Statistics and Probability Letters 80 (2010) 1015–1021

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

On an asymptotic distribution of dependent random variables on a 3-dimensional lattice[☆]

Danielle J. Harvey[a,*], Qian Weng[b], Laurel A. Beckett[a]

[a] Division of Biostatistics, Department of Public Health Sciences, School of Medicine, University of California, Davis, CA 95616, USA

[b] Department of Statistics, University of California, Davis, CA 95616, USA

ARTICLE INFO

Article history:
Received 30 June 2009
Received in revised form 18 February 2010
Accepted 22 February 2010
Available online 6 March 2010

Keywords:
Central limit theorem
Spatial data
Dependent data

ABSTRACT

We define conditions under which sums of dependent spatial data will be approximately normally distributed. A theorem on the asymptotic distribution of a sum of dependent random variables defined on a 3-dimensional lattice is presented. Examples are also presented.

1. Introduction

Imaging techniques provide non-invasive tools to track the development and progression of chronic pathologic processes. In early phase clinical trials, imaging may be used on animals to evaluate the usefulness of potential new therapeutic drugs. For example, the efficacy of a new cancer drug may be assessed by measuring the reduction in tumor size as seen on an image, negating the need to sacrifice the animals or at least extending the time until the sacrifice is conducted and allowing for multiple assessments on the same animal. Human imaging studies are also useful in the study of cancer as well as neurologic disorders such as dementia and schizophrenia where abnormalities of tissue are known to occur. Imaging enables the study of these abnormalities without the need of surgery, and is particularly useful in the context of neurodegenerative diseases where slicing into the brain is an unlikely strategy. Therefore, imaging has become a critical tool for the study of these chronic processes and the evaluation of new treatments for those conditions. However, imaging is incredibly expensive, so studies tend to be of small or moderate size.

Each image, itself, is an extremely rich data source. The images are broken down into hundreds of thousands if not millions of volume elements, or voxels, each of which contains information about the region or tissue being studied. These voxels may be thought of as data points on a 3-dimensional lattice. Due to the underlying anatomy and biology of the disease process, these data are likely to be highly correlated locally, with the correlation decreasing as the Euclidean distance between points increases. There is, therefore, a need to perform sensible data reduction in such a way that under reasonable conditions, the reduced summary measures will yield approximately normally distributed measures for use in statistical analyses. Demonstration of the approximate normality of summary data from high-dimensional imaging would support the use of standard linear model techniques for analysis even with small numbers of patients or animals, for example, use of two-sample t-tests to compare two treatments in small pre-clinical experiments.

☆ This work was supported by NIH/NIA grant # 5 P30 AG010129.

* Corresponding author. Tel.: +1 530 752 8036; fax: +1 530 752 3239.
E-mail addresses: djharvey@ucdavis.edu (D.J. Harvey), wqhb@yahoo.com (Q. Weng), labeckett@ucdavis.edu (L.A. Beckett).

0167-7152/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2010.02.016

Asymptotic distributions of sums of dependent data in one dimension have been studied extensively in the literature, using various dependence schemes. Many researchers have proved asymptotic normality for m-dependent random variables, in which random variables m units apart in sequence are assumed to be independent. Hoeffding and Robbins (1948) assumed a fixed value for m while Berk (1973) and Romano and Wolf (2000) allowed for m to change with the sample size and to grow infinitely large. Serfling (1968) examined asymptotic properties of sums of random variables under less stringent dependence structures.

Sajjan (2000) developed a central limit theorem for a stationary dependent process on the 2-dimensional lattice. For the lattice case, the idea of m-dependence is extended to (m_1, m_2)-dependence. Christofides and Mavrikiou (2003) also focused on the two-dimensional case, but relaxed the assumption of stationarity. In doing so, they made strong assumptions that were difficult to meet in practice. They also defined a dependence structure for the lattice, called ρ-radius dependence, that will be discussed below.

Use of linear combinations of data located on a lattice offers promise for the study of spatially distributed pathological processes, which may be observed through structural magnetic resonance images (MRI) or positron emission tomography (PET) scans as well as other imaging modalities. We present notation followed by a theorem describing conditions necessary for asymptotic normality of the sum of variables located on a 3-dimensional lattice. We provide the proof of the theorem as well as several examples for which the conditions of the theorem hold or where the theorem might be useful.

2. Theoretical considerations

Christofides and Mavrikiou (2003) defined the following concept of local dependence appropriate for the issue considered here:

Definition. For a positive integer r let $\mathbf{N}^r$ denote the r-dimensional positive integer lattice and let $\{X_{\mathbf{i}}, \mathbf{i} \in \mathbf{N}^r\}$ be an array of random variables defined on a common probability space $(\Omega, \mathcal{A}, P)$. Let $\rho \geq 0$. The random variables $\{X_{\mathbf{i}}, \mathbf{i} \in \mathbf{N}^r\}$ are said to be *ρ-radius dependent* if $X_{\mathbf{i_1}}$ and $X_{\mathbf{i_2}}$ are independent whenever $d(\mathbf{i_1}, \mathbf{i_2}) > \rho$, where $d(\mathbf{i_1}, \mathbf{i_2})$ is the Euclidean distance between $\mathbf{i}_1$ and $\mathbf{i}_2$.

Let $\{X_{\mathbf{i}}, \mathbf{i} = (i_1, i_2, i_3) \leq (n_1, n_2, n_3)\}$ be an array of ρ-radius dependent three-dimensionally indexed random variables. $X_{\mathbf{i}}$ represent either individual values or weighted values, and we are interested in asymptotic properties of their sum. n_1 defines the vertical dimension (back to front), n_2 defines the horizontal dimension (left to right) and n_3 defines the spatial dimension (bottom to top).

Let $\rho^* = \lceil \rho \rceil$, the smallest integer greater than or equal to ρ.

Let ν_n be a positive integer greater than ρ^*, which is allowed to change with $n = (n_1, n_2, n_3)$.

Let $T_{i_1,i_2,i_3} = \{(j_1, j_2, j_3) : i_k - \nu_n \leq j_k \leq i_k + \nu_n, k = 1, 2, 3\}$ so that T_{i_1,i_2,i_3} is the $(2\nu_n + 1) \times (2\nu_n + 1) \times (2\nu_n + 1)$ cube whose center is the point (i_1, i_2, i_3).

Let

$$S^{(T)}_{i_1,i_2,i_3} = \sum_{(j_1,j_2,j_3) \in T_{i_1,i_2,i_3}} X_{j_1,j_2,j_3},$$

the total (or weighted total) in the cube.

The lattice may be divided into a set of independent cubes and borders separating the cubes, beginning in the back left corner of the bottom layer. The first cube is centered at (k_n, k_n, k_n) where $k_n = \nu_n + 1$. Cube centers will be regularly spaced at intervals of $\lambda_n = 2\nu_n + \rho^* + 1$, where λ_n represents the width of the cube and the adjacent border. Because the variables X_{i_1,i_2,i_3} are ρ-radius dependent and the cubes are separated by a border of width $\rho^* > \rho$, the cube sums are independent. We assume that $n_i + \rho^*$ may be evenly divided by λ_n, so that there are no partial cubes in the ith dimension ($i = 1$: back to front, $i = 2$: left to right, $i = 3$: bottom to top). The cubes are centered at $(i_1, i_2, i_3) = (k_n + j_1\lambda_n, k_n + j_2\lambda_n, k_n + j_3\lambda_n)$, where $0 \leq j_i \leq D_i - 1$ and D_i is the number of cubes in each dimension.

The regions of points that do not belong to any of the cubes may be divided into seven zones for each cube (Fig. 1). Three regions are the "flat" rectangular areas adjacent to the right, front, and top of a cube, but not symbolically extending past the edges of the cube. Next we identify three strips adjacent to the edges but not extending past the corners of the cube. Finally, the remaining region is the small cube adjacent to the top right front corner.

For each cube, T_k where $k = (i_1, i_2, i_3)$, define $S^{(B)}_k$ to be the sum of the points in the seven boundary regions surrounding it.

The proof of the theorem relies on taking limits in such a fashion that the normalized sum of the sums of the random variables in the cubes is asymptotically normal, by standard central limit theorem arguments, while the normalized sum of the random variables in the boundary regions goes to zero in probability.

2.1. Theorem

The main theorem for the asymptotic distribution of the sum of random variables located on a spatial lattice is given below.

Theorem 1. *Let $\{X_{i_1,i_2,i_3}, (i_1, i_2, i_3) \leq (n_1, n_2, n_3)\}$ be an array of ρ-radius dependent three dimensionally indexed random variables. Without loss of generality, assume that these variables have mean equal to zero. Let ν_n be a positive integer greater*