Asymptotics for Statistical Distances Based on Voronoi Tessellations

R. Jiménez1 and J. E. Yukich2

Received November 2, 2000; revised October 19, 2001

We obtain an information-type inequality and a strong law for a wide class of statistical distances between empirical estimates and random measures based on Voronoi tessellations. This extends some basic results in the asymptotic theory of sample spacings, when the cells of the Voronoi tessellation are interpreted as \(d\)-dimensional spacings.

KEY WORDS: Voronoi tessellations; statistical distances; multi-dimensional spacings; entropy; \(\phi\)-divergence.

1. INTRODUCTION

Voronoi tessellations generated by random sets of points are of general interest and have been used in many diverse fields. We refer to the survey of Aurenhammer (1991) and the texts of Okabe et al. (1992), Möller (1994), and Stoyan et al. (1995) for a complete treatment. In this paper such tessellations are used as an adaptive scheme to compare probability densities on \(\mathbb{R}^d\), \(d \geq 1\).

We study the a.s. behavior of a wide class of statistical distances between empirical estimates and random measures based on Voronoi tessellations. We obtain an information-type inequality and a strong law which together provide a general asymptotic method for characterizing multivariate densities. When \(d = 1\) the results describe the asymptotics of sample spacings. Thus we extend one-dimensional methods based on spacings to multi-dimensional methods based on Voronoi tessellations.
Given a set of points \( \mathcal{P} := \{x_1, \ldots, x_n \} \subset \mathbb{R}^d \) and a Borel subset \( B \) of \( \mathbb{R}^d \), consider for any \( x_i \in \mathcal{P} \cap B \) the locus of points closer to \( x_i \) than to any other point of \( \mathcal{P} \cap B \). The intersection of this set of points with \( B \) is denoted by \( C_i(B) := C_i(B; x_1, \ldots, x_n) \), that is

\[
C_i(B; x_1, \ldots, x_n) := \{ y \in B : \|y - x_i\| \leq \|y - x_j\|, \forall x_j \in \mathcal{P} \cap B \} \tag{1.1}
\]

where \( \| \cdot \| \) denotes the Euclidean distance. If \( x_i \notin B \) then we define \( C_i(B) = \emptyset \). Thus, \( \{ C_i(B), 1 \leq i \leq n \} \) is a partition of \( B \) which is called the Voronoi tessellation of \( B \) generated by \( \mathcal{P} \) and is denoted by \( \mathcal{V}(B; \mathcal{P}) \). It is understood that if \( \mathcal{P} \cap B = \emptyset \) then \( \mathcal{V}(B; \mathcal{P}) = B \).

An attractive feature of Voronoi tessellations which greatly facilitates our study is a monotonicity property analogous to that found in one dimensional spacings: the cell around a given point decreases in size as the number of points increases. In other words, for all \( 1 \leq i \leq n \) we have

\[
C_i(B; x_1, \ldots, x_n) \subset C_i(B; x_1, \ldots, x_n, y_1, \ldots, y_k) \text{ for any points } x_1, \ldots, x_n, y_1, \ldots, y_k.
\]

Throughout \( X_1, X_2, \ldots \) are independent random variables in \( \mathbb{R}^d \) with common probability density \( f \), \( g \) is an arbitrary nonnegative integrable function on \( B \), and \( \lambda \) denotes Lebesgue measure on \( \mathbb{R}^d \).

**Definition 1.1.** For each \( n \geq 1 \), we define for \( 1 \leq i \leq n \) the Voronoi sample spacings

\[
D_{i,n}(B) := D_i(B; X_1, \ldots, X_n) := D_{X_i}(B; X_1, \ldots, X_n) := \lambda(C_i(B; X_1, \ldots, X_n))
\tag{1.2}
\]

and the transformed Voronoi spacings

\[
D'_{i,n}(B) := D'_i(B; X_1, \ldots, X_n) := \int_{C_i(B; X_1, \ldots, X_n)} g(x) \lambda(dx)
\tag{1.3}
\]

For all \( 1 \leq i \leq n \) we have \( D'_{i,n}(B; x_1, \ldots, x_n) \leq D'_{i,n}(B; x_1, \ldots, x_n, y_1, \ldots, y_k) \) for any functions \( 0 \leq g_1 \leq g_2 \). We will measure the discrepancy on \( B \) between \( g \) and the sample density \( f \) by comparing the transformed spacings \( \{ D'_{i,n}(B), 1 \leq i \leq n \} \) with \( \{ D'_{i,n}(B), 1 \leq i \leq n \} \).

Given a strictly convex function \( \phi : \mathbb{R}^+ \to \mathbb{R} \), Csiszár’s (1978) \( \phi \)-divergence between two nonnegative \( n \)-dimensional vectors \( p := (p_1, \ldots, p_n) \) and \( q := (q_1, \ldots, q_n) \) is

\[
I_\phi(p, q) := \sum_{i=1}^n q_i \phi \left( \frac{p_i}{q_i} \right)
\]