RELATIONSHIP BETWEEN THE LEIMKUHLER CURVE AND RELIABILITY MEASURE CONCEPTS IN DOUBLE TRUNCATED VARIABLES

Vahideh Asghari, Gholamreza Mohtashami Borzadaran and Hadi Jabbari

•

Department of Statistics, Ferdowsi University of Mashhad, Mashhad-Iran
vahideh.asghari@mail.um.ac.ir grmohtashami@um.ac.ir jabbarinh@um.ac.ir

Abstract

This paper investigates the application of Leimkuhler curve and doubly truncated distributions in informetrics. Leimkuhler curve, ranking sources in descending order, emerges as a key tool for identifying efficient information sources. The study introduces a random variable representing the age of cited articles, influencing the probability distribution in retrospective citation analysis. Reliability measures, including mean residual life function and mean past residual life function are employed to analyze engineering and reliability aspects in informometric data. Truncation in probability distributions, particularly the doubly truncated distribution, is explored, revealing its broad applicability. The relationship between the Leimkuhler curve and truncated distributions will also be examined.

Keywords: leimkuhler curve, mean residual life function, mean past life function, double truncated random variable, risk measures

1. Introduction

The Leimkuhler curve and Lorenz curve serve as valuable tools in both information processing and economics. In economics, they are utilized to graphically represent the cumulative distribution of productivity versus resources. Moreover, they find application in analyzing the concentration of bibliometric distributions within the field of information sciences. The key distinction between the Lorenz [23] curve and the Leimkuhler curve lies in their ranking order: the Lorenz curve ranks sources (or individuals) in ascending order of productivity (income), while the Leimkuhler curve ranks them in descending order. In informetrics, where the focus often lies on identifying the most efficient sources of information, the Leimkuhler curve (LKC) serves as the equivalent graphical representation (see Burrell [9, 10, 13]). Its general definition can be found in Sarabia's work [28], and Balakrishnan et al. [3] have highlighted the relationships between the reliability function and the Leimkuhler curve. In retrospective citation studies within informetrics, interest is drawn to the age at which an article is cited, referring to the elapsed time from its publication to inclusion in the examined collection.

To conduct such studies, a single random variable X indicating article age determines the probability distribution of X. Burrell's research [12] has linked the data types reported in retrospective citation analysis with reliability models. Hazard rate, mean residual life function, reversed hazard rate, mean past life function, and vitality function are commonly employed tools for analyzing engineering and reliability aspects (Barlow and Proschan [4]). Truncation in probability distributions often arises in studies like reliability analysis, where unit failure may only be observed within specific time frames.(Abdul Sathar and Nair [1]) The broader utility of truncated distributions has been explored in many references such as Bernardic and Candel [8], Belzunce et al. [6], Kupka and Loo [21] , Ato and Bernardic [2], Coffey and Muller [14] and Nair et al. [25] were analyze truncated data in various disciplines, necessitating the examination of truncated versions of the standard distribution, particularly in relation to reliability issues and economic inequality.

The doubly truncated distribution encompasses right truncated, left truncated, and non-truncated distributions as special cases. Notably, Belzunce et al. [6] have identified properties of concentration truncated distribution curves, while Behdani et al. [5] have explored the properties and applications of doubly truncated distributions in income inequality.

The subsequent sections of the paper are organized as follows: preliminary information is presented in section 2. Section 3 covers measures of reliability and risk related to the Leimkuhler curve, section 4 specifies the relationship between the mean past life function and the Leimkuhler curve, section 5 explores the Leimkuhler curve of doubly truncated distributions, section 6 discusses the relation between the geometric vitality function and the Leimkuhler curve, and finally, section 7 presents some conclusions.

2. Preliminaries

Let X be a non-negative random variable with finite and positive mean $E(X) = \mu$. The distribution function and survival function of X are symbolized by F and $\bar{F} = 1 - F$, respectively. The quantile function is defined as $F^{-1}(t) = \inf\{x : F(x) \geq t, t \in (0,1)\}$. The Lorenz curve, introduced by Lorenz, is a widely used graphical tool for illustrating and examining size distribution and wealth. For a random income variable X, the Lorenz curve is defined as:

$$L(p) = \frac{\int_0^p F^{-1}(t)dt}{\int_0^1 F^{-1}(t)dt}, \quad 0 \leq p \leq 1. \tag{1}$$

Here, the function $L(p)$ represents the cumulative percentage of total income earned by the lowest $100p\%$ earners. This paper presents the main result in the form of the Leimkuhler curve $K(p)$, as proposed by Sarabia [28] . The Leimkuhler curve is defined as:

$$K(p) = \frac{\int_{1-p}^1 F^{-1}(t)dt}{\int_0^1 F^{-1}(t)dt} = \frac{\int_{F^{-1}(1-p)}^{\infty} tf(t)dt}{\int_0^1 F^{-1}(t)dt}, \quad 0 \leq p \leq 1. \tag{2}$$

This curve indicates the share of total productivity returning to sources with productivity $100p\%$ greater. The Leimkuhler curve is essentially an inverted image of the Lorenz curve reflected along the diagonal line at 45 degrees.

The definitions of the Lorenz and Leimkuhler curves, $L(p)$ and $K(p)$ respectively, imply that these curves are linked by the relationship:

$$K(p) + L(1-p) = 1. \tag{3}$$

It is evident that the Leimkuhler curve acts as a distribution function, exhibiting continuity on $[0,1]$, with a second derivative $K^{''}(p) \leq 0$, $K(0) = 0$, $K(1) = 1$, $K^{'}(1^-) \geq 0$ and $K(p) \geq p$, among other trivial properties. The Gini index G, representing the area between the Leimkuhler and Lorenz curves, serves as a measure of income inequality. The Gini coefficient theoretically ranges from 0 (complete equality) to 1 (complete inequality), expressed as:

$$G = \int_0^1 [K(p) - L(p)]dp = 1 - \frac{\int_0^\infty \bar{F}(x)^2 dx}{E(x)}. \tag{4}$$

A low Gini index suggests a more equitable distribution of productivity, whereas a high index indicates a more unequal distribution. To illustrate, let's consider a classical Pareto distribution with a distribution function.

$$F(x) = 1 - (\frac{x}{\sigma})^{-\alpha}, \ x \geq \sigma \tag{5}$$

where $\sigma > 0$ is a scale parameter and $\alpha > 0$ is a shape parameter. The Lorenz and the Leimkuhler curves of the classical Pareto distribution are each given by

$$L(p;\alpha) = 1 - (1-p)^{1-\frac{1}{\alpha}}, \ 0 \le p \le 1,$$

$$K(p;\alpha) = p^{1-\frac{1}{\alpha}}, \ 0 \le p \le 1. \tag{6}$$

Using relation (4), the Gini index of the classical Pareto distribution is

$$G(\alpha) = \frac{1}{2\alpha - 1}.$$

Figure 1 illustrates how the Lorenz curve, Leimkuhler curve, and Gini index are visualized for a Pareto distribution. The Gini coefficient is represented by the ratio of the area between the Lorenz curve and the line of equality to the total area under the line of equality. In this plot, the Gini coefficient is indicated by the relative size of the yellow-shaded area, demonstrating that the classical Pareto distribution for $\alpha = 2$ has significant inequality.

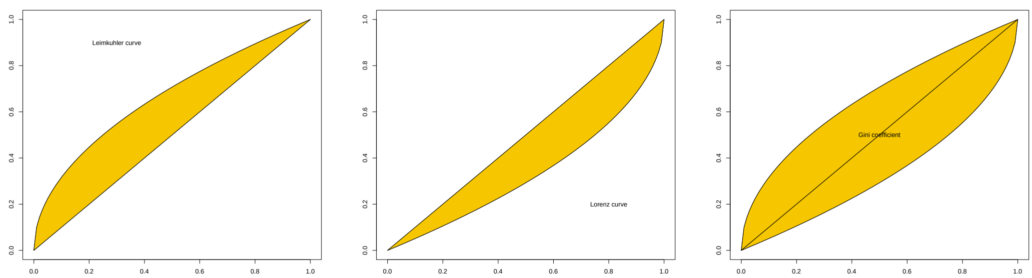

Figure 1: *Plot of the Lorenz and Leimkuhler curves and Gini index of the classical Pareto distribution for $\alpha = 2$.*

In prospective studies, there exists a time span until the initial citation occurs, as discussed by Burrell [11]. Assuming X represents a continuous random variable denoting the age of an object, the random variable $X(t,\infty) = \{X - t \mid X > t\}$ signifies the remaining lifespan of an entity aged t. The anticipated additional lifespan of an item, provided it has endured until time t, forms a function dependent on t, referred to as the mean residual life (MRL) function. This function, introduced by Knight [19], represents the average remaining life of a component that has survived until time t, and is given by:

$$m(t) = E(X - t \mid X > t) = \frac{1}{\bar{F}(t)} \int_t^\infty (x-t) dF(x). \tag{7}$$

In retrospective studies, we define the random variable $X(.,t) = \{t - X \mid X < t\}$, termed the inactivity time or elapsed lifetime of X. This variable signifies the duration since the expiration of a unit with a lifetime of at most t. We express this as the mean past life (MPL) function:

$$m^*(t) = E(t - X \mid X < t) = \frac{1}{F(t)} \int_0^t (t-x) dF(x). \tag{8}$$

In many practical situations, lifespan information is only available between two points in time, which necessitates the examination of reliability measures under the condition of truncated random variables. In reliability theory and survival analysis, often only individuals whose event time falls within a certain time interval are observed, and we have information on the lifetime between two points in time. Doubly truncated variables are the most general case, as they include right-truncated, left-truncated and non-truncated variables.

If X denotes the lifetime of a unit, then the random variable is $X(t_1, t_2) = \{X - t_1 | t_1 < X < t_2\}$ means remaining lifetime truncated twice. Note that the random variable, $X(t_1, \infty)$ is the special case of $X(t_1, t_2)$ when t_2 tends to ∞. Also, doubly truncated past lifetime is the random variable $X^*(t_1, t_2) = (t_2 - X | t_1 < X < t_2)$), which is special case $t_1 = 0$, it is the past lifetime random variable $X(., t)$.

The mean residual life function for a doubly truncated variable is defined as:

$$m(t_1, t_2) = E(X - t_1 | t_1 < X < t_2) = \frac{1}{F(t_2) - F(t_1)} \int_{t_1}^{t_2} (x - t_1) dF(x). \tag{9}$$

Similarly, the mean past life function for a doubly truncated variable can be written as:

$$m^*(t_1, t_2) = E(t_2 - X | t_1 < X < t_2) = \frac{1}{F(t_2) - F(t_1)} \int_{t_1}^{t_2} (t_2 - x) dF(x). \tag{10}$$

In engineering and other branches related to reliability, $\nu(t) = E(X | X > t)$ is referred to as the vitality function and $\nu^*(t) = E(X | X < t))$ is termed the past vitality function. Additionally, the vitality function for a doubly truncated variable is defined as:

$$\nu(t_1, t_2) = E(X \mid t_1 < X < t_2) = \frac{1}{F(t_2) - F(t_1)} \int_{t_1}^{t_2} x dF(x) \tag{11}$$

which is related to MRL and MPL functions via

$$\nu(t_1, t_2) = m(t_1, t_2) + t_1, \tag{12}$$

$$= t_2 - m^*(t_1, t_2). \tag{13}$$

3. Leimkuhler curve and reliability measures

As is commonly understood, the hazard rate function $r(t) = \dfrac{f(t)}{\bar{F}(x)}$ of F indicates whether the random variable X exhibits an increasing failure rate (IFR) or a decreasing failure rate (DFR) if $r(x)$ is respectively an increasing or decreasing function on the interval $(0, \infty)$.

Barlow and Proschan [4] demonstrated that if F is $IFR(DFR)$, then under certain conditions:

$$\begin{cases} \bar{F}(t) \geq (\leq) e^{-at}, & t \leq F^{-1}(p) \\ \bar{F}(t) \leq (\geq) e^{-at}, & t \geq F^{-1}(p) \end{cases}$$

where $a = -\dfrac{\ln(1-p)}{F^{-1}(p)}$.

Proposition 1. Let X be a non-negative random variable with distribution function F. If X is $IFR(DFR)$, then

$$K(p) \leq (\geq) \mu^{-1} [pF^{-1}(1-p)(1 - (\ln p)^{-1})]. \tag{14}$$

Proof. By using (2), we obtain:

$$\begin{aligned} K(p) &= \frac{1}{\mu} \int_{F^{-1}(1-p)}^{\infty} t f(t) dt \\ &= \frac{1}{\mu} [pF^{-1}(1-p) + \int_{F^{-1}(1-p)}^{\infty} \bar{F}(t) dt] \\ &\leq (\geq) \frac{1}{\mu} [pF^{-1}(1-p) + \int_{F^{-1}(1-p)}^{\infty} e^{\frac{t \ln p}{F^{-1}(1-p)}} dt] \\ &= \mu^{-1} [pF^{-1}(1-p)(1 - (\ln p)^{-1})]. \end{aligned}$$

This result is thus derived. ■

The following assertions are also noticeable:

- The stop loss transformation

$$\begin{aligned}\pi(t) = E(X-t)_+ &= \int_t^\infty \bar{F}(x)dx \\ &= -t\bar{F}(t) + \int_t^\infty xf(x)dx \\ &= -t\bar{F}(t) + \int_{F^{-1}(F(t))}^\infty xf(x)dx \\ &= -t\bar{F}(t) + \mu K(\bar{F}(t)),\end{aligned}$$

  where $(X-t)_+$ means $X \geq t$. On taking $t = F^{-1}(p)$ then

$$K(p) = \frac{1}{\mu}(\pi_X(F^{-1}(p)) + pF^{-1}(1-p)).$$

- The total time on test transformation (TTT)

$$\begin{aligned}T(p) &= \int_0^{F^{-1}(p)} \bar{F}(t)dt \\ &= (1-p)F^{-1}(p) + \mu - \mu K(1-p),\end{aligned}$$

  leading to

$$K(p) = \frac{1}{\mu}[pF^{-1}(1-p) - T(1-p)] + 1.$$

- $e(t) = E(\frac{X}{t} \mid X > t)$), representing the expected proportion of income up to t for incomes greater than t, can be expressed in terms of LKC

$$\begin{aligned}e(t) &= \frac{1}{t\bar{F}(t)} \int_t^\infty xf(x)dx \\ &= \frac{\mu K(\bar{F}(t))}{t\bar{F}(t)}.\end{aligned}$$

- The vertical diameter inequality index introduced in Eliazar [15] is given by

$$\begin{aligned}\varepsilon_{Vdiam} &= E(\frac{X}{\mu} \mid X \leq median) \\ &= \frac{\int_0^{median} \frac{x}{\mu} f(x)dx}{F(median)} \\ &= \frac{2}{\mu}[\mu - \int_{median}^\infty xf(x)dx] \\ &= 2[1 - K(\frac{1}{2})].\end{aligned}$$

In insurance and risk, the value at risk (VaR) for risk of X with distribution F is defined as $VaR(x;p) = F^{-1}(p) = \inf\{x \in R \mid F(x) \geq p\};\ p \in (0,1)$. Some of the measures related to risks based on VaR as mentioned in Belzunce et al. [7] arranged in table 1.

Example 1. Let X have a classical Pareto distribution (5) with Leimkuhler curve (6), $F^{-1}(x) =$

Table 1: *Some risk measures*

Name of measure	formula	LKC	reliability function
Tail value at risk	$TVaR(X;p)=\frac{1}{1-p}\int_p^1 F^{-1}(u)du$	$\frac{\mu}{1-p}K(1-p)$	$\nu(F^{-1}(p))$
Conditional tail expectation	$CTE(X;p)=E(X\mid X>F^{-1}(p))$	$\frac{\mu}{1-p}K(1-p)$	$\nu(F^{-1}(p))$
Conditional value at risk	$CVaR(X;p)=E(X-F^{-1}(p)\mid X>F^{-1}(p))$	$\frac{\mu}{1-p}K(1-p)-F^{-1}(p)$	$m(F^{-1}(p))$
Expected shortfall	$ES(X;p)=E(X-F^{-1}(p))_+^1$	$\mu K(1-p)-(1-p)F^{-1}(p)$	$\pi(F^{-1}(p))$
Expected proportional shortfall	$EPS(X;p)=E[(\frac{X-F^{-1}(p)}{F^{-1}(p)})_+]$	$\frac{\mu K(1-p)-(1-p)F^{-1}(p)}{F^{-1}(p)}$	$\frac{\pi(F^{-1}(p))}{F^{-1}(p)}$

[1] $(x)_+=\begin{cases}0 & x<0\\ x & x\geq 0\end{cases}$

$\sigma(1-x)^{\frac{-1}{\alpha}}$ and $\mu=\frac{\alpha\sigma}{1-\alpha}$, $\alpha>1$, based on Table 1 we have

$$\begin{aligned}TVaR(X;p)&=\frac{\alpha\sigma}{\alpha-1}(1-p)^{\frac{-1}{\alpha}},\\ CTE(X;p)&=\frac{\alpha\sigma}{\alpha-1}(1-p)^{\frac{-1}{\alpha}},\\ CVaR(X;p)&=\sigma(1-p)^{\frac{-1}{\alpha}},\\ ES(X;p)&=\sigma(1-p)^{1-\frac{1}{\alpha}},\\ EPS(X;p)&=1-p.\end{aligned}$$

4. Mean past life function and Leimkuhler curve

This section commences by delineating the connection between the Leimkuhler curve and the mean residual life function through a theorem:

Theorem 1. (Balakrishnan et al. [3]) Let X be a random variable with cumulative distribution function F , $\bar{F}(x)=1-F(x)=P(X>x)$, Leimkuhler curve $K(p)$ and expectation μ. There exists a relationship between the Leimkuhler curve and the MRL function expressed as follows:

$$m(t)=\frac{\mu}{\bar{F}(t)}K[\bar{F}(t)]-t,\quad t>0. \tag{15}$$

The MPL function and LKC can also be connected similarly to the theorem:

Theorem 2. Let X be a random variable with cumulative distribution function F, survival function $\bar{F}(x)=1-F(x)$, Leimkuhler curve $K(p)$ and expectation μ. The relationship between the Leimkuhler curve and the mean past life function is described as follows:

$$m^*(t)=t-\frac{\mu}{F(t)}[1-K(\bar{F}(t))],\quad t>0. \tag{16}$$

Proof. Beginning with the definition of the MPL function as (8), we express:

$$\begin{aligned}m^*(t)&=t-\frac{1}{F(t)}[\mu-\int_t^\infty xf(x)dx]\\ &=t-\frac{1}{F(t)}[\mu-\mu K(\bar{F}(t))]\\ &=t-\frac{\mu}{F(t)}[1-K(\bar{F}(t))].\end{aligned}$$

■

5. Double Truncated Distributions and Leimkuhler Curve

Truncated data holds significant importance in statistical analysis, representing variables that have been limited or constrained due to specific selection criteria. When observations falling outside certain ranges or conditions are disregarded in analysis, truncated variables emerge. Examples of this include distributions like doubly truncated exponential, normal, and Cauchy distributions. A truncated variable undergoes restriction to a defined range or set of conditions, leading to alterations in its probability density function (PDF) and cumulative distribution function (CDF) compared to the untruncated variable.

Let's consider a random variable X with PDF $f(x)$ and CDF $F(x)$. We want to find the PDF and CDF of the truncated variable

$$X_{(t_1,t_2)} = \{X|t_1 < X < t_2\}.$$

The PDF and CDF of the truncated variable can be expressed as:

$$f(x|t_1,t_2) = \frac{f(x)}{F(t_2) - F(t_1)}, \quad x > 0, \quad t_1, t_2 > 0. \tag{17}$$

$$F(x|t_1,t_2) = \begin{cases} 0 & x < t_1 \\ \dfrac{F(x) - F(t_1)}{F(t_2) - F(t_1)} & t_1 \leq x \leq t_2 \\ 1 & x > t_2. \end{cases} \tag{18}$$

The quantile and the survival function of $X_{(t_1,t_2)}$ is

$$F^{-1}(p|t_1,t_2) = F^{-1}(pF(t_2) + (1-p)F(t_1)),$$

$$\bar{F}(x|t_1,t_2) = \frac{\bar{F}(x) - \bar{F}(t_2)}{\bar{F}(t_1) - \bar{F}(t_2)}.$$

The LKC of $X_{(t_1,t_2)}$ has the following form:

$$K(p|t_1,t_2) = \frac{K[p(F(t_2) - F(t_1)) + \bar{F}(t_2)] - K(\bar{F}(t_2))}{K(\bar{F}(t_1)) - K(\bar{F}(t_2))}, \quad 0 \leq p \leq 1.$$

The Gini index for the double truncation is given by:

$$2\int_0^1 K(p|t_1,t_2)dp - 1.$$

Truncation to the right is a special case of double truncation when $t_1 \to 0$. This is evident from the following:

$$F^{-1}(p|t_2) = F^{-1}(pF(t_2)),$$

and its Leimkuhler curve is:

$$K(p|t_2) = \frac{K[pF(t_2) + \bar{F}(t_2)] - K(\bar{F}(t_2))}{1 - K(\bar{F}(t_2))}, \quad 0 \leq p \leq 1.$$

When $t_2 \to \infty$ in the doubly truncated distribution, we have a left truncated distribution:

$$F^{-1}(p|t_1) = F^{-1}(p\bar{F}(t_1) + F(t_1)),$$

and is the Leimkuhler curve

$$K(p|t_1) = \frac{K(p\bar{F}(t_1))}{K(\bar{F}(t_1))}, \quad 0 \leq p \leq 1.$$

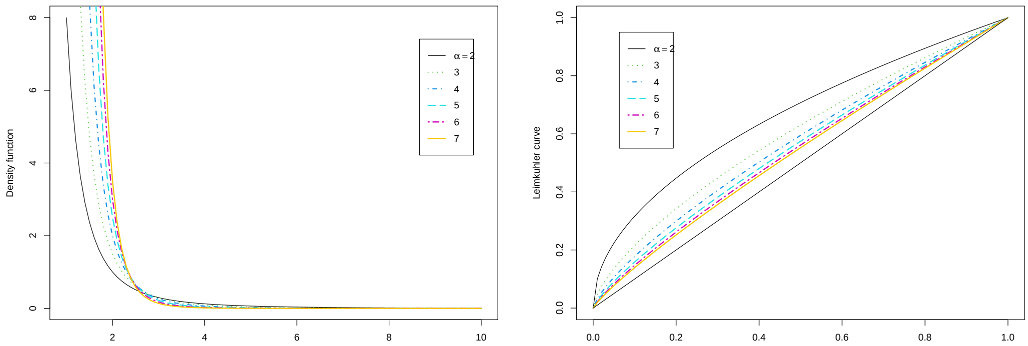

Figure 2: *The density function and Leimkuhler curve for Pareto (2, α)*

The density function and the Leimkuhler curve of the Pareto distribution with various parameters ($\alpha = 2, 3, \ldots, 7$) are depicted in Figure 2. It is evident that the density function of the Pareto distribution consistently decreases. Therefore, it proves advantageous in modeling distributions of high or moderate productivity. Pareto Leimkuhler curves never intersect, for $X_i \sim Pareto(\sigma, \alpha), i = 1, 2$

$$X_1 \leq_{LKC} X_2 \Longleftrightarrow \alpha_1 \leq \alpha_2.$$

Figure 3 illustrates the density function and Leimkuhler curves for the original Pareto distribution (O), left-truncated (L), right-truncated (R), and double-truncated (D) for Pareto (1, 3), $t_1 = 3, t_2 = 6$. It's noteworthy that the Leimkuhler curve of the left-truncated distribution remains unaffected by the truncation point.

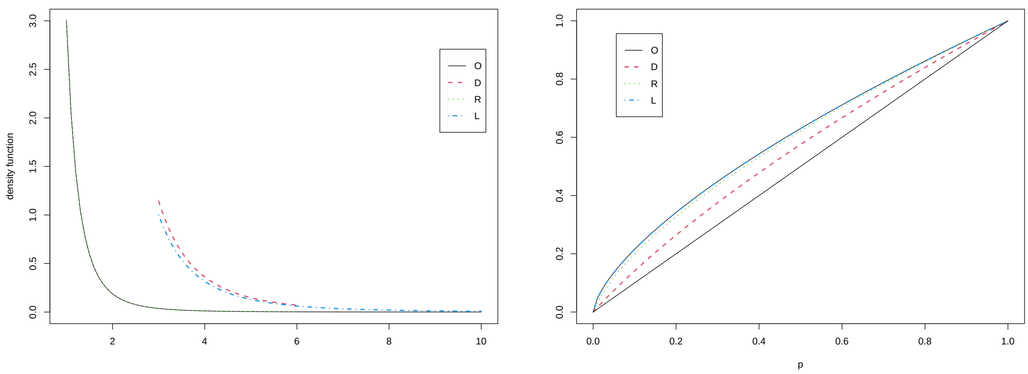

Figure 3: *Density function and Leimkuhler curve for the truncated and original Pareto distribution for Pareto (2,3),* $t_1 = 3, t_2 = 6$

The following theorem, an extended version by Balakrishnan et al. [3], pertains to the doubly truncated case:

Theorem 3. Let X be a non-negative random variable with cumulative distribution function $F(x)$, survival function $\bar{F}(x) = 1 - F(x)$, Leimkuhler curve $K(p)$ and expectation μ. Then, the relations between the Leimkuhler curve and the mean residual (past) lifetime of a doubly truncated random variable are given by:

$$m(t_1, t_2) = \frac{\mu(K[\bar{F}(t_1)] - K[\bar{F}(t_2)])}{\bar{F}(t_1) - \bar{F}(t_2)} - t_1, \tag{19}$$

$$m^*(t_1,t_2) = t_2 - \frac{\mu(K[\bar{F}(t_1)] - K[\bar{F}(t_2)])}{\bar{F}(t_1) - \bar{F}(t_2)}, \tag{20}$$

$$\nu(t_1,t_2) = \frac{\mu(K[\bar{F}(t_1)] - K[\bar{F}(t_2)])}{\bar{F}(t_1) - \bar{F}(t_2)}, \tag{21}$$

which holds for all $0 < t_1 < t_2$.

Proof. Via (9)

$$\begin{aligned} m(t_1,t_2) = E(X - t_1 \mid t_1 < X < t_2) &= \int_{t_1}^{t_2} \frac{(x-t_1)f(x)}{F(t_2)-F(t_1)} dx \\ &= \frac{1}{F(t_2)-F(t_1)}\left(\int_{t_1}^{\infty} xf(x)dx - \int_{t_2}^{\infty} xf(x)dx\right) - t_1. \end{aligned}$$

Taking $t_1 = F^{-1}(1-p_1)$ and $t_2 = F^{-1}(1-p_2)$ we can express
$m(t_1,t_2) = \dfrac{1}{p_1 - p_2}[\int_{F^{-1}(1-p_1)}^{\infty} xf(x)dx - \int_{F^{-1}(1-p_2)}^{\infty} xf(x)dx] - F^{-1}(1-p_1)$,
By changing variables to $z_1 = F^{-1}(1-p_1)$ and $z_2 = F^{-1}(1-p_2)$), we derive the relation in (19). Similarly, utilizing (10) $m^*(t_1,t_2)$ can be expressed as:

$$\begin{aligned} m^*(t_1,t_2) = E(t_2 - X \mid t_1 < X < t_2) &= \int_{t_1}^{t_2} \frac{(t_2-x)f(x)}{F(t_2)-F(t_1)} dx \\ &= t_2 - \frac{1}{F(t_2)-F(t_1)}\left(\int_{t_1}^{\infty} xf(x)dx - \int_{t_2}^{\infty} xf(x)dx\right). \end{aligned}$$

By performing a similar change of variables, we arrive at the relation in (20) Finally, using relation (12) we establish the relationship in (21). ■

Theorem 22 is a special case of Theorem 3 when $t_2 \to \infty$. Additionally, when $t_1 = 0$ in Theorem 3, we obtain $m^*(t)$ as presented in Theorem 2.

Example 2. Let X be random variable with classical Pareto distribution (5) and Leimkuhler curve (6), the density function and cumulative distribution function for the right-truncated and doubly truncated variable (with $t_1 = \sigma$ and $t_2 = t$) are given by:

$$f(x \mid \sigma < X < t) = \frac{\alpha\sigma x^{-(1+\alpha)}}{1 - \left(\frac{t}{\sigma}\right)^{-\alpha}},$$

$$F(x \mid \sigma \leq X \leq t) = \frac{1 - \left(\frac{x}{\sigma}\right)^{-\alpha}}{1 - \left(\frac{t}{\sigma}\right)^{-\alpha}}, \quad \alpha > 0.$$

Now, if we assume $\sigma = 1$, $t = \beta + 1$, and $\alpha = 1$, the cumulative distribution function of the right-truncated and doubly truncated Pareto variable is the same as that of the Bradford distribution (see Leimkuhler [22]):

$$f(x \mid 1 < X < 1+\beta) = \frac{1+\beta}{\beta x^2}, \quad 1 < x < 1+\beta,$$

$$F(x \mid 1 < X < 1+\beta) = \frac{1+\beta}{\beta}\left(1 - \frac{1}{x}\right).$$

The mean of the truncated Pareto random variable is equal to the survival function of this variable:

$$\mu = \frac{1+\beta}{\beta}\ln(1+\beta).$$

Using equation (21), we have:

$$\nu(1,(1+\beta)) = \frac{1}{F(1+\beta) - F(1)} \int_1^{1+\beta} \frac{1}{x} dx = \frac{1+\beta}{\beta}\ln(1+\beta) = \mu.$$

6. Geometric Vitality Functions and their Links with Leimkuhler Curve

In recent years, special attention has been given to various forms of conditional distribution functions. The function

$$\phi(t_1, t_2) = E(h(x)|t_1 < X < t_2), \tag{22}$$

is called the geometric vitality function. This function is similar to the survival function and is used in the analysis of lifetime data. The goal is to establish the relationship between this function and the Limkohler curve, which we derive in the following theorem. If $\{h(X)|t_1 < X < t_2\}$ is an increasing (decreasing) function, then
$\{h(X)|t_1 < X < t_2\} = \{Y|h^{-1}(t_1) < Y < h^{-1}(t_2)\} \quad (\{Y|h^{-1}(t_2) < Y < h^{-1}(t_1)\})$.

Theorem 4. Assume that the random variable X has a distribution F, and the function $\{h(X)|t_1 < X < t_2\}$ is a continuous and increasing function with respect to X. Then

$$\phi(t_1, t_2) = \frac{\mu_g(y)[K(\bar{F}(t_1)) - K(\bar{F}(t_2))]}{F(t_2) - F(t_1)}$$

where $\mu_g(z) = \int_{h(a)}^{h(b)} h(z)f(z)dz, \quad z \in (a, b)$.

Proof. Assuming $Y = h(X)$, the proof is as follows: $Y = h(X)$ implies $G(y) = F(h^{-1}(y))$, and

$$\begin{aligned}
\phi(t_1, t_2) &= E(h(X)|t_1 < X < t_2) \\
&= E(Y|t_1 < h^{-1}(y) < t_2) \\
&= E(Y|h(t_1) < Y < h(t_2)) \\
&= \frac{\int_{h(t_1)}^{h(t_2)} yg(y)dy}{G(h(t_2)) - G(h(t_1))} \\
&= \frac{\mu_g(y)[K(\bar{G}(h(t_1))) - K(\bar{G}(h(t_2)))]}{G(h(t_2)) - G(h(t_1))} \\
&= \frac{\mu_g(y)[K(\bar{F}(t_1)) - K(\bar{F}(t_2))]}{F(t_2) - F(t_1)}.
\end{aligned}$$

■

Various special cases of Theorem 4 are noteworthy:

(i) When h is decreasing, then

$$\phi(t_1, t_2) = \frac{\mu_g^*(y)[K(\bar{F}(t_2)) - K(\bar{F}(t_1))]}{F(t_1) - F(t_2)}, \tag{23}$$

where $\mu_g^*(z) = \int_{h(b)}^{h(a)} h(z)f(z)dz, \quad z \in (a, b)$.

(ii) When $h(x) = x - t_1$, it results in a doubly truncated residual life denoted as $X(t_1, t_2) = \{X - t_1|t_1 < X < t_2\}$ as discussed in Sankaran and Sunoj [29], indicating its association with the Leimkuhler curve. As $t_2 \to \infty$, , the connection between the Mean Residual Life (MRL) function and Leimkuhler curve becomes evident.

(iii) The MPL function for doubly truncated variables, as outlined in Khorashadizadeh et al. [18] and Ruiz and Navarro [26, 27] for $h(x) = t_2 - x$ in Theorem 4 , reveals a specific case where $t = 0$ leads to the MPL function. This relationship becomes apparent when put into relation (9) , demonstrating their connection to the Leimkuhler curve.

(iv) The doubly truncated geometric vitality is defined as

$$\phi(t_1, t_2) = E(\ln X|t_1 < X < t_2), \tag{24}$$

while a special case of Theorem 4, through $h(x) = \ln x$ is also the link with LKC specified. When $t_2 \to \infty$, tends to infinity, it leads to (10) in Nair and Rajesh [24] illustrating the relationship between geometric vitality and the Leimkuhler curve.

(v) When $h(x) = -\log \frac{f(x)}{\bar{F}(t_1) - \bar{F}(t_2)}$, then Theorem 4 leads to doubly truncated dynamic entropy, as discussed in Khorashadizadeh et al. [17].

7. Conclusions

This paper has demonstrated the versatility and significance of the Leimkuhler curve across various domains, including information processing, economics, and reliability analysis. The integration of reliability measures, such as the mean past life function, with the Leimkuhler curve, adds depth to the analysis by connecting informetric data with engineering principles, providing a novel approach to understanding data longevity and efficiency. The exploration of doubly truncated distributions further expands the applicability of the Leimkuhler curve, showing its relevance in contexts where data is naturally bounded or limited. Additionally, the study of geometric vitality functions and their relationship with the Leimkuhler curve highlights the broader implications of these mathematical tools, especially in areas related to resource distribution and productivity analysis. By bridging concepts from different fields, this paper underscores the Leimkuhler curve's potential as a comprehensive tool for both theoretical and applied research.

References

[1] Abdul Sathar EI and Nair KRM (2002). *On truncated versions of certain measures of inequality and stability*. Diss, Department of Statistics, Faculty of Science.

[2] Ato, J. C. and Bernadic, M. (2011). *The characterization of the symmetric Lorenz curves by doubly truncated mean function*. Communications in Statistics-Theory and Methods, 40(18), 3269-3280.

[3] Balakrishnan, N., Sarabia, J. M. and Kolev, N. (2010). *A simple relation between the Leimkuhler curve and the mean residual life*. Journal of Informetrics, 4(4), 602-607.

[4] Barlow, R. E. and Proschan, F. (1996). *Mathematical theory of reliability*.(Vol. 17). Siam.

[5] Behdani, Z., Mohtashami Borzadaran, G. R. and Gildeh, B. S. (2020). *Some properties of double truncated distributions and their application in view of income inequality*. Computational Statistics, 35(1), 359-378.

[6] Belzunce, F., Candel, J. and Ruiz, J. M. (1995). *Ordering of truncated distributions through concentration curves*. Sankhyā: The Indian Journal of Statistics, Series A, 375-383.

[7] Belzunce, F., Pinar, J. F., Ruiz, J. M. and Sordo, M. A. (2012). *Comparison of risks based on the expected proportional shortfall*. Insurance: Mathematics and Economics, 51(2), 292-302.

[8] Bernadic, M. and Candel, J. (2012). *The doubly truncated function of indices on discrete distributions*. Statistical Papers, 53(1), 177-193.

[9] Burrell, Q. (1991). *The Bradford distribution and the Gini index*. Scientometrics, 21(2), 181-194.

[10] Burrell, Q. L. (1992). *The Gini index and the Leimkuhler curve for bibliometric processes*. Information processing and management, 28(1), 19-33.

[11] Burrel, Q. L. (2001). *Stochastic modelling of the first-citation distribution*. Scientometrics, 52(1), 3-12.

[12] Burrell, Q. (2002). *Modelling citation age data: Simple graphical methods from reliability theory*. Scientometrics, 55(2), 273-285.

[13] Burrell, Q. L. (2005). *Symmetry and other transformation features of Lorenz/Leimkuhler representations of informetric data*. Information processing and management, 41(6), 1317-1329.

[14] Coffey, C. S. and Muller, K. E. (2000). *Properties of doubly-truncated gamma variables*. Communications in Statistics-Theory and Methods, 29(4), 851-857.

[15] Eliazar, I. (2018). *A tour of inequality*. Annals of Physics, 389, 306-332.

[16] Gupta, R. C. and Bradley, D. M. (2003). *Representing the mean residual life in terms of the failure rate*. Mathematical and Computer Modelling, 37(12-13), 1271-1280.

[17] Khorashadizadeh, M., Rezaei Roknabadi, A. H. and Mohtashami Borzadaran, G. R. (2013). *Doubly truncated (interval) cumulative residual and past entropy*. Statistics and Probability Letters, 83(5), 1464-1471.

[18] Khorashadizadeh, M., Rezaei Roknabadi, A. H. and Mohtashami Borzadaran, G. R. (2017). *Some characterization results based on doubly truncated reversed residual lifetime random variable*. REVSTATStatistical Journal, 15(1), 89-105.

[19] Knight, W. R. (1959). *Exponential and subexponential distributions in statistical life testing* (Doctoral dissertation, University of Toronto).

[20] Kotz, S. and Shanbhag, D. N. (1980). *Some new approaches to probability distributions*. Advances in Applied Probability, 12(4), 903-921.

[21] Kupka, J. and Loo, S. (1989). *The hazard and vitality measures of ageing*. Journal of applied probability, 26(3), 532-542.

[22] Leimkuhler, F. F. (1967). *The bradford distribution*. Journal of documentation, 23(3), 197-207.

[23] Lorenz, M. O. (1905). *Methods of measuring the concentration of wealth*. Publications of the American statistical association, 9(70), 209-219.

[24] Nair, K. R. M. and Rajesh, G. (2000). *Geometric vitality function and its applications to reliability*. IAPQR Transactions, 25(1), 1-8.

[25] Nair, N. U., Sankaran, P. G. and Vineshkumar, B. (2012). *Characterization of distributions by properties of truncated Gini index and mean difference*. Metron, 70(2-3), 173-191.

[26] Ruiz, J. M. and Navarro, J. (1995). *Characterization of discrete distributions using expected values*. Statistical Papers, 36(1), 237-252.

[27] Ruiz, J. M. and Navarro, J. (1996). *Characterizations based on conditional expectations of the doubled truncated distribution*. Annals of the Institute of Statistical Mathematics, 48(3), 563-572.

[28] Sarabia, J. M. (2008). *A general definition of the Leimkuhler curve*. Journal of Informetrics, 2(2), 156-163.

[29] Sankaran, P. G. and Sunoj, S. M. (2004). *Identification of models using failure rate and mean residual life of doubly truncated random variables*. Statistical Papers, 45(1), 97-109.