Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 138 (2008) 2544–2551
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Characterizations based on Rényi entropy of order statistics and record values

S. Baratpour, J. Ahmadi*[1], N.R. Arghami[1]

Department of Statistics, School of Mathematical Sciences, Ferdowsi University of Mashhad, P.O. Box 91775-1159, Mashhad, Iran

Received 24 November 2005; received in revised form 20 October 2007; accepted 30 October 2007
Available online 1 February 2008

Abstract

Two different distributions may have equal Rényi entropy; thus a distribution cannot be identified by its Rényi entropy. In this paper, we explore properties of the Rényi entropy of order statistics. Several characterizations are established based on the Rényi entropy of order statistics and record values. These include characterizations of a distribution on the basis of the differences between Rényi entropies of sequences of order statistics and the parent distribution.

MSC: primary 62G30; secondary 62E10; 62B10; 94A17

Keywords: Hazard rate function; Reversed hazard rate function; Minimal repair; Series (parallel) system; Laguerre polynomial; Minkowski inequality

1. Introduction

Suppose that $X_1, \ldots, X_n$ are independent and identically distributed (iid) observations from an absolutely continuous cumulative distribution function (cdf) $F(x)$ and probability density function (pdf) $f(x)$. The order statistics of the sample are defined by the arrangement of $X_1, \ldots, X_n$ from the smallest to the largest, denoted as $X_{1:n} \leqslant X_{2:n} \leqslant \cdots \leqslant X_{n:n}$. These statistics have been used in a wide range of problems, including robust statistical estimation, detection of outliers, characterization of probability distributions and goodness-of-fit tests, entropy estimation, analysis of censored samples, reliability analysis, quality control and strength of materials; for more details, see Arnold et al. (1992), David and Nagaraja (2003) and references therein.

Let $X_1, X_2, \ldots$ be a sequence of iid random variables having an absolutely continuous cdf $F(x)$ and pdf $f(x)$. An observation X_j is called an upper record value if its value exceeds that of all previous observations. Thus, X_j is an upper record if $X_j > X_i$ for every $i < j$. Record data arise in a wide variety of practical situations. Examples include industrial stress testing, meteorological analysis, hydrology, seismology, sporting and athletic events, and oil and mining surveys. Properties of record data have been studied extensively in the literature. Interested readers may refer to the books by Arnold et al. (1998) and Nevzorov (2001).

* Corresponding author.
E-mail addresses: baratpur@math.um.ac.ir (S. Baratpour), ahmadi-j@ferdowsi.um.ac.ir (J. Ahmadi), arghami@math.um.ac.ir (N.R. Arghami).
[1] Member of Ordered and Spatial Data Center of Excellence of Ferdowsi University of Mashhad.

0378-3758/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2007.10.024

In reliability theory, order statistics and record values are used for statistical modeling. The $(n-m+1)$th order statistics in a sample of size n represents the life length of an m-out of-n system. Record values are used in shock models and *minimal repair systems* (see Kamps, 1994). Several authors have studied the subject of characterization of F based on the properties of order statistics and record values. The papers of Huang (1975), Nagaraja (1988), Nagaraja and Nevzorov (1997), Stepanov (1994), Balakrishnan and Balasubramanian (1995), Balakrishnan and Stepanov (2004), Abu-Youssef (2003), Park and Zheng (2004), Hofmann et al. (2005) and Raqab and Awad (2000) contain characterizations based on order statistics and record values.

The Shannon entropy of a random variable X is a mathematical measure of information which measures the average reduction of uncertainty of X. The Rényi entropy is a generalization of Shannon entropy and is known to be of importance in cryptography (see Cachin, 1997), resolution in time–frequency (see Knockaert, 2000). Since, for a given α, two different distributions may have the same Rényi entropy, a distribution cannot be determined by its Rényi entropy. We study conditions under which the Rényi entropy of order statistics and record values can uniquely determine the parent distribution F.

The rest of this paper is organized as follows. Section 2 contains some preliminaries. In Section 3, we present some characterizations based on the Rényi entropy of a sequence of order statistics; also we characterize the exponential model based on the difference between Rényi entropy of the first order statistic and the parent distribution. In Section 4, we show that F can be uniquely determined by the equality of Rényi entropy of record values.

2. Preliminaries

The entropy of order α or Rényi entropy of a distribution (Rényi, 1961) is defined as

$$\begin{aligned} H_\alpha(X) &= \frac{1}{1-\alpha}\log\int_{-\infty}^{+\infty} f^\alpha(x)\,\mathrm{d}x \\ &= \frac{1}{1-\alpha}\log E[f(X)]^{\alpha-1} \\ &= \frac{1}{1-\alpha}\log E_{f_X,\alpha}[r_X^{\alpha-1}(X)] - \frac{\log\alpha}{1-\alpha}, \end{aligned} \tag{1}$$

where $\alpha>0$, $\alpha\neq 1$, and $r_X(t)=f(t)/\bar{F}(t)$, $t>0$, is the hazard rate function of X, $\bar{F}(t)=1-F(t)$, and $E_{f_X,\alpha}$ denotes the expectation with respect to the density function

$$f_{X,\alpha}(x) = -\frac{\mathrm{d}\bar{F}^\alpha(x)}{\mathrm{d}x} = \alpha\bar{F}^{\alpha-1}(x)f(x), \quad \alpha>0.$$

It can be easily shown that $\lim_{\alpha\to1} H_\alpha(X)=H(X)$, where

$$H(X) = -\int_{-\infty}^{+\infty} f(x)\log f(x)\,\mathrm{d}x$$

is commonly referred to as the entropy or *Shannon information measure* of X. The properties and virtues of $H(X)$ have been thoroughly investigated by Shannon (1948). A relatively recent reference for Shannon entropy is Cover and Thomas (1991).

Let θ be the parameter of one of the following families:

(i) location $F_\theta(x)=F_0(x-\theta)$, θ real;
(ii) scale $F_\theta(x)=F_0(\theta x)$, $\theta>0$.

Then in the location case the Rényi entropy is free of θ and for the case of scale family, it is a function of $-\log\theta$. This is also confirmed by Table 1 borrowed from Song (2001) which contains $H_\alpha(X)$ for some common distributions, e.g. exponential, Pareto, normal, Weibull and beta distributions. In that table $B(.,.)$ is the complete beta function. Nadarajah and Zografos (2003) also derived analytical formulas for Rényi entropy for 26 flexible families of univariate continuous distributions.