Kullback-Leibler Information in View of an Extended Version of $k$-Records

Ahmadi Mosayeb$^a$, Mohtashami Borzadaran G.R.$^1$.$^a$

$^a$Department of Statistics, OSDCE, Ferdowsi University of Mashhad

Abstract

This paper introduces an extended version of $k$-records. Kullback-Leibler (K-L) information between two generalized distributions arising from $k$-records is derived; subsequently, it is shown that K-L information does not depend on the baseline distribution. The behavior of K-L information for order statistics and $k$-records, is studied. The exact expressions for K-L information between distributions of order statistics and upper (lower) $k$-records are obtained and some special cases are provided.

Keywords: Order statistics, $k$-records, Kullback-Leibler information.

1. Introduction

Statistical analysis is widely used in science to achieve a better insight to the studies of phenomena. Many observational studies lack a well defined sampling frame for the selection of sampling units, particularly in observational studies of humans, wildlife, insects, fisheries and plants. When an investigator registers an observation by nature according to a certain stochastic model, the registered observation will not have the original distribution unless every observation is given an equal chance of being registered. Weighted distribution theory provides a unifying approach for correction of biases that exist in unequally weighted sample data. The concept of weighted distribution, introduced by Fisher (1934) and Rao (1965), has been discussed in many situations. Particular practical views in a wide variety of probability and statistics areas have an important role in statistical inference. Many phenomena in nature can be expressed via the study of the relations between weighted distributions and the parent distributions in various aspects of statistical analysis. We refer the reader to Patil and Rao (1977, 1978) and Rao (1985) for a survey of statistical applications due to weighted distributions (especially data relating to human population ecology). From Gupta and Keating (1986) to Bartoszewicz (2009) and their references, the idea of the weighted distributions aspects and special cases developed into various applications and properties. It is therefore important to study the structure of the distributions due to weighted random variables with respect to the original distribution. Among weights, proportional hazard and proportional reversed hazard models, order statistics, record values and $k$-records are those concentrated in ordered data analysis. The obvious question is, “Can we obtain distributions arising extension of the above weights?” We are interested in concentrating of ordered data such as versions of order statistics and record values.

The origin of order statistics as a case of the ordered data back to Wilks (1948) with many different applications in the area of statistics and its related criteria. After 1948, many books on various aspects of the order statistics have been written such as the book by David and Nagaraja (2003) and Arnold

$^1$Corresponding author: Department of Statistics, Ordered and Spatial Data Center of Excellence, Ferdowsi University of Mashhad, Mashhad, Iran. E-mail: grmohtashami@um.ac.ir
Ahmadi Mosayeb, Mohtashami Borzadaran G.R. (2008) which are popular books. Many ordered data such as records and $k$-records have also progressed and some of their references can be found in Arnold et al. (1998); subsequently, numerous papers about this subject have been published. The study of record values is in many ways parallel to order statistics but in a slightly different meaning. Wong and Chen (1990) examined the entropy of ordered sequence of random variables and order statistics. Applying the idea of the K-L information as an extended version of entropy to ordered data cases are focused on here.

Information theory today is considered a very fundamental field that can connect to other areas such as thermodynamics, communication theory, probability theory, statistics and computer science. After Shannon (1948), Kullback and Leibler (1951) defined and studied the properties of measures of information which is now referred to as K-L information measure as an extension of the Shannon entropy. K-L information expresses the distance between any arbitrary probability distribution and a true distribution (like data, observations or a precise distribution). This indicates that it can be used to determine “how far away” a probability distribution is from another distribution. Note that it is not the same as a distance in calculus and we cannot call it a metric on the space of probability distribution. In Bayesian statistics, the K-L information can be used as a measure of information gain in moving from prior to posterior distribution. For more details on information theory see Cover and Thomas (2006). It becomes meaningful when the approximation of the value of K-L information measure is near zero. The K-L information of two distributions near zero means that the two distributions are approximately the same.

This paper presents K-L information for an extended version of $k$-records data and its properties for some special cases. The cases of records introduced by MirMostafee and Ahmadi (2012) and families of distributions arising from order statistics obtained by Jones (2004) are special cases for our model. We have obtained results in view of the K-L information for two models of this generalized version and also achieved various special cases and some of their numerical studies.

The link between order statistics and record data has attracted the attention of many researchers such as Gupta and Ahsanullah (2004), Nevzorov (2000), Gupta (1984), Balakrishnan et al. (2009) and Ahmadi and Balakrishnan (2012). As we know, based on Shaked and Shanthikumar (1994, p.496) and Ahmadi and Arghami (2001), one application of records arises in studying the lifetime component with a minimal repairs process; however, the $k^{th}$ order statistic can be represented as the lifetime of $(n-k+1)$-out-of-$n$ systems. Therefore, finding the information of the lifetime of $k$-out-of-$n$ systems can be useful for the prediction of next minimal repair. For more details see, Ahmadi et al. (2010). It is necessary to know when we can say the distribution of order statistics and record values are almost identical; subsequently, we ask “Can we approximate records distribution with order statistics distribution and vice versa?” To answer this question, we have applied for two generalized versions of $k$-records and concentrate on the analysis of the K-L information between records and order statistics distributions which our achievements here are valuable. The results obtained via the state that “the K-L information between two distributions have amounts of near zero”. A numerical study for special cases and their importance and motivations (the other direction of this paper) are also achieved.

2. Preliminaries

As mentioned in Section 1, we are interested in finding the K-L information between extended versions of ordered data. With this in mind, first, we recall some well-known basic definitions and preliminaries.