Available online at www.sciencedirect.com
ScienceDirect
Journal of Informetrics 2 (2008) 156–163
Journal of INFORMETRICS An International Journal
www.elsevier.com/locate/joi

A general definition of the Leimkuhler curve

José María Sarabia *

Department of Economics, University of Cantabria, Avda. de los Castros s/n, 39005 Santander, Spain

Received 15 October 2007; received in revised form 16 January 2008; accepted 17 January 2008

Abstract

In this paper, we provide a general definition of the Leimkuhler curve in terms of the theoretical cumulative distribution function. The definition applies to discrete, continuous and mixed random variables. Several examples are given to illustrate the use of the formula.

Keywords: Productivity; Cumulative distribution function; Incomplete beta function; Lorenz curve

1. Introduction

The Lorenz and Leimkuhler curves are two important instruments in informetrics (Burrell, 1991, 2006, 2007; Egghe, 2005; Rousseau, 2007). These curves plot the cumulative proportion of total productivity against the cumulative proportion of sources. The difference between the two constructions is that for the Lorenz curve in economics, one arranges the sources in increasing order of productivity while for the Leimkuhler curves in informetrics they are arranged in decreasing order.

In this paper, we provide a general definition of the Leimkuhler curve in terms of the theoretical cumulative distribution function (cdf). The definition applies to discrete, continuous and mixed random variables.

The contents of this paper are as follows. In Section 2, we propose a general definition of the Leimkuhler curve. Section 3 shows the equivalence with the classical definition. Finally, several illustrative examples are presented in Section 4, including exponential, Singh–Maddala, power, arcsine, lognormal, mixed Pareto and geometric distributions.

2. The general definition

Let X be a random variable, which denotes the productivity of a randomly chosen source. We assume that X has cumulative distribution function $F_X(x)$, which represents the proportion of the population with productivity less than or equal to x.

Now, let $\mathcal{L}$ be the class of all non-negative random variables with positive finite expectations. For a random variable X in $\mathcal{L}$ with cumulative distribution function F_X we define its inverse distribution function by

$$F_X^{-1}(y) = \inf\{x \; : \; F_X(x) \geq y\}. \tag{1}$$

* Corresponding author. Tel.: +34 942 201635; fax: +34 942 201603.
E-mail address: sarabiaj@unican.es.

1751-1577/$ – see front matter
doi:10.1016/j.joi.2008.01.002

The mathematical expectation of X is

$$\mu_X = \int_0^\infty x\,\mathrm{d}F_X(x) = \int_0^1 F_X^{-1}(y)\mathrm{d}y,$$

after making the change of variable $y = F_X(x)$, and where the last expression is given in terms of the inverse of the cumulative distribution function. Next definition makes use of the inverse of the cdf, as in the Gastwirth's (1971) definition of the Lorenz curve. However, the Gastwirth's formulation has not previously been used in an informetrics context.

Definition 1. Let $X \in \mathcal{L}$ with cumulative distribution function F_X and inverse distribution function F_X^{-1}. The Leimkuhler curve K_X corresponding to X is defined by

$$K_X(p) = \frac{1}{\mu_X}\int_{1-p}^1 F_X^{-1}(y)\mathrm{d}y, \quad 0 \le p \le 1. \tag{2}$$

This definition contains the cases of discrete, continuous and mixed random variables. In formula (2)F_X^{-1} is piecewise continuous and the integrals can be assumed to be ordinary Riemann integrals.

From definition (2) we can show that a Leimkuhler curve will be a continuous, non-decreasing concave function that is differentiable almost everywhere in [0, 1] with $K_X(0) = 0$ and $K_X(1) = 1$.

The Leimkhuler curve determines the distribution of X up to a scale factor transformation. This is true since $F_X^{-1}(x) = \mu_X K'_X(1-x)$, almost everywhere and F_X^{-1} will determine F_X.

There is a simple relation between the Leimkuhler curves of the random variables X and $Y = X + \lambda$, where $\lambda \ge 0$.

Proposition 1. *Let X be random variable in $\mathcal{L}$ with mathematical expectation μ_X and Leimkuhler curve $K_X(p)$. Then, the Leimkuhler curve of the random variable $Y = X + \lambda$ is,*

$$K_Y(p) = \frac{\lambda p + \mu_X K_X(p)}{\lambda + \mu_X}, \tag{3}$$

which is a convex linear combination of the Leimkuhler curves p and $K_X(p)$.

Proof. Let $F_X(\cdot)$ be the cdf of X. Then, $F_Y(y) = F_X(y-\lambda)$, $F_Y^{-1}(y) = \lambda + F_X^{-1}(y)$ and $E(Y) = \lambda + \mu_X$. Now, using (2)

$$K_Y(p) = \frac{1}{\mu_Y}\int_{1-p}^1 F_Y^{-1}(y)\mathrm{d}y = \frac{\lambda p + \int_{1-p}^1 F_X^{-1}(y)\mathrm{d}y}{\lambda + \mu_X},$$

which is (3). □

Finally, we include the relation between the Lorenz and Leimkuhler curves. If $L_X(p)$ is a Lorenz curve, it is direct to show that,

$$K_X(p) = 1 - L_X(1-p), \quad 0 \le p \le 1.$$

3. Equivalence with the classical definition

The usual definition of Leimkuhler curve (see for example Burrell, 1991, 1992) is based on two equations. First one solve for x,

$$p = 1 - F_X(x) = \int_x^\infty f_X(y)\mathrm{d}y, \tag{4}$$

and then considers

$$K_X(p) = \psi_X(x) = \frac{1}{\mu_X}\int_x^\infty y f_X(y)\mathrm{d}y. \tag{5}$$