2.1. *(i)* $X \overset{disp}{\leq} Y \Rightarrow \xi J(T^X) \geq \xi J(T^Y)$.
(ii) $\inf_{v\in A_1} \frac{h^2(v)}{v^2} \geq \sup_{v\in A_2} \frac{h^2(v)}{v^2}$ *and* $\xi J(X) \geq \xi J(Y)$ *, then* $\xi J(T^X) \geq \xi J(T^Y)$*, where* A_1 *and* A_2 *are defined in Proposition 5.1.2.*

Proposition 5.2.2. *Suppose* $X_1, X_2, \cdots, X_n$ *are the component lifetimes with common cdf* F *of a mixed system* T *with signature* **s***. Then*

$$B_3\xi J(X) \leq \xi J(T) \leq B_4\xi J(X),$$

where $B_3 = \sup_{v\in[0,1]} \frac{h^2(v)}{v^2}$ *and* $B_4 = \inf_{v\in[0,1]} \frac{h^2(v)}{v^2}$.

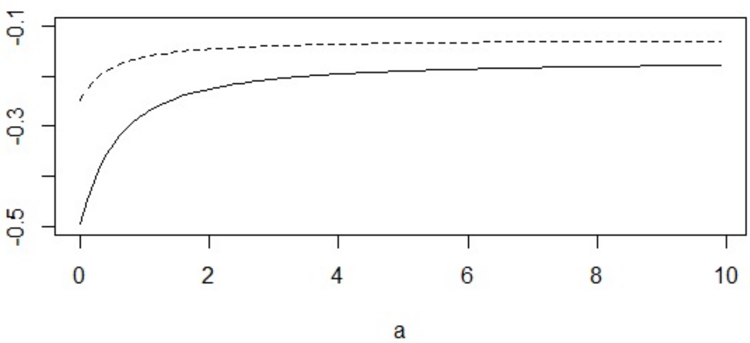
Proposition 5.2.3. *The CREx of a mixed system with n d.i.d. components satisfies*

$$-\frac{1}{2m}\int_0^1 h^2(v)dv \leq \xi J(T) \leq -\frac{1}{2M}\int_0^1 h^2(v)dv,$$

where $m = \inf_{x\in A} f(x)$, $M = \sup_{x\in A} f(x)$, f *is the common pdf of component lifetimes and* A *is the support of* f.

Example 5.2.3. *Consider the series system with d.i.d. components given in Example 5.2.2. Here for different values of a, we present the upper bound given in Proposition 5.2.3 along with* $\xi J(T)$ *is Figure 5.4. It is observed that as the dependency parameter increases the bound performs better.*

Fig. 5.4: *Upper bounds of CREx for different values of a.*



5.3 Applications in reliability engineering

We provide applications of CREx of mixed systems in engineering reliability problems. In reliability engineering, one important problem is to measure the complexity of systems. Also comparisons among various systems is a useful area in system reliability where it is often required to choose the better system among different systems that can not be comparable by usual stochastic orderings. In this section, we provide two applications on measuring complexity and comparing systems based on CREx of mixed systems consisting of iid components. For these purposes, we propose a Jensen type divergence measure and a new ordering of systems and obtain some interesting results associated with them. Also we discuss some applications involving system redundancy.

5.3.1 Jensen-Cumulative residual extropy divergence and complexity of systems

As stated earlier, one of the important applications of information measures in reliability engineering is to measure the complexity of the system. To address this issue, Asadi et al. (2016) proposed the Jensen-Shannon (JS) divergence between the system T and $X_{1:n},\cdots,X_{n:n}$ as

$$JS(T:X_{1:n},\cdots,X_{n:n}) = H(T) - \sum_{i=1}^{n} s_i\, H(X_{i:n}), \tag{5.9}$$

where $H(T)$ is the Shannon entropy of T. This measure compares the system entropy with its component entropies and it is zero for the k-out-of-n systems. This property helps us to study the complexity of systems as higher values of $JS(T:X_{1:n},\cdots,X_{n:n})$ will imply that the n-component system T is more complex than the k-out-of-n systems consisting of same type of components. Analogous to JS divergence, Qiu et al. (2019) defined the Jensen-Extropy (JE) divergence between system T and $X_{1:n},\cdots,X_{n:n}$ as

$$JE(T:X_{1:n},\cdots,X_{n:n}) = J(T) - \sum_{i=1}^{n} s_i\, J(X_{i:n}), \tag{5.10}$$

where $J(T)$ is the extropy of T. By analogy of Eq. (5.9) and (5.10), we propose the Jensen-Cumulative residual extropy (JCREx) divergence in terms of the CREx function. The JCREx divergence between T and $X_{1:n},\cdots,X_{n:n}$ is defined as

$$JCREx\,(T:X_{1:n},\cdots,X_{n:n}) = \xi J(T) - \sum_{i=1}^{n} s_i\, \xi J(X_{i:n}). \tag{5.11}$$

We can express Eq. (5.11) as

$$JCREx\,(T : X_{1:n}, \cdots, X_{n:n}) = -\frac{1}{2}\int_0^1 \frac{G_V^2(v) - \sum_{i=1}^n s_i\, G_{i:n}^2(v)}{f(S^{-1}(v))dv} dv. \tag{5.12}$$

Like JS and JE divergence measures, JCREx divergence is non-negative and from Proposition 5.1.5 we can see that $JCREx\,(T : X_{1:n}, \cdots, X_{n:n}) = 0$ for k-out-of-n systems. JCREx divergence measures the complexity of a system in comparison with a k-out-of-n system having the same number of iid components. In the following proposition, we compare the JCREx divergence of two systems having the same structure but different components.

Proposition 5.3.1. *Consider two mixed systems with lifetime T^X and T^Y having same signature* **s** *with components lifetimes $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_n$, respectively. If $X \leq_{disp} Y$, then*

$$JCREx\,(T^X : X_{1:n}, \cdots, X_{n:n}) \geq JCREx\,(T^Y : Y_{1:n}, \cdots, Y_{n:n}).$$

Proof. Proof follows using the definition of dispersive ordering in Eq. (5.13). □

Table 5.2: *JCREx divergence and lower bounds in terms of k-out-of-n systems.*

System	$\sum_{i=1}^n s_i J_S(X_{i:n})$	JCREx	System	$\sum_{i=1}^n s_i J_S(X_{i:n})$	JCREx
Group 1			Group 5		
1)	-0.0625	0	13)	-0.1518	0.0089
2)	-0.7244	0	14)	-0.6351	0.0268
3)	-0.3673	0			
4)	-0.1815	0			
Group 2			Group 6		
5)	-0.2125	0.0089	15)	-0.1220	0.0119
6)	-0.3363	0.0089	16)	-0.5458	0.0357
Group 3			Group 7		
7)	-0.2434	0.0143	17)	-0.1827	0.0214
8)	-0.2434	0.0143	18)	-0.4256	0.0416
9)	-0.3054	0.0143			
10)	-0.3054	0.0143			
Group 4			Group 8		
11)	-0.2744	0.0161	19)	-0.2446	0.0357
12)	-0.2744	0.0161	20)	-0.3637	0.0607

Now we propose a relative CREx measure, analogous to the relative extropy measure of Lad et al. (2015), which will be used for comparisons between two systems. This measure is a scale transformation of the energy distance between two non-negative rvs. Also this is related to quadratic distance between two sfs. The relative CREx between two non-negative continuous rvs X and Y with sfs S_1 and S_2 is defined as

$$R(X:Y) = \frac{1}{2}\int_0^{+\infty} (S_1(x) - S_2(x))^2\, dx. \tag{5.13}$$

The JCREx divergence can be expressed in terms of the relative CREx measure.

Proposition 5.3.2. *For a system T with iid components $X_1, X_3, \cdots, X_n$, the JCREx divergence between the system T and its ordered component lifetimes $X_{1:n}, \cdots, X_{n:n}$ is given by*

$$JCREx\,(T : X_{1:n}, \cdots, X_{n:n}) = \sum_{i=1}^{n} s_i\, R(T : X_{i:n}).$$

Proof. From Eq. (5.13) we have,

$$\begin{aligned}
\sum_{i=0}^{n} s_i\, R(T:X_{i:n}) &= \frac{1}{2}\sum_{i=0}^{n} s_i \int_0^{+\infty} (S_T(x) - S_{i:n}(x))^2 dx \\
&= \frac{1}{2}\sum_{i=0}^{n} s_i \int_0^{+\infty} \left(\sum_{i=0}^{n} s_i\, S_{i:n}(x) - S_{i:n}(x)\right)^2 dx \\
&= \frac{1}{2}\sum_{i=0}^{n} s_i \int_0^{1} \frac{(G_V(v) - G_{i:n}(v))^2}{f(S^{-1}(v))} dv \\
&= \frac{1}{2}\sum_{i=0}^{n} s_i \int_0^{1} \frac{[G_V^2(v) - 2G_V(v)G_{i:n}(v) + G_{i:n}^2(v)]}{f(S^{-1}(v))} dv \\
&= -\frac{1}{2}\int_0^{1} \frac{G_V^2(v) - \sum_{i=0}^{n} s_i\, G_{i:n}^2(v)}{f(S^{-1}(v))} dv \\
&= JCREx\,(T : X_{1:n}, \cdots, X_{n:n}).
\end{aligned}$$

□

The JCREx divergence for the set of systems in Table 5.1 are presented in Table 5.2. The set of systems have four iid components having standard exponential lifetimes. Asadi et al. (2016) divided the 20 systems into 8 groups in accordance with the dualities of the system signatures. Systems with dual signatures are put in the same group. In Table 5.2 it is observed that Group 1 contains k-out-of-4 systems, for $k = 1(1)4$, with JCREx divergence zero. These systems are the least complex systems. From Table 5.2 we also observe that as the complexity of the systems increases, the JCREx divergence also increases. Like JE divergence as pointed out in Qiu et al. (2019), JCREx divergence for Groups 2-4 are the same for the different structures with dual signatures in each Group. However, from Groups 5-8 the JCREx divergence for the structures with dual signatures in the same Group are different.

One advantage of the JCREx divergence measure is that, we can compare system complexity of systems consisting of different numbers of iid components. Consider the following examples illustrating this fact.

Example 5.3.1. *Consider the 3-components series-parallel system* $T_{sp} = \min\{X_1, \max\{X_2, X_3\}\}$ *and the parallel-series system* $T_{ps} = \max\{\min\{X_1, X_2\}, X_3\}$. *The series-parallel system has the signature* $\left(\frac{1}{3}, \frac{2}{3}, 0\right)$, *CREx is* $J(T_{sp})$ = *-0.1381 and its lower bound is* $\sum_{i=1}^{n} s_i J(X_{i:n})$ = *-0.1476. The JCREx divergence between* T_{sp} *and its components is* $JCREx(T_{sp} : X_{1:3}, X_{2:3}, X_{3:3})$= *0.0095. The parallel-series system has the signature* $\left(0, \frac{2}{3}, \frac{1}{3}\right)$ *and the values of the corresponding information measures and bound are* $J(T_{ps})$ = *-0.2214,* $\sum_{i=1}^{n} s_i J(X_{i:n})$ = *-0.2309 and* $JCREx(T_{sp} : X_{1:3}, X_{2:3}, X_{3:3})$ = *0.0095. Obviously, the parallel-series system is the better system and this is also evident from the fact that* $J(T_{ps}) < J(T_{sp})$. *The complexity of these two systems are the same. From Table 5.2, it is observed that systems of Group 2 and the first system of Group 5 are less complex than the 3-components series-parallel and parallel-series systems. So, increasing the number of components of a system does not necessarily imply that the system will become more complex. Complexity also depends on the structure of the systems.*

Fig. 5.5: *Structure functions of 3-component series-parallel and parallel-series systems.*

(a) *3-component series-parallel system.* **(b)** *3-component parallel-series system*

5.3.2 A new ordering of systems

Sometimes pairwise comparisons of systems by usual stochastic order may not be possible due to the complicated structure of the systems (Kochar et al., 1999). So instead of pairwise comparisons of systems, Toomaj et al. (2017) compared systems by determining how close (far) the structure of the system is towards the structure of the parallel (series) system. Since for any system T we have $X_{1:n} \overset{st}{\leq} T \overset{st}{\leq} X_{n:n}$, so one can compare two systems by observing how close the system is towards the structure of the parallel system and how far the system is towards the structure of the series system. Toomaj et al. (2017) considered this type of ordering of systems using cumulative residual KL divergence measure.

We propose a new measure to compare the performance of two systems based on the relative CREx defined in Eq. (5.13). Consider the following lemma which will be helpful in the development of the discrimination measure for comparison of the systems.

Lemma 5.3.1. *Let X_1, X_2 and X_3 be three rvs with sfs S_1, S_2 and S_3, respectively. If $X_1 \overset{st}{\leq} X_2 \overset{st}{\leq} X_3$ then $R(X_1, X_2) \leq R(X_1, X_3)$ and $R(X_2, X_3) \leq R(X_1, X_3)$.*

Proof. $X_1 \overset{st}{\leq} X_2 \overset{st}{\leq} X_3$ implies $S_1(t) \leq S_2(t) \leq S_3(t)$, $\forall t > 0$. So we have,

$$(S_2(t) - S_1(t))^2 \leq (S_3(t) - S_1(t))^2, \forall t > 0.$$

Also, we have, $\frac{S_1(t)}{S_3(t)} \leq \frac{S_2(t)}{S_3(t)} \leq 1$. Using the fact that $(x-1)^2$ is decreasing in (0,1), we get

$$(S_2(t) - S_3(t))^2 \leq (S_1(t) - S_3(t))^2, \forall t > 0.$$

The results follows by integration. □

Remark 5.3.1. *Since $X_1 \overset{st}{\leq} X_2 \overset{st}{\leq} X_3$, from Lemma 5.3.1 we have, for any mixed system T having n iid component lifetimes $X_1, \cdots, X_n$, $R(T, X_{i:n}) \leq R(X_{1:n}, X_{n:n})$.*

Now we define the discrimination information based on relative CREx. Discrimination information for a mixed system T is defined as

$$\Delta(T) = \frac{R(T, X_{1:n}) - R(T, X_{n:n})}{R(X_{1:n}, X_{n:n})}.$$

Clearly $-1 \leq \Delta(T) \leq 1$, where $\Delta(T) = 1$ iff T is a parallel system and $\Delta(T) = -1$ iff T is a series system. So when $\Delta(T)$ is close to 1, the system T is close to a parallel system and when $\Delta(T)$ is close to -1, the system T is close to a series system. Now we define an ordering between two systems based on $\Delta(\cdot)$. It is easy to see that, $\Delta(T)$ = 0 if T is the uniform mixture of series and parallel systems defined in Example 5.1.2.

Now we define an ordering between two systems based on $\Delta(\cdot)$ for comparing various systems which are not comparable by conventional methods.

Definition 5.3.1. *For two mixed systems with lifetimes T^X and T^Y, T^X is said to be less preferable than T^Y, denoted by $T^X \overset{\Delta}{\leq} T^Y$, if $\Delta(T^X) \leq \Delta(T^Y)$.*

Example 5.3.2. *Consider two coherent systems with four iid standard exponentially distributed components with system lifetimes* $T_1 = \min\{X_1, \max\{X_2, X_3, X_4\}\}$ *and* $T_2 = \max\{X_1, \min\{X_2, X_3, X_4\}\}$. *These two systems are not comparable by usual stochastic order. Now* $\Delta(T_1) = -0.3392$ *and* $\Delta(T_2) = -0.0044$ *and therefore* $T_1 \overset{\Delta}{\leq} T_2$. *So the system* T_2 *is closer to the parallel system* $X_{4:4}$ *than system* T_1. *As we have mentioned,* $\Delta(T)$ *= 0 for the uniform mixture of series and parallel systems. Now* $\Delta(T_2) = -0.0044$, *which is very close to zero so the system* T_2 *can be effectively approximated by the uniform mixture of series and parallel systems. Series and parallel systems are easily understood by the practitioners and they are least complex systems. So working with their mixture will be much easier than working with other complex systems. Therefore, the proposed discrimination method* $\Delta(T)$ *may be used to approximate various complex systems with relatively less complex mixed systems. Although we introduced* $\Delta(T)$ *for systems comparison purposes, it can also be used for potential application of mixed reliability systems.*

In the following result we show that stochastic ordering among the system signatures implies the Delta ordering among systems.

Proposition 5.3.3. *Let* T_1 *and* T_2 *be the lifetimes of two mixed systems, each consisting of n iid components with respective signatures* $\mathbf{s_1}$ *and* $\mathbf{s_2}$. *If* $\mathbf{s_1} \overset{st}{\leq} \mathbf{s_2}$ *then* $T_1 \overset{\Delta}{\leq} T_2$.

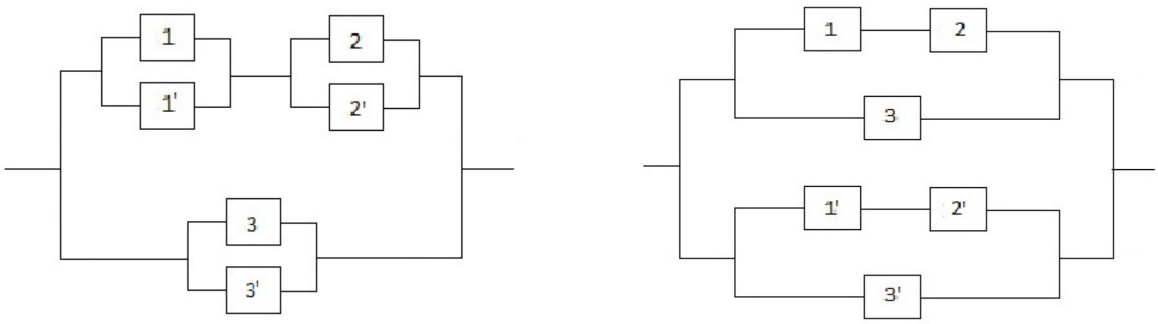
Proof. $\mathbf{s_1} \overset{st}{\leq} \mathbf{s_2}$ implies $X_{1:n} \overset{st}{\leq} T_1 \overset{st}{\leq} T_2 \overset{st}{\leq} X_{n:n}$ (Navarro et al., 2008). From Lemma 5.3.1 we have, $R(T_1, X_{1:n}) \leq R(T_2, X_{1:n})$ and $R(T_2, X_{n:n}) \leq R(T_1, X_{n:n})$. Hence the proof. □

Note that for two mixed systems T_1 and T_2 consisting of iid components $(X_1, \cdots, X_n)$, $T_1 \overset{st}{\leq} T_2$ implies $T_1 \overset{\Delta}{\leq} T_2$ and if $T_1 \overset{st}{=} T_2$ then $T_1 \overset{\Delta}{=} T_2$. So we can use Δ order two compare two systems when usual stochastic orders can not be applicable.

5.3.3 Application in redundancy allocation

In reliability engineering, a common way to enhance reliability of a system is to build redundancy into it. One way to achieve this is to attach spare components in parallel to each component of the system. This is called redundancy at component level. Another method is to add the same system parallel to the original system. This is called redundancy at system level. Redundancy at component level is better than redundancy at system level (Barlow and Proschan, 1975). Adding redundancy to a system will increase the reliability of the system but the cost will also increase. Consider the 3-component parallel-series system $T_{ps} = \max\{\min\{X_1, X_2\}, X_3\}$. The structure function of the system is given in Figures 5.5b. The redundant systems corresponding to the system T_{ps} are shown in Figure 5.6a and 5.6b,

Fig. 5.6: *Redundant systems of parallel-series system at component and system level*



(a) *Redundancy at component level.* **(b)** *Redundancy at system level.*

respectively. Now we study information properties of these redundant systems. We denote the redundant systems at component and system levels by T_{ps}^{cr} and T_{ps}^{sr}, respectively. The signature vector of T_{ps}^{cr} is $\left(0, \frac{2}{30}, \frac{4}{30}, \frac{8}{30}, \frac{16}{30}, 0\right)$. The CREx of T_{ps}^{cr} is -0.2444, the lower bound is $\sum_{i=1}^{n} s_i J(X_{i:n})$ = -0.2578 and the JCREx divergence between T_{ps}^{cr} and its components is 0.0134. The signature of T_{ps}^{sr} is $\left(0, \frac{2}{30}, \frac{7}{30}, \frac{13}{30}, \frac{8}{30}, 0\right)$, CREx is -0.2183, lower bound is -0.2303 and the JCREx divergence is 0.0119. Since CREx of T_{ps}^{cr} is less than that of T_{ps}^{sr}, we conclude that redundancy of components level is better than redundancy at system level. However, JCREx divergence for T_{ps}^{cr} is higher than JCREx for T_{ps}^{sr}. So, the system T_{ps}^{cr} is more complex than T_{ps}^{sr}. Clearly, the cost of constructing and running a more complex system is higher than a less complex one. Also redundancy at system level is easier to build since it is only required to add two same systems in parallel. A reliability engineer needs to decide to what extent redundancy is needed for reliability improvement while also keeping the cost as reasonable as possible. Another problem is how many redundant components one should add to the systems. Here we add one redundant component to each of the components and one redundant system to the system redundancy. One can add multiple components and systems in parallel as well. This type of redundancy is used in situations where high reliability is required. This is an interesting problem worth investigating. Also in component level redundancy, often the number of spare components are not sufficient for adding redundant components to each of the components. In those situations, a reliability engineer needs to find out which components are more important to the system and add redundant components to it so that the reliability is maximized.

In terms of information theory, adding redundancy will decrease the CREx of the system and increase the complexity of the system as well as the cost. So one needs to minimize the CREx of the system by adding redundancy subject to an upper bound of the JCREx divergence.

5.4 Estimation of CREx

In this Section, we propose three non-parametric estimators for CREx measures, study their properties and compare their performances by evaluating their MSEs. Let $X_1, X_2, \cdots, X_n$ be a rs drawn from a continuous distribution having cdf F and let $X_{1:n}, X_{2:n}, \cdots, X_{n:n}$ be the corresponding order statistics. We propose the first estimator of CREx measure analogously to the estimator of GWSE defined in Eq. (2.21) as

$$\xi J^1_{F_n}(X) = -\frac{1}{2}\sum_{i=0}^{n-1}(X_{(i+1):n} - X_{i:n})\left(1-\frac{i}{n}\right)^2. \tag{5.14}$$

This estimator is a modification of the estimator proposed by Jahanshahi et al. (2020). The proposed estimator is consistent.

Next we suggest another estimator of the CREx measure. First, consider the following theorem which will be useful in defining the estimator.

Theorem 5.4.1. *For a non-negative, continuous rv X with cdf F and sf S,*

$$\xi J(X) = -\int_0^{+\infty} xS(x)dF(x). \tag{5.15}$$

Proof. We have,

$$\begin{aligned}
-\int_0^{+\infty} xS(x)dF(x) &= -\int_0^{+\infty}\left(\int_0^x dv\right)S(x)dF(x)\\
&= -\int_0^{+\infty}\left(\int_v^{+\infty} S(x)dF(x)\right)dv\\
&= -\frac{1}{2}\int_0^{+\infty} S^2(v)dv.
\end{aligned}$$

Hence the proof. □

Using Theorem 5.4.1, we propose the second estimator of CREx as

$$\begin{aligned}
\xi J^2_{F_n}(X) &= -\int_0^{+\infty} xS_n(x)dF_n(x)\\
&= -\frac{1}{n}\sum_{i=1}^{n-1}\left(1-\frac{i}{n}\right)X_{i:n}.
\end{aligned} \tag{5.16}$$

In the following theorem we study the consistency and asymptotic normality of $\xi J^2_{F_n}(X)$.

Theorem 5.4.2. *Let $X_1, X_2, \cdots, X_n$ be a rs from a non-negative, continuous distribution with finite second moment. Then, $\sqrt{n}\left(\xi J^2_{F_n}(X) - \xi J_F(X)\right)$ is asymptotically normally distributed with mean zero and variance σ^2_F, where*

$$\sigma^2_F = \int_0^{+\infty}\int_0^{+\infty} [F(min(u,v)) - F(u)F(v)]\left(1-\frac{u}{n}\right)\left(1-\frac{v}{n}\right) dudv. \tag{5.17}$$

Proof. Proof is similar to that of Theorems 2 and 3 of Stigler (1974), hence it is omitted. □

Note that, σ^2_F contains unknown parameters. A consistent estimator of σ^2_F can be obtained as

$$\hat{\sigma}^2_F = \sum_{i=1}^{n-1}\sum_{i=1}^{n-1}\left(min\left(\frac{i}{n},\frac{j}{n}\right) - \frac{i}{n}\frac{j}{n}\right)\left(1-\frac{i}{n}\right)\left(1-\frac{j}{n}\right)(X_{(i+1):n} - X_{i:n})(X_{(j+1):n} - X_{j:n}). \tag{5.18}$$

An approximate 100(1-α)% confidence interval for $\xi J^2_{F_n}(X)$ can be computed as

$$\xi J^2_{F_n}(X) \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\sigma}^2_F}{n}},$$

where Z_α is the upper-α point of the standard normal distribution.

Next estimator is based on the Kernel function (Parzen, 1962). Kernel estimation of $f(\cdot)$ is given by

$$\hat{f}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right),$$

where h_n is the bandwidth parameter. The kernel function K satisfies the following properties:

1. $K(v) \geq 0$ for all v.
2. $\int_{-\infty}^{+\infty} K(v)dv = 1$.
3. $K(\cdot)$ is symmetric about zero.
4. $K(\cdot)$ satisfies the Lipschitz condition, that there exists a positive constant M such that $\mid K(v) - K(w) \mid \leq M \mid v - w \mid$.

Kernel estimator for CREx is defined as

$$\xi J^3_{F_n}(X) = -\frac{1}{2}\int_0^{+\infty} \hat{S}^2_n(x)dx, \tag{5.19}$$

where $\hat{S}_n(x) = \int_x^{+\infty} \hat{f}_n(w)dw$. From the consistency of kernel, it is obvious that $\xi J^3_{F_n}(X)$ is consistent. Now we compare the performance of these estimators.

Table 5.3: *Bias and MSEs for standard exponential distribution.*

n	$\xi J^1_{F_n}(X)$		$\xi J^2_{F_n}(X)$		$\xi J^3_{F_n}(X)$	
	Bias	MSE	Bias	MSE	Bias	MSE
10	-0.0249	0.0101	0.0243	0.0079	-0.0214	0.0093
20	-0.0121	0.0047	0.0113	0.0040	-0.0088	0.0041
30	-0.0088	0.0030	0.0080	0.0027	-0.0074	0.0027
40	-0.0065	0.0022	0.0062	0.0021	-0.0035	0.0021
50	-0.0056	0.0018	0.0052	0.0016	-0.0031	0.0016
100	-0.0025	0.0009	0.0022	0.0008	-0.0021	0.0008

Table 5.4: *Bias and MSEs for standard uniform distribution.*

n	$\xi J^1_{F_n}(X)$		$\xi J^2_{F_n}(X)$		$\xi J^3_{F_n}(X)$	
	Bias	MSE	Bias	MSE	Bias	MSE
10	-0.0088	0.0024	0.0170	0.0021	-0.0049	0.0021
20	-0.0040	0.0011	0.0083	0.0011	-0.0088	0.0010
30	-0.0024	0.0007	0.0051	0.0007	-0.0074	0.0007
40	-0.0020	0.0006	0.0039	0.0005	-0.0035	0.0006
50	-0.0015	0.0004	0.0037	0.0004	-0.0031	0.0004
100	-0.0009	0.0002	0.0017	0.0002	-0.0021	0.0002

5.4.1 Comparison of estimators

We conduct a simulation study to check the performance of the estimators by means of MSE. We consider exponential distribution with mean 1 and standard uniform distribution. We generate rs of size 10, 20, 30, 40, 50 and 100 from each reference distribution and calculate the Bias and MSE using 10000 replications. The true value of CREx for standard exponential distribution is $-\frac{1}{4}$ and for standard uniform distribution is $-\frac{1}{6}$. The results are provided in Tables 5.3 and 5.4, respectively. From the tables we observe that, as sample size increases, the MSE of all the estimators decreases. All the estimators perform similarly for uniform distribution in terms of MSE. For exponential distribution, $\xi J^2_{F_n}(X)$ performs better than other two estimators and $\xi J^3_{F_n}(X)$ performs better than $\xi J^1_{F_n}(X)$. So it may be useful to use $\xi J^2_{F_n}(X)$ for estimating CREx because it is very easy to calculate and its null distribution is normal, whose parameter can easily be estimated from a data set using Eq. (5.18).

5.4.2 Data Analysis

We analyse a real data set for illustration. The data set is given in Grubbs (1971) that we studied in Chapter 2, where we observed that exponential distribution with cdf
$F_\theta(x) = 1 - e^{-\frac{x}{\theta}}, x > 0,\ \theta > 0$, fits the data well. The mle of θ is $\hat{\theta} = \frac{1}{\bar{X}} = 0.001$. Parametric

estimate of $\xi J_F(X)$ is $-\frac{1}{4\hat{\theta}}$ = -249.4868. The non-parametric estimates are $\xi J^1_{F_n}(X)$ = -299.392, $\xi J^2_{F_n}(X)$ = -273.1302 and $\xi J^3_{F_n}(X)$ = -287.2928. We see that $\xi J^2_{F_n}(X)$ is closer to the parametric estimate than the other two and kernel based estimator $\xi J^3_{F_n}(X)$ is closer to the true value than $\xi J^1_{F_n}(X)$. We noticed this same pattern in Table 5.3 as well.

5.5 Testing equality between two distribution functions

In this section, we provide an application in hypothesis testing problems. Making use of the asymptotic normality of $\xi J^2_{F_n}(X)$, we provide a test statistic to perform goodness-of-fit tests among two distributions. Jahanshahi et al. (2020) showed that for two rvs X_1 and X_2, $X_1 \overset{st}{\leq} X_2 \Longrightarrow \xi J(X_1) \leq \xi J(X_2)$. Now it is known that, if $X_1 \overset{st}{\leq} X_2$ and $X_2 \overset{st}{\leq} X_1$, then $X_1 \overset{st}{=} X_2$ i.e. $F_{X_1}(v) = F_{X_2}(v)$, $\forall v$. So it is quite obvious that, for two non-negative, continuous rvs X_1 and X_2 having cdfs (sfs) F and G, $\xi J(X_1) = \xi J(X_2)$ implies $F(v) = G(v)$, $\forall v > 0$. And consequently, $\xi J^2_{F_n}(X) = \xi J^2_{G_n}(X)$ also implies $F(v) = G(v)$, $\forall v > 0$.

Let $U_1, U_2, \cdots, U_{n_1}$ and $V_1, V_2, \cdots, V_{n_2}$ be two independent rs from non-negative, continuous distributions with cdfs F and G, respectively. We want to test the hypothesis

$$H_0 : F(u) = G(u) \quad vs. \quad H_1 : F(u) \neq G(u).$$

Consider the following theorem which will be used to develop the test statistic.

Theorem 5.5.1. *Let $U_1, U_2, \cdots, U_{n_1}$ and $V_1, V_2, \cdots, V_{n_2}$ be two rs from non-negative continuous distributions with cdfs F and G, respectively. Also let $\xi J^2_{F_n}(U)$ and $\xi J^2_{G_n}(V)$ be the empirical estimators of $\xi J(U)$ and $\xi J(V)$, respectively. Both U and V have a finite second moment. Consider the difference,*
$\Delta(U,V) = \xi J(U) - \xi J(V)$ and $\Delta_N(U,V) = \xi J^2_{F_n}(U) - \xi J^2_{G_n}(V)$. Then,

$$\sqrt{N}\left(\Delta_N(U,V) - \Delta(U,V)\right) \xrightarrow{d} N\left(0, \frac{\sigma^2(F)}{\tau} + \frac{\sigma^2(G)}{1-\tau}\right),$$

where $N = n_1 + n_2$ and as $\min\{n_1, n_2\} \to +\infty$ we have $\frac{n_1}{N} \to \tau$, $\frac{n_2}{N} \to 1-\tau$.

Proof. Proof follows by using Theorem 5.4.2 and additive property of normal distribution. □

Now we can define the test statistic as

$$Z = \frac{\sqrt{N}\Delta_N(U,V)}{\sqrt{\frac{\sigma^2(F)}{\tau} + \frac{\sigma^2(G)}{1-\tau}}}.$$

Note that $\Delta(U,V) = 0$ under H_0. Since Z contains unknown parameter so we will estimate the variances by (5.18) and thus obtain the estimated test statistic as

$$\tilde{Z} = \frac{\sqrt{N}\Delta_N(U,V)}{\sqrt{\frac{\hat{\sigma}^2(F)}{\tau} + \frac{\hat{\sigma}^2(G)}{1-\tau}}}.$$

Reject the null hypothesis at significance level α if $|\tilde{Z}| > Z_{\frac{\alpha}{2}}$.

5.5.1 Simulation Study

A simulation study is conducted to assess the performance of the proposed test. We compare the power of the test with that of Kolmogorov-Smirnov (KS) and Wilcoxon rank sum (W) tests. For reference distributions, we consider exponential, Weibull and gamma.

Table 5.5: *Power of the test of equality between two exponential distributions.*

λ_1	λ_2	$\tilde{Z}$	KS	W
5	1	0.996	1	1
	2	0.999	0.999	1
	3	0.856	0.766	0.876
	4	0.266	0.205	0.273
	5	0.054	0.049	0.051
	6	0.203	0.154	0.195
	7	0.499	0.416	0.543
	8	0.792	0.687	0.838
	9	0.939	0.891	0.945

Table 5.6: *Power of the test of equality between two Weibull distributions.*

λ_1	λ_2	$\tilde{Z}$	KS	W
5	1	0.99	1	0.575
	2	0.948	0.975	0.318
	3	0.595	0.446	0.158
	4	0.168	0.084	0.063
	5	0.050	0.045	0.051
	6	0.139	0.062	0.066
	7	0.339	0.163	0.082
	8	0.593	0.351	0.126
	9	0.756	0.562	0.171

For each family of distributions, we take rs of size $n_1 = n_2 = 100$ and test the hypothesis whether the two samples come from the same distribution. For the Exp(λ) distribution, we

Table 5.7: *Power of the test of equality between two gamma distributions.*

λ_1	λ_2	$\tilde{Z}$	KS	W
5	1	1	1	1
	2	1	1	1
	3	1	1	1
	4	0.933	0.825	0.932
	5	0.044	0.040	0.049
	6	0.879	0.730	0.862
	7	1	1	1
	8	1	1	1
	9	1	1	1

take $\lambda = 5$ of the first sample and varied λ of the second sample from 1 to 9. For Weibull and gamma distributions, the shape parameter of the first sample is fixed at 5 and the shape parameter of the second sample varies from 1 to 9. With no loss of generality, the scale parameters of Weibull and gamma distribution are taken to be 1. We obtain power of the tests at 5% significance level from 5000 replications and provide them in Tables 5.5, 5.6 and 5.7, respectively. From the tables, we observe that the proposed test performs better than KS and Wilcoxon tests for Weibull distribution. For exponential and gamma distributions, all tests perform similarly. However, when the two distributions are close (i.e. $\lambda_2 = 4, 6$), the proposed test has higher power than the other two tests. All the tests attain the significance level under the null hypothesis.

5.6 Discussions

We studied CREx of mixed systems with identically distributed components and obtained some bounds. We computed the CREx of coherent systems with iid components having standard exponentially distributed lifetimes and observed that the parallel system has the minimum CREx and the series system has the maximum CREx. This is quite obvious since for two rvs the one with minimum extropy is better than the other. Also we compare various systems with the same structure but different components using CREx and analyze CREx for systems having d.i.d. components. We developed a discrimination information that measures how close (far) a system is towards a parallel (series) system and this discrimination information can be used to compare two systems (consisting of iid components) when usual stochastic order is not possible. We also proposed a divergence measure based on CREx of a mixed system which measures the complexity of the system, i.e how much the system is more complex than the k-out-of-n system.

We provided numerous applications in comparing between systems, measuring system complexity and in redundancy allocation problems. A future problem that we discussed is to allocate redundant components to minimize the CREx of the respective system subject to a suitable upper bound of the system complexity measure. Various problems can arise in this direction such as: How many redundant components need to apply to minimize the system CREx subject to a given budget. When the spare components stock is low, then which components should have redundant components in order to minimize the CREx of the respective systems. More work is needed in this direction.

We proposed two estimators of CREx measure and study their asymptotic properties. We compared these estimators with kernel based estimators. We found that the best estimator, in terms of MSE, is very easy to calculate and asymptotically normally distributed. Consistent estimator of the variance of the asymptotic normal distribution is obtained. Using the asymptotic normality of the proposed estimator that has minimum MSE, we constructed an equality test between two distributions. We compared the power of the proposed test with that of Kolmogorov-Smirnov and Wilcoxon rank sum test and found that our test performed better than the other two tests. For future work, we can extend this test for random right censored data.

Chapter 6

On some weighted generalized extropy measures with applications

Concept of weighted entropy was introduced in the literature five decades ago by Guiasu (1971). However, only in recent years, considerable attention has been given towards the analysis and development of various weighted cumulative information measures following the works of Misagh et al. (2011). The recent developments in this area are dealing with weighted information measures related to extropy measure. Balakrishnan et al. (2022) have proposed the weighted extropy measure, defined in Eq.(1.14), studied various properties and also defined dynamic versions (residual and past) of weighted extropy measure. For more details on dynamic extropy measures see Sathar and Nair (2021a). Recently, Sathar and Nair (2021b) have proposed the weighted version of CREx measure which they termed as weighted survival extropy (WSEx) measure. They have also introduced dynamic WSEx (DWSEx) measure, studied some properties and obtained its non-parametric estimator. For a rv X, the WSEx is defined as

$$\xi J^{w}(X) = -\frac{1}{2}\int_{0}^{+\infty} xS^{2}(x)dx \tag{6.1}$$

and the DWSEx measure can be defined as

$$\xi J^{w}(X;t) = -\frac{1}{2S^{2}(t)}\int_{t}^{+\infty} xS^{2}(x)dx. \tag{6.2}$$

Note that $\xi J^{w}(X;0) = \xi J^{w}(X)$. It is important to note that these measures are non-positive quantities. In actuarial study, information measures are often used as risk measures for portfolio optimization problems. Risk measures are always positive quantities and in order to use extropy and related measures as risk measures, some modifications have to be made. Jahanshahi et al. (2020) used $|\xi J(X)|$ for this purpose. Extropy and its related measures are all negative. These measures are popular because of their useful applications in various fields which we discussed in detail. This is the reason that these measures are widely used although they are not positive. If we do not want to work with negative measures such as CREx, then we can use negative CREx $(-\xi J(X))$ or the absolute value of CREx $(|\xi J(X)|)$ instead. Note that $|\xi J(X)|$ is basically $-\xi J(X)$, since $\xi J(X)$ is non-positive. The mathematical properties will remain the same (inequalities will reverse) and higher value of $-\xi J(X)$ means more uncertainty. Using this notion of negative CREx, Tahmasebi and Toomaj (2022) have proposed a new information measure called negative cumulative extropy (NCEx) measure. The NCEx for a rv X is defined as

$$\mathscr{C}(X) = \frac{1}{2}\int_0^{+\infty} \left[1 - F^2(x)\right] dx. \tag{6.3}$$

We can also express NCEx measure in terms of the expected value of $X_{2:2}$. Now $X_{2:2} = \max\{X_1, X_2\}$ is the lifetime of a 2-component parallel system with cdf $F^2(x)$. So we have

$$E(X_{2:2}) = \int_0^{+\infty} S_{X_{2:2}}(x)dx = \int_0^{+\infty} [1 - F^2(X)]dx = 2\mathscr{C}(X).$$

Now if we have an even number of iid sample $X_1, X_2, \cdots, X_n$ then, $\max\{X_1, X_2\}$, $\max\{X_3, X_4\}$, $\cdots, \max\{X_{n-1}, X_n\}$ is a sample from $X_{2:2}$ and this can be used to estimate NCEx measure.

From an application point of view, NCEx is a very important measure since it has applications in various areas. For example, Tahmasebi and Toomaj (2022) used NCEx as an alternative risk measure and also provided applications in system reliability. Noughabi (2021) developed a goodness-of-fit test for uniform distribution using NCEx measure.

In this Chapter, we study some important properties of WSEx measure such as existence, convolution property and some bounds. Next we propose a generalization of WSEx called weighted extended survival extropy (WESEx) along with its dynamic version and obtain some interesting results. Also, we introduce weighted negative cumulative extropy (WNCEx) measure and obtain some properties. Estimations and applications of these proposed measures are the main focus of this chapter. We propose edf based non-parametric estimators for WESEx, Dynamic WESEx and WNCEx measures for iid observations and obtain some asymptotic results of these estimators. Also we propose a recursive kernel based estimator for WSEx measure when the underlying sample satisfies α - mixing dependent condition (Rosenblatt, 1956). Extensive simulation studies are carried out to assess the performance of the proposed estimators. Several examples are provided and real data sets are analyzed. Two applications of WSEx measure are considered in model discrimination problems and in financial risk analysis. Using the estimator of WNCEx measure, a goodness-of-fit test is developed for uniform distribution. The power of the proposed test is compared with tests based on entropy, NCEx and some popular existing tests. The proposed test performs well.

The rest of this chapter is organised as follows. Some new properties of WSEx measure are considered in Section 6.1. The WESEx measure and its dynamic version are proposed and their properties are studied in Section 6.2. The WNCEx measure is introduced in Section 6.3. Non-parametric estimations of these proposed measures are studied in detail in Section 6.4. Applications of these measures are considered in Section 6.5. Finally, some concluding remarks are made in Section 6.6.

6.1 Weighted survival extropy

In this section, we study some interesting properties of the WSEx measure. Note that WSEx is a measure of information which takes into account the realizations of the rvs. For a given data set, if $\xi J^w(X_1) \leq \xi J^w(X_2)$ then X_1 is considered better than X_2. We calculate WSEx for some well known distribution and report them in Table 6.1.

Table 6.1: *WSEx for some distributions.*

Distributions		$\xi J^w(X)$
Exponential:	$F(x) = 1 - e^{-\lambda x}$; $x > 0, \lambda > 0$	$-\frac{1}{8\lambda^2}$
Power:	$F(x) = 1 - x^{\alpha}$, $0 < x < 1$, $\alpha > 0$	$-\frac{\alpha^2}{4(\alpha+1)(\alpha+2)}$
Finite Range:	$F(x) = (1 - ax)^b$, $x \in (0, \frac{1}{a})$, $a, b > 0$	$-\frac{1}{4a^2(b+1)(2b+1)}$
Rayleigh:	$F(x) = 1 - e^{-\frac{x^2}{2\sigma^2}}$, $x > 0$, $\sigma > 0$	$-\frac{\sigma^2}{4}$

In the following theorem, we study the existence of WSEx measure i.e. we provide conditions for WSEx being finite.

Theorem 6.1.1. *For a non-negative continuous rv X, WSEx will be finite if for some $p > 1$, $E(X^p) < +\infty$.*

Proof. We have

$$\int_0^{+\infty} xS^2(x)dx = \int_0^1 xS^2(x)dx + \int_1^{+\infty} xS^2(x)dx$$

Using Markov's inequality we have,

$$\begin{aligned}\int_0^{+\infty} xS^2(x)dx &\leq \int_0^1 xdx + \int_1^{+\infty} x\left(\frac{E(X^p)}{x^p}\right)^2 dx \\ &= \frac{1}{2} + (E(X^p))^2 \int_1^{+\infty} x^{1-2p}dx.\end{aligned}$$

Note that for $p > 1$, $\int_1^{+\infty} x^{1-2p}dx$ is finite. Hence the result. □

Next, we consider the effect of location and scale transformation on WSEx in the following lemma. This result will be used later in deriving some results and also in the application of WSEx measure.

Lemma 6.1.1. *Let X be a non-negative continuous rv and $Z = cX + d$, $c > 0, d \geq 0$. Then $\xi J^w(Z) = c^2\xi J^w(X) + cd\xi J(X)$.*

Proof. The proof follows using $S_{cX+d}(z) = S_X(\frac{z-d}{c})$, $x \in \mathbb{R}^+$. □

Lemma 6.1.1 shows that WSEx is a shift-dependent measure and it is not position free. Now we provide a bound of WSEx for convolution of two independent rvs.

Theorem 6.1.2. *For two non-negative, continuous and independent rvs X_1 and X_2 with respective cdfs F_1 and F_2 and sfs S_1 and S_2,*

$$\xi J^w(X_1+X_2) \geq \max\left\{\xi J^w(X_1) + E(X_2)\xi J(X_1) - \frac{E(X_2^2)}{4},\ \xi J^w(X_2) + E(X_1)\xi J(X_2) - \frac{E(X_1^2)}{4}\right\}.$$

Proof. Since X_1 and X_2 are independent, we have

$$P(X_1+X_2 > t) = \int_0^{+\infty} S_1(t-x_2)dF_2(x_2).$$

Using Jensen's inequality to the convex function $[P(X_1+X_2 > t)]^2$ we get,

$$[P(X_1+X_2 > t)]^2 = \left[\int_0^{+\infty} S_1(t-x_2)dF_2(x_2)\right]^2 \leq \int_0^{+\infty} S_1^2(t-x_2)dF_2(x_2). \tag{6.4}$$

Multiplying both sides by $-\frac{t}{2}$ and integrating with respect to t from 0 to $+\infty$ we have

$$\begin{aligned}&- \frac{1}{2}\int_0^{+\infty} t\,[P(X_1+X_2 > t)]^2\,dt \\ &\geq -\frac{1}{2}\int_0^{+\infty}\int_0^{+\infty} tS_1^2(t-x_2)dt\,dF_2(x_2) \\ &= -\frac{1}{2}\int_0^{+\infty}\left[\int_0^{x_2} tS_1^2(t-x_2)dt + \int_{x_2}^{+\infty} tS_1^2(t-x_2)dt\right]dF_2(x_2)\end{aligned}$$

$$= -\frac{1}{2}\int_0^{+\infty}\left[\int_0^{x_2} t dt + \int_0^{+\infty}(u+x_2)S_1^2(u)du\right] dF_2(x_2).$$

$$\Rightarrow \xi J^w(X_1+X_2) \geq -\frac{E(X_2^2)}{4} + \xi J^w(X_1) + E(X_2)\xi J(X_1)$$

Similarly, it can be proved that, $\xi J^w(X_1+X_2) \geq -\frac{E(X_1^2)}{4} + \xi J^w(X_2) + E(X_1)\xi J(X_2)$. Hence the result. □

The following Proposition gives upper and lower bounds for WSEx of a rv X in terms of WCRE and second raw moment of X.

Proposition 6.1.1. *For a non-negative, continuous rv X,*

$$-\frac{E(X^2)}{4} \leq \xi J^w(X) \leq \frac{1}{2}\left[CRE^w(X) - \frac{E(X^2)}{2}\right].$$

Proof. The proof of lower bound follows using the fact that

$$-\frac{1}{2}\int_0^{+\infty} xS^2(x)dx \geq -\frac{1}{2}\int_0^{+\infty} xS(x)dx.$$

The upper bound follows applying the inequality $\log x \leq x-1$, $\forall x > 0$ in $CRE^w(X)$ and after some simplifications. □

Consider the following examples which illustrate the bounds defined in Proposition 6.1.1.

Example 6.1.1. *Suppose X follows exponential distribution with mean $\frac{1}{\lambda}$ then $\xi J^w(X) = -\frac{1}{8\lambda^2}$, $\xi^w(X) = E(X^2) = \frac{2}{\lambda^2}$. So from Proposition 6.1.1 we have, $-\frac{1}{2\lambda^2} < \xi J^w(X) < \frac{1}{2\lambda^2}$.*

Example 6.1.2. *Suppose X has the Rayleigh distribution with cdf given in Table 6.1. Then we have, $\xi J^w(X) = -\frac{\sigma^2}{4}$, $\xi^w(X) = \sigma^2$ and $E(X^2) = 2\sigma^2$. So we have $\xi J^w(X) = -\frac{\sigma^2}{4} > -\frac{E(X^2)}{4} = -\frac{\sigma^2}{2}$ and $\xi J^w(X) < 0$.*

In the following lemma we compare WSEx of X, X_θ and θX where X_θ and X satisfy the PHRM with proportionality constant θ.

Lemma 6.1.2. *Suppose X_θ and X are independent and satisfy the PHRM. Then,*

$$\xi J^w(X_\theta) \geq \xi J^w(X) \geq \xi J^w(\theta X), if\, \theta \geq 1.$$

The inequality will reverse for $0 < \theta < 1$.

Proof. For $\theta > 1$,

$$\xi J^w(X_\theta) = -\frac{1}{2}\int_0^{+\infty} xS^\theta(x)dx \geq -\frac{1}{2}\int_0^{+\infty} xS^2(x)dx.$$

From Lemma 6.1.1 we have, $\xi J^w(\theta X) = \theta^2 \xi J^w(X) \leq \xi J^w(X)$. Hence the result.
Proof for $0 < \theta < 1$ can be obtained similarly. □

Example 6.1.3. *Suppose X has the Rayleigh distribution with cdf given in Table 6.1. Then for $\theta \geq 1$ we have, $\xi J^w(X_\theta) = -\frac{\sigma^2}{4\theta} \geq \xi J^w(X) = -\frac{\sigma^2}{4} \geq \xi J^w(\theta X) = -\frac{\theta^2\sigma^2}{4}$.*

6.2 Weighted extended survival extropy and its dynamic version

In this section, we introduce two new information measures called weighted extended survival extropy (WESEx) and dynamic weighted extended survival extropy (DWESEx) measures. Here we consider WESEx in the sense that we take the weight as a non-negative continuous function of x. The WESEx is a generalized information measure which contains SEx and WSEx measures as special cases.

Definition 6.2.1. *For a non-negative, continuous rv X, the WESEx is defined as*

$$\xi J^\varepsilon(X) = -\frac{1}{2}\int_0^{+\infty} \varepsilon(x)S^2(x)dx, \tag{6.5}$$

where $\varepsilon(\cdot)$ is a non-negative function of x.

From the WESEx measure, we obtain SEx if $\varepsilon(x) = 1$ and WSEx if $\varepsilon(x) = x$. In practice, the choice of $\varepsilon(x)$ will depend on the data. The weight function should be continuous in its arguments. If the observations are large then fraction weights such as $\varepsilon(x) = 1 - e^{-x}$, will be a good choice. The different choices of $\varepsilon(x)$ for a given data set will have an impact on the MSE of the estimators of WESEx. It is ideal to choose weights such that the estimators of WESEx measure have minimum MSE. An empirical study will be required to determine the proper weights.

Now we define DWESEx measure of a rv X which is the WESEx of the residual lifetime $[X - t|X > t]$.

Definition 6.2.2. *Dynamic version of WESEx for a non-negative continuous rv X is defined as*

$$\xi J^\varepsilon(X;t) = -\frac{1}{2S^2(t)}\int_t^{+\infty} \varepsilon(x)S^2(x)dx. \tag{6.6}$$

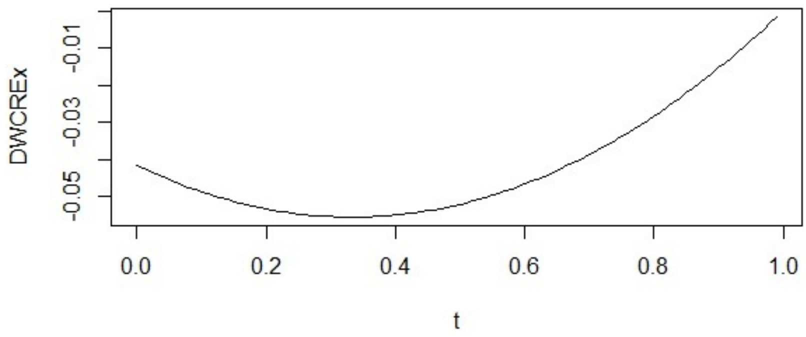
If we take $\varepsilon(x) = x$ then $\xi J^{\varepsilon}(X;t)$ reduces to $\xi J^{w}(X;t)$. Now we provide some examples of DWSEx measure.

Example 6.2.1. (i) *Let X be distributed as* $U(0,a)$ *then* $\xi J^{w}(X;t) = -\frac{(a-t)(a+3t)}{24}$.

(ii) *If X has the exponential distribution with mean* $\frac{1}{\lambda}$, *then* $\xi J^{w}(X;t) = -\left(\frac{t}{4\lambda} + \frac{1}{8\lambda^2}\right)$. *Sathar and Nair (2021b) showed that dynamic survival extropy of exponential distribution is constant, which is not the case for DWSEx. The DWSEx for exponential distribution is a linear function of time t and is decreasing in t.*

(iii) *Suppose X has Rayleigh distribution with cdf given in Table 6.1. Then* $\xi J^{w}(X;t) = -\frac{\sigma^2}{4}$, *is a constant.*

Fig. 6.1: *DWSEx for* $U(0,1)$ *distribution as a function of* t.



We plot the DWSEx for $U(0,1)$ distribution as a function of t in Figure 6.1. From the figure it is observed that DWSEx decreases and then increases. So DWSEx for $U(0,1)$ distribution is not monotone. Next we define two new orderings based on WESEx and DWESEx measures.

Definition 6.2.3. *X_1 is smaller than X_2 in WESEx order, denoted by,* $X_1 \overset{WESEx}{\leq} X_2$, *if*

$$\xi J^{\varepsilon}(X_1) \geq \xi J^{\varepsilon}(X_2).$$

Definition 6.2.4. *X_1 is smaller than X_2 in DWESEx order, denoted by,* $X_1 \overset{DWESEx}{\leq} X_2$, *if for all* $t > 0$

$$\xi J^{\varepsilon}(X_1;t) \geq \xi J^{\varepsilon}(X_2;t).$$

In the following propositions, we provide some basic properties of these measures. Most of the proof follows from the definition of WESEx and DWESEx measures. Proofs of the other results are discussed in detail in terms of other weighted information measures that we proposed earlier. So we omit the proofs of these propositions.

Proposition 6.2.1. *WESEx has the following properties:*

(i) *If* $\varepsilon(x) = f(x)$ *i.e. the pdf of the rv* X *then* $\xi J^{\varepsilon}(X) = -\frac{1}{6}$.

(ii) *If* $\varepsilon(x) = \lambda_F(x)$ *i.e. the hr function of the rv* X *then* $\xi J^{\varepsilon}(X) = -\frac{1}{4}$.

(iii) *The inequality* $0 \geq \xi J^{\varepsilon}(X) \geq -\frac{1}{2}m_F^{\varepsilon}(0)$ *holds, where* $m_F^{\varepsilon}(0) = \int_0^{+\infty} \varepsilon(x)S(x)dx$.

(iv) *For two rvs* X_1 *and* X_2, $X_1 \overset{st}{\leq} X_2 \Longrightarrow X_1 \overset{WESEx}{\leq} X_2$.

Proposition 6.2.2. *The following properties hold for DWESEx measure:*

(i) $\xi J^{\varepsilon}(X;0) = \xi J^{\varepsilon}(X)$.

(ii) *If* $\varepsilon(x) = \lambda_F(x)$ *i.e. the hr function of the rv* X *then*

$$\xi J^{\varepsilon}(X;t) = -\frac{1}{2}\int_t^{+\infty} \frac{f^2(x)}{S^2(t)}dx = J(X;t),$$

where $J(X;t)$ *is the dynamic extropy of* X *proposed by Qiu and Jia (2018b).*

(iii) *The inequality* $0 \geq \xi J^{\varepsilon}(X;t) \geq -\frac{1}{2}m_F^{\varepsilon}(t)$ *holds, where* $m_F^{\varepsilon}(t) = \int_t^{+\infty} \varepsilon(x)\frac{S(x)}{S(t)}dx$ *is the weighted extended mrl of* X.

(iv) *For two rvs* X_1 *and* X_2, $X_1 \overset{hr}{\leq} X_2 \Longrightarrow X_1 \overset{DWESEx}{\leq} X_2\ \forall t > 0$.

(v) *DWESEx is increasing (decreasing) in* t, *iff,*

$$\xi J^{\varepsilon}(X;t) \geq (\leq) -\frac{\varepsilon(t)}{4\lambda_F(t)},\ \forall t > 0,$$

where $\lambda_F(t)$ *is the hr function of* X.

(vi) *DWESEx uniquely determines the underlying distribution of the rv.*

Proposition 6.2.3. *For a non-negative continuous rv* X, *let* $\varepsilon(cX+d) = c\varepsilon(X)+d$, *where* $c > 0$ *and* $d \geq 0$. *Then*

(i) $\xi J^{\varepsilon}(cX+d) = c^2\xi J^{\varepsilon}(X) + cd\xi J(X)$,

(ii) $\xi J^{\varepsilon}(cX+d;t) = c^2 \xi J^{\varepsilon}(X;\frac{t-d}{c}) + cd\xi J(X;\frac{t-d}{c})$.

Following theorem addresses the closure of DWESEx order under scale transformation.

Theorem 6.2.1. *Let X_1 and X_2 be two non-negative continuous rvs with $X_1 \overset{DWESEx}{\geq (\leq)} X_2$ and $\varepsilon(cX+d) = c\varepsilon(X)+d$. Let $Z_1 = c_1X_1$ and $Z_2 = c_2X_2$, where $c_1, c_2 > 0$. Then $Z_1 \overset{DWESEx}{\geq (\leq)} Z_2$, if $\xi J^{\varepsilon}(X;t)$ is decreasing in $t > 0$ and $c_1 \leq (\geq)\, c_2$.*

Proof. Suppose $c_1 \leq c_2$. Since $\xi J^{\varepsilon}(X;t)$ is decreasing in t, we have $\xi J^{\varepsilon}(X_1;\frac{t}{c_1}) \leq \xi J^{\varepsilon}(X_1;\frac{t}{c_2})$. Again, $\xi J^{\varepsilon}(X_1;\frac{t}{c_2}) \leq \xi J^{\varepsilon}(X_2;\frac{t}{c_2})$ as $X_1 \overset{DWSEx}{\geq} X_2$. Combining these two inequalities and using part (ii) of Proposition 6.2.3, we get

$$\xi J^{\varepsilon}(Z_1;t) = c_1^2\, \xi J^{\varepsilon}\left(X_1;\frac{t}{c_1}\right) \leq c_2^2\, \xi J^{\varepsilon}\left(X_2;\frac{t}{c_2}\right) = \xi J^{\varepsilon}(Z_2;t).$$

Similarly, when $c_1 \geq c_2$ and $X_1 \overset{DWSEx}{\leq} X_2$, we can prove that $Z_1 \overset{DWSEx}{\leq} Z_2$. □

Corollary 6.2.1.1. *If $\varepsilon(x) = x$ then the above result can be interpreted in terms of WSEx measure.*

Corollary 6.2.1.2. *Let X_1 and X_2 be two non-negative continuous rvs with $X_1 \overset{DWESEx}{\geq (\leq)} X_2$ and $\varepsilon(x) = 1$. Let $Z_1 = c_1X_1 + d_1$ and $Z_2 = c_2X_2 + d_2$, where $c_1, c_2 > 0$ and $d_1, d_2 \geq 0$. Then $Z_1 \overset{DWESEx}{\geq (\leq)} Z_2$, if $\xi J^{\varepsilon}(X;t)$ is decreasing in $t > 0$ and $c_1 \leq (\geq)\, c_2$.*

6.2.1 Generalized inequalities

Now we consider some generalized inequalities associated with the WESEx measure. Also we study inequalities related to WESEx for some popular choices of $\varepsilon(X)$. First we provide some important inequalities which will be used to obtain various results.
For a non-negative continuous function $f \in L^p(0,+\infty)$, Hardy inequality reads

$$\int_0^{+\infty} \left(\frac{1}{x}\int_0^x f(v)dv\right)^p dx \leq \left(\frac{p}{p-1}\right)^p \int_0^{+\infty} f^p(x)dx, \quad p > 1. \tag{6.7}$$

For details see Hardy (1920). For probabilistic proof of Hardy's inequality see Walker (2015). Kaijser et al. (2002) proposed a generalized Hardy-Knopp type inequality for positive functions f as

$$\int_0^{+\infty} \left(\frac{1}{x}\int_0^x f(t)dt\right) \frac{dx}{x} \leq \int_0^{+\infty} \Psi(f(x))\frac{dx}{x}, \tag{6.8}$$

where Ψ is a convex function in $(0,+\infty)$. Čižmešija et al. (2003) generalized inequality (6.8) which is given in the following theorem.

Theorem 6.2.2. *(Čižmešija et al., 2003) Suppose $0<b<+\infty$, $u:(0,b)\to\mathbb{R}$ is a non-negative function such that the function $x:\to\frac{u(x)}{x^2}$ is locally integrable in $(0,b)$ and the function v is defined by*

$$v(u)=t\int_t^b \frac{u(x)}{x^2}dx,\quad t\in(0,b).$$

If the real valued function Ψ is convex on (a,c), where $-\infty<a<c<+\infty$, then the inequality

$$\int_0^b \Psi\left(\frac{1}{x}\int_0^x f(t)dt\right)\frac{dx}{x}\leq\int_0^b v(x)\Psi(f(x))\frac{dx}{x} \tag{6.9}$$

holds for all integrable functions $f:(0,b)\to\mathbb{R}$, such that $f(x)\in(a,c)$ $\forall x\in(0,b)$.

Now we propose generalized inequalities in terms of WESEx measure for different choices of the weight function.

Theorem 6.2.3. *Let X be a non-negative continuous rv with sf S. If $\varepsilon(x)=\frac{1}{x}$ then*

$$\xi J^{\varepsilon}(X)\leq-\frac{1}{2}\int_0^{+\infty}\left(\frac{1}{x}\int_0^x S(t)dt\right)\frac{dx}{x}.$$

Proof. If $\varepsilon(x)=\frac{1}{x}$ then $\xi J^{\varepsilon}(X)=-\frac{1}{2}\int_0^{+\infty}S^2(x)\frac{dx}{x}$. From inequality (6.9) we can write $\int_0^{+\infty}S^2(x)\frac{dx}{x}\geq\int_0^{+\infty}\left(\frac{1}{x}\int_0^x S(t)dt\right)\frac{dx}{x}$. Hence the result. □

Theorem 6.2.4. *Let X be a non-negative continuous rv with sf S. If $\varepsilon(x)=S^m(x)$ then*

$$\xi J^{\varepsilon}(X)\leq-\frac{1}{2}\left(\frac{m+1}{m+2}\right)^{m+2}\int_0^{+\infty}\left(\frac{1}{x}\int_0^x S(t)dt\right)^{2+m}dx.$$

Proof. If $\varepsilon(x)=S^m(x)$ then $\xi J^{\varepsilon}(X)=-\frac{1}{2}\int_0^{+\infty}S^{2+m}(x)dx$. Now applying Hardy's inequality and after some calculation, the result follows. □

Corollary 6.2.4.1. *If $m=0$ then we have $\xi J(X)\leq-\frac{1}{8}\int_0^{+\infty}\left(\frac{1}{x}\int_0^x S(t)dt\right)dx$. This inequality for survival extropy measure is obtained by Goodarzi and Amini (2021).*

Theorem 6.2.5. *For a non-negative continuous rv X with sf S, if $\varepsilon(x)=m_F(x)$ then $\xi J^{\varepsilon}(X)\leq-\frac{1}{2}\int_0^{+\infty}S(x)\left(\int_0^x S(t)dt\right)dx$, where $m_F(t)=\int_t^{+\infty}\frac{S(x)}{S(t)}dx$ is the MRL of X.*

Proof. If $\varepsilon(X)=m_F(x)$ then, we have

$$\xi J^{\varepsilon}(X)=-\frac{1}{2}\int_0^{+\infty}xm_F(x)S(x)S(x)\frac{dx}{x}. \tag{6.10}$$

Consider the function $u(x) = x^2 S(x)$ then from Theorem 6.2.2 we get $v(x) = x\, m_F(x)S(x)$. Now applying Theorem 6.2.2 in Eq. (6.10) we obtain the required result. □

Theorem 6.2.6. *For a non-negative continuous rv X, the following inequalities holds:*

(i) $\xi J^{\varepsilon}(X) \leq -\frac{1}{2}\exp(H(X) + E[\log \varepsilon(X)] - 2)$,

(ii) $\xi J^{\varepsilon}(X) \geq J(X)\exp(2H(X) + E(\log \varepsilon(X)) - 2)$,

(iii) $\xi J^{\varepsilon}(X) \leq -\frac{1}{2}\exp\left(\log E(X) - \frac{CRE(X)}{E(X)} + \frac{1}{E(X)}\int_0^{+\infty} S(x)\log \varepsilon(x)dx\right)$.

Proof. (i) From log-sum inequality we have,

$$\int_0^{+\infty} f(x)\log\frac{f(x)}{\varepsilon(x)S^2(x)}dx \geq \int_0^{+\infty} f(x)dx\, \log\frac{\int_0^{+\infty} f(x)dx}{\int_0^{+\infty}\varepsilon(x)S^2(x)dx} = \log\frac{1}{-2\xi J^{\varepsilon}(X)}. \tag{6.11}$$

After some simplifications, $\int_0^{+\infty} f(x)\log\frac{f(x)}{\varepsilon(x)S^2(x)}dx$ becomes $-H(X) - E[\log\varepsilon(X)] + 2$. Hence the result.

(ii) Using log-sum inequality we get

$$\int_0^{+\infty} f(x)\log\frac{f^2(x)}{\varepsilon(x)S^2(x)}dx \geq \log\frac{J(X)}{\xi J^{\varepsilon}(X)}. \tag{6.12}$$

After some algebraic simplifications, the term $\int_0^{+\infty} f(x)\log\frac{f^2(x)}{\varepsilon(x)S^2(x)}dx$ reduces to $-2H(X) - E[\log\varepsilon(X)] + 2$ and the result follows from Eq. (6.12).

(iii) Proof follows by applying log-sum inequality in $\int_0^{+\infty} S(x)\log\frac{S(x)}{\varepsilon(x)S^2(x)}dx$ and proceeding similarly as (i). □

6.3 Weighted negative cumulative extropy

In this section, we propose weighted negative cumulative extropy (WNCEx) measure and obtain some properties. First we define weighted negative survival extropy (WNSEx) measure.

Definition 6.3.1. *For a non-negative absolutely continuous rv X with sf S, the weighted negative survival extropy (WNSEx) is defined as*

$$\mathscr{J}^{w}(X) = \frac{1}{2}\int_0^{+\infty} xS^2(x)dx. \tag{6.13}$$

Note that WNSEx is nothing but the negative of WSEx measure and naturally it possesses similar properties like WSEx. Now we define WNCEx measure.

Definition 6.3.2. *Let X be a non-negative absolutely continuous rv with cdf F. Then, weighted negative cumulative extropy (WNCEx) of X is defined as*

$$\mathscr{C}^w(X) = \frac{1}{2}\int_0^{+\infty} x\left[1 - F^2(x)\right] dx. \tag{6.14}$$

Table 6.2: *WNCEx for some distributions.*

Distributions	$F(x)$	$\mathscr{C}^w(X)$
Uniform	$\frac{x}{a};\ 0 < x < a$	$\frac{a^2}{8}$
Exponential	$1 - e^{-\lambda x};\ x > 0,\ \lambda > 0$	$\frac{7}{8\lambda^2}$
Rayleigh	$1 - e^{-\lambda x^2};\ x \geq 0,\ \lambda > 0$	$\frac{3}{8\lambda}$
Power	$(\beta x)^\alpha;\ 0 < x < \frac{1}{\beta},\ \alpha,\ \beta > 0$	$\frac{\alpha}{4(\alpha+1)\beta^2}$
Pareto I	$1 - \left(\frac{k}{x}\right)^a;\ 0 < x < k,\ k > 0$	$\frac{k^2(2-3a)}{4(a-1)(a-2)},\ a > 2$
Weibull	$1 - e^{-(\lambda x)^\alpha};\ x > 0,\ \alpha,\ \lambda > 0$	$\frac{\Gamma\left(\frac{2}{\alpha}\right)\left(2^{\frac{\alpha+2}{\alpha}} - 1\right)}{2^{\frac{\alpha+2}{2}}\alpha\beta^2}$
Pareto II	$1 - (1 + px)^{-q};\ x \geq 0,\ p > 0, q > 0$	$\frac{1}{p^2(q-1)(q-2)} - \frac{1}{4p^2(q-1)(2q-1)},\ q > 2$

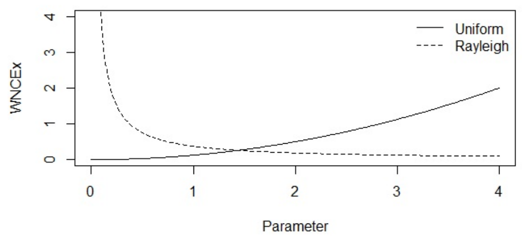
The WNCEx for some popular distributions are provided in Table 6.2. We plot the WNCEx of uniform and Rayleigh distributions provided in Table 6.2 and present them in Figure 6.2a. The parameters of the distributions vary from 0.10 to 4. From the figure it is observed that WNCEx of uniform distribution increases as the parameter a increases and WNCEx for Rayleigh distribution decreases as λ increases. Also we plot WNCEx for power distributions for β = 0.2, 0.3, 0.4 and 0.5 and, α varies from 0.10 to 4. The plots are provided in Figure 6.2b and it is found that WNCEx for power distribution is increasing in α for fixed β and decreasing in β for fixed α.

Now we study some important properties of WNCEx measure. Consider the following lemma that shows the effect of linear transformation of X on WNCEx.

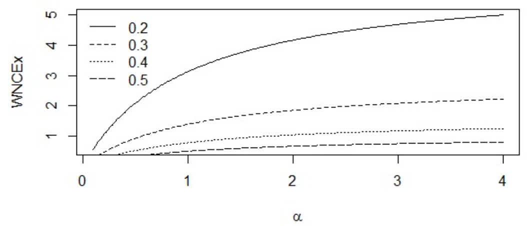
Lemma 6.3.1. *Consider the linear transformation $Y = cX + d,\ c > 0, d \geq 0$. Then*

$$\mathscr{C}^w(Y) = c^2\mathscr{C}^w(X) + cd\mathscr{C}(X) + \frac{d^2}{4}.$$

Fig. 6.2: *WNCEx for some distributions given in Table 6.2.*



(a) *WNCEx for uniform and Rayleigh distributions for 0.10$\leq a, \lambda \leq 4$.*



(b) *WNCEx for power distributions for 0.10$\leq \alpha \leq 4$.*

Proof. We have

$$F_{aX+b}(x) = \begin{cases} 0, & \text{if } x < d, \\ F_X\left(\frac{x-d}{c}\right), & \text{if } x \geq d. \end{cases}$$

Using this in Eq. (6.14), we get

$$\begin{aligned} \mathscr{C}^w(X) &= \frac{1}{2}\left[\frac{b^2}{2} + \int_b^{+\infty} x\left[1 - F_X^2\left(\frac{x-d}{c}\right)\right]dx\right] \\ &= \frac{1}{2}\int_0^{+\infty} c(cx+d)\left[1 - F_X^2(x)\right]dx + \frac{d^2}{4} \\ &= c^2\mathscr{C}^w(X) + cd\mathscr{C}(X) + \frac{d^2}{4}. \end{aligned}$$

Hence the proof. □

Remark 6.3.1. *For NCEx measure, the effect of linear transformation is expressed through the following relation:*

$$\mathscr{C}(X) = c\mathscr{C}(X) + \frac{d}{2}.$$

Next we study WNCEx for convolution of two independent rvs. The following theorem states that WNCEx for convoluted rv is greater than that of either.

Theorem 6.3.1. *Let X and Y be two non-negative, independent and continuous random variables having cdfs F and G, respectively. Then*

$$\mathscr{C}^w(X+Y) \geq max\left[\mathscr{C}^w(X), \mathscr{C}^w(Y)\right].$$

Proof. Proof is similar to Theorem 2.6 of Tahmasebi and Toomaj (2022), hence omitted. □

Corollary 6.3.1.1. *Let $X_1, X_2, \cdots, X_n$ be independent, non-negative and continuous rvs with*

WNCEx $\mathscr{C}^w(X_k)$, $k = 1(1)n$, *respectively. Then*

$$\mathscr{C}^w\left(\sum_{k=1}^{n} X_k\right) \geq \max\{\mathscr{C}^w(X_1), \cdots, \mathscr{C}^w(X_n)\}.$$

The following proposition provides an alternative representation of WNCEx which will be useful to prove some important results.

Proposition 6.3.1. *For a non-negative continuous rv X with cdf F and sf S,*

$$\mathscr{C}^w(X) = \frac{1}{2}\left[\frac{1}{2}E(X^2) + \int_0^{+\infty} xF(x)S(x)dx\right]. \tag{6.15}$$

Proposition 6.3.2. *For a non-negative continuous rv X, sum of WNCEx and WNCREx is equal to $\frac{1}{2}E(X^2)$.*

Proof. Using Eqs. (6.13) and (6.15), it follows that

$$\begin{aligned}
\mathscr{C}^w(X) + \mathscr{J}^w(X) &= \frac{1}{2}\left[\frac{1}{2}E(X^2) + \int_0^{+\infty} x[1-S(x)]\,S(x)dx + \int_0^{+\infty} xS^2(x)dx\right] \\
&= \frac{1}{2}\left[\frac{1}{2}E(X^2) + \frac{1}{2}E(X^2)\right] \\
&= \frac{1}{2}E(X^2).
\end{aligned}$$

□

An important property of WNCEx is that it can be expressed in terms of both WMRL and WMPL of the rv. Consider the following definitions.

Theorem 6.3.2. *Suppose for a non-negative continuous rv X, WMRL and WNCEx are $m^w(x)$ and $\mathscr{C}^w(X)$, respectively. Then the following identity hold.*

$$\mathscr{C}^w(X) = \frac{1}{2}\left[\frac{1}{2}E(X^2) + E[m^w(X)S(X)]\right].$$

Proof. From Eq. (6.15) and changing order of integration, one can obtain

$$\begin{aligned}
\mathscr{C}^w(X) &= \frac{1}{2}\left[\frac{1}{2}E(X^2) + \int_0^{+\infty}\left(\int_0^{x} f(v)dv\right) xS(x)dx\right] \\
&= \frac{1}{2}\left[\frac{1}{2}E(X^2) + \int_0^{+\infty} f(v)\int_v^{+\infty} xS(x)dx\,dv\right] \\
&= \frac{1}{2}\left[\frac{1}{2}E(X^2) + E[m^w(X)S(X)]\right].
\end{aligned}$$

Hence the proof. □

Example 6.3.1. *If X has the exponential distribution with cdf $F(x) = 1 - e^{-\lambda x}$; $x > 0, \lambda > 0$. Then, we have $E[m^w(X)S(X)] = \frac{3}{4\lambda^2}$ and $\frac{1}{2}E(X^2) = \frac{1}{\lambda^2}$. Therefore,*

$$\mathscr{C}^w(X) = \frac{1}{2}\left[\frac{1}{2}E(X^2) + E[m^w(X)S(X)]\right] = \frac{7}{8\lambda^2}.$$

Theorem 6.3.3. *Suppose for a non-negative continuous rv X, WMPL and WNCEx are $\mu^w(x)$ and $\mathscr{C}^w(X)$, respectively. Then the following identity holds.*

$$\mathscr{C}^w(X) = \frac{1}{2}\left[\frac{1}{2}E(X^2) + E[\mu^w(X)F(X)]\right].$$

Proof. Proceeding along the same lines as in the proof of Theorem 6.3.2, we have from Eq. (6.15)

$$\begin{aligned}
\mathscr{C}^w(X) &= \frac{1}{2}\left[\frac{1}{2}E(X^2) + \int_0^{+\infty}\left(\int_x^{+\infty} f(v)dv\right) xF(x)dx\right] \\
&= \frac{1}{2}\left[\frac{1}{2}E(X^2) + \int_0^{+\infty} f(v)\int_0^{v} xF(x)dx\, dv\right] \\
&= \frac{1}{2}\left[\frac{1}{2}E(X^2) + E(\mu^w(X)F(X))\right].
\end{aligned}$$

Hence the proof. □

Example 6.3.2. *Suppose X follows power distribution with cdf $F(x) = x^\alpha$; $0 < x < 1$, $\alpha > 0$. Then, we have $E(X^2) = \frac{\alpha}{2(\alpha+2)}$, $E[\mu^w(X)F(X)] = \frac{\alpha}{2(\alpha+1)(\alpha+2)}$ and*

$$\mathscr{C}^w(X) = \frac{1}{2}\left[\frac{1}{2}E(X^2) + E(\mu^w(X)F(X))\right] = \frac{\alpha}{4(\alpha+1)}.$$

Proposition 6.3.3. *Let X_θ and X satisfy the PRHM with proportionality parameter θ then, for $\theta \geq (\leq)1$,*

$$\mathscr{C}^w(X_\theta) \geq (\leq)\mathscr{C}^w(X).$$

Proof. Proof follows using $F^{2\theta}(x) \geq (\leq)F^2(x)$ for $\theta \leq (\geq)1$. □

Proposition 6.3.4. *Let X_1 and X_2 be two non-negative rvs with cdfs F and G, respectively. If $X_1 \overset{st}{\leq} X_2$ then,*

$$\mathscr{C}^w(X_1) \leq \mathscr{C}^w(X_2).$$

Proof. Proof follows using the definition of stochastic order and WNCEx. □

Example 6.3.3. *Suppose X_1 and X_2 follows the exponential distribution with cdf $F(x) =$*

$1-e^{-2x}$; $x>0$, *and* $G(x)=1-e^{-x}$; $x>0$, *respectively. Then we have* $\mathscr{C}^w(X_1)=\frac{7}{32}$ *and* $\mathscr{C}^w(X_2)=\frac{7}{8}$. *Hence* $X_1 \overset{st}{\leq} X_2 \implies \mathscr{C}^w(X_1) \leq \mathscr{C}^w(X_2)$.

In the next theorem we provide an upper bound of WNCEx in terms of Shannon entropy.

Theorem 6.3.4. *Let X be a non-negative continuous rv having pdf f and cdf F, respectively. Then,*

$$\mathscr{C}^w(X) \geq 2e^{-2}\exp[H(X)+E(\log(X))].$$

Proof. Using the log-sum inequality, we have

$$\begin{aligned}\int_0^{+\infty} f(x)\log\frac{f(x)}{x[1-F^2(x)]}dx &\geq \int_0^{+\infty} f(x)dx\,\log\frac{\int_0^{+\infty} f(x)dx}{\int_0^{+\infty} x[1-F^2(x)]dx}\\ &= \log\frac{1}{2\mathscr{C}^w(X)}.\end{aligned}\tag{6.16}$$

After some calculation, $\int_0^{+\infty} f(x)\log\frac{f(x)}{x(1-F^2(x))}dx$ reduces to

$$-H(X)-E[\log X]+2-2\log 2.$$

Therefore, from Eq. (6.16) we get

$$-H(X)-E[\log X]+2-2\log 2 \geq \log\frac{1}{2\mathscr{C}^w(X)}.$$

The result follows after some simplification. □

6.4 Non-parametric Estimation

In this section, we study non-parametric estimations of the proposed measures. First, we consider estimations of WESEx and WNCEx measures based on the edf function for iid data sets. We study asymptotic properties of these estimators and also evaluate their performance by simulation. Finally, we propose a recursive kernel based non-parametric estimator of WSEx measure when the underlying observations are not independent. Various real data sets are also analyzed.

6.4.1 Estimation of WESEx and related measures for iid observations

Here we propose non-parametric estimators for WESEx and dynamic WESEx measures. As special cases, we obtain non-parametric estimators for SEx and WSEx measures as well. Some examples are provided for illustrative purposes. Let $X_1, X_2, \cdots, X_n$ be a rs drawn from a continuous distribution with cdf F and let $X_{1:n}, X_{2:n}, \cdots, X_{n:n}$ be the corresponding order statistics. Suppose $F_n(x)$ be the edf of F. Then a non-parametric estimator for WESEx can be defined as

$$\begin{aligned} \xi J^{\varepsilon}(F_n) &= -\frac{1}{2}\int_0^{+\infty} \varepsilon(x) S_n^2(x) dx \\ &= -\frac{1}{2}\sum_{i=1}^{n-1}\int_{X_{i:n}}^{X_{(i+1):n}} \varepsilon(x)\left(1-\frac{i}{n}\right)^2 dx \\ &= -\frac{1}{2}\sum_{i=1}^{n-1}[\eta(X_{(i+1):n}) - \eta(X_{i:n})]\left(1-\frac{i}{n}\right)^2, \end{aligned} \tag{6.17}$$

where $\eta(x) = \int_0^x \varepsilon(u)du$.

Example 6.4.1. *Let $X_1, X_2, \cdots, X_n$ be a rs drawn from a continuous distribution with cdf F and pdf f. Suppose the weight function is $\varepsilon(x) = f(x)$ then we have,*

$$\xi J^{\varepsilon}(F_n) = -\frac{1}{2}\sum_{i=1}^{n-1}[F(X_{(i+1):n}) - F(X_{i:n})]\left(1-\frac{i}{n}\right)^2.$$

Let $Z_i = F(X_{(i+1):n}) - F(X_{i:n})$, $i = 1, 2, \cdots, n-1$. Then Z_i's are independent and have Beta (1,n) distribution. So the mean and variance of $\xi J^{\varepsilon}(F_n)$ are given by,

$$E[\xi J^{\varepsilon}(F_n)] = -\frac{1}{2}\sum_{i=1}^{n-1}\frac{1}{n+1}\left(1-\frac{i}{n}\right)^2$$

and

$$V[\xi J^{\varepsilon}(F_n)] = \frac{1}{4}\sum_{i=1}^{n-1}\frac{n}{(n+1)^2(n+2)}\left(1-\frac{i}{n}\right)^4.$$

Table 6.3: *Mean and variance of $\xi J^{\varepsilon}(F_n)$ with density as weight function.*

n	$E[\xi J^{\varepsilon}(F_n)]$	$V[\xi J^{\varepsilon}(F_n)]$	n	$E[\xi J^{\varepsilon}(F_n)]$	$V[\xi J^{\varepsilon}(F_n)]$
10	-0.1295	0.0026	40	-0.1565	0.0011
20	-0.1470	0.0018	50	-0.1585	0.0009
30	-0.1533	0.0013	100	-0.1625	0.0004

We calculate the means and variances of the estimator for the WESEx measure with density as weight function for n =10, 20, 30, 40, 50 and 100 and show them in Table 6.3. Note that, mean and variance depends only on the sample size. The true value of WESEx in this case is -0.1667. The mean approaches to the true value as sample size increases and variances approaches to zero.

Now if we take $\varepsilon(x) = 1$ in $\xi J^{\varepsilon}(F_n)$ then we get the non-parametric estimator for the SEx proposed by Jahanshahi et al. (2020), which is given by,

$$\xi J(F_n) = -\frac{1}{2}\sum_{i=1}^{n-1}(X_{(i+1):n} - X_{i:n})\left(1-\frac{i}{n}\right)^2. \tag{6.18}$$

If we take $\varepsilon(x) = x$ then we will get the non-parametric estimator for the WSEx as

$$\begin{aligned}\xi J^{w}(F_n) &= -\frac{1}{4}\sum_{i=1}^{n-1}(X^2_{(i+1):n} - X^2_{i:n})\left(1-\frac{i}{n}\right)^2 \\ &= -\frac{1}{4}\sum_{i=1}^{n-1}U_i\left(1-\frac{i}{n}\right)^2, \end{aligned}\tag{6.19}$$

where $U_i = (X^2_{(i+1):n} - X^2_{i:n})$, $i = 1, \cdots, n-1$.

Example 6.4.2. *Suppose $X_1, X_2, \cdots, X_n$ follows Rayleigh distribution with pdf $f(x) = 2\lambda x e^{-\lambda x^2}$, $x > 0$, $\lambda > 0$, then $U_i = (X^2_{(i+1):n} - X^2_{i:n})$ are independent and exponentially distributed with mean $\frac{1}{(n-i)\lambda}$, for $i = 1, \cdots, n-1$. Therefore, the mean and variance of $\xi J^{w}(F_n)$ for Rayleigh distribution are given by,*

$$\begin{aligned}E[\xi J^{w}(F_n)] &= -\frac{1}{4}\sum_{i=1}^{n-1}E(U_i)\left(1-\frac{i}{n}\right)^2 \\ &= -\frac{1}{8\lambda}\left(1-\frac{1}{n}\right)\end{aligned}$$

and

$$\begin{aligned}V[\xi J^{w}(F_n)] &= \frac{1}{16}\sum_{i=1}^{n-1}V(U_i)\left(1-\frac{i}{n}\right)^4 \\ &= \frac{1}{96\lambda^2 n}\left(1-\frac{1}{n}\right)\left(2-\frac{1}{n}\right).\end{aligned}$$

The true value of WSEx is $-\frac{1}{8\lambda}$ and $\lim_{n\to+\infty} E[\xi J^{w}(F_n)] = -\frac{1}{8\lambda}$.

We provide the numerical values for mean and variance of WSEx for Rayleigh distribution with λ=1 for n =10, 20, 30 40, 50 and 100 in Table 6.4. The mean closes to the true value and the variance decreases as sample size increases.

The following theorem address the convergence of $\xi J^{\varepsilon}(F_n)$ measure.

Table 6.4: *Mean and variance of $\xi J^w(F_n)$ for Rayleigh distribution.*

n	$E[\xi J^w(F_n)]$	$V[\xi J^w(F_n)]$	n	$E[\xi J^w(F_n)]$	$V[\xi J^w(F_n)]$
10	-0.1125	0.0018	40	-0.1218	0.0005
20	-0.1187	0.0010	50	-0.1225	0.0004
30	-0.1208	0.0007	100	-0.1237	0.0002

Theorem 6.4.1. *Let $X \in L^p$ for some $p > 2$ then $\xi J^{\varepsilon}(F_n) \to \xi J^{\varepsilon}(X)$ a.s.*

Proof. We have to prove that $\int_0^{+\infty} \varepsilon(x) S_n^2(x) dx \to \int_0^{+\infty} \varepsilon(x) S^2(x) dx$ a.s.
We have,

$$\int_0^{+\infty} \varepsilon(x) S_n^2(x) dx = \int_0^1 \varepsilon(x) S_n^2(x) dx + \int_1^{+\infty} \varepsilon(x) S_n^2(x) dx.$$

The first integral converges to $\int_0^1 \varepsilon(x) S^2(x) dx$ as $n \to +\infty$ by Glivenco-Cantelli and dominated convergence theorem. Now from Rao et al. (2004) we have for $x \in [1, +\infty]$,

$$S_n(x) \leq x^{-p} \left(\sup_n \frac{1}{n} \sum_{i=1}^{n} X_i^p \right).$$

Therefore, by dominated convergence theorem we have

$$\int_1^{+\infty} \varepsilon(x) S_n^2(x) dx \to \int_1^{+\infty} \varepsilon(x) S^2(x) dx.$$

Hence the result. □

Next we establish asymptotic normality of $\xi J^w(F_n)$ measure when the observations come from a Rayleigh distribution.

Theorem 6.4.2. *Let $X_1, X_2, \cdots, X_n$ be a rs from a Rayleigh distribution with pdf $f(x) = 2\lambda x e^{-\lambda x^2}$, $x > 0$, $\lambda > 0$, then*

$$\frac{\xi J^w(F_n) - E[\xi J^w(F_n)]}{[Var(\xi J^w(F_n))]^{1/2}} \xrightarrow{d} N(0,1).$$

Proof. Note that $\xi J^w(F_n)$ can be expressed as $\xi J^w(F_n) = \sum_{i=1}^{n-1} Z_i$, where $Z_i = -\frac{1}{4} U_i \left(1 - \frac{i}{n}\right)^2$. Now $U_1, \cdots, U_n$ are independent exponential rvs with mean $\frac{1}{\lambda(n-i)}$. So the mean and variance of Z_i can be obtained as

$$E(Z_i) = -\frac{1}{4\lambda n} \left(1 - \frac{i}{n}\right) \text{ and } Var(Z_i) = \frac{1}{16\lambda^2 n^2} \left(1 - \frac{i}{n}\right)^2.$$

For any exponentially distributed rv Z_i, $E[|Z_i - E(Z_i)|^3] = 2e^{-1}(6-e)[E(Z_i)]^3$ (Di Crescenzo and Longobardi, 2009). Denote $A^n_{i,p} = E[|Z_i - E(Z_i)|^p]$, then for large n we have,

$$\sum_{i=1}^{n} A^n_{i,2} = \sum_{i=1}^{n} E[|Z_i - E(Z_i)|^2] = \frac{1}{16\lambda^2 n^2}\sum_{i=1}^{n}\left(1-\frac{i}{n}\right)^2 \approx \frac{C_1}{16\lambda^2 n},$$

where $C_1 = \int_0^1 (1-x)^2 dx = \frac{1}{3}$.

$$\text{Again } \sum_{i=1}^{n} A^n_{i,3} = \sum_{i=1}^{n} E[|Z_i - E(Z_i)|^3] = \frac{2(6-e)}{64e\lambda^3 n^3}\sum_{i=1}^{n}\left(1-\frac{i}{n}\right)^3 \approx \frac{(6-e)C_2}{32e\lambda^3 n^2},$$

where $C_2 = \int_0^1 (1-x)^3 dx = \frac{1}{4}$. Now for some constant C,

$$\frac{(\sum_{i=1}^{n} A^n_{i,3})^{1/3}}{(\sum_{i=1}^{n} A^n_{i,2})^{1/2}} \approx Cn^{-\frac{1}{6}} \to 0, \quad \text{as } n \to +\infty.$$

So Lyapunov's condition for CLT is satisfied. Hence the proof. □

Similar result also holds for SEx measure when a rs comes from exponential distribution.

Theorem 6.4.3. *Let $X_1, X_2, \cdots, X_n$ be a rs from the exponential distribution with mean $\frac{1}{\lambda}$, then*

$$\frac{\xi J(F_n) - E(\xi J(F_n))}{(Var(\xi J(F_n)))^{1/2}} \xrightarrow{d} N(0,1).$$

Next we study non-parametric estimation of dynamic DWESEx measure. The estimator for DWESEx measure is defined as

$$\xi J^{\varepsilon}(F_n;t) = -\frac{1}{2}\int_t^{+\infty} \varepsilon(x)\left(\frac{S_n(x)}{S_n(t)}\right)^2 dx. \tag{6.20}$$

Suppose the sample values that are greater than t are $X_{j:n}, \cdots, X_{n:n}$ then Eq. (6.20) reduces to

$$\xi J^{\varepsilon}(F_n;t) = -\frac{1}{2}\sum_{i=j}^{n-1}[\eta(X_{(i+1):n}) - \eta(X_{i:n})]\left(\frac{n-i}{n-j+1}\right)^2. \tag{6.21}$$

In the following theorem we show that $\xi J^{\varepsilon}(F_n;t)$ converges to $\xi J^{\varepsilon}(X;t)$ a.s.

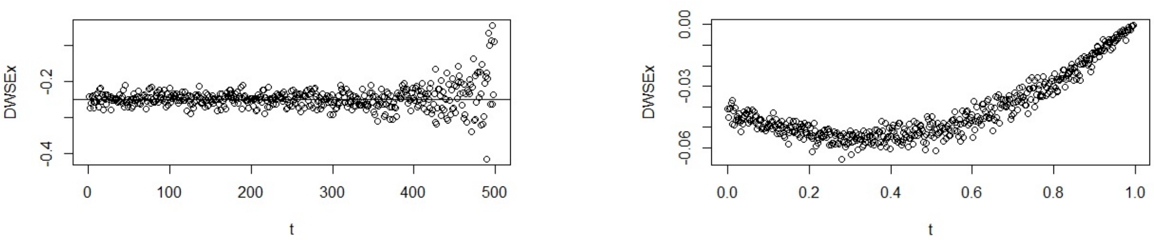
Theorem 6.4.4. *Let $X \in L^p$ for some $p > 2$ then $\xi J^{\varepsilon}(F_n;t) \to \xi J^{\varepsilon}(X;t)$ a.s.*

Proof.

$$
\begin{aligned}
-\xi J^{\varepsilon}(F_n;t) &= \frac{1}{2}\int_t^{+\infty} \varepsilon(x)\left(\frac{S_n(x)}{S_n(t)}\right)^2 dx \\
&= \frac{1}{2S_n^2(t)}\left[\int_0^{+\infty} \varepsilon(x)S_n^2(x)dx - \int_0^t \varepsilon(x)S_n^2(x)dx\right]. \qquad (6.22)
\end{aligned}
$$

From Theorem 6.4.1 we have, $\frac{1}{2}\int_0^{+\infty}\varepsilon(x)S_n^2(x)dx \to \frac{1}{2}\int_0^{+\infty}\varepsilon(x)S^2(x)dx$ a.s. Again, using dominated convergence theorem we have, $\frac{1}{2}\int_0^t \varepsilon(x)S_n^2(x)dx \to \frac{1}{2}\int_0^t \varepsilon(x)S^2(x)dx$ a.s. Also $S_n(t) \to S(t)$ a.s. Then using these in Eq. (6.22) the result follows. □

Fig. 6.3: *Non-parametric estimates of DWSEx for the Rayleigh distribution with parameter 1 and for the $U(0,1)$ distribution for different values of t.*



(a) *Empirical DWSEx for Rayleigh distribution.*

(b) *Empirical DWSEx for U(0,1) distribution.*

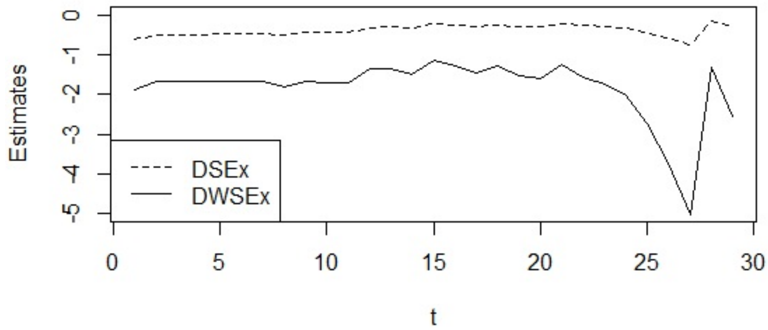
We simulate a rs of size 500 from the Rayleigh distribution with parameter 1 and from $U(0,1)$ distribution and calculate the non-parametric estimate of DWSEx for different values of t and plot them in Figure 6.3. We know that for Rayleigh distribution DWSEx is constant, which is -0.25. From Figure 6.1 it is observed that for $U(0,1)$ distribution DWSEx is not monotone, first it decreases then increases. The empirical DWSEx behaves similarly. From Figure 6.3 we see that for Rayleigh distribution most of the points lie around the line x = -0.25 and for $U(0,1)$ the non-parametric estimate of DWSEx behaves similar to that in Figure 6.1.

Data Analysis

We analyze the average daily wind speeds data of Best et al. (2010) for illustration. The mle of Rayleigh distribution for this data set is σ^2 = 10.1095. The parametric estimate of WSEx for Rayleigh distribution is -2.527 and the estimate of WSEx is -1.879. The parametric and non-parametric estimates of SEx for Rayleigh distribution are -1.409 and -0.616. So we see that WSEx is less than SEx which means that WSEx measures more information than

SEx. We also obtain the non-parametric estimates of the dynamic versions of WSEx and SEx measure based on this data and plot them in Figure 6.4.

Fig. 6.4: *Plots of non-parametric estimates of DWSEx and DSEx for the Rayleigh distribution for average wind speed data.*



6.4.2 Non-parametric estimation of WNCEx measure

In this section we propose an estimator for WNCEx measure analogous to that of WESEx measure and study its performance by simulation. The non-parametric estimator for WNCEx is defined as

$$\begin{aligned}\mathscr{C}^w(\hat{F}_n) &= \frac{1}{2}\int_0^\infty x[1-\hat{F}_n^2(x)]dx \\ &= \frac{1}{2}\sum_{k=1}^{n-1}\int_{X_{k:n}}^{X_{(k+1):n}} x\left[1-\left(\frac{k}{n}\right)^2\right]dx \\ &= \frac{1}{2}\sum_{k=1}^{n-1}\frac{X_{(k+1):n}^2-X_{k:n}^2}{2}\left[1-\left(\frac{k}{n}\right)^2\right] \\ &= \frac{1}{2}\sum_{k=1}^{n-1} Z_{k+1}\left[1-\left(\frac{k}{n}\right)^2\right], \end{aligned} \tag{6.23}$$

where $Z_{k+1}=\frac{X_{(k+1):n}^2-X_{k:n}^2}{2}$, $k=1,2,\cdots,n-1$.

Example 6.4.3. *Let $X_1,X_2,\cdots,X_n$ be a rs from a distribution with pdf $f(x)=2x$, $0<x<1$. Then X^2 has standard uniform distribution. Further $Z_{k+1}=\frac{X_{(k+1):n}^2-X_{k:n}^2}{2}$, $k=1,2,\cdots,n-1$, follows beta distribution with mean $\frac{1}{2(n+1)}$ and variance $\frac{n}{4(n+1)^2(n+2)}$. The mean and* *variance of $\mathscr{C}^w(\hat{F}_n)$ is given by*

$$\begin{aligned} E\left(\mathscr{C}^w(\hat{F}_n)\right) &= \frac{1}{2}\sum_{k=1}^{n-1} E[Z_{k+1}]\left[1-\left(\frac{k}{n}\right)^2\right] \\ &= \frac{1}{4(n+1)}\sum_{k=1}^{n-1}\left[1-\left(\frac{k}{n}\right)^2\right] \end{aligned}$$

and

$$\begin{aligned} Var\left(\mathscr{C}^w(\hat{F}_n)\right) &= \frac{1}{4}\sum_{k=1}^{n-1} Var[Z_{k+1}]\left[1-\left(\frac{k}{n}\right)^2\right]^2 \\ &= \frac{n}{16(n+1)^2(n+2)}\sum_{k=1}^{n-1}\left[1-\left(\frac{k}{n}\right)^2\right]^2. \end{aligned}$$

Lemma 6.4.1. *The estimator $\mathscr{C}^w(\hat{F}_n)$ is consistent.*

Proof. We can write,

$$\mathscr{C}^w(X) = \frac{1}{2}\int_0^1 x\left(1-F_n^2(x)\right)dx + \frac{1}{2}\int_1^{+\infty} x\left(1-F_n^2(x)\right)dx.$$

Now proceeding along the same line as in Theorem 6.4.1 we have $\mathscr{C}^w(\hat{F}_n) \to \mathscr{C}^w(X)$ as $n \to +\infty$. Hence the proof. □

Analogous to the WSEx measure, we can obtain a CLT for $\mathscr{C}^w(\hat{F}_n)$ when rs comes from Rayleigh distribution.

Theorem 6.4.5. *Let $X_1, X_2, \cdots, X_n$ be a rs from a Rayleigh distribution with pdf $f(x) = 2\lambda x e^{-\lambda x^2}$; $x > 0$, $\lambda > 0$. Then,*

$$\frac{\mathscr{C}^w(\hat{F}_n) - E\left[\mathscr{C}^w(\hat{F}_n)\right]}{\sqrt{Var\left[\mathscr{C}^w(\hat{F}_n)\right]}} \xrightarrow{d} N(0,1).$$

Proof. Proof follows proceeding along the same line as Theorem 6.4.2. □

We have found that the estimators for weighted entropy measures are asymptotically normally distributed when the sample comes from Rayleigh distribution. These results will be useful for testing goodness-of-fit for Rayleigh distribution. Due to the nature of these weighted measures, tests based on these measures will be less likely to be influenced by extreme observations.

We conduct a simulation study to assess the performance of the proposed estimator by means of bias and MSE. We generate 10000 samples from an exponential distribution with λ = 1, 2 and 3 and for n = 10, 20, 30, 40, 50 and 100 and calculate the bias and MSE. The results reported in Table 6.5. From Table 6.5, it is observed that as sample size increases bias and MSE decreases. Also bias and MSE decreases as λ increases.

Table 6.5: *Bias and MSE of $\mathscr{C}^w(\hat{F}_n)$ for exponential distribution with λ = 1 and 2 and for various n.*

	λ=1		λ=2		λ=3	
n	Bias	MSE	Bias	MSE	Bias	MSE
10	0.04240	0.38169	0.01035	0.02404	0.00445	0.00492
15	0.02026	0.27832	0.00601	0.01724	0.00332	0.00320
20	0.01394	0.20504	0.00496	0.01281	0.00267	0.00240
25	0.01283	0.16138	0.00370	0.01001	0.00165	0.00202
30	0.01193	0.13787	0.00349	0.00856	0.00120	0.00170

We analyse a real data set for illustrative purposes. The data represents failure times of 23 deep-groove ball bearings. Rayleigh distribution with cdf $F(x,\lambda) = 1 - e^{-\lambda x^2}$; $x,\ \lambda > 0$, provides a good fit for this data set, see Raqab (2002). The observations are: 0.1788, 0.2892, 0.33, 0.4152, 0.4212, 0.4560, 0.4848, 0.5184, 0.5196, 0.5412, 0.5556, 0.6780, 0.6864, 0.6864, 0.6888, 0.8412, 0.9312, 0.9864, 1.0512, 1.0584, 1.2792, 1.2804, 1.7340.
The mle of λ is $\hat{\lambda}$ = 1.5242. The parametric estimate of WNCEx is 0.2460 and $\mathscr{C}^w(\hat{F}_n)$ = 0.2380.

6.4.3 Recursive kernel estimation of WSEx for dependent observations

We considered Kernel based estimation of CREx measure in Chapter 6 for iid observation. Here we study estimation of WSEx measure when the underlying sample need not be independent. We consider recursive kernel function for this purpose. The kernel (Parzen, 1962) estimation of $f(x)$ is given by

$$f_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right),$$

where h_n is the bandwidth parameter. The kernel function K satisfies the following properties:

1. $K(v) \geq 0$ for all v.
2. $\int_{-\infty}^{+\infty} K(v)dv = 1$.
3. $K(\cdot)$ is symmetric about zero.

4. $K(\cdot)$ satisfies the Lipschitz condition, i.e. there exists a positive constant M such that $\mid K(v) - K(w) \mid \leq M \mid v - w \mid$.

Recursive kernel estimation of $f(x)$ was introduced by Wolverton and Wagner (1969) as

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i}K\left(\frac{x - X_i}{h_i}\right),$$

where K is a kernel of order s and $\{h_i\}$ is a sequence of real numbers satisfying
1. $\lim_{n\to+\infty} h_n = 0$;
2. $\lim_{n\to+\infty} nh_n = +\infty$;
3. $\lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n}\left(\frac{h_i}{h_n}\right)^l = \beta_l < +\infty,\ l = 1, 2, \cdots, s+1$;
4. $\lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n}\left(\frac{h_n}{h_i}\right)^l = \theta_l < +\infty,\ 1 \leq l < 2$.

The function $\hat{f}_n(x)$ has the following recursive property

$$\hat{f}_n(x) = \frac{n-1}{n}\hat{f}_{n-1}(x) + \frac{1}{nh_n}K\left(\frac{x - X_n}{h_n}\right).$$

Using this recursive kernel function, we define a non-parametric estimator for WSEx measure under α-mixing dependent condition. The α-mixing condition, also known as the strong mixing condition, was introduced by Rosenblatt (1956). Consider the following definition.

Definition 6.4.1. *Let $(\Omega, \mathscr{A}, \mathscr{P})$ be a probability space and for $-\infty < L < U < +\infty$, $\mathscr{A}_L^U$ is the σ-field of events generated by the rvs $\{X_i,\ L < i < U\}$. Let Q, R be two σ-fields belong to $\mathscr{A}$. Then the stationary process $\{X_i\}$ is said to follow α-mixing condition if for $Q \in \mathscr{A}_{-\infty}^{i}$ and $R \in \mathscr{A}_{i+n}^{\infty}$,*

$$sup|\mathscr{P}(Q \cap R) - \mathscr{P}(Q)\mathscr{P}(R)| = a(n) \to 0 \text{ as } n \to +\infty.$$

This mixing condition has many applications and various stochastic and time series processes follow the α-mixing condition, see, for example, Bradley (2007) and Doukhan (2012). For weakly stationary process under α-mixing setup, the bias and variance of the recursive kernel estimator are given by

$$Bias(\hat{f}_n(x)) \approx \frac{h_n^s C_s}{s!} f^{(s)}(x)\beta_s$$

and

$$Var(\hat{f}_n(x)) \approx \frac{\theta_1 C_k}{nh_n} f(x),$$

where $C_s = \int_{-\infty}^{+\infty} w^s K(w)dw$, $C_k = \int_{-\infty}^{+\infty} K^2(w)dw$ and $f^{(s)}(x)$ is the s-th derivative of f with respect to x. See Masry (1986) for further details. Now we define the recursive kernel based estimator of WSEx measure.

Definition 6.4.2. *Let $X_1, X_2, \cdots, X_n$ be identically distributed rvs not necessarily independent. A non-parametric estimator of WSEx is defined as*

$$\xi J^w(\hat{S}_n) = -\frac{1}{2}\int_0^{+\infty} x\hat{S}_n^2(x)dx, \tag{6.24}$$

where $\hat{S}_n(x) = \int_x^{+\infty} \hat{f}_n(w)dw$.

Maya et al. (2021) provided the bias and variance of $\hat{S}_n(x)$ as

$$Bias(\hat{S}_n(x)) \approx \frac{h_n^s C_s \beta_s}{s!}\int_x^{+\infty} f^{(s)}(w)dw \tag{6.25}$$

and

$$Var(\hat{S}_n(x)) \approx \frac{\theta_1 C_k S(x)}{nh_n}. \tag{6.26}$$

Now we study the consistency and asymptotic normality of the proposed estimator.

Theorem 6.4.6. *The recursive kernel based estimator $\xi J^w(\hat{S}_n)$ is consistent.*

Proof. Applying Taylor's expansion we get,

$$\hat{S}_n^2(x) \approx S^2(x) + 2S(x)(\hat{S}_n(x) - S(x)).$$

Now from Eq. (6.24), we obtain the bias and variance of $\xi J^w(\hat{S}_n)$ as,

$$\begin{aligned} Bias(\xi J^w(\hat{S}_n)) &\approx -\frac{1}{2}\int_0^{+\infty} 2x\, S(x)Bias(\hat{S}_n(x))dx \\ &= -\frac{h_n^s C_s \beta_s}{s!}\int_0^{+\infty} xS(x)\left(\int_x^{+\infty} f^{(s)}(w)dw\right)dx \end{aligned} \tag{6.27}$$

and

$$\begin{aligned} Var(\xi J^w(\hat{S}_n)) &\approx \frac{1}{4}\int_0^{+\infty} 4x^2 S^2(x)Var(\xi J^w(\hat{S}_n))dx \\ &= \frac{\theta_1 C_k}{nh_n}\int_0^{+\infty} x^2 S^3(x)dx. \end{aligned} \tag{6.28}$$

□

It is easy to see that, as $n \to +\infty$, bias and variance of $\xi J^w(\hat{S}_n)$ reduce to zero. So the MSE of $\xi J^w(\hat{S}_n)$ also approaches to zero as $n \to +\infty$. Hence $\xi J^w(\hat{S}_n)$ is consistent.

Theorem 6.4.7. *Let $\xi J^w(\hat{S}_n)$ be a recursive kernel estimator of $\xi J^w(X)$. Then as $n \to +\infty$,*

$$(nh_n)^{\frac{1}{2}} \left(\frac{\xi J^w(\hat{S}_n) - \xi J^w(x)}{\hat{\sigma}} \right) \xrightarrow{d} N(0,1).$$

Proof. Proof follows using the asymptotic normality of $\hat{f}_n(x)$. □

Simulation study and data analysis

We conduct a simulation study to assess the performance of the proposed kernel based estimator. We consider standard exponential distribution and generate $\{X_i\}$ from AR(1) process i.e. $X_i = \rho X_{i-1} + \varepsilon_i$, with correlation coefficient (ρ) 0.10, 0.20 and 0.30, respectively. We calculated Bias and MSE for sample size 50, 100 and 200 and provided them in Table 6.6. From Table 6.6, we see that as sample size increases MSE decreases and when correlation coefficient of AR(1) model increases, MSE also increases.

Table 6.6: *Bias and MSEs of $\xi J^w(\hat{S}_n)$.*

ρ	$n = 50$		$n = 100$		$n = 200$	
	Bias	MSE	Bias	MSE	Bias	MSE
0.10	-0.04320	0.00431	-0.03916	0.00266	-0.03780	0.00200
0.20	-0.08692	0.01146	-0.08631	0.00932	-0.08620	0.00839
0.30	-0.1590	0.03202	-0.15770	0.02817	-0.15582	0.02601

For illustration, we analyze real data relating to relief times of 20 patients receiving an analgesic. This data set was reported by Gross and Clark (1975):
1.1, 1.4, 1.3, 1.7, 1.9, 1.8, 1.6, 2.2, 1.7, 4.1, 1.8, 1.5, 1.2, 1.4, 3.0, 1.7, 2.3, 1.6, 2.0.
This data can be fitted by a 3-parameter Weibull distribution with shape, scale and location parameters 0.9924, 0.8755 and 1.0568, respectively. The true value of $\xi J^w(\hat{S}_n)$ is -0.3270. Bias and MSE have been calculated from 100 Bootstrap samples of size 20 as 0.1869 and 0.0542, respectively.

6.5 Applications

In this section, we consider three applications in different fields namely, model discrimination, quantitative risk analysis and uniformity test. This section has been divided into three subsections. First two subsections consist of the application of WSEx in model discrimination and as risk measure. In the final subsection we study WNCEx based goodness-of-fit tests for uniform distribution.

6.5.1 Discriminating between distributions

In reliability analysis one important problem is to choose the correct lifetime model for a given data set. One popular method is to use ratio of maximized likelihood (RML) to discriminate the distributions (Gupta and Kundu, 2003). According to Burnham and Anderson (2004), for a given data set the best fitted model is the one that contains maximum uncertainty. Now both SEx and WSEx are uncertainty measures and lower values of them indicate higher uncertainty. Using SEx and WSEx measures, we can discriminate between two distributions for a given data set. Consider the groove ball bearings data studied by Gupta and Kundu (2003) provided below:
17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.80, 51.84, 51.96, 54.12, 55.56, 67.80, 68.64, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04 and 173.40.
They showed that both Weibull (WE) and generalized exponential (GE) distribution fit the data set reasonably well. Weibull distribution has the cdf $F_{WE}(y,\beta,\theta) = 1-e^{-(\theta y)^{\beta}}$, $y,\beta,\theta > 0$ and GE distribution has the cdf $F_{GE}(y,\alpha,\lambda) = (1-e^{-(\lambda y)})^{\alpha}$, $y,\alpha,\lambda > 0$. Now the mles of the parameters of WE distribution are $\hat{\beta}$ = 2.1031 and $\hat{\theta}$ = 0.0122 and the log-likelihood (LL) value is -113.69. The mles for the parameters of GE distribution are $\hat{\alpha}$ = 5.2589 and $\hat{\lambda}$ = 0.0314 and LL is -112.98. GE distribution is a better fit since it has higher LL value.

Now let us consider the SEx and WSEx measures for both fitted distributions. We calculate SEx and WSEx and provide them in Table 6.7 along with LL values, Akaike information criterion (AIC) and Bayesian information criterion (BIC).

Table 6.7: *Discrimination measures for fitted WE and GE distributions.*

Discrimination measures	WE($\hat{\beta},\hat{\theta}$)	GE($\hat{\alpha},\hat{\lambda}$)	Difference (WE-GE)
SEx	-26.1068	-26.7731	0.6663
WSEx	-851.7	-856.2	4.5
LL	-113.69	-112.98	-0.71
AIC	231.38	229.96	1.42
BIC	233.65	232.23	1.42

From the table it is observed that, both SEx and WSEx for GE distribution are less than the ones of WE distribution. So according to SEx and WSEx measures, GE distribution is a better fit. AIC and BIC also support this claim. Here also we see that WSEx measures more information than SEx as it measures both quantitative and qualitative characteristics of information. However, the difference between RML and SEx measures for the distributions is negligible. But WSEx clearly suggests that GE distribution fits the data better than WE distribution. So in situations like this, we can use WSEx for model selection.

6.5.2 Risk measure

Use of cumulative entropy measures as an alternative risk measure has gained quite popularity in recent years. Yang (2012) first studied CRE as an alternative to standard deviation (s.d.) for heavy tailed distributions and Psarrakos and Toomaj (2017) studied generalized CRE as a risk measure. For more works on cumulative entropy as a risk measure see Tahmasebi and Parsa (2019), Tahmasebi and Toomaj (2022). Weighted cumulative entropy measures have also been used as an alternative risk measure in recent years. Recently, Kayal (2018) considered weighted CRE as a risk measure. We consider WSEx as a risk measure and apply it for analysing stock return data.

Let X be a rv describing payoff. Positive values of X indicates gain and negative values of X indicates loss. Let Ω be the set of all values of X. A risk measure μ is a mapping $\mu : \Omega \to R$ that satisfies some of the following properties (Ramsay, 1995):

- Sub-additivity. For $X,Y \in \Omega$, $\mu(X+Y) \leq \mu(X) + \mu(Y)$.
- Consistency. For $X \in \Omega$ and $a \in R$, $\mu(X+a) = \mu(X)$.
- Positive Homogeneity. $X \in \Omega$ and $c > 0$, $\mu(cX) = c\mu(X)$.
- Monotonicity. For $X,Y \in \Omega$ and $X \leq Y$, $\mu(X) \geq \mu(Y)$.
- Convexity. For $X,Y \in \Omega$ and $\lambda \in [0,1]$, $\mu(\lambda X + (1-\lambda)Y) \leq \lambda\mu(X) + (1-\lambda)\mu(Y)$.

Jahanshahi et al. (2020) used $\mid \xi J(X) \mid$ as a risk measure which possesses consistency and positive homogeneity properties of risk measure. Under the assumption of usual stochastic order, $\mid \xi J(X) \mid$ also possesses monotonicity property. We consider $\mid \xi J^w(X) \mid$ as an alternative risk measure and compare it with the s.d., variance and $\mid \xi J(X) \mid$. Now $\mid \xi J^w(X) \mid$ being a shift-dependent measure, it does not satisfy the consistency property of the risk measure. In order to satisfy positive homogeneity properties, we consider $\delta(X) = \sqrt{\mid \xi J^w(X) \mid}$ as a risk measure. Now from Lemma 6.1.1, we see that $\delta(X)$ satisfies the positive homogeneity property. From Proposition 6.2.1 (iv), we find that $\delta(X)$ possesses monotonicity property under usual stochastic ordering.

Jahanshahi et al. (2020) showed that the sub-additivity property for $|\xi J(X)|$ can hold under some specific assumptions. They proved that for two non-negative independent rvs X and Y with same right-end support $u_x = u_y < +\infty$ and if X and Y have log-concave densities, then $\xi J(X+Y) \geq \xi J(X) + \xi J(Y)$. This follows from the fact that, under the above assumptions, $X \overset{st}{\geq} X+Y$ and $Y \overset{st}{\geq} X+Y$. From Proposition 6.2.1 (iv) we get $\xi J^w(X+Y) \geq \xi J^w(X)(\xi J^w(Y))$. Since $\xi J^w(X)$ is always negative, we have, $\xi J^w(X+Y) \geq \xi J^w(X) + \xi J^w(Y)$. Therefore,

$$
\begin{aligned}
|\xi J^w(X+Y)| &\leq |\xi J^w(X)| + |\xi J^w(Y)| \\
\Rightarrow \delta(X+Y) &\leq \sqrt{|\xi J^w(X)| + |\xi J^w(Y)|} \leq \delta(X) + \delta(Y).
\end{aligned}
$$

So $\delta(X)$ also satisfies sub-additivity property as $|\xi J(X)|$. Consider the following example.

Example 6.5.1. *Suppose X has a distribution with sf $S(x) = (1-x)^3$, $0 < x < 1$. Then s.d.(X) = 0.1936, $|\xi J(X)|$ = 0.0714 and $\delta(X)$ = 0.0945. So $|\xi J(X)| < \delta(X) <$s.d.(X).*

Table 6.8: *Variance, s.d., $|\xi J(X)|$ and $\delta(X)$ of normalized log-rate of returns of each months in 2015 from BSE SENSEX.*

Month	s.d.(X)	$Var(X)$	$\|\xi J(X)\|$	$\delta(X)$
January	0.2147	0.0461	0.2376	0.2581
February	0.2458	0.0604	0.1922	0.2170
March	0.2388	0.0570	0.1840	0.2098
April	0.2939	0.0864	0.2298	0.2611
May	0.2632	0.0693	0.2304	0.2608
June	0.2559	0.0655	0.2338	0.2622
July	0.2696	0.0727	0.2359	0.2667
August	0.2089	0.0436	0.3077	0.3270
September	0.3022	0.0914	0.2024	0.2424
October	0.2640	0.0697	0.0801	0.1165
November	0.2703	0.0731	0.1668	0.2015
December	0.3420	0.1170	0.1358	0.1904

Now we analyze a real life data set. The data on daily BSE SENSEX are collected from Yahoo Finance website for the period from 1^{st} January to 31^{st} December 2015. Let Z_t be the closing price of day t. The log-rate of return is defined as $R_t = \log \frac{Z_t}{Z_{t-1}}$. Consider the normalizing transformation of R_t as $X_t = \frac{R_t - min(R_t)}{max(R_t) - min(R_t)}$. Here X denotes the normalized log-rate of returns. We obtain the variance, s.d., $|\xi J(X)|$ and $\delta(X)$ for each month in the year 2015. The results are provided in Table 6.8.

From Table 6.8 it is observed that $\delta(X)$ is greater than $|\xi J(X)|$ for all the months. When $|\xi J(X)|$ increases $\delta(X)$ also increases. High values of the risk measures imply high volatility in the market. Note that $\delta(X)$ measures more variability than $\mid \xi J(X) \mid$. In terms of $\delta(X)$ and $\mid \xi J(X) \mid$, the month of August has the highest volatility while in terms of s.d.(X), the month of December contains highest volatility in the market. We can use $\delta(X)$ and $\mid \xi J(X) \mid$ as alternative risk measures for heavy tailed distributions for which s.d. is not a suitable risk measure.

6.5.3 Test of uniformity

Uniform distribution is the simplest model among all probability distributions. Perhaps the most important application of uniform distribution is random number generation. The cdf of any distribution is uniformly distributed in the interval [0,1]. This property has been widely used to simulate rs from various distributions. So the problem of testing uniformity is of high importance among statisticians and it has been studied quite extensively in the literature. Stephens (1974) studied uniformity test using edf based statistics and Dudewicz and Van Der Meulen (1981) proposed entropy based uniformity test. They found that entropy based tests perform better than several other popular tests. Recently, Noughabi (2021, 2022) applied CRE and NCEx to test uniformity and compare the power of the test with several alternatives. Motivated by their work, we proposed a uniformity test using the estimator of the WNCEx measure and compared the performance of our test with the test based on NCEx and other popular tests.

Suppose $X_1, X_2, \cdots, X_n$ is a rs from a continuous distribution with cdf F and pdf f concentrated on [0,1], i.e. $f(x) = 0$ if $x \notin [0,1]$. Also let $X_{1:n}, X_{2:n}, \cdots, X_{n:n}$ be the corresponding order statistics. We want to test whether the sample comes from a uniform distribution, denoted by U(0,1). So the hypothesis is

$$H_0 : f(x) \sim U(0,1) \quad vs. \quad H_1 : f(x) \nsim U(0,1).$$

Consider the non-parametric estimator of WNCEx in Eq. (6.23)

$$\mathscr{C}^w(\hat{F}_n) = \frac{1}{2} \sum_{k=1}^{n-1} \frac{X_{(k+1):n}^2 - X_{k:n}^2}{2} \left[1 - \left(\frac{k}{n}\right)^2\right].$$

Consider the following lemma which will be useful in developing the test.

Lemma 6.5.1. *Let $X_1, X_2, \cdots, X_n$ be a rs drawn from a continuous distribution concentrated on [0,1]. Then,* $0 \leq \mathscr{C}^w(\hat{F}_n) \leq \frac{1}{4}$.

Table 6.9: *Critical values of* $\mathscr{C}^w(\hat{F}_n)$.

n	$\mathscr{T}_{0.025,n}$	$\mathscr{T}_{0.975,n}$	$\mathscr{T}_{0.05,n}$	$\mathscr{T}_{0.95,n}$	n	$\mathscr{T}_{0.025,n}$	$\mathscr{T}_{0.975,n}$	$\mathscr{T}_{0.05,n}$	$\mathscr{T}_{0.95,n}$
4	0.02118	0.17785	0.02983	0.16697	27	0.08748	0.15628	0.09227	0.15139
5	0.03088	0.17607	0.04062	0.16653	28	0.08769	0.15616	0.09298	0.15089
6	0.04021	0.17467	0.04814	0.16539	29	0.08825	0.15589	0.09414	0.15073
7	0.04551	0.17417	0.05547	0.16499	30	0.08971	0.15507	0.09493	0.15043
8	0.05085	0.17308	0.06201	0.16449	32	0.09105	0.15500	0.09551	0.14980
9	0.05620	0.17103	0.06594	0.16304	34	0.09164	0.15432	0.09654	0.14933
10	0.05914	0.16945	0.06821	0.16271	36	0.09220	0.15363	0.09776	0.14857
11	0.06217	0.16910	0.07295	0.16164	38	0.09340	0.15221	0.09804	0.14832
12	0.06681	0.16632	0.07454	0.16108	40	0.09397	0.15177	0.09899	0.14743
13	0.06865	0.16623	0.07667	0.16016	45	0.09541	0.15055	0.10108	0.14633
14	0.06899	0.16549	0.07867	0.15891	50	0.09791	0.14934	0.10171	0.14526
15	0.07284	0.16535	0.08059	0.15864	55	0.09917	0.14820	0.10297	0.14492
16	0.07514	0.16476	0.08252	0.15706	60	0.09974	0.14783	0.10405	0.14374
17	0.07695	0.16424	0.08346	0.15636	65	0.10074	0.14659	0.10509	0.14349
18	0.07844	0.16195	0.08441	0.15566	70	0.10221	0.14589	0.10610	0.14267
19	0.07435	0.16182	0.08638	0.15548	75	0.10389	0.14544	0.10621	0.14229
20	0.08007	0.16052	0.08744	0.15491	80	0.10397	0.14455	0.10697	0.14131
21	0.08147	0.16025	0.08851	0.15433	85	0.10424	0.14420	0.10760	0.14090
22	0.08215	0.15952	0.08995	0.15412	90	0.10452	0.14369	0.10783	0.14043
23	0.08425	0.15912	0.09099	0.15308	95	0.10545	0.14345	0.10860	0.14040
24	0.08455	0.15821	0.09091	0.15297	100	0.10546	0.14283	0.10885	0.13960
25	0.08562	0.15766	0.09179	0.15250	110	0.10688	0.14171	0.10965	0.13908
26	0.08667	0.15720	0.09181	0.15159	120	0.10798	0.14126	0.11063	0.13869

Proof. The function $g(v) = \frac{1-v^2}{2}$, $0 \le v \le 1$ has maximum value $\frac{1}{2}$. So we have,

$$\mathscr{C}^w(\hat{F}_n) \le \frac{1}{4}\sum_{k=1}^{n-1}\left(X^2_{(k+1):n} - X^2_{k:n}\right) \le \frac{X^2_{n:n} - X^2_{1:n}}{4} \le \frac{1}{4}.$$

Hence the proof. □

For U(0,1) distribution, $\mathscr{C}^w$ = 0.125 which lies exactly between (0, 0.25). So we can use $\mathscr{C}^w(\hat{F}_n)$ as a test statistic. We will reject the null hypothesis if $\mathscr{C}^w(\hat{F}_n)$ is large or small. The critical region for a sample of sixe n and significance level α is

$$\mathscr{C}^w(\hat{F}_n) \le T_{\frac{\alpha}{2},n} \ or\ \mathscr{C}^w(\hat{F}_n) \ge T_{1-\frac{\alpha}{2},n},$$

where $T_{\alpha,n}$($T_{1-\alpha,n}$) is the lower (upper) α% quantile point of the distribution of $\mathscr{C}^w(\hat{F}_n)$.

Lemma 6.5.2. *The test based on* $\mathscr{C}^w(\hat{F}_n)$ *is consistent.*

Proof. From Lemma 6.4.1 we know $\mathscr{C}^w(\hat{F}_n) \to \mathscr{C}^w(X)$ a.s. Therefore, under H_0, $\mathscr{C}^w(\hat{F}_n)$

converges to the true value 0.125. Hence the result. □

The exact distribution of $\mathscr{C}^w(\hat{F}_n)$ is intractable so we obtain the critical points by Monte-Carlo simulation. We generate 10000 samples from U(0,1) distribution and compute critical points at 5% and 10% level of significance for different sample sizes and present them in Table 6.9. To obtain the power of the proposed test, we consider the following alternatives given in Stephens (1974). These alternative distributions are specifically developed for testing uniformity. Stephens argues that, alternatives A and B provide points close to 0 and 1, respectively and alternative C gives two points close to 0 and 1.

$$
\begin{aligned}
A_j : F(z) &= 1-(1-z)^j, \quad 0 \le z \le 1 \; (j = 1.5,\, 2) \\
B_j : F(z) &= \begin{cases} 2^{j-1} z^j, & 0 \le z \le 0.5 \\ & \qquad (j = 1.5,\, 2,\, 3) \\ 1-2^{j-1}(1-z)^j, & 0.5 \le z \le 1 \end{cases} \\
C_j : F(z) &= \begin{cases} 0.5-2^{j-1}(0.5-z)^j, & 0 \le z \le 0.5 \\ & \qquad (j = 1.5,\, 2) \\ 0.5+2^{j-1}(0.5-z)^j, & 0.5 \le z \le 1 \end{cases}
\end{aligned}
$$

The performance of our test is compared with some omnibus tests such as Kolmogorov-Smirnov (KS), Cramer-von Mises (CvM), Kuiper and Anderson-Darling (AD). These tests are widely popular among practitioners across various fields. Also we compare the performance of our test with some tests that are specially developed for testing uniformity. D'Agostino and Stephens (1986) discussed various directed tests for testing uniformity. For power comparison we use two tests based on order statistics and sample entropy. The first test statistic is $T = \frac{1}{n}\sum_{i=1}^{n} v_i^2$ where $v_i = U_{(i)} - \frac{i}{n+1}$ and $U_{(1)}, U_{(2)}, \cdots, U_{(n)}$ are the order statistics of the U(0,1) distribution. We reject the null hypothesis for large values of T. The second statistics is based on the estimation of entropy introduced by Vasicek (1976). This test statistic was used for uniformity test by Dudewicz and Van Der Meulen (1981). The test statistic is defined as

$$H_{mn} = \frac{1}{n}\sum_{k=1}^{n} \log[\frac{n}{2m}(X_{(k+m)} - X_{(k-m)})],$$

where m is a positive integer less than $\frac{n}{2}$. Under H_0, H_{mn} converges to zero and it is less than zero otherwise. So the null hypothesis is rejected for small values for H_{mn}. Note that this test is based on higher order sample spacings. The test is called the ENT test and Dudewicz and Van Der Meulen (1981) found that the ENT test performed significantly better than many popular tests. We use another uniformity test for power comparison which was introduced recently by Noughabi (2021). The test is based on the non-parametric estimator of NCEx measure which is defined as

$$\mathscr{C}\mathscr{J}(\hat{F}_n) = \frac{1}{2}\sum_{k=1}^{n-1}\left(T_{(k+1)} - T_{(k)}\right)\left[1 - \left(\frac{k}{n}\right)^2\right].$$

This test is based on sample spacings. The null hypothesis is rejected if $\mathscr{C}\mathscr{J}(\hat{F}_n)$ is large or small.

Table 6.10: *Power of the test for various sample sizes.*

n	Alternatives	$\mathscr{C}\mathscr{J}^w(\hat{F}_n)$	$\mathscr{C}\mathscr{J}(\hat{F}_n)$	T	ENT	KS	CvM	Kuiper	AD
10	$A_{1.5}$	0.2375	0.0658	0.2162	0.1420	0.1489	0.1660	0.0962	0.1608
	A_2	0.4924	0.1328	0.4728	0.2893	0.3762	0.4378	0.2303	0.4091
	$B_{1.5}$	0.1423	0.1188	0.1268	0.1884	0.0421	0.0330	0.1272	0.0200
	B_2	0.2421	0.3188	0.2253	0.4363	0.0398	0.0226	0.2962	0.0081
	B_3	0.3722	0.7505	0.4504	0.7958	0.0908	0.0486	0.7061	0.0182
	$C_{1.5}$	0.0458	0.1244	0.0218	0.0302	0.1137	0.0978	0.1329	0.1354
	C_2	0.0780	0.2524	0.0126	0.0330	0.2003	0.1491	0.2875	0.2180
20	$A_{1.5}$	0.4356	0.0823	0.3573	0.2411	0.2841	0.3184	0.1712	0.3065
	A_2	0.8183	0.2208	0.7502	0.6207	0.7038	0.7726	0.4668	0.7480
	$B_{1.5}$	0.1584	0.2683	0.1525	0.3039	0.0555	0.0502	0.2256	0.0258
	B_2	0.2850	0.7025	0.3579	0.7115	0.1176	0.0972	0.5804	0.1003
	B_3	0.4623	0.9887	0.7552	0.9905	0.4201	0.5089	0.9828	0.5639
	$C_{1.5}$	0.0586	0.2037	0.0383	0.0589	0.1455	0.1224	0.2355	0.1633
	C_2	0.1158	0.4265	0.0958	0.1461	0.3092	0.2538	0.6006	0.3846
30	$A_{1.5}$	0.5924	0.0908	0.4784	0.3268	0.3988	0.4713	0.2356	0.4667
	A_2	0.9389	0.2929	0.8835	0.8201	0.8617	0.9192	0.6660	0.9173
	$B_{1.5}$	0.1907	0.4048	0.1846	0.4047	0.0788	0.0585	0.3226	0.0600
	B_2	0.3482	0.8873	0.4947	0.8849	0.2413	0.2568	0.8044	0.3011
	B_3	0.5560	1	0.9128	0.9991	0.7320	0.8915	0.9969	0.9259
	$C_{1.5}$	0.0814	0.2756	0.0547	0.1112	0.1836	0.1397	0.3184	0.2022
	C_2	0.1730	0.5914	0.2125	0.3577	0.4403	0.4016	0.8129	0.5296

We calculate power for n = 10, 20 and 30 and present the results in Table 6.10. The proposed tests perform better than other tests for A_j alternatives. Tests best on WNCEx performs better when alternative distributions have points close to 0 or 1. This phenomenon is quite natural because weighted information measures put importance to the observed values of the rvs. So in dealing with inferential problems involving extreme observations of tail probabilities, weighted information measures can be useful instead of non-weighted information measures.

Remark 6.5.1. *Consider the problem of goodness-of-fit tests for U(a,b) distribution. The U(a,b) has the pdf $f(t) = \frac{1}{b-a}$; $a < t < b$. If the parameters (a,b) are known then the transformation $V = \frac{U-a}{b-a}$ gives a random sample from U(0,1) distribution. So we can easily apply these tests by transforming to standard uniform samples. The power of the tests will not be effected for testing U(a,b) distribution when (a,b) are known.*

If the parameters are not known then we will estimate them from the data using the maximum likelihood estimators. Suppose $U_1, U_2, \cdots, U_n$ are a random sample of size n from the U(a,b) distribution. The maximum likelihood estimators for a and b are the smallest and the largest order statistics $U_{(1)}$ and $U_{(n)}$, respectively. Now the transformation $V_i = \frac{U_{(i+1)}-U_{(1)}}{U_{(n)}-U_{(1)}}$, $i = 1, 2, \cdots, n-2$ yields a random sample of size $n-2$ from U(0,1) distribution. So this transformation converts the problem of testing U(a,b) from a sample of size n, to a problem of testing U(0,1) from a sample of size $n-2$. So we can apply the proposed test for testing uniformity with unknown parameters as well. Here the sample size reduces from n to $n-2$. Therefore, the power of the test may be slightly low for small sample sizes. However, for large or moderately large samples, the power will be more or less the same.

6.6 Conclusion

In this chapter, we considered weighted survival extropy measure and further generalized this measure by taking a non-negative continuous function as weight function instead of X. This measure is called weighted extended survival extropy measure. Also we proposed its dynamic version and studied various properties of these generalized information measures by considering different weight functions. Also, we introduced weighted negative cumulative extropy measure and studied various properties. Non-parametric estimations of these proposed measures are studied in detail. First we considered estimation of these measures when the underlying observations are iid. Also we proposed a recursive kernel based estimation for weighted survival extropy measure when sample obeys α-mixing dependent condition. The performance of these estimators are assessed by simulation and real data sets are also analyzed for illustrations.

We proposed two potential applications of weighted survival extropy measure in model discrimination and quantitative risk analysis. As an application of weighted negative cumulative extropy measure, we developed a uniformity test. The power of the proposed test is compared with some omnibus tests and with some specific tests of uniformity. The proposed test performed better than the other tests when alternative distribution has observations closer to the smallest extreme point.

Chapter 7

Application of cumulative entropy measures in life testing

APPLICATIONS of information measures in life-testing mainly focused on developing goodness-of-fit tests for various lifetime models under complete as well as censored data. These topics have been addressed in the literature by many authors. Park (2005) first developed an entropy based exponentiality test under Type-II censoring and Balakrishnan et al. (2007) extended this to the progressive Type-II censoring case. Goodness-of-fit tests under censored data are also developed using divergence measures (CRKL and CKL) that are based on cumulative entropy measures. Using CRKL information measure, Park and Lim (2015) proposed an exponentiality test for Type-II censored data and, Baratpour and Rad (2016) studied exponentiality tests for progressive Type-II censored data using both CRKL and CKL measures. In Chapter 4, we have proposed new KL type information measures WCRKL and WCKL based on WCRE and WCE measures and obtained an exponentiality test for complete, Type-I and Type-II censored data.

In life-testing, censored life tests are considered due to time and cost constraints. An important problem in life-testing is to design the censoring experiment. Some questions that frequently arise in life-testing are "How long an experimenter should run the experiment?" and "How many units are required for the life-tests and how many units to be censored?". An increase in the number of units and test time will result in increasing the total information of the experiment. Increasing information implies that the precision of estimating parameters associated with the life-test will rise. But this will also increase the cost of running the experiment which is not ideal in any practical scenario. Optimal design of life-testing experiments are usually obtained by minimizing variance or equivalently maximizing information of the life-test and by minimizing the total cost of the experiment. In this Chapter, we introduce design criteria based on cumulative entropy measures for progressive type-II censored experiments.

In many instances, it is required to remove the surviving units from a life-test before the end of the test due to constraints in testing facilities and getting quick information on lifetimes. Type I, Type II and hybrid censoring schemes are the common censoring schemes used in life-tests. However, these censoring schemes do not allow the removal of surviving units during the test. Progressive censoring allows removal of surviving units before the end of the test (Cohen, 1963, 1965). Progressive Type I and Progressive Type II (PCII) censoring schemes are the common progressive censoring schemes. This flexibility of removing items during the experiment makes progressive censoring very effective but it also increases the number of possible censoring schemes. So the problem of choosing optimal progressive censoring schemes has gained a lot of attention in the literature. A PCII censoring scheme is described as follows. Suppose n units are put on a test and a pre-fixed number m of failures is allowed. Let $R_1, R_2, \cdots, R_m$ be prefixed integers such that $R_1 + R_2 + \cdots + R_m = n - m$. When first failure occurs, R_1 of the remaining $n-1$ surviving units are randomly removed from the test. Then R_2 of the remaining $n - R_1 - 2$ units are randomly removed at the time of second failure. Finally, at the time of the m-th failure, all the remaining $R_m = n - m - \sum_{i=1}^{m-1} R_i$ units are removed from the test. The failure times are denoted by $X_{1:m:n}, \cdots, X_{m:m:n}$. Note that these are known as progressively type-II censored order statistics (PCOS). For details on PCII censoring scheme and data, see Balakrishnan and Aggarwala (2000) and Balakrishnan and Cramer (2014). Complete sample observations are obtained when $m = n$ and $R_i = 0, \forall i$. If $R_i = 0$ for $i = 1, \cdots, m-1$ and $R_m = n - m$, it reduces to the Type-II censoring scheme. A schematic representation of PCII censoring scheme in presented in Figure 7.1. In a PCII censored experiment, a reliability engineer needs to choose the

Fig. 7.1: *Schematic representation of progressive Type-II censoring scheme.*

design parameters of the experiment $(n, m, R_1, \cdots, R_m)$ beforehand. An important problem in life-testing is designing the experiment. For example, in PCII censored experiments, the choice of the design parameters will affect the reliability of the products and thus, it is important to choose the design parameters appropriately. Optimal PCII censoring schemes usually obtained by optimizing some design criteria set by the experimenter. Usually, design criteria are developed based on the information and cost of the experiment. One wants to maximize the overall information contained in the experiment or equivalently reduce the variability. The most commonly used criteria for this purpose are A and D-optimality criteria which minimize the trace and the determinant of the variance-covariance matrix of the estimated model parameters, respectively. For example, see Ng et al. (2004) and Dahmen et al. (2012). Recently, Pradhan and Kundu (2009, 2013) obtained optimal PCII censoring schemes by minimizing a new variability measure $\int_0^1 Var(\log \hat{T}_p)dp$, where $\hat{T}_p$ is the MLE of the p-th quantile of the underlying distribution. This measure is independent of p. Another approach of determining optimal design is to choose the design that minimizes the total cost of the experiment. Bhattacharya et al. (2014) first developed optimal PCII censored design by minimizing the total cost of the experiment.

However, these criteria have some disadvantages. The A and D-optimality criteria are not scale invariant. The quantile based criterion of Pradhan and Kundu (2009, 2013) is an important design criterion as it is not influenced by the extreme observations and also it is scale invariant. One drawback of this criterion is that it is based on the asymptotic result. It is important to note that for fixed values of n and m, there exists $\binom{n-1}{m-1}$ number of different censoring schemes. Therefore the number of progressive censoring schemes becomes very large even for moderate n and m. So from a practical point of view, it is not ideal to use large or moderately large samples for PCII censored experiments. If sample size is not large then using asymptotic results based criteria may not be accurate. To overcome this problem, we propose cumulative entropy based design criteria that are independent of asymptotic results.

In this chapter, we consider optimal design of PCII censored experiment using criteria based on cumulative entropy measures i.e. CRE and CE measures. As we have mentioned earlier, CRE (CE) is an alternative information measure and the large value of CRE (CE) of an experiment means more information contained in that experiment. We obtain optimal design by maximizing the joint CRE (CE) of PCII censored experiment and obtain a constraint design of the experiment by maximizing the joint CRE subject to a cost constraint. A new design strategy is implemented following the procedure of Bhattacharya (2020), called compound optimal design (COD), in which we optimize two competitive criteria simultaneously in order to achieve a trade off between them. The rest of the chapter is organised as follows.

We discuss information measures for consecutive order statistics and PCOS in Section 7.1. We propose design criteria and study maximum cumulative information designs of PCII censored experiments in Section 7.2. Maximum cumulative information design subject to a cost constraint is discussed in Section 7.3. The COD is implemented in Section 7.4. Some concluding remarks are made in Section 7.5.

7.1 Cumulative entropy measures of consecutive order statistics and PCOS

In this section, we discuss entropy, CRE and CE measures for consecutive order statistics and PCOS. We will use these results to construct design criteria for optimum design of PCII censored experiments. Park (2005) obtained the expression of joint entropy of first r order statistics as

$$H_{1\cdots r:n} = -(\log n + \cdots + \log(n-r+1)) + r - n\int_0^{+\infty}(1-F_{r:n-1}(x))f(x)\log h(x)dx,$$

where $h(x) = \frac{f(x)}{S(x)}$ is the hazard rate of X. Extending this result to the PCII censoring case, Balakrishnan et al. (2007) provided the expression of joint entropy of PCOS as

$$H_{1\cdots m:m:n} = -\log c + m - \int_0^{+\infty}\sum_{i=1}^{m} f_{X_{i:m:n}}\log h(x)dx,$$

where $c = n(n-R_1-1)\cdots\left(n-\sum_{i=1}^{m-1}R_i - m+1\right)$. However, as it is mentioned earlier that entropy for continuous rvs can be negative and this also true for entropy of PCOS. So the CRE and CE measures become useful in this context. Recently, Park and Kim (2014) studied CRE of first r order statistics and provided a single integral representation as

$$CRE_{1\cdots r:n} = -n\int_0^{+\infty} S_{r:n-1}(x)S(x)\log S(x)dx,$$

where $S_{r:n-1}$ is the sf of the rth order statistic of a sample of size $n-1$. Abo-Eleneen et al. (2018) obtained joint CRE of PCOS as

$$CRE_{1\cdots m:m:n} = -\int_0^{+\infty}\frac{1}{h(x)}\log S(x)\sum_{i=1}^{m} f_{X_{i:m:n}}(x)dx, \tag{7.1}$$

where $f_{X_{i:m:n}}$ is the pdf of $X_{i:m:n}$. However, CE for order statistics has not been studied in the literature. In the next theorem, we provide expression of CE measure for consecutive last $(n-s+1)$ order statistics.

Theorem 7.1.1. *The joint cumulative entropy $CE_{s\cdots m:n}$ for the last $(n-s+1)$ order statistics is given by*

$$CE_{s\cdots n:n} = -n\int_0^{+\infty} F(x)F_{s-1:n-1}(x)\log F(x)dx.$$

Proof. CE for the largest order statistic is given by

$$CE_{n:n} = -\int_0^{+\infty} F_{n:n}(x)\log F_{n:n}(x)dx.$$

After some algebraic simplification, it can be expressed as

$$CE_{n:n} = -\int_0^{+\infty} \frac{1}{r(x)} f_{n:n}(x)\log F(x)dx, \tag{7.2}$$

where $r(x) = \frac{f(x)}{F(x)}$ is the rhr of X. Using the decomposition property, we have

$$CE_{s\cdots n:n} = CE_{n:n} + CE_{n-1:n|n:n} + \cdots + CE_{s:n|s+1:n}, \tag{7.3}$$

where $CE_{i:n|i+1:n}$ is the conditional CE of the i-th order statistic given $(i+1)$-th order statistic. Following the result given in Arnold et al. (2008) (p. 23), the i-th order statistic given $(i+1)$-th order statistic can be treated as the largest order statistic of a sample of size i from a distribution having cdf $\left[\frac{F(x)}{F(x_{i+1})}\right]^i$, $x < x_{i+1}$. Therefore, from Eq. (7.2), we have

$$CE_{i:n|i+1:n} = -\int_0^{+\infty} \frac{1}{r(x)} f_{i:n}(x)\log F(x)dx. \tag{7.4}$$

Using (7.4) in (7.3), we get

$$\begin{aligned}
CE_{s\cdots n:n} &= -\int_0^{=\infty} \frac{1}{r(x)} \sum_{i=s}^{n} f_{i:n}(x)\log F(x)dx \\
&= -\int_0^{+\infty} \sum_{i=s}^{n} \frac{n!}{(i-1)!(n-i)!} F^i(x)S^{n-i}(x)\log F(x)dx \\
&= -n\int_0^{+\infty} F(x)\sum_{i=s}^{n} \binom{n-1}{i-1} F^{i-1}(x)S^{n-i}(x)\log F(x)dx \\
&= -n\int_0^{+\infty} F(x)F_{s-1:n-1}(x)\log F(x)dx.
\end{aligned}$$

□

In general, reverse Markovian property does not hold for PCOS as the lifetimes of the removed items R_i, $i = 1(1)m$ might as well be greater than x_{i+1}. So like $CRE_{1\cdots m:m:n}$, CE for PCOS can not be represented in a single integral formation. However, for the special cases when the lifetimes of R_i, $i = 1(1)m$ is less than x_{i+1}, then we can express CE for PCOS in a simplified expression which can be used for optimization problems.

Theorem 7.1.2. *The cumulative entropy for progressive type-II censored order statistics can be expressed as*

$$CE_{1\cdots m:m:n} = -\int_0^{+\infty} \frac{1}{r(x)} \log F(x) \sum_{i=1}^{m} f_{X_{i:m:n}}(x)dx, \tag{7.5}$$

provided the lifetimes of the removed items R_i, $i = 1(1)m$ is less than x_{i+1}.

Proof. Using decomposition property, we have

$$CE_{1\cdots m:m:n} = CE_{m:m:n} + CE_{m-1:m:n|m:m:n} + CE_{m-2:m:n|m-1:m:n} + \cdots + CE_{1:m:n|2:m:n}, \tag{7.6}$$

where $CE_{i:m:n|i+1:m:n}$ is the conditional CE of the i-th PCII censored order statistic given $(i+1)$-th PCII censored order statistic $X_{i+1:m:n} = x_{i+1}$. Let $F_{i:m:n|i+1:m:n}$ be the conditional cdf of $X_{i:m:n}|X_{i+1:m:n} = x_{i+1}$. Then $F_{i:m:n|i+1:m:n}$ has the same distribution as the largest order statistic of a sample of size $\sum_{j=1}^{i} R_j + i = n - \sum_{j=i+1}^{m} R_j - m + 1$ with cdf $\left[\frac{F(x)}{F(x_{i+1})}\right]$, $x < x_{i+1}$. Therefore, proceeding in the same way as Theorem 7.1.1, we readily have

$$F_{i:m:n|i+1:m:n}(x|x_{i+1}) = \left[\frac{F(x)}{F(x_{i+1})}\right]^{(\sum_{j=1}^{i} R_j + i)}, \; x < x_{i+1}.$$

Proceeding with the similar arguments as in Theorem 7.1.1, From Eq. (7.2) we get

$$CE_{i:m:n|i+1:m:n} = -\int_0^{+\infty} \frac{1}{r(x)} f_{X_{i:m:n}}(x) \log F(x)dx. \tag{7.7}$$

The theorem follows upon replacing (7.7) in (7.6). □

The expressions of $CE_{1\cdots m:m:n}(X)$ and $CE_{1\cdots m:m:n}(X)$ can be further simplified by using the expression of $f_{X_{i:m:n}}(x)$ provided by Balakrishnan and Aggarwala (2000). Note that, $f_{X_{i:m:n}} = c_{i-1} \sum_{j=1}^{i} a_{j,i} (1-F(x))^{\gamma_j - 1} f(x)$, $-\infty < x < +\infty$, $1 \le i \le m$, where $\gamma_i = m - i + 1 + \sum_{j=i}^{m} R_j$, $c_{i-1} = \prod_{j=1}^{i} \gamma_j$ and $a_{j,i} = \prod_{k=1, k \neq j}^{i} \frac{1}{\gamma_k - \gamma_j}$, $1 \le j \le i \le m$. Therefore, by taking $F(x) = v$, one can express $CRE_{1\cdots m:m:n}$ as

$$CRE_{1\cdots m:m:n} = \sum_{i=1}^{m} c_{i-1} \sum_{j=1}^{i} a_{j,i} \int_0^1 v^{\gamma_j} (\log v) \left(\frac{d}{dv}\left(F^{-1}(1-v)\right)\right) dv. \tag{7.8}$$

Similarly, $CE_{1\cdots m:m:n}$ can be expressed as

$$CE_{1\cdots m:m:n} = -\sum_{i=1}^{m} c_{i-1} \sum_{j=1}^{i} a_{j,i} \int_0^1 v(1-v)^{\gamma_j - 1} \log v \left(\frac{d}{dv}\left(F^{-1}(v)\right)\right) dv. \tag{7.9}$$

Example 7.1.1. *Let X follows Weibull distribution with cdf given by*

$$F(x) = 1 - e^{-(\lambda x)^{\alpha}},\ x > 0\ ,\alpha, \lambda > 0,$$

where α *and* λ *are the shape and scale parameters, respectively. Then, CRE and CE for PCOS can be computed as*

$$CRE_{1\cdots m:m:n} = -\frac{1}{\alpha\lambda}\sum_{i=1}^{m} c_{i-1}\sum_{j=1}^{i} a_{j,i}\int_0^1 v^{\gamma_j - 1}(\log v)(-\log v)^{\frac{1}{\alpha}-1}dv$$

and

$$CE_{1\cdots m:m:n} = -\frac{1}{\alpha\lambda}\sum_{i=1}^{m} c_{i-1}\sum_{j=1}^{i} a_{j,i}\int_0^1 v(1-v)^{\gamma_j - 2}\log v\,[-\log(1-v)]^{\frac{1}{\alpha}-1}\,dv.$$

In the next section, we obtain optimal design for PCII censored experiment by maximizing design criterion based on $CRE_{1\cdots m:m:n}$ and $CE_{1\cdots m:m:n}$ measures for the Weibull lifetime distribution.

7.2 Determination of optimal censoring schemes using cumulative entropy measures

In this section, we study optimal designs of PCII censored experiments by maximizing CRE and CE measures of PCOS. This is also equivalent to minimizing the variability of the experiment. We consider the Weibull distribution for illustrations. The $CRE_{1\cdots m:m:n}$ and $CE_{1\cdots m:m:n}$ measures represent overall information of a PCII censored experiment but they are not scale-invariant. So we propose two scale-invariant design criteria as follows

$$\phi_A(\mathscr{R}) = \frac{CRE_{1\cdots m:m:n}}{E[X_{1:m:n}]} \tag{7.10}$$

and

$$\phi_B(\mathscr{R}) = \frac{CE_{1\cdots m:m:n}}{E[X_{1:m:n}]}. \tag{7.11}$$

Note that $E[X_{1:m:n}]$ is the expected value of the first failure of a PCII censoring experiment. Also $E[X_{1:m:n}] = E[X_{1:n}]$, where $X_{1:n}$ is the first order statistic of a sample of size n. For Weibull distribution

$$E[X_{1:m:n}] = \frac{\Gamma\left(1+\frac{1}{\alpha}\right)}{\lambda n^{\frac{1}{\alpha}}}.$$

Lemma 7.2.1. *The criteria $\phi_A(\mathscr{R})$ and $\phi_B(\mathscr{R})$ defined in Eq. (7.10) and (7.11) are scale-invariant.*

Proof. Suppose that $Y = cX$, with c being a non-zero constant. Then it is easy to verify that $E[Y_{1:m:n}] = cE[X_{1:m:n}]$. Now, using the relation $F_{cX}(x) = F_X\left(\frac{x}{c}\right)$, $x \in \mathbb{R}$, and after some algebraic calculations, we have

$$CRE_{1\cdots m:m:n}(Y) = c\, CRE_{1\cdots m:m:n}(X).$$

Therefore, $\phi_A(\mathscr{R})$ is scale-invariant. Similarly, we can show that $\phi_B(\mathscr{R})$ is scale-invariant. Hence the result. □

Table 7.1: *Optimal designs for Weibull distribution by maximizing $\phi_A(\mathscr{R})$.*

n	m	$(\alpha,\lambda) = (2,1)$		$(\alpha,\lambda) = (3,1)$	
		Optimal Scheme $(\mathscr{R}^*)$	$\phi_A(\mathscr{R}^*)$	Optimal Scheme $(\mathscr{R}^*)$	$\phi_A(\mathscr{R}^*)$
10	5	(1,1,1,1,1)	5.5902	(1,1,1,1,1)	2.85
15	5	(10,0,0,0,0)	8.6564	(10,0,0,0,0)	3.7613
15	10	(1*5,0*5)	17.69	(1*5,0*5)	7.6925
20	5	(15,0,0,0,0)	9.8098	(15,0,0,0,0)	4.0729
20	10	(1*10)	15.8114	(1*10)	7.1814
20	15	(1*5,0*10)	32.1189	(1*5,0*10)	13.1371
25	5	(20,0,0,0,0)	10.8334	(20,0,0,0,0)	4.3390
25	10	(15,0*9)	23.75	(15,0*9)	9.3460
25	15	(1*10,0*5)	32.3230	(1*10,0*5)	13.1427
25	20	(1*5,0*15)	48.7319	(1*5,0*15)	19.1104

In planning the experiment, it is very difficult to search the optimal scheme out of all possible censoring schemes even for a moderate size of n and m. Addressing this problem, Bhattacharya et al. (2016) proposed a variable neighbourhood search (VNS) algorithm that provides optimal or near optimal schemes within a reasonable computation time. VNS is an excellent tool to perform discrete optimization problems under progressive censoring setup. Now, for fixed n and m, we obtain optimal schemes by maximizing $\phi_A(\mathscr{R})$ and $\phi_B(\mathscr{R})$ using the VNS algorithm. The optimal schemes for different (α,λ) and for various combinations of (n,m) are reported in Tables 7.1 and 7.2, respectively. The notation $a*b$ refers to a is repeated b times. For instance, corresponding to $(n,m) = (15,10)$, the term $(1*5,0*5)$ refers to $(1,1,1,1,1,0,0,0,0,0)$. From the tables it is observed that as sample size n increases $\phi_A(\mathscr{R})$ and $\phi_B(\mathscr{R})$ also increases. This is because as n increases the duration of PCII experiment also increases and, consequently, information increases resulting in smaller variance.

Table 7.2: *Optimal designs for Weibull distribution by maximizing $\phi_B(\mathscr{R})$.*

n	m	$(\alpha,\lambda) = (0.5,1)$		$(\alpha,\lambda) = (2,1)$	
		Optimal Scheme ($\mathscr{R}^*$)	$\phi_B(\mathscr{R}^*)$	Optimal Scheme ($\mathscr{R}^*$)	$\phi_B(\mathscr{R}^*)$
10	5	(1,1,1,1,1)	177.533	(0,0,0,0,5)	7.164
15	5	(10,0,0,0,0)	819.676	(4,0,0,0,6)	8.647
15	10	(1*5,0*5)	1634.716	(0*9,5)	17.562
20	5	(15,0,0,0,0)	1430.132	(10,0,0,0,5)	9.8624
20	10	(1*10)	1420.264	(0*9,10)	20.335
20	15	(1*5,0*10)	4755.537	(0*14,5)	30.239
25	5	(20,0,0,0,0)	2209.845	(15,0,0,0,5)	10.919
25	10	(15,0*9)	4965.943	(5,0*0,10)	22.598
25	15	(1*10,0*5)	6049.293	(0*14,10)	34.178
25	20	(1*5,0*15)	10200.93	(0*19,5)	44.829

7.3 Constraint optimal design

In industrial setup, a reliability engineer has to design certain experiments within a given budget. It is not always feasible to design PCII censored life-testing experiments by considering maximum information principle only. Since the maximum information design will also have the highest cost associated with it. A reasonable practical approach is to design PCII censored experiments subject to a pre-fixed cost constraint, see Bhattacharya et al. (2016). Motivated from their work, we have proposed a constraint optimal design. First, we give a brief discussion of the total cost associated with a PCII experiment defined as

$$C_0 + C_f m + C_t E[X_{m:m:n}],$$

where C_t is the cost per unit duration, C_f is the cost per unit failure and C_0 is fixed cost independent of design parameters. The quantity $E[X_{m:m:n}]$ is interpreted as the expected duration of the experiment and, for the Weibull distribution, it is given by

$$E[X_{m:m:n}] = \frac{1}{\lambda}\Gamma\left(1+\frac{1}{\alpha}\right) c_{m-1} \sum_{j=1}^{m} \frac{a_{j,m}}{\gamma_j^{1+\frac{1}{\alpha}}}.$$

Therefore, the constraint design problem can be formulated as follows:

$$\begin{aligned} &\text{Maximize} \quad \phi_A(\mathscr{R}) = \frac{CRE_{1\cdots m:m:n}}{E[X_{1:m:n}]} \\ &\text{subject to} \quad C_0 + C_f m + C_t E[X_{m:m:n}] \le C_b, \end{aligned} \tag{7.12}$$

where C_b is the pre defined budget cost. For solving this constraint optimization problem,

Table 7.3: *Optimal solutions for constraint design problem with m = 5 and* (C_0, C_f, C_t) = *(20, 10, 50).*

Parameters (α,λ)	n	m	C_b	$\mathscr{R}^*$	$\phi_A(\mathscr{R}^*)$
(2,1)	15	5	120	(8, 0, 0, 0, 2)	6.741
			140	(9, 1, 0, 0, 0)	8.409
			160	(10, 0, 0, 0, 0)	8.656
	20	5	120	(13, 0, 0, 0, 2)	7.567
			140	(14, 1, 0, 0, 0)	9.52
			160	(15, 0, 0, 0, 0)	9.81
	25	5	120	(18, 0, 0, 0, 2)	8.320
			140	(19, 1, 0, 0, 0)	10.506
			160	(20, 0, 0, 0, 0)	10.833
(1,1)	15	5	120	(7, 1, 0, 0, 2)	35.071
			160	(7, 2, 0, 1, 0)	49.821
			180	(10, 0, 0, 0, 0)	65.00
	20	5	120	(12, 1, 0, 0, 2)	45.095
			160	(13, 1, 0, 1, 0)	66.667
			180	(15, 0, 0, 0, 0)	85.00
	25	5	120	(18, 0, 0, 0, 2)	57.50
			160	(18, 1, 0, 1, 0)	82.083
			180	(20, 0, 0, 0, 0)	105.00
(0.5,1)	15	5	120	(6, 1, 0, 1, 2)	368.00
			160	(8, 0, 0, 1, 1)	617.00
			200	(9, 0, 0, 0, 1)	839.125
	20	5	120	(12, 0, 0, 1, 2)	659.116
			160	(13, 0, 0, 1, 1)	1058.873
			200	(14, 0, 0, 0, 1)	1447.778
	25	5	120	(17, 0, 0, 1, 2)	1001.684
			160	(18, 0, 0, 1, 1)	1620.556
			200	(19, 0, 0, 0, 1)	2222.569

we first find those set of PCII censoring schemes ($\tilde{\mathscr{R}}$) such that the condition

$$C_0 + C_f m + C_t E[X_{m:m:n}] \leq C_b$$

is satisfied. Then using the VNS algorithm, we can find the optimal schemes from the set of $\tilde{\mathscr{R}}$, that maximizes $\phi(\mathscr{R})$. We obtain optimal schemes for WE (2,1), WE (1,1) and WE

(0.5,1) with (C_0, C_f, C_t) = (20, 10, 50) and for various choices of C_b. We choose n = 15, 20, 25 and m = 5 and report the results in Table 7.3. From the table it is observed that for fixed n and m, as the budget increases, the overall information of the PCII experiment also increases, as expected.

7.4 Compound optimal design

Recently, Bhattacharya (2020) developed a compound optimal design strategy under the PCII experiment by simultaneously optimizing two competitive criteria. Suppose $\phi_1(\mathscr{R})$ and $\phi_2(\mathscr{R})$ are two competitive design criteria and $\mathscr{R}_1^*$ and $\mathscr{R}_2^*$ are the corresponding optimal designs, respectively. Then, the compound criterion maximizes

$$\Psi(\mathscr{R}|\lambda) = \lambda\Psi_1(\mathscr{R}) + (1-\lambda)\Psi_2(\mathscr{R}),$$

where the functions $\Psi_i(\mathscr{R}) = \frac{\phi_i(\mathscr{R}^*)}{\phi_i(\mathscr{R})}$, $i = 1,2$, are the relative efficiencies of the criterion. Note that, these relative efficiencies lie between 0 and 1. For detailed discussion on the fundamental properties of the compound optimal design, see Bhattacharya (2020).

From Tables 7.1 and 7.2, it is observed that as the information of a PCII censored experiment increases, the duration of the experiment also increases. This implies that a PCII censored experiment with high information will result in a high cost associated with the experiment. So there are two competitive criteria. The aim is to choose an experiment such that the information is maximized and the cost is minimized simultaneously. This is the situation where the COD optimizes the two criteria simultaneously. First, we obtain COD by simultaneously optimizing the total cost of the experiment and the criterion $\phi_A(\mathscr{R})$. Therefore, the design criteria are

$$\phi_1(\mathscr{R}) = C_0 + C_f m + C_t E[X_{m:m:n}],$$

and

$$\phi_2(\mathscr{R}) = \frac{1}{\phi_A(\mathscr{R})}.$$

To obtain COD, first we need to find the respective single objective optimal designs i.e. ϕ_1-optimal design and ϕ_2-optimal design. Bhattacharya et al. (2016) showed that, for Weibull distribution with parameters (α, λ), conventional type-II censoring $(0,0,\cdots,n-m)$ minimizes the total cost $\phi_1(\mathscr{R})$. To obtain ϕ_2-optimal design, we use the VNS algorithm. Now for WE (2,1) with (C_0, C_f, C_t) = (100, 10, 50) we obtain compound optimal design for (n,m) = (15,11), (20,14), (25,10) and (25,20). From Figure 7.2, we get the approximate λ values as 0.66, 0.68, 0.69 and 0.66, respectively and corresponding compound optimal design schemes are (1*3, 0*7, 1), (5, 0*12, 1), (14, 0*8, 1) and (1*4, 0*15, 1).

Table 7.4: *Comparison between ϕ_1 and ϕ_2-optimal and compound optimal designs under Weibull distribution with (α, λ) = (2,1).*

(n,m)	ϕ_1-optimal design			ϕ_2-optimal design			compound optimal design		
	Scheme	Cost	CRE	Scheme	Cost	CRE	Scheme	Cost	CRE
(15,11)	(0*10,4)	264.86	16.31	(1*4,0*7)	293.94	20.22	(1*3,0*7,1)	280.70	18.86
(20,14)	(0*13,6)	293.04	23.10	(1*6,0*8)	327	29.27	(5,0*12,1)	314.86	28.58
(25,10)	(0*9,15)	234.83	13.13	(15,0*9)	282.86	23.75	(14,0*8,1)	268.97	21.57
(25,20)	(0*19,5)	361.44	40.48	(1*5,0*15)	392.98	48.73	(1*4,0*15,1)	380.14	46.37

A comparative study is presented in Table 7.4. From the table, we observe that the ϕ_1-optimal design has the minimum cost and minimum information; and the ϕ_2-optimal design has the maximum information and highest cost among the three designs. As expected, the cost and the CRE values corresponding to COD lie in between the other two designs, which can be interpreted as a trade-off between them.

Similarly, using the CE measure of PCOS, we can obtain COD for PCII censored experiments. Since the total cost of the experiment and the CE are competitive criteria, hence we can take the compound design criteria as

$$\phi_1(\mathscr{R}) = C_0 + C_f m + C_t E[X_{m:m:n}]$$

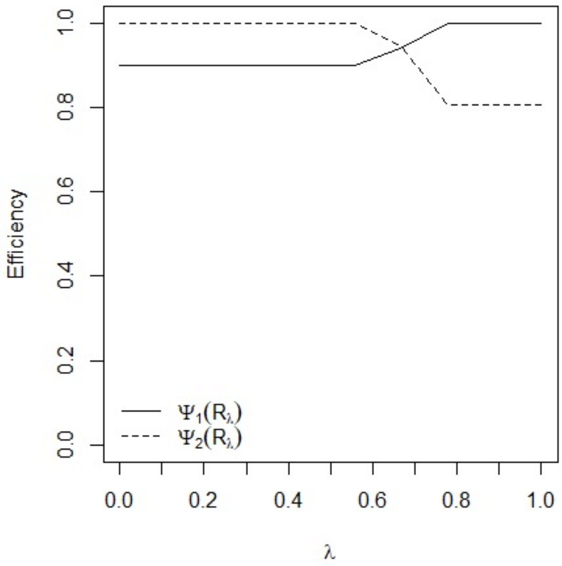
and

$$\phi_2(\mathscr{R}) = \frac{1}{\phi_B(\mathscr{R})}.$$

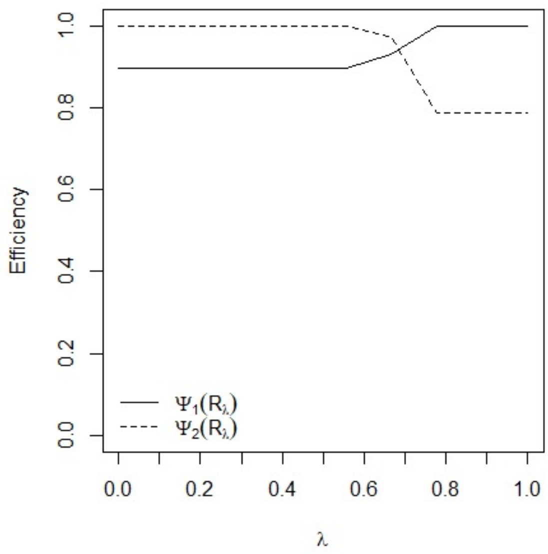
Now for WE (0.5,1) distribution with (C_0, C_f, C_t) = (100,10,50) we obtain compound optimal design for (n,m) = (15,5), (20,5), (20,14) and (25,20). From Figure 7.3 the approximate λ values are calculated as 0.61, 0.60, 0.50 and 0.44, respectively and corresponding compound optimal design schemes are (1,0,0,0,9), (8,0,0,0,7), (2,0*12,4) and (3,0*18,2).

The findings are reported in Table 7.5 and it is observed that the ϕ_1-optimal design has the minimum cost and minimum CE; and the ϕ_2-optimal design has the maximum CE and highest cost among the three designs. This is evident from the fact that if the CE of the PCII scheme increases (decreases), the overall cost of the experiment will also increase (decrease). Like the CRE case, the cost and CE value corresponding to COD lie in between the other two designs, which implies that the trade-off is achieved.

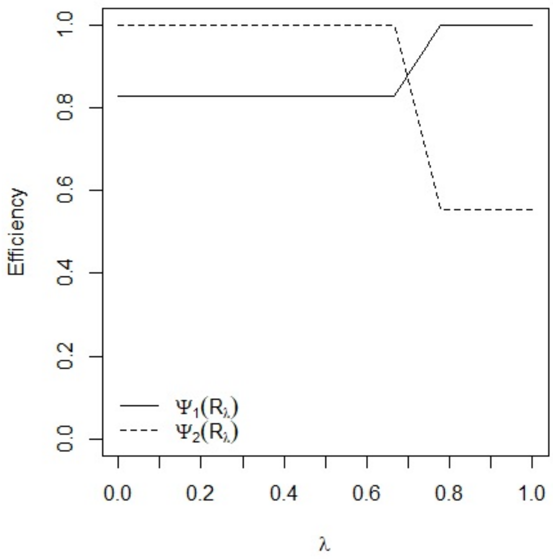
Fig. 7.2: *Plotting of relative efficiencies of the compound design (CRE and cost) versus values of λ in [0,1].*



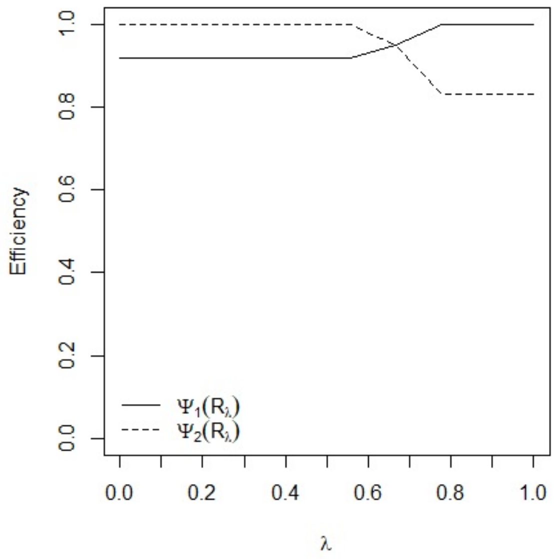
(a) *n = 15* and *m = 11*



(b) *n = 20* and *m = 14*



(c) *n = 25* and *m = 10*



(d) *n = 25* and *m = 20*

Table 7.5: *Comparison between ϕ_1 and ϕ_2-optimal and compound optimal designs under Weibull distribution with (α,λ) = (0.5,1).*

(n,m)	ϕ_1-optimal design			ϕ_2-optimal design			compound optimal design		
	Scheme	Cost	CE	scheme	Cost	CE	Scheme	Cost	CE
(15,5)	(0*4,10)	159.11	120.461	(10,0*4)	452.528	819.676	(1,0*3,9)	160.508	131.726
(20,5)	(0*4,15)	154.692	132.370	(15,0*4)	448.861	1430.132	(8,0*3,7)	163.814	266.789
(20,14)	(0*13,6)	311.104	1956.355	(1*6,0*8)	811.14	4255.305	(2,0*12,4)	347.207	2411.236
(25,20)	(0*19,5)	424.553	5681.97	(1*5,0*15)	1031.58	10200.93	(3,0*18,2)	556.291	7700.36

Remark 7.4.1. *The budget cost C_b is pre-fixed and most of the time there isn't any specific guidance available to choose the appropriate amount required for experimentation. If the budget is small then it is likely impossible to design certain experiments due to cost constraints. As a consequence, usually a high budget is preferred on the cost. This often leads to wastage of money which is not ideal in any business. Using COD, we can determine the budget cost C_b for the constraint design problem. Bhattacharya (2020) studied the equivalence between compound and constraint optimal designs and, by making use of this method, we can determine C_b for the constraint problem.*

We can equivalently formulated the constraint problem defined in (7.12), in terms of the relative efficiencies, as

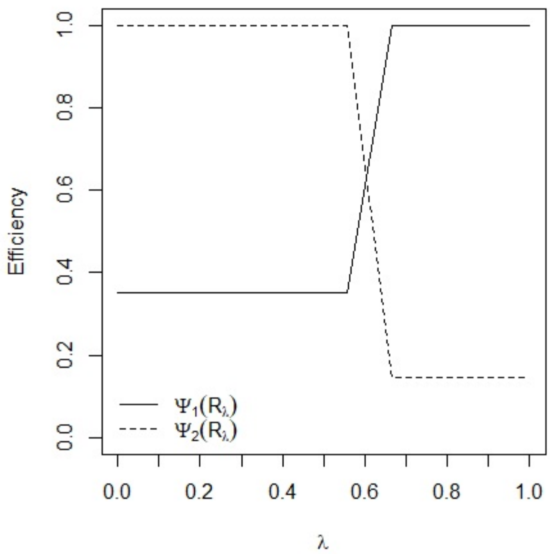
$$\begin{aligned} &\textit{Maximize} \quad \Psi_2(\mathscr{R}) \\ &\textit{subject to} \quad \Psi_1(\mathscr{R}) \geq \varepsilon, \end{aligned}$$

*where Ψ_1 and Ψ_2 are the relative efficiencies and $\varepsilon = \frac{\phi_1(\mathscr{R}_1^*)}{C_b}$. Now for WE (2,1) with (n,m) = (15, 11), (C_0, C_f, C_t) = (10, 10, 50) and $C_b = 200$, the corresponding constraint optimal design is $\mathscr{R}$ = (1*3,0*6,1,0). The equivalent COD (see Lemma 3 in Bhattacharya (2020)) is $\mathscr{R}_\lambda$ = (1*3,0*7,1) where λ = 0.57. So we have, $\Psi_1(\mathscr{R}_\lambda)$ = 0.9170 and the adjusted budget constraint is $C_b = \phi_1(\mathscr{R}_1^*)/0.9170 = 190.6963$, which is less than the corresponding budget of constraint problem in 7.12. This established the fact that it is beneficial to compute the bound C_b (or equivalently ε) for the constraint optimization problem from the corresponding equivalent COD problem. As a matter of fact, the experimenter no need to specify the bound beforehand, instead it can be approximated from the solution of COD.*

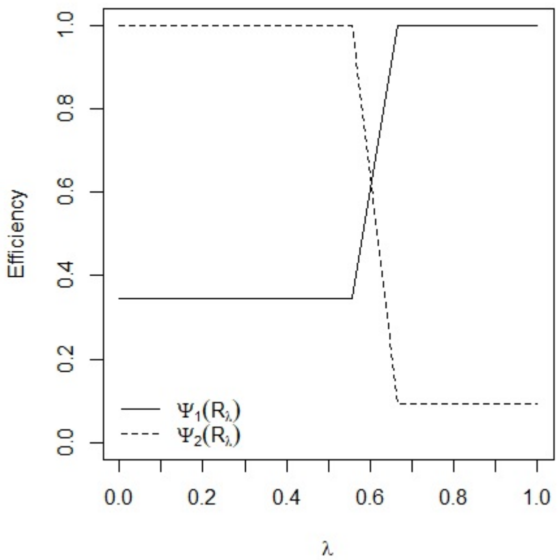
7.4.1 Data Analysis

In this section, we analyse a real data set. The progressively type-II censored data is presented by Viveros and Balakrishnan (1994) from a data set of failure times of an insulating

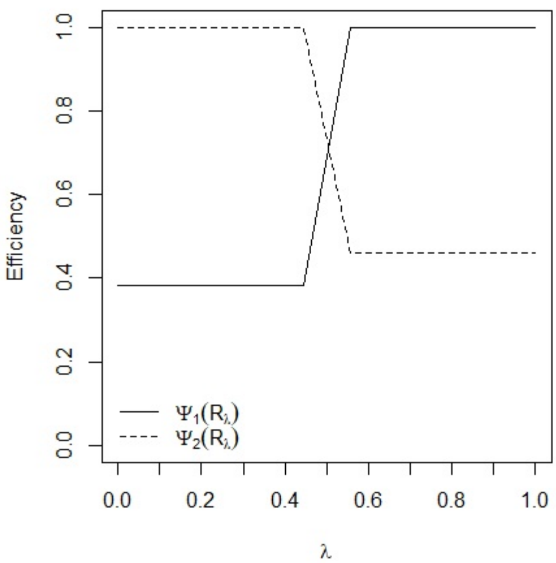
Fig. 7.3: *Plotting of relative efficiencies of the compound design (CE and cost) versus values of λ in [0,1].*



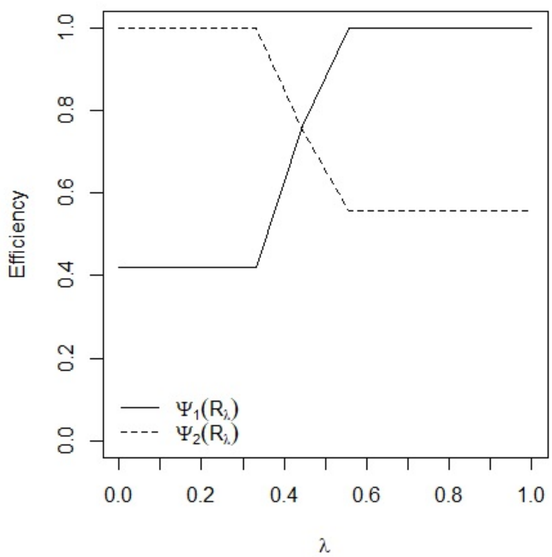
(a) *n=15 and m=5*



(b) *n=20 and m=5*



(c) *n=20 and m=14*



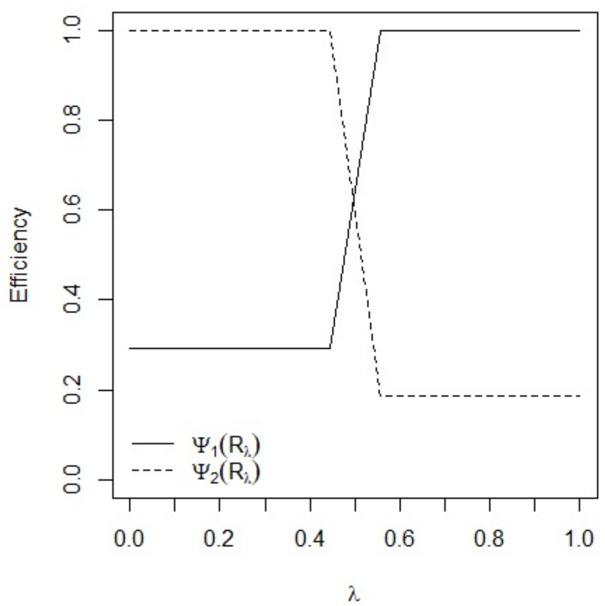
(d) *n=25 and m=20*

Table 7.6: *ϕ_1 and ϕ_2-optimal and compound optimal designs under Weibull distribution with (α,λ) = (0.974, 0.108) and (n,m) = (19,8).*

ϕ_1-optimal design			ϕ_2-optimal design			compound optimal design		
Scheme	Cost	CE	scheme	Cost	CE	Scheme	Cost	CE
(0*7,11)	205.99	64.89	(11,0*7)	316.127	100.90	(10,0*6,1)	270.083	94.74

fluid in an accelerated test conducted by Nelson (1982), Table 6.1, p. 228. The progressively censored data is generated from n = 19 observations recorded at 34kV in Nelson's Table 6.1 with m = 9 number of failures using censoring scheme $\mathscr{R}$ = (0, 0, 3, 0, 3, 0, 0, 5). The data is presented in table 7.6. Weibull distribution provides a good fit to this data with shape and scale parameters 0.974 and 0.108, respectively. Now, we apply compound optimal design for Weibull distribution with parameters (α,λ) = (0.974, 0.108) and (n,m) = (19,8). We plot the ϕ_1 and ϕ_2 efficiencies in Figure 7.4. From the figure, we find that the approximate value of λ is 0.49. The corresponding numerical results are provided in Table 7.6 and it is observed that the corresponding ϕ_1, ϕ_2-optimal designs and compound optimal design are (0*7, 11), (11, 0*7) and (10, 0*6, 1), respectively. The CE value for ϕ_1-optimal design is low and the cost is high for the ϕ_2-optimal design, whereas the COD is a trade-off between them, as expected.

Fig. 7.4: *Efficiency plot under Weibull distribution with parameters (α,λ) = (0.974, 0.108).*



7.5 Conclusion

In this Chapter, we proposed new criteria based on cumulative entropy measures that represent the overall information of a PCII censored experiment. Maximizing these criteria, we obtained optimal progressive censoring schemes. Maximizing information will increase the total cost of the experiment, so we implement a constraint optimal design in which we maximized the information of the PCII experiment subject to a cost constraint. In the constraint design, we found that as the budget amount increases the information of the experiment also increases. This is practical because increasing information of an experiment means increasing the duration of that experiment. Hence, the cost associated with the experiment will also increase. Also we studied compound optimal design that simultaneously maximizes the information criterion and minimizes the total cost of the experiment. These cumulative entropy based design criteria are very easy to calculate and they do not rely on asymptotic results thus the designs based on these measures are also independent of asymptotic results.

Chapter 8

Conclusions and future work

THIS thesis considers development of some weighted cumulative information measures along with their inference and applications. Throughout the course of this thesis, various weighted information measures are proposed and numerous properties such as bounds, monotonicity, convolution and relationship with some well known information measure are studied in detail. Aging classes are proposed using the dynamic information measures and some characterization results for Rayleigh and power distributions are obtained. In the following we discuss some problems as part of future work.

Weighted extended information measures

We propose a weighted extended survival extropy measure by taking $\varepsilon(x)$, a continuous function of x as the weight function. Analogously we can define extended versions of the various weighted information measures that we have introduced. For example, generalized weighted extended survival entropy (GWESE) measure of order (θ_1, θ_2) can be defined as

$$\xi^{\varepsilon}_{\theta_1,\theta_2}(X) = \frac{1}{\theta_2 - \theta_1} \log \int_0^{+\infty} \varepsilon(x) S^{\theta_1+\theta_2-1}(x) dx,\ \theta_2 \geq 1,\ \theta_2 - 1\ < \theta_1 <\ \theta_2.$$

As a future problem, one can study the properties of $\xi^{\varepsilon}_{\theta_1,\theta_2}(X)$ for different choices of the weight functions.

Double truncated weighted information measures

In many situations, we only have informations betweeen two time points i.e. the random variable (rv) is double truncated. In reliability, double truncated residual rv is defined as $X_{t_1,t_2} = (X - t_1 | t_1 \leq X \leq t_2)$ and this reduces to the residual rv when $t+2 \to +\infty$. One can study our proposed information measures for double truncated rvs. For $\{(t_1,t_2) : S(t_1) > S(t_2)\}$, the GWSE for double truncated (interval) rv can be defined as

$$\xi^{\varepsilon}_{\theta_1,\theta_2}(X;t_1,t_2) = \frac{1}{\theta_2 - \theta_1} \log \int_{t_1}^{t_2} x \left(\frac{S(x)}{S(t_1) - S(t_2)} \right)^{\theta_1+\theta_2-1} dx,\ \theta_2 \geq 1,\ \theta_2 - 1\ < \theta_1 <\ \theta_2.$$

In similar manner, we can also define other measures for double truncated set up as well. These areas can be explored in future work.

Estimations of these measures are an important issue that have been discussed in detail. Non-parametric estimators based on edf, L- statistics and kernel function are proposed and their performances are compared in terms of MSE. A recursive kernel based non-parametric estimator for weighted survival extropy measure is considered for identically distributed observations which may not be independent. Various real data also analyzed for illustrative purposes.

One future problem that can be considered is developing estimators of the proposed information measures using U-Statistics. It will be interesting to see how U-Statistics based estimators perform compared to the proposed estimators. Non-parametric Bayesian estimations of the proposed measures is another problem of interest. Also estimations of the weighted extended and double truncated information measures is an interesting problem worth studying.

Application in life-testing

Another key aspect of this thesis is the applications of proposed measures as well as some existing measures such as cumulative residual entropy (CRE) and cumulative residual extropy (CREx) measures. We studied CREx for mixed reliability systems and proposed two applications in system complexity analysis and comparison between systems. Also using the L- statistics based estimator of CREx measure, we construct a test of equality between two distributions. We develop goodness-of-fit tests for complete, type-I and type-II censored data and uniformity tests using various proposed measures and also provide potential applications in risk analysis and model discrimination problems. Based on the CRE measure, we proposed a design criterion for optimally selecting a progressive type-II (PCII) censored experiment. This design is independent of asymptotic results. As a future problem, this asymptotic free design criterion can be used to determine optimum reliability acceptance sampling plans (RASP) for PC-II censored experiments.

A RASP can be expressed as follows: Suppose L is a lower specification limit and an item with lifetime less than L is nonconforming i.e. unacceptable. Let μ and σ are the location and scale parameters of a lifetime distribution belonging to the location-scale family. Lieberman and Resnikoff (1955) define a criterion for a lot to be accepted it

$$\hat{\mu} - l\hat{\sigma},$$

where $\hat{\mu}$ and $\hat{\sigma}$ are the maximum likelihood estimates of μ and σ, respectively. Now using CRE based criterion, a RSAM for PCII can be formulated as

$$\begin{aligned}
&\underset{n,m,\mathscr{R}}{\text{maximize}}\ \frac{CRE_{1\cdots m:m:n}}{E[X_{1:m:n}]},\\
&\text{subject to}\quad C_{\mathrm{f}}m + C_{\mathrm{t}}E[X_{m:m:n}] \leq C_{\mathrm{b}},\\
&\left[\frac{u_{\alpha} - u_{1-\beta}}{z_{p_{\alpha}} - z_{p_{\beta}}}\right]^2 [\mathscr{I}^{11}(\theta) + l^2\mathscr{I}^{22}(\theta) - 2l\mathscr{I}^{12}(\theta)] = 1.
\end{aligned} \tag{8.1}$$

The Eq. (8.1) is the solution of the sample size n (Ng et al., 2004). We will undertake this problem in future study. The CRE based RASP under Bayesian environment will also be considered.

References

Abbasnejad, M. (2011). Some characterization results based on dynamic survival and failure entropies. *Communications for Statistical Applications and Methods*, 18(6):787–798. 8, 22

Abbasnejad, M., Arghami, N. R., Morgenthaler, S., and Borzadaran, G. M. (2010). On the dynamic survival entropy. *Statistics & probability letters*, 80(23-24):1962–1971. 8

Abo-Eleneen, Z., Almohaimeed, B., and Ng, H. (2018). On cumulative residual entropy of progressively censored order statistics. *Statistics & Probability Letters*, 139:47–52. 19, 140

Alizadeh Noughabi, H. (2010). A new estimator of entropy and its application in testing normality. *Journal of Statistical Computation and Simulation*, 80(10):1151–1162. 10

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (2008). *A first course in order statistics*. SIAM. 73, 141

Asadi, M., Ebrahimi, N., Soofi, E. S., and Zohrevand, Y. (2016). Jensen–Shannon information of the coherent system lifetime. *Reliability Engineering & System Safety*, 156:244–255. 11, 88, 90

Asadi, M. and Zohrevand, Y. (2007). On the dynamic cumulative residual entropy. *Journal of Statistical Planning and Inference*, 137(6):1931–1941. 8, 22

Balakrishnan, N. and Aggarwala, R. (2000). *Progressive censoring: theory, methods, and applications*. Springer Science & Business Media. 138, 142

Balakrishnan, N., Buono, F., and Longobardi, M. (2022). On weighted extropies. *Communications in Statistics-Theory and Methods*, 51(18):6250–6267. 9, 102

Balakrishnan, N. and Cramer, E. (2014). The art of progressive censoring. *Statistics for industry and technology*. 138

Balakrishnan, N., Rad, A. H., and Arghami, N. R. (2007). Testing exponentiality based on Kullback-Leibler information with progressively Type-II censored data. *IEEE Transactions on Reliability*, 56(2):301–307. 11, 137, 140

Baratpour, S. and Rad, A. H. (2012). Testing goodness-of-fit for exponential distribution based on cumulative residual entropy. *Communications in Statistics-Theory and Methods*, 41(8):1387–1396. 10, 42, 60, 67, 70

Baratpour, S. and Rad, A. H. (2016). Exponentiality test based on the progressive type II censoring via cumulative entropy. *Communications in Statistics-Simulation and Computation*, 45(7):2625–2637. 11, 137

Barlow, R. E. and Proschan, F. (1975). *Statistical theory of reliability and life testing: probability models*. Holt, Rinehart and Winston: New York. 12, 77, 93

Beck, C. (2009). Generalised information and entropy measures in physics. *Contemporary Physics*, 50(4):495–510. 6

Belis, M. and Guiasu, S. (1968). A quantitative-qualitative measure of information in cybernetic systems (corresp.). *IEEE Transactions on Information Theory*, 14(4):593–594. 7

Best, D. J., Rayner, J. C., and Thas, O. (2010). Easily applied tests of fit for the Rayleigh distribution. *Sankhya B*, 72(2):254–263. 122

Bhattacharya, R. (2020). Implementation of compound optimal design strategy in censored life-testing experiment. *Test*, 29(4):1029–1050. 20, 139, 147, 150

Bhattacharya, R., Pradhan, B., and Dewanji, A. (2014). Optimum life testing plans in presence of hybrid censoring: a cost function approach. *Applied Stochastic Models in Business and Industry*, 30(5):519–528. 139

Bhattacharya, R., Pradhan, B., and Dewanji, A. (2016). On optimum life-testing plans under Type-II progressive censoring scheme using variable neighborhood search algorithm. *Test*, 25:309–330. 144, 145, 147

Boland, P. J. and Samaniego, F. J. (2004). The signature of a coherent system and its applications in reliability. In *Mathematical reliability: An expository perspective*, pages 3–30. Springer. 77

Bradley, R. C. (2007). *Introduction to strong mixing conditions*. Kendrick press. 126

Burnham, K. and Anderson, D. (2004). Model selection and multi-model inference. *Second. NY: Springer-Verlag*, 63(2020):10. 5, 129

Calì, C., Longobardi, M., and Ahmadi, J. (2017). Some properties of cumulative Tsallis entropy. *Physica A: Statistical Mechanics and its Applications*, 486:1012–1021. 46

Cartwright, J. (2014). Roll over, boltzmann. *Physics World*, 27(05):31. 45

Chamany, A. and Baratpour, S. (2014). A dynamic discrimination information based on cumulative residual entropy and its properties. *Communications in Statistics-Theory and Methods*, 43(6):1041–1049. 61

Čižmešija, A., Pečarić, J., and Persson, L.-E. (2003). On strengthened Hardy and Pólya–Knopp's inequalities. *Journal of Approximation Theory*, 125(1):74–84. 111

Cohen, A. C. (1963). Progressively censored samples in life testing. *Technometrics*, 5(3):327–339. 138

Cohen, A. C. (1965). Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples. *Technometrics*, 7(4):579–588. 138

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202. 25

Crzcgorzewski, P. and Wirczorkowski, R. (1999). Entropy-based goodness-of-fit test for exponentiality. *Communications in Statistics-Theory and Methods*, 28(5):1183–1202. 10

D'Agostino, R. and Stephens, M. A. (1986). *Goodness-of-fit-techniques*. Marcel Dekker. 134

Dahmen, K., Burkschat, M., and Cramer, E. (2012). A-and d-optimal progressive Type-II censoring designs based on Fisher information. *Journal of Statistical Computation and Simulation*, 82(6):879–905. 139

Daneshi, S., Nezakati, A., and Tahmasebi, S. (2019). On weighted cumulative past (residual) inaccuracy for record values. *Journal of Inequalities and Applications*, 2019(1):1–22. 62, 63, 64, 65

David, H. A. and Nagaraja, H. N. (2004). *Order statistics*. John Wiley & Sons. 73

Di Crescenzo, A. and Longobardi, M. (2002). Entropy-based measure of uncertainty in past lifetime distributions. *Journal of Applied probability*, 39(2):434–440. 8

Di Crescenzo, A. and Longobardi, M. (2009). On cumulative entropies. *Journal of Statistical Planning and Inference*, 139(12):4072–4087. 7, 8, 22, 57, 121

Di Crescenzo, A. and Longobardi, M. (2015). Some properties and applications of cumulative Kullback-Leibler information. *Applied Stochastic Models in Business and Industry*, 31(6):875–891. 60

Di Crescenzo, A. and Toomaj, A. (2022). Weighted mean inactivity time function with applications. *Mathematics*, 10(16):2828. 13

Dong, X. (2016). The gravity dual of Rényi entropy. *Nature communications*, 7(1):1–6. 10

Doukhan, P. (2012). *Mixing: properties and examples*, volume 85. Springer Science & Business Media. 126

Dudewicz, E. J. and Van Der Meulen, E. C. (1981). Entropy-based tests of uniformity. *Journal of the American Statistical Association*, 76(376):967–974. 10, 132, 134, 135

Ebrahimi, N. (1996). How to measure uncertainty in the residual life time distribution. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 48–56. 8

Ebrahimi, N., Habibullah, M., and Soofi, E. S. (1992). Testing exponentiality based on Kullback-Leibler information. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(3):739–748. 10, 42, 67, 69

Esteban, M., Castellanos, M., Morales, D., and Vajda, I. (2001). Monte carlo comparison of four normality tests using different entropy estimates. *Communications in Statistics-Simulation and computation*, 30(4):761–785. 10

Gokhale, D. (1983). On entropy-based goodness-of-fit tests. *Computational Statistics & Data Analysis*, 1:157–165. 10

Goodarzi, F. and Amini, M. (2021). Reliability and expectation bounds based on Hardy's inequality. *Communications in Statistics-Theory and Methods; DOI:10.1080/03610926.2021.1966037*. 111

Gross, A. J. and Clark, V. A. (1975). *Survival distributions: reliability applications in the biomedical sciences*, volume 11. Wiley New York. 128

Grubbs, F. E. (1971). Approximate fiducial bounds on reliability for the two parameter negative exponential distribution. *Technometrics*, 13(4):873–876. 42, 69, 97

Guiasu, S. (1971). Weighted entropy. *Reports on Mathematical Physics*, 2(3):165–179. 7, 102

Guiasu, S. (1986). Grouping data by using the weighted entropy. *Journal of Statistical Planning and Inference*, 15:63–69. 7

Gupta, R. C., Gupta, P. L., and Gupta, R. D. (1998). Modeling failure time data by lehman alternatives. *Communications in Statistics-Theory and methods*, 27(4):887–904. 29

Gupta, R. D. and Kundu, D. (2003). Discriminating between Weibull and generalized exponential distributions. *Computational statistics & data analysis*, 43(2):179–196. 129

Gupta, R. D. and Nanda, A. K. (2001). Some results on reversed hazard rate ordering. *Communications in Statistics-theory and Methods*, 30(11):2447–2457. 12

Hardy, G. H. (1920). Note on a theorem of Hilbert. *Mathematische Zeitschrift*, 6(3):314–317. 110

Hashempour, M., Kazemi, M., and Tahmasebi, S. (2022). On weighted cumulative residual extropy: characterization, estimation and testing. *Statistics*, 56(3):681–698. 10

Hazeb, R., Bayoud, H. A., and Raqab, M. Z. (2021a). Kernel and CDF-Based Estimation of Extropy and Entropy from Progressively Type-II Censoring with Application for Goodness of Fit Problems. *Stochastics and Quality Control*, 36(1):73–83. 11

Hazeb, R., Raqab, M. Z., and Bayoud, H. A. (2021b). Non-parametric estimation of the extropy and the entropy measures based on progressive type-II censored data with testing uniformity. *Journal of Statistical Computation and Simulation*, 91(11):2178–2210. 11

Jahanshahi, S., Zarei, H., and Khammar, A. (2020). On cumulative residual extropy. *Probability in the Engineering and Informational Sciences*, 34(4):605–625. 9, 10, 76, 95, 98, 103, 119, 130

Jaynes, E. T. (1968). Prior probabilities. *IEEE Transactions on systems science and cybernetics*, 4(3):227–241. 5

Jaynes, E. T. (1982). On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, 70(9):939–952. 5

Jose, J. and Sathar, E. A. (2019). Residual extropy of k-record values. *Statistics & Probability Letters*, 146:1–6. 9

Kaijser, S., Persson, L.-E., and Öberg, A. (2002). On Carleman and Knopp's inequalities. *Journal of Approximation Theory*, 117(1):140–151. 110

Kamari, O. and Buono, F. (2021). On extropy of past lifetime distribution. *Ricerche di Matematica*, 70(2):505–515. 9

Kapur, J. (1967). Generalized entropies of order α and type β. In *The Maths Seminar*, volume 4, pages 78–94. 23

Kattumannil, S. K., Sreedevi, E., and Balakrishnan, N. (2022). A generalized measure of cumulative residual entropy. *Entropy*, 24(4):444. 23, 47

Kayal, S. (2015). On generalized dynamic survival and failure entropies of order (α, β). *Statistics & Probability Letters*, 96:123–132. 8, 22, 23

Kayal, S. (2018). On weighted generalized cumulative residual entropy of order n. *Methodology and Computing in Applied Probability*, 20(2):487–503. 8, 130

Kayal, S. (2019). On a generalized entropy of mixed systems. *Journal of Statistics and Management Systems*, 22(6):1183–1198. 77

Khammar, A. and Jahanshahi, S. (2018). On weighted cumulative residual Tsallis entropy and its dynamic version. *Physica A: Statistical Mechanics and its Applications*, 491:678–692. 8, 46

Kochar, S., Mukerjee, H., and Samaniego, F. J. (1999). The "signature" of a coherent system and its application to comparisons among systems. *Naval Research Logistics (NRL)*, 46(5):507–523. 91

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86. 10

Kumar, V. and Taneja, H. (2011). Some characterization results on generalized cumulative residual entropy measure. *Statistics & probability letters*, 81(8):1072–1077. 8

Kundu, C. and Ghosh, A. (2017). Inequalities involving expectations of selected functions in reliability theory to characterize distributions. *Communications in Statistics-Theory and Methods*, 46(17):8468–8478. 12

Kundu, C. and Singh, S. (2020). On generalized interval entropy. *Communications in Statistics-Theory and Methods*, 49(8):1989–2007. 23

Lad, F., Sanfilippo, G., and Agro, G. (2015). Extropy: Complementary dual of entropy. *Statistical Science*, 30(1):40–58. 9, 78, 89

Lawless, J. F. (2011). *Statistical models and methods for lifetime data*. John Wiley & Sons. 70

Li, B. and Zhang, R. (2021). A new mean-variance-entropy model for uncertain portfolio optimization with liquidity and diversification. *Chaos, Solitons & Fractals*, 146:110842. 10

Lieberman, G. J. and Resnikoff, G. J. (1955). Sampling plans for inspection by variables. *Journal of the American Statistical Association*, 50(270):457–516. 156

Masry, E. (1986). Recursive probability density estimation for weakly dependent stationary processes. *IEEE Transactions on Information Theory*, 32(2):254–267. 127

Mathai, A. and Haubold, H. J. (2007). Pathway model, superstatistics, Tsallis statistics, and a generalized measure of entropy. *Physica A: Statistical Mechanics and its Applications*, 375(1):110–122. 5

Maya, R., Irshad, M., and Archana, K. (2021). Recursive and non-recursive kernel estimation of negative cumulative residual extropy under α-mixing dependence condition. *Ricerche di Matematica*, pages 1–21. 127

Mehrali, Y. and Asadi, M. (2021). Parameter Estimation Based on Cumulative Kullback–Leibler Divergence. *REVSTAT-Statistical Journal*, 19(1):111–130. 10

Mercurio, P. J., Wu, Y., and Xie, H. (2020). An entropy-based approach to portfolio optimization. *Entropy*, 22(3):332. 10

Mirali, M. and Baratpour, S. (2017). Dynamic version of weighted cumulative residual entropy. *Communications in Statistics-Theory and Methods*, 46(22):11047–11059. 8, 63

Mirali, M., Baratpour, S., and Fakoor, V. (2017). On weighted cumulative residual entropy. *Communications in Statistics-Theory and Methods*, 46(6):2857–2869. 8

Misagh, F., Panahi, Y., Yari, G., and Shahi, R. (2011). Weighted cumulative entropy and its estimation. In *2011 IEEE International Conference on Quality and Reliability*, pages 477–480. IEEE. 7, 13, 16, 47, 50, 65, 102

Mittal, D. (1975). On some functional equations concerning entropy, directed divergence and inaccuracy. *Metrika*, 22(1):35–45. 23

Moharana, R. and Kayal, S. (2019). On Weighted Kullback–Leibler Divergence for Doubly Truncated Random Variables. *REVSTAT-Statistical Journal*, 17(3):297–320. 61

Nair, R. D. and Sathar, E. (2020). On dynamic failure extropy. *Journal of the Indian Society for Probability and Statistics*, 21(2):287–313. 9

Nanda, A. K. and Shaked, M. (2001). The hazard rate and the reversed hazard rate orders, with applications to order statistics. *Annals of the Institute of Statistical Mathematics*, 53(4):853–864. 12

Navarro, J., del Águila, Y., Sordo, M. A., and Suárez-Llorens, A. (2013). Stochastic ordering properties for systems with dependent identically distributed components. *Applied Stochastic Models in Business and Industry*, 29(3):264–278. 85

Navarro, J., Samaniego, F. J., Balakrishnan, N., and Bhattacharya, D. (2008). On the application and extension of system signatures in engineering reliability. *Naval Research Logistics (NRL)*, 55(4):313–327. 18, 93

Nelson, W. (1982). *Lifetime data analysis*. Wiley, New York. 152

Nelson, W. B. (2003). *Applied life data analysis*, volume 521. John Wiley & Sons. 74

Ng, H., Chan, P. S., and Balakrishnan, N. (2004). Optimal progressive censoring plans for the Weibull distribution. *Technometrics*, 46(4):470–481. 139, 156

Nishioka, T. (2014). The gravity dual of supersymmetric Rényi entropy. *Journal of High Energy Physics*, 2014(7):1–11. 10

Noughabi, H. A. (2017). Testing exponentiality based on Kullback—Leibler information for progressively Type II censored data. *Communications in Statistics-Simulation and Computation*, 46(10):7624–7638. 11

Noughabi, H. A. (2021). Testing uniformity based on negative cumulative extropy. *Communications in Statistics-Theory and Methods; DOI:10.1080/03610926.2021.2001015*. 103, 132, 135

Noughabi, H. A. (2022). Cumulative residual entropy applied to testing uniformity. *Communications in Statistics-Theory and Methods*, 51(12):4151–4161. 132

Noughabi, H. A. and Chahkandi, M. (2018). Testing the validity of the exponential model for hybrid Type-I censored data. *Communications in Statistics-Theory and Methods*, 47(23):5770–5778. 11

Noughabi, H. A. and Jarrahiferiz, J. (2022). Extropy of order statistics applied to testing symmetry. *Communications in Statistics-Simulation and Computation*, 51(6):3389–3399. 10

Noughabi, M. S., Borzadaran, G. M., and Roknabadi, A. R. (2013). On the reliability properties of some weighted models of bathtub shaped hazard rate distributions. *Probability in the Engineering and Informational Sciences*, 27(1):125–140. 12

Pakgohar, A., Habibirad, A., and Yousefzadeh, F. (2020). Goodness of fit test using Lin-Wong divergence based on Type-I censored data. *Communications in Statistics-Simulation and Computation*, 49(9):2485–2504. 11

Pakyari, R. and Balakrishnan, N. (2013). Testing exponentiality based on Type-I censored data. *Journal of Statistical Computation and Simulation*, 83(12):2369–2378. 73

Park, S. (2005). Testing exponentiality based on the Kullback-Leibler information with the type II censored data. *IEEE Transactions on reliability*, 54(1):22–26. 11, 71, 137, 140

Park, S. and Kim, I. (2014). On cumulative residual entropy of order statistics. *Statistics & Probability Letters*, 94:170–175. 140

Park, S. and Lim, J. (2015). On censored cumulative residual Kullback-Leibler information and goodness-of-fit test with type ii censored data. *Statistical Papers*, 56(1):247–256. 11, 71, 137

Park, S. and Pakyari, R. (2015). Cumulative residual Kullback–Leibler information with the progressively Type-II censored data. *Statistics & Probability Letters*, 106:287–294. 11

Parzen, E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076. 96, 125

Pharwaha, A. P. S. and Singh, B. (2009). Shannon and non-shannon measures of entropy for statistical texture feature extraction in digitized mammograms. In *Proceedings of the world congress on engineering and computer science*, volume 2, pages 20–22. 23

Pradhan, B. and Kundu, D. (2009). On progressively censored generalized exponential distribution. *Test*, 18:497–515. 139

Pradhan, B. and Kundu, D. (2013). Inference and optimal censoring schemes for progressively censored Birnbaum–Saunders distribution. *Journal of Statistical Planning and Inference*, 143(6):1098–1108. 139

Prescott, P. (1976). On a test for normality based on sample entropy. *Journal of the Royal Statistical Society: Series B (Methodological)*, 38(3):254–256. 10

Psarrakos, G. and Navarro, J. (2013). Generalized cumulative residual entropy and record values. *Metrika*, 76(5):623–640. 8

Psarrakos, G. and Toomaj, A. (2017). On the generalized cumulative residual entropy with applications in actuarial science. *Journal of computational and applied mathematics*, 309:186–199. 36, 130

Pyke, R. (1965). Spacings. *Journal of the Royal Statistical Society: Series B (Methodological)*, 27(3):395–436. 57

Qiu, G. (2017). The extropy of order statistics and record values. *Statistics & Probability Letters*, 120:52–60. 9

Qiu, G. and Jia, K. (2018a). Extropy estimators with applications in testing uniformity. *Journal of Nonparametric Statistics*, 30(1):182–196. 10

Qiu, G. and Jia, K. (2018b). The residual extropy of order statistics. *Statistics & Probability Letters*, 133:15–22. 9, 109

Qiu, G., Wang, L., and Wang, X. (2019). On extropy properties of mixed systems. *Probability in the Engineering and Informational Sciences*, 33(3):471–486. 11, 77, 78, 88, 90

Rahimi, S., Tahmasebi, S., and Lak, F. (2020). Extended cumulative entropy based on kth lower record values for the coherent systems lifetime. *Journal of Inequalities and Applications*, 2020(1):1–22. 77

Rajesh, G. and Sunoj, S. (2019). Some properties of cumulative Tsallis entropy of order α. *Statistical papers*, 60(3):933–943. 45, 46

Ramsay, C. M. (1995). Loading gross premiums for risk without using utility theory. *Insurance Mathematics and Economics*, 2(16):169–170. 130

Rao, M., Chen, Y., Vemuri, B. C., and Wang, F. (2004). Cumulative residual entropy: a new measure of information. *IEEE transactions on Information Theory*, 50(6):1220–1228. 6, 9, 120

Raqab, M. Z. (2002). Inferences for generalized exponential distribution based on record statistics. *Journal of statistical planning and inference*, 104(2):339–350. 125

Rényi, A. (1961). On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, volume 4, pages 547–562. University of California Press. 5

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the national Academy of Sciences*, 42(1):43–47. 103, 126

Samaniego, F. J. (2007). *System signatures and their applications in engineering reliability*, volume 110. Springer Science & Business Media. 18, 77, 79

Saridakis, E. N., Bamba, K., Myrzakulov, R., and Anagnostopoulos, F. K. (2018). Holographic dark energy through Tsallis entropy. *Journal of Cosmology and Astroparticle Physics*, 2018(12):012. 10

Sathar, E. A. and Nair, R. D. (2021a). On dynamic weighted extropy. *Journal of Computational and Applied Mathematics*, 393:113507. 9, 102

Sathar, E. A. and Nair, R. D. (2021b). A study on weighted dynamic survival and failure extropies. *Communications in Statistics-Theory and Methods*, pages 1–20. 9, 102, 108

Sathar, E. A. and Nair R, D. (2021). On dynamic survival extropy. *Communications in Statistics-Theory and Methods*, 50(6):1295–1313. 9

Sati, M. M. and Gupta, N. (2015). Some characterization results on dynamic cumulative residual Tsallis entropy. *Journal of probability and statistics*, 2015. 8, 45

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic orders*. Springer. 14

Shaked, M. and Suarez-Llorens, A. (2003). On the comparison of reliability experiments based on the convolution order. *Journal of the American Statistical Association*, 98(463):693–702. 83

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423. 4, 5

Sharma, B. D. and Taneja, I. J. (1975). Entropy of type (α, β) and other generalized measures in information theory. *Metrika*, 22:205–215. 23

Stephens, M. A. (1974). Edf statistics for goodness of fit and some comparisons. *Journal of the American statistical Association*, 69(347):730–737. 132, 134

Stigler, S. M. (1974). Linear functions of order statistics with smooth weight functions. *The Annals of Statistics*, pages 676–693. 96

Sunoj, S. and Sankaran, P. (2012). Quantile based entropy function. *Statistics & Probability Letters*, 82(6):1049–1053. 8

Tahmasebi, S. (2020). Weighted extensions of generalized cumulative residual entropy and their applications. *Communications in Statistics-Theory and Methods*, 49(21):5196–5219. 8

Tahmasebi, S. and Parsa, H. (2019). Notes on cumulative entropy as a risk measure. *Stochastics and Quality Control*, 34(1):1–7. 130

Tahmasebi, S. and Toomaj, A. (2022). On negative cumulative extropy with applications. *Communications in Statistics-Theory and Methods*, 51(15):5025–5047. 103, 114, 130

Taufer, E. (2002). On entropy based tests for exponentiality. *Communications in Statistics-Simulation and Computation*, 31(2):189–200. 10

Toomaj, A. (2017). On the effect of dependency in information properties of series and parallel systems. *Statistical Methods & Applications*, 26(3):419–435. 77

Toomaj, A. and Atabay, H. A. (2022). Some new findings on the cumulative residual Tsallis entropy. *Journal of Computational and Applied Mathematics*, 400:113669. 46

Toomaj, A., Chahkandi, M., and Balakrishnan, N. (2021). On the information properties of working used systems using dynamic signature. *Applied Stochastic Models in Business and Industry*, 37(2):318–341. 11, 77

Toomaj, A. and Di Crescenzo, A. (2020). Connections between weighted generalized cumulative residual entropy and variance. *Mathematics*, 8(7):1072. 13

Toomaj, A., Di Crescenzo, A., and Doostparast, M. (2018). Some results on information properties of coherent systems. *Applied Stochastic Models in Business and Industry*, 34(2):128–143. 77

Toomaj, A. and Doostparast, M. (2014). A note on signature-based expressions for the entropy of mixed r-out-of-n systems. *Naval Research Logistics (NRL)*, 61(3):202–206. 77

Toomaj, A. and Doostparast, M. (2016). On the Kullback Leibler information for mixed systems. *International Journal of Systems Science*, 47(10):2458–2465. 77

Toomaj, A., Sunoj, S., and Navarro, J. (2017). Some properties of the cumulative residual entropy of coherent and mixed systems. *Journal of Applied Probability*, 54(2):379–393. 11, 77, 78, 91, 92

Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of statistical physics*, 52(1):479–487. 5

Tsallis, C., Gell-Mann, M., and Sato, Y. (2005). Asymptotically scale-invariant occupancy of phase space makes the entropy sq extensive. *Proceedings of the National Academy of Sciences*, 102(43):15377–15382. 6

Ullah, A. (1996). Entropy, divergence and distance measures with econometric applications. *Journal of Statistical Planning and Inference*, 49(1):137–162. 23

Varma, R. (1966). Generalizations of Renyi's entropy of order α. *Journal of Mathematical Sciences*, 1(7):34–48. 5

Vasicek, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society: Series B (Methodological)*, 38(1):54–59. 10, 67, 134

Viveros, R. and Balakrishnan, N. (1994). Interval estimation of parameters of life from progressively censored data. *Technometrics*, 36(1):84–91. 150

Walker, S. G. (2015). A probabilistic proof of the Hardy inequality. *Statistics & Probability Letters*, 103:6–7. 110

Wilk, G. and Włodarczyk, Z. (2008). Example of a possible interpretation of Tsallis entropy. *Physica A: Statistical Mechanics and its Applications*, 387(19-20):4809–4813. 6

Wolverton, C. and Wagner, T. (1969). Asymptotically optimal discriminant functions for pattern classification. *IEEE Transactions on Information Theory*, 15(2):258–265. 126

Xiong, H., Shang, P., and Zhang, Y. (2019). Fractional cumulative residual entropy. *Communications in Nonlinear Science and Numerical Simulation*, 78:104879. 8

Xiong, P., Zhuang, W., and Qiu, G. (2021). Testing symmetry based on the extropy of record values. *Journal of Nonparametric Statistics*, 33(1):134–155. 10

Xiong, P., Zhuang, W., and Qiu, G. (2022). Testing exponentiality based on the extropy of record values. *Journal of Applied Statistics*, 49(4):782–802. 10, 70

Yang, L. (2012). Study on cumulative residual entropy and variance as risk measure. In *2012 Fifth International Conference on Business Intelligence and Financial Engineering*, pages 210–213. IEEE. 130

Yousaf, F., Ali, S., and Shah, I. (2019). Statistical inference for the Chen distribution based on upper record values. *Annals of Data Science*, 6(4):831–851. 69, 70

Zamanzade, E. and Arghami, N. R. (2012). Testing normality based on new entropy estimators. *Journal of Statistical Computation and Simulation*, 82(11):1701–1713. 10

Zografos, K. and Nadarajah, S. (2005). Survival exponential entropies. *IEEE transactions on information theory*, 51(3):1239–1246. 22