Journal of Statistical Planning and Inference 169 (2016) 34–42

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

On normal variance–mean mixtures as limit laws for statistics with random sample sizes

V.Yu. Korolev [a,b], A.I. Zeifman [b,c,d,*]

[a] Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Russian Federation
[b] Institute of Informatics Problems, FRC CSC RAS, Russian Federation
[c] Vologda State University, Russian Federation
[d] Institute of Socio-Economic Development of Territories, RAS, Russian Federation

ARTICLE INFO

Article history:
Received 12 March 2014
Received in revised form 23 July 2015
Accepted 27 July 2015
Available online 1 August 2015

MSC:
60F05
60G50
62E20

Keywords:
Random sequence
Random index
Transfer theorem
Samples with random sizes
Normal variance–mean mixture

ABSTRACT

We prove a general transfer theorem for multivariate random sequences with independent random indexes in the double array limit setting. We also prove its partial inverse providing necessary and sufficient conditions for the convergence of randomly indexed random sequences. Special attention is paid to the case where the elements of the basic double array are formed as statistics constructed from samples with random sizes. Under rather natural conditions we prove the theorem on convergence of the distributions of such statistics to multivariate normal variance–mean mixtures and, in particular, to multivariate generalized hyperbolic laws.

1. Introduction

In classical problems of mathematical statistics, the size of the available sample, i.e., the number of available observations, is traditionally assumed to be deterministic. In the asymptotic settings it plays the role of infinitely increasing *known* parameter. At the same time, in practice very often the data to be analyzed is collected or registered during a certain period of time and the flow of informative events each of which brings a next observation forms a random point process. Therefore, the number of available observations is unknown till the end of the process of their registration and also must be treated as a (random) observation. For example, this is so in insurance statistics where during different accounting periods different numbers of insurance events (insurance claims and/or insurance contracts) occur and in high-frequency financial statistics where the number of events in a limit order book during a time unit essentially depends on the intensity of order flows. Moreover, contemporary statistical procedures of insurance and financial mathematics do take this circumstance into consideration as one of possible ways of dealing with heavy tails. However, in other fields such as medical statistics or quality control this approach has not become conventional yet although the number of patients with a certain disease varies from month to month due to seasonal factors or from year to year due to some epidemic reasons and the number of failed

* Corresponding author at: Vologda State University, Russian Federation.
E-mail addresses: victoryukorolev@yandex.ru (V.Yu. Korolev), a_zeifman@mail.ru (A.I. Zeifman).

http://dx.doi.org/10.1016/j.jspi.2015.07.007

items varies from lot to lot. In these cases the number of available observations as well as the observations themselves are unknown beforehand and should be treated as random to avoid underestimation of risks or error probabilities.

Therefore it is quite reasonable to study the asymptotic behavior of general statistics constructed from samples with random sizes for the purpose of construction of suitable and reasonable asymptotic approximations. As this is so, to obtain non-trivial asymptotic distributions in limit theorems of probability theory and mathematical statistics, an appropriate centering and normalization of random variables and vectors under consideration must be used. It should be especially noted that to obtain reasonable approximation to the distribution of the basic statistics, both centering and normalizing values should be non-random. Otherwise the approximate distribution becomes random itself and, for example, the problem of evaluation of quantiles or significance levels becomes senseless.

In asymptotic settings, statistics constructed from samples with random sizes are special cases of random sequences with random indices. The randomness of indices usually leads to that the limit distributions for the corresponding random sequences are heavy-tailed even in the situations where the distributions of non-randomly indexed random sequences are asymptotically normal see, e.g., Bening and Korolev (2002, 2005), Gnedenko and Korolev (1996). For example, if a statistic which is asymptotically normal in the traditional sense is constructed on the basis of a sample with random size having negative binomial distribution, then instead of the expected normal law, the Student distribution appears as an asymptotic law for this statistic.

The literature on random sequences with random indexes is extensive, see, e.g., the references above and references therein.

Although the mathematical theory of random sequences with random indexes is well-developed, there still remain some unsolved problems. For example, convenient conditions for the convergence of the distributions of general statistics constructed from samples with random sizes to normal variance–mean mixtures have not been found yet. The desire to fill this theoretical gap is quite natural. Another motivation for this research is practical and is as follows. In applied probability there is a convention, apparently going back to the book Gnedenko and Kolmogorov (1954), according to which a model distribution is reasonable and/or justified enough only if it is an *asymptotic approximation*, that is, there exist a more or less simple limit setting and the corresponding limit theorem in which the model under consideration is a limit distribution. The existence of such a setting may bring a deeper insight into the phenomena under consideration than just fitting a more or less convenient model.

General normal variance–mean mixtures are examples of such convenient models widely used to describe observed statistical regularities in many fields. In particular, in 1977–78 O. Barndorff-Nielsen (Barndorff-Nielsen, 1977, 1978) introduced the class of *generalized hyperbolic distributions* as a class of special univariate variance–mean mixtures of normal laws in which the mixing is carried out in one parameter since location and scale parameters of the mixed normal distribution are directly linked. The range of applications of generalized hyperbolic distributions varies from the theory of turbulence or particle size description to financial mathematics, see Barndorff-Nielsen et al. (1982). Multivariate generalized hyperbolic distributions were introduced in the seminal paper (Barndorff-Nielsen, 1977) mentioned above as a natural generalization of the univariate case. They were further investigated in Blæsild (1981) and Blæsild and Jensen (1981). It is a convention to explain such a good adequacy of generalized hyperbolic models by that they possess many parameters to be suitably adjusted. But actually, it would be considerably more reasonable to explain this phenomenon by limit theorems yielding the possibility of the use of generalized hyperbolic distributions as convenient *asymptotic* approximations.

The main results presented in this paper deal with the description of conditions which provide the convergence of the distributions of statistics constructed from samples with random sizes to multivariate normal variance–mean mixtures, in particular, to multivariate generalized hyperbolic laws. The conditions presented below are formulated in terms of the asymptotic behavior of random sample sizes and have the 'if and only if' form. This circumstance is proved to be very promising and constructive. For example, in Korolev et al. (2015) a problem of construction of suitable approximations to the distribution of the so-called order flow imbalance process in high-frequency trading systems was considered. It was empirically shown that generalized hyperbolic distributions are very likely models for that. But these distributions are variance–mean mixtures with *one* mixing parameter. By means of one-dimensional limit theorems for random sums, this fact directly leads to theoretical understanding that the intensities of buy and sell orders actually must be proportional to *one and the same* random process reflecting general market agitation. This theoretical inference concerning the flows intensities then found its statistical proof, see Korolev et al. (2015). In other words, the 'if and only if' character of the presented conditions makes testing goodness-of-fit of financial data with generalized hyperbolic models *equivalent* to testing goodness-of-fit of the corresponding flow intensities (i.e., volatilities) with generalized inverse Gaussian models, which is much simpler.

In the present paper both structural and multivariate generalizations of some results of Korolev et al. (2015) to general statistics are presented. The paper is organized as follows. Basic notation is introduced in Section 2. Here an auxiliary result on the asymptotic rapprochement of the distributions of randomly indexed random sequences with special scale–location mixtures is proved. In Section 3 of the present paper we prove a general transfer theorem for random sequences with independent random indexes in the double array limit setting. We also prove its partial inverse providing necessary and sufficient conditions for the convergence of randomly indexed random indexes. Following the lines of Korolev (1993), we first formulate a general result improving some results of Korolev (1993) and Bening and Korolev (2002) by removing some superfluous assumptions and relaxing some conditions. Special attention is paid to the case where the elements of the basic double array are formed as statistics constructed from samples with random sizes. This case is considered in Section 4 where