Title

The maximum spacing estimation for multivariate observations

Journal

Journal of Statistical Planning and Inference, 129(1-2)

ISSN

0378-3758

Authors

Ranneby, Bo
Jammalamadaka, S Rao
Teterukovskiy, Alex

Publication Date

2005-02-01

DOI

10.1016/j.jspi.2004.06.059

Peer reviewed

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Statistical Planning and Inference 129 (2005) 427–446
journal of statistical planning and inference
www.elsevier.com/locate/jspi

The maximum spacing estimation for multivariate observations

Bo Ranneby[a,*], S. Rao Jammalamadaka[b], Alex Teterukovskiy[a]

[a] *Centre of Biostochastics, The Swedish University of Agricultural Sciences, Umeas 90183, Sweden*
[b] *Department of Statistics and Applied Probability, University of California at Santa Barbara, USA*

Available online 25 August 2004

Abstract

For independently and identically distributed (i.i.d.) univariate observations a new estimation method, the maximum spacing (MSP) method, was defined in Ranneby (Scand. J. Statist. 11 (1984) 93) and independently by Cheng and Amin (J. Roy. Statist. Soc. B 45 (1983) 394). The idea behind the method, as described by Ranneby (Scand. J. Statist. 11 (1984) 93), is to approximate the Kullback–Leibler information so each contribution is bounded from above. In the present paper the MSP-method is extended to multivariate observations. Since we do not have any natural order relation in $\mathscr{R}^d$ when $d > 1$ the approach has to be modified. Essentially, there are two different approaches, the geometric or probabilistic counterpart to the univariate case. If we to each observation attach its Dirichlet cell, the geometrical correspondence is obtained. The probabilistic counterpart would be to use the nearest neighbor balls. This, as the random variable, giving the probability for the nearest neighbor ball, is distributed as the minimum of $(n-1)$ i.i.d. uniformly distributed variables on the interval (0, 1), regardless of the dimension d. Both approaches are discussed in the present paper.

Keywords: Estimation; Spacings; Consistency; Multivariate observations

1. Introduction

For independently and identically distributed (i.i.d.) univariate observations, a new estimation method called the "Maximum Spacings (MSP)" method, was developed in Ranneby (1984) and independently by Cheng and Amin (1983). The idea behind the method,

* Corresponding author. Tel.: +46-90-786-8236; fax: +46-90-778-8164.
E-mail address: bo.ranneby@sekon.slu.se (B. Ranneby).

0378-3758/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2004.06.059

as described by Ranneby (1984), is to approximate the Kullback–Leibler information so each contribution is bounded from above. The estimation method obtained from this approximation is called the maximum spacing method and it works also in situations when the ML-method breaks down. The method is also discussed in Titterington (1985), where he states that "in principle, of course, it would be possible to treat multivariate data by grouping: although definition of the multinomial cells would be more awkward."

In the present paper, our goal is to extend the MSP-method to multivariate observations. Since we do not have any natural order relation in $\mathscr{R}^d$ when $d > 1$, we have to modify the approach. Essentially, there are two different approaches to choose between, namely the geometric and the probabilistic counterparts to the univariate spacings. Let $\xi_1, \xi_2, \ldots, \xi_n$ be a sequence of independent and identically distributed d-dimensional random vectors with true distribution P_0 and define the nearest neighbor distance to the point ξ_i, namely

$$R_n(i) = \min_{j \neq i} |\xi_i - \xi_j|$$

and let

$$B(x, r) = \{y : |x - y| \leqslant r\}$$

denote the ball of radius r with center at x.

The Dirichlet cells $V_n(\xi_i)$ attached to each observation ξ_i may be interpreted as the geometrical correspondence. The Dirichlet cell $V_n(\xi_i)$ surrounding ξ_i consists of all points $x \in \mathscr{R}^d$ which are closer to ξ_i than to any other observation, and the Dirichlet cells split $\mathscr{R}^d$ into n identically distributed random sets. The main advantage with this approach is that the probabilities for the Dirichlet cells always add up to one. The Dirichlet cells are cumbersome to handle, both from a practical and theoretical point of view. The probabilistic counterpart to univariate spacings would be to use the nearest neighbor balls, as the random variable $P_0(B(\xi_i, R_n(i)))$ is distributed as the minimum of $(n-1)$ i.i.d. uniformly distributed variables on the interval $(0, 1)$, regardless of the dimension d. The latter approach is our main focus in this paper, but the geometric approach is also discussed. Goodness of fit tests based on nearest neighbor balls have been considered earlier by Bickel and Breiman (1983) as well as Jammalamadaka and Zhou (1993) but our goal here is estimation and not testing hypotheses.

2. Definitions

In this section the definitions of the two different extensions of the MSP-method to the multivariate case will be given.

2.1. MSP based on NN-balls

Let $\xi_1, \xi_2, \ldots, \xi_n$ be i.i.d. random vectors with an absolutely continuous distribution P_0 with density function $g(x)$ and suppose that we assign a model with density functions $\{f(x, \theta), \theta \in \Theta\}$, where $\Theta \subset R^q$. Define,

$$z_i(n, \theta) = n P_\theta(B(\xi_i, R_n(i))),$$

$$\bar{z}(n,\theta) = \frac{1}{n}\sum_{i=1}^{n} z_i(n,\theta),$$

$I(A) =$ indicator function of the set A.

A natural generalization of the univariate definition of the spacing function is to define it in the multivariate case as

$$\frac{1}{n}\sum_{i=1}^{n} \log z_i(n,\theta).$$

However, there are serious shortcomings of this approach, mainly that under some probability measures the sum of the probabilities of the nearest neighbor balls may be too large (as for instance when P_θ has the same location as P_0 but much smaller variance). As a consequence there is no guarantee that the estimator will be consistent.

To overcome this problem, we will normalize the probabilities for the nearest neighbor balls when the sum of their probabilities exceeds one. When the sum is less than one we let the remaining probability enter the spacing function in the same way as the probabilities for the nearest neighbor balls. This leads us to the following definition of the spacing function $S_n(\theta)$

$$S_n(\theta) = \frac{1}{n}\sum \log(z_i(n,\theta)) + \left(\frac{1}{n}\log(1-\bar{z}(n,\theta))\right) I(\bar{z}(n,\theta) \leqslant 1) - I(\bar{z}(n,\theta) > 1)\log \bar{z}(n,\theta).$$

Remark 1. This means that when the sum of the probabilities exceeds one the spacing function is defined as

$$S_n(\theta) = \frac{1}{n}\sum_i \log\left[\frac{nP_\theta(B(\xi_i, R_n(i)))}{\sum_j P_\theta(B(\xi_j, R_n(j)))}\right].$$

Definition. The parameter value which maximizes $S_n(\theta)$ is called the maximum spacing estimate (MSP-estimate) of θ.

Remark 2. If sup $S_n(\theta)$ is not attained for any θ belonging to the admissible set Θ, we define the MSP-estimate $\hat{\theta}_n$ as any point belonging to the set Θ and satisfying

$$S_n(\hat{\theta}_n) \geqslant \frac{\log c_n}{n} + \sup_{\theta\in\Theta} S_n(\theta).$$

In this expression $0 < c_n < 1$ and $c_n \to 1$ as $n \to \infty$.

For mixtures of continuous distributions it happens that the likelihood function tends to infinity for certain parameter combinations and then the ML-method breaks down.

Example 1. Let $\xi_1, \xi_2, \ldots$ be i.i.d. observations from a mixture of two bivariate normal distributions. The density function $f(x, y, \theta)$ is given by

$$f(x,y,\theta) = ph(x,y,\mu_1,\mu_2,\sigma_1,\sigma_2,\rho_1) + (1-p)h(x,y,\mu_3,\mu_4,\sigma_3,\sigma_4,\rho_2)$$

where h is the density function for a bivariate normal distribution with parameters indicated by the notation. Say that we have observations $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. If we put $\mu_1 = x_1$, $\mu_2 = y_1$ and let σ_1 (or σ_2) go to zero, then the likelihood function tends to infinity. Consequently, the ML-method is not suitable. The maximum spacing estimate obtained by maximizing $S_n(\theta)$, defined above will be consistent.

Remark 3. It may be argued that if the σ_i's are bounded away from zero by some small number the ML method performs well. Theoretically that is true but quite frequently the numerical maximization breaks down, as a consequence of the unboundedness of the likelihood function when the σ_i's are not bounded away from zero.

2.2. MSP based on Dirichlet tesselation

Before we give an alternative definition of the spacing function, consider the following definitions.

Given an open set $\Omega \subset \mathbb{R}^d$, the set $\{V_i\}_{i=1}^n$ is called a *tesselation* of Ω if $V_i \cap V_j = \emptyset$ for $i \neq j$ and $\bigcup_{i=1}^n \bar{V}_i = \bar{\Omega}$. Let $|\cdot|$ denote the Euclidean norm on $\mathbb{R}^d$. Given a set of points $\{\xi_i\}_{i=1}^n$ belonging to $\bar{\Omega}$, the Dirichlet cell $V(\xi_i)$ corresponding to the point ξ_i is defined by

$$V(\xi_i) = \{y \in \Omega : |y - \xi_i| \leqslant |y - \xi_j|, \text{ for } j = 1, \ldots, n, \ j \neq i\}.$$

The probabilities of the Dirichlet cells, of course, always add up to the probability of Ω.

Now we consider the following alternative definition of the spacing function based on the Dirichlet tesselation. Let

$$v_i(n, \theta) = n P_\theta(V(\xi_i)).$$

The spacing function $S_n^*(\theta)$ is defined as follows:

$$S_n^*(\theta) = \frac{1}{n} \sum \log(v_i(n, \theta)).$$

The MSP-estimate of θ is now defined as the maximizer of $S_n^*(\theta)$.

This approach has been used in Ranneby (1996) and it is also discussed in Jimenez and Yukich (2002).

3. Consistency of MSP based on NN-balls

Before the main results will be stated some results of independent interest will be given.

3.1. Preliminaries

Let $\xi_1, \xi_2, \ldots, \xi_n$ be a sequence of independent d-dimensional random vectors. Then, for each fixed i, we can make the transformation $P_0(B(\xi_i, |\xi_i - \xi_j|)), \ j \neq i$. The $(n-1)$ random variables are not only uniformly distributed, but they are also mutually independent. As the following proposition shows, it is also possible to let the random vectors $\xi_1, \xi_2, \ldots, \xi_n$ have different distributions.

Proposition 1. *Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent random variables with the respective distributions $P_j(\cdot)$, $j=1, 2, \ldots, n$ which are absolutely continuous w.r.t. Lebesgue measure. Then for each fixed i it holds that the random variables $P_j(B(\xi_i, |\xi_i - \xi_j|))$, $j \neq i$, $j = 1, 2, \ldots, n$ have the same distribution as the joint distribution of $(n-1)$ independent and uniformly distributed (on the interval (0,1)) random variables.*

Proof.

$$Pr(P_j(B(\xi_i, |\xi_i - \xi_j|)) > \alpha_j,\ j \neq i) = \int Pr(P_j(B(\xi_i, |\xi_i - \xi_j|)) > \alpha_j,\ j \neq i | \xi_i = x)\, \mathrm{d}P_i(x).$$

Given that $\xi_i = x$, the random variables $P_j(B(x, |x - \xi_j|))$ will be independent. Thus we obtain

$$\begin{aligned} &Pr(P_j(B(\xi_i, |\xi_i - \xi_j|)) > \alpha_j,\ j \neq i) \\ &= \int \prod_{j \neq i} Pr(P_j(B(x, |x - \xi_j|)) > \alpha_j) \mathrm{d}P_i(x) \\ &= \int \prod_{j \neq i} Pr(|\xi_j - x| > r(j, x, \alpha_j))\, \mathrm{d}P_i(x), \end{aligned} \tag{1}$$

where $r(j, x, \alpha_j)$ is chosen so that $P_j(|\xi_j - x| \leqslant r(j, x, \alpha_j)) = \alpha_j$.

Such numbers always exist because $P_j(|\xi_j - x| \leqslant \beta)$ is a continuous function of β. The definition of $r(j, x, \alpha_j)$ implies that

$$Pr(|\xi_j - x| > r(j, x, \alpha_j)) = 1 - \alpha_j.$$

By inserting the right side of this expression into (1) we get

$$Pr(P_j(B(\xi_i, |\xi_i - \xi_j|)) > \alpha_j,\ j \neq i) = \prod_{j \neq i}(1 - \alpha_j),$$

which is what was to be proved. □

Because of Proposition 1 the moments of $n \min_{j \neq i} P_j(B(\xi_i, |\xi_i - \xi_j|))$ are easily calculated, see e.g. Reiss (1989, p. 45), giving us the following corollary.

Corollary 1. *Under the assumptions in Proposition* 1 *it holds that*

$$E\left[n \min_{j \neq i} P_j(B(\xi_i, |\xi_i - \xi_j|))\right] \to 1,$$

$$\mathrm{Var}\left[n \min_{j \neq i} P_j(B(\xi_i, |\xi_i - \xi_j|))\right] \to 1,$$

$$E\left[(n \min_{j \neq i} P_j(B(\xi_i, |\xi_i - \xi_j|)))^3\right] \to 6.$$

As a direct consequence of results in Pyke (1965) or Rényi (1953) we have the following lemma.

Lemma 1. *Under the assumptions in Proposition* 1 *it holds that the random variable* $\min_{j\neq i} P_j(B(\xi_i, |\xi_i - \xi_j|))$ *has the same distribution as* $Y_1/\sum_{j=1}^n Y_j$, *where* $Y_1, Y_2, \ldots, Y_n$ *are i.i.d. random variables with an exponential distribution with mean* 1. *Furthermore,* $n \min_{j\neq i} P_j(B(\xi_i, |\xi_i - \xi_j|))$ *converges in distribution to an exponential distribution with mean* 1.

Lemma 2. *Under the assumptions of Proposition* 1 *it holds that*

$$E\left[\log n \min_{j\neq i} P_j(B(\xi_i, |\xi_i - \xi_j|))\right] \to -\gamma,$$

where γ *is Euler's constant and equals* $0.57\cdots$,

$$\mathrm{Var}\left[\log n \min_{j\neq i} P_j(B(\xi_i, |\xi_i - \xi_j|))\right] \to \frac{\pi^2}{6} - 1$$

and

$$E\left[\left|\log n \min_{j\neq i} P_j(B(\xi_i, |\xi_i - \xi_j|))\right|^3\right] \to \int |\log x|^3 \mathrm{e}^{-x}\,\mathrm{d}x < \infty.$$

Proof. It follows from Lemma 1 that $n \min_{j\neq i} P_j(B(\xi_i, |\xi_i - \xi_j|))$ will have the same distribution as $Y_1/\bar{Y}$, where $\bar{Y} \to 1$ almost surely. Thus we will be done if we can show that the sequence $|\log \bar{Y}|^3$ is uniformly integrable. Jensen's inequality gives that $\log \bar{Y} \geqslant \frac{1}{n}\sum_{i=1}^n \log Y_i$. Since we also have $\log \bar{Y} \leqslant \bar{Y}$, we get

$$|\log \bar{Y}| \leqslant \left|\frac{1}{n}\sum_{i=1}^n \log Y_i\right| + |\bar{Y}|.$$

This result shows us that $|\log \bar{Y}|^3$ is uniformly integrable. Since $\bar{Y} \to 1$ a.s. we get that

$$E\left[\log n Y_1 \Big/ \sum Y_j\right] \to E(\log Y_1) = -\gamma,$$

$$\mathrm{Var}\left[\log n Y_1 \Big/ \sum Y_j\right] \to \mathrm{Var}(\log Y_1) = \frac{\pi^2}{6} - 1,$$

and

$$E\left[\left|\log n Y_1 \Big/ \sum Y_j\right|^3\right] \to E(|\log Y_1|^3) < \infty. \qquad \square$$

The calculations of $E(\log Y_1)$ and $\mathrm{Var}(\log Y_1)$ may be found for example in Darling (1953).

In the following we will assume that the random vectors $\xi_1, \xi_2, \ldots, \xi_n$ have the same distribution P_0 with density function $g(x)$. In the rest of the paper, we use the following notation (see also Section 4 of Jammalamadaka and Janson (1986)):

$$\|B(x,r)\| = \text{volume of the ball } B(x,r) = c_d r^d,$$

where $c_d = \pi^{d/2}/\Gamma(d/2+1)$.

Proposition 2. *Let $\xi_1, \xi_2, \ldots, \xi_n$ be i.i.d. random vectors with an absolutely continuous distribution P_0 with density function $g(x)$. Then $(\xi_i, \eta_i(n))$, where*

$$\eta_i(n) = n\|B(\xi_i, R_n(i))\| = nc_d R_n^d(i),$$

converges in distribution to (X,Y) where X has density function $g(x)$ and Y given $X = x$ has an exponential distribution with parameter $g(x)$.

Proof. As the volume of the ball $B(x,r)$ exceeds y/n if and only if none of the variables ξ_j, $j \neq i$, $j \leqslant n$ falls in the ball $B(x, V^{-1}(y/n))$. (Here $V^{-1}(y/n)$ denotes the radius giving the volume y/n, i.e. $r = (y/n)^{1/d} c_d^{-1/d}$). Thus

$$Pr(n\|B(\xi_i, R_n(i)\| > y | \xi_i = x) = (1 - P_0(\xi_1 \in B(x, V^{-1}(y/n))))^{n-1}.$$

As

$$\frac{P_0(\xi_1 \in B(x, V^{-1}(y/n))}{y/n} \to g(x),$$

(see Mattila, 1995, p. 36), it follows that

$$(1 - P_0(\xi_1 \in B(x, V^{-1}(y/n))))^{n-1} \to \mathrm{e}^{-yg(x)}. \qquad \square$$

Next we prove that $(\xi_i, \eta_i(n))$ and $(\xi_j, \eta_j(n))$ are asymptotically independent.

Proposition 3. *When n tends to infinity it holds that*

$$Pr(\eta_i(n) > y_i, \eta_j(n) > y_j | \xi_i = x_i, \xi_j = x_j) \to \mathrm{e}^{-y_i g(x_i)} \mathrm{e}^{-y_j g(x_j)}$$

Proof. Since the random variables $(\xi_i, \eta_i(n))$, $i = 1, 2, \ldots, n$ are exchangeable it is sufficient to prove the proposition for $i = 1$ and $j = 2$. Instead of proving the convergence for $\eta_1(n)$ and $\eta_2(n)$ we shall prove it for $\tilde{\eta}_1(n)$ and $\tilde{\eta}_2(n)$, where

$$\tilde{\eta}_k(n) = \min_{j \geqslant 3} nc_d |\xi_k - \xi_j|^d \quad \text{for } k = 1, 2.$$

By symmetry $\tilde{\eta}_k(n)$ and $\eta_k(n)$ differs only on a set having probability $1/(n-1)$. Thus

$$\tilde{\eta}_k(n) - \eta_k(n) \xrightarrow{p} 0, \quad k = 1, 2$$

which implies that $(\eta_1(n), \eta_2(n))$ and $(\tilde{\eta}_1(n), \tilde{\eta}_2(n))$ have the same limit distribution. By conditioning on $\xi_1 = x_1$ and $\xi_2 = x_2$ we get

$$Pr(\tilde{\eta}_1(n) > y_1, \tilde{\eta}_2(n) > y_2 | \xi_1 = x_1, \xi_2 = x_2) = Pr(E(x_1, n) > y_1, E(x_2, n) > y_2),$$

where

$$E(x_i, n) = n \| B(x_i, \min_{j \geqslant 3} |x_i - \xi_j|) \|, \quad i = 1, 2.$$

The event $\{E(x_i, n) > y_i\}$ occurs if all ξ_j, $j \geqslant 3$ falls outside the ball $B(x_i, V^{-1}(y_i/n)) = B(x_i)$.

Thus

$$Pr(\tilde{\eta}_1(n) > y_1, \tilde{\eta}_2(n) > y_2 | \xi_1 = x_1, \xi_2 = x_2) = (1 - Pr(B(x_1) \cup B(x_2)))^{n-2}.$$

But, as $x_1 \neq x_2$, $B(x_1)$ and $B(x_2)$ are disjoint if n is sufficiently large so

$$\lim_{n \to \infty} (1 - Pr(B(x_1) \cup B(x_2)))^{n-2} = e^{-y_i g(x_i)} e^{-y_j g(x_j)}. \qquad \square$$

As the set $\{\xi_1 = \xi_2\}$ has probability zero this gives us the asymptotic independence of $(\xi_i, \eta_i(n))$ and $(\xi_j, \eta_j(n))$.

Proposition 4. *Let the distribution P_0 of the sequence $\xi_1, \xi_2, \ldots, \xi_n$ of independent random vectors be absolutely continuous w.r.t. Lebesgue measure. Then it holds that*

$$\frac{1}{n} \sum_{i=1}^{n} \log n P_0(B(\xi_i, R_n(i))) \xrightarrow{p} -\gamma$$

as n tends to infinity.

Proof. Define

$$E_i(n) = n P_0(B(\xi_i, R_n(i)).$$

The exchangeability of $(\xi_i, \eta_i(n))$ gives

$$E\left[\frac{1}{n} \sum_{i=1}^{n} \log E_i(n)\right] = E(\log E_1(n)),$$

and

$$\mathrm{Var}\left[\frac{1}{n} \sum_{i=1}^{n} \log E_i(n)\right] = \frac{\mathrm{Var}(\log E_1(n))}{n} + 2 \frac{n^2 - n}{n^2} \mathrm{Cov}(\log E_1(n), \log E_2(n)).$$

Since, see Lemma 2

$$E(\log E_1(n)) \to -\gamma$$

and

$$\mathrm{Var}(\log E_1(n)) \to \frac{\pi^2}{6} - 1 < \infty,$$

an application of Chebychev's inequality will give the desired result if we can show that $\mathrm{Cov}(\log E_1(n), \log E_2(n)) \to 0$ as $n \to \infty$. Since $E_1(n)$ and $E_2(n)$ are asymptotically independent we will be through if we show that the sequence $\{\log E_1(n) \log E_2(n)\}_{n=1}^{\infty}$ is uniformly integrable. We have

$$E^2|\log E_1(n) \log E_2(n)|^{1.5} \leqslant E|\log E_1(n)|^3 E|\log E_2(n)|^3.$$

The right-hand side of this expression converges to $(\int |\log x|^3 \mathrm{e}^{-x}\,\mathrm{d}x)^2$, which is finite, giving us

$$\sup_n E|\log E_1(n) \log E_2(n)|^{1+0.5} < \infty.$$

Consequently, the sequence $\{\log E_1(n) \log E_2(n)\}_{n=1}^{\infty}$ has to be uniformly integrable, which completes the proof. □

Proposition 5. *Let the distribution P_0 of the sequence $\xi_1, \xi_2, \ldots, \xi_n$ of independent random vectors be absolutely continuous w.r.t. Lebesque measure. Then it holds that*

$$\frac{1}{n}\sum_{i=1}^{n} n P_0(B(\xi_i, R_n(i))) \xrightarrow{p} 1$$

as n tends to infinity.

Proof. Follows in a similar way from Corollary 1 and Proposition 3. □

3.2. Consistency

To prove the consistency of the MSP-estimate, we need some kind of continuity and identifiability condition. If we allow the distributions for parameter values close to each other to be too different, we cannot expect our estimation procedure to produce consistent estimators. This is clearly demonstrated by an example in Basu (1955). Our continuity condition is inspired by the Arzela–Ascoli theorem.

Define

$$z(n, \theta, x, y) = nB(x, r_n) \quad \text{where } r_n = (c_d^{-1} y/n)^{1/d},$$

$P(x, y)$=the distribution function of (X, Y)
with density function $p(x, y)$ defined by

$$p(x, y) = g^2(x) \exp(-yg(x)),\ y > 0.$$

Condition C1. Let (X, Y) have the distribution $P(x, y)$. For each $\varepsilon > 0$ and $\eta > 0$ there exists an integer m, sets $K_j \subset R^{d+1}$, $j = 1, 2, \ldots, m$, a partition of Θ into disjoint sets $\Theta_1, \Theta_2, \ldots, \Theta_m$ and parameter values $\psi_j \in \Theta_j$, $j = 1, 2, \ldots, m$, such that for each $j = 1, 2, \ldots, m$,

(i) the boundary ∂K_j of the set K_j has Lebesgue measure zero,
(ii) $P((X, Y) \in K_j) > 1 - \eta$,
(iii) $\sup_{\theta \in \Theta_j} |z(n, \theta, x, y) - z(n, \psi_j, x, y)| < \varepsilon$ for all $(x, y) \in K_j$ and for all $n \geqslant N(\varepsilon, \eta)$.

Remark 4. All or some of the sets $K_1, K_2, \ldots, K_m$ may be equal.

The mixture distribution in Example 1 satisfies Condition C1. The technique used in Ranneby (1984) to verify Condition C1 is also applicable in the multivariate situations. It follows that the continuity condition C1 usually is satisfied.

The identifiability condition we are going to use is called, according to the terminology in Rao (1973), a strong identifiability condition. However, when weak identifiability conditions are used, these conditions are usually used in combination with other conditions implying that a strong identifiability condition is satisfied.

Let $T(M, \theta)$ denote the expected value of $\max(-M, \log Y f_\theta(X))$, where (X, Y) has the distribution $P(x, y)$.

Condition C2. For each $\delta > 0$ there exists a constant $M_1 = M_1(\delta)$ such that

$$\sup_{\theta \in B^c(\theta^o, \delta)} T(M_1, \theta) < T(\theta^o) = E(\log Y g(X)).$$

Remark 5. Using the results in Corollary 2.5 in Ranneby (1984) it is easily seen that the identifiability condition C2 is satisfied if the density functions $f_\theta(x)$ are continuous functions of θ for almost all x and the weak identifiability condition are satisfied.

Theorem 1. *Let $\xi_1, \xi_2, \ldots, \xi_n$ be a sequence of i.i.d. random vectors in $\mathscr{R}^d$ with distribution P_θ and density function $f_\theta(x)$, where θ belongs to an admissible set Θ. Suppose that Conditions C1 and C2 are satisfied. Then the MSP-estimate $\hat{\theta}_n$ converges in probability to the true parameter value θ^o.*

Before we prove the theorem we establish two lemmas and introduce some notations. Let

$$t_M(x) = \max(-M, \log x),$$

$$h_{M,N}(x) = \min(N, t_M(x)),$$

$$a_N(x) = \max(0, \log x - N),$$

$$H_n(M, N, \theta) = \frac{1}{n} \sum_{i=1}^{n} h_{M,N}(z_i(n, \theta)),$$

$$T(M,\theta)=\int t_M(yf_\theta(x))\,\mathrm{d}P(x,y),$$

$$H(M,N,\theta)=\int h_{M,N}(yf_\theta(x))\,\mathrm{d}P(x,y),$$

$$A(N,\theta)=\int a_N(yf_\theta(x))\,\mathrm{d}P(x,y).$$

Lemma 3. *$H_n(M,N,\theta)$ converges in probability to $H(M,N,\theta)$. Further if Condition C*1 *is satisfied then the convergence is uniform in θ.*

Proof. We begin with the pointwise convergence of $H_n(M,N,\theta)$. The random variables $z_i(n,\theta)$ are exchangeable. Thus

$$E(H_n(M,N,\theta))=E(h_{M,N}(z_i(n,\theta)))$$

and

$$\mathrm{Var}(H_n(M,N,\theta))=\frac{1}{n^2}\mathrm{Var}(h_{M,N}(z_1(n,\theta))) + \frac{n^2-n}{n^2}\mathrm{Cov}(h_{M,N}(z_1(n,\theta)),h_{M,N}(z_2(n,\theta))).$$

Next we show the convergence of

$$E[h_{M,N}(z_1(n,\theta))]=\int h_{M,N}(z(n,\theta,x,y))\,\mathrm{d}P_n(x,y),$$

where P_n denotes the distribution of $(\xi_i,\eta_i(n))$.

As $h_{M,N}$ is a bounded continuous function and since

$$z(n,\theta,x,y)\to yf_\theta(x),\quad n\to\infty$$

and

$$P_n(x,y)\to P(x,y),\quad n\to\infty$$

it follows from Lebesgue Dominated Convergence Theorem that

$$E(h_{M,N}(z_i(n,\theta)))\to\int h_{M,N}(yf_\theta(x))\,\mathrm{d}P(x,y)=H(M,N,\theta).$$

As $(\xi_1,\eta_1(n))$ and $(\xi_2,\eta_2(n))$ are asymptotically independent, see Proposition 3, it follows that

$$E(h_{M,N}(z_1(n,\theta))h_{M,N}(z_2(n,\theta)))\to H(M,N,\theta)^2$$

so that

$$\mathrm{Cov}(h_{M,N}(z_1(n,\theta)),h_{M,N}(z_2(n,\theta)))\to 0\quad \text{as } n\to\infty.$$

Thus

$$E(H_n(M, N, \theta)) \to H(M, N, \theta),$$

and

$$\mathrm{Var}(H_n(M, N, \theta)) \to 0$$

implying that $H_n(M, N, \theta)$ converges in probability to $H(M, N, \theta)$ as $n \to \infty$.

Now we continue with the uniform convergence of $H_n(M, N, \theta)$.

Choose the sets K_j in Condition C1 such that $P(K_j) > 1 - \frac{\varepsilon}{16\max(M,N)}$. Furthermore, choose $\Theta_1, \Theta_2, \ldots, \Theta_m$ and $\psi_1, \psi_2, \ldots, \psi_m$ such that

$$\sup_{\theta\in\Theta_j} |z(n, \theta, x, y) - z(n, \psi_j, x, y)| < \frac{\varepsilon \mathrm{e}^{-M}}{8},$$

for all $(x, y) \in K_j$, $j = 1, 2, \ldots, m$. Write for $\theta \in \Theta_j$,

$$H(M, N, \theta) = \lim_{n\to\infty} \int_{K_j} h_{M,N}(z(n, \theta, x, y))\, \mathrm{d}P(x, y) + \lim_{n\to\infty} \int_{K_j^c} h_{M,N}(z(n, \theta, x, y))\, \mathrm{d}P(x, y).$$

Let $\theta \in \Theta_j$. We have, for n sufficiently large,

$$|h_{M,N}(z(n, \theta, x, y)) - h_{M,N}(z(n, \psi_j, x, y))| \leqslant \begin{cases} \varepsilon/8 & \text{on } K_j, \\ 2\max(M, N) & \text{on } K_j^c. \end{cases} \tag{2}$$

Thus

$$\sup_{\theta\in\Theta_j} |H(M, N, \theta) - H(M, N, \psi_j)| < \frac{\varepsilon}{8} + \frac{\varepsilon}{8} = \frac{\varepsilon}{4}. \tag{3}$$

We have

$$\begin{aligned} &|H_n(M, N, \theta) - H_n(M, N, \psi_j)| \\ &\quad < \frac{1}{n}\sum |h_{M,N}(z_i(n, \theta)) - h_{M,N}(z_i(n, \psi_j)| I(\xi_i, \eta_i(n) \in K_j) \\ &\qquad + \frac{1}{n}\sum |h_{M,N}(z_i(n, \theta)) - h_{M,N}(z_i(n, \psi_j)| I(\xi_i, \eta_i(n) \in K_j^c). \end{aligned}$$

Since the boundary ∂K_j has P-measure zero we get

$$\frac{1}{n}\sum I((\xi_i, \eta_i(n)) \in K_j^c) \to_p P((X, Y) \in K_j^c) < \frac{\varepsilon}{16\max(M, N)}$$

and then it follows from (2) that

$$P\left(|H_n(M, N, \theta) - H_n(M, N, \psi_j)| < \frac{\varepsilon}{4}\right) \to 1 \quad \text{as } n \to \infty. \tag{4}$$

Combining (3) and (4) we get that

$$H_n(M, N, \theta) \xrightarrow{p} H(M, N, \theta)$$

uniformly in θ. □

Lemma 4. *Define*

$$\tilde{z}_i(n, \theta) = \begin{cases} z_i(n, \theta) & if\ \bar{z}(n, \theta) \leqslant 1, \\ z_i(n, \theta)/\bar{z}(n, \theta) & if\ \bar{z}(n, \theta) > 1. \end{cases}$$

Then the random function

$$A_n(N, \theta) = \frac{1}{n} \sum_{i=1}^{n} \max(0, \log \tilde{z}_i(n, \theta) - N)$$

converges to zero for all elementary events, uniformly in n and θ as $N \to \infty$.

Proof. Obviously, $\sum \tilde{z}_i(n, \theta) \leqslant n$. The rest of the proof is only a slight modification of Lemma 2 in Ranneby and Ekström (1997). □

Proof of Theorem 1. Recall the definition of $S_n(\theta)$ as

$$\begin{aligned} S_n(\theta) = & \frac{1}{n} \sum \log z_i(n, \theta) \\ & + \frac{1}{n} \log(1 - \bar{z}(n, \theta)) I(\bar{z}(n, \theta) \leqslant 1) - \log \bar{z}(n, \theta) I(\bar{z}(n, \theta) > 1). \end{aligned}$$

We have

$$\begin{aligned} S_n(\theta) & \leqslant \frac{1}{n} \sum \log z_i(n, \theta) - \log \bar{z}(n, \theta) I(\bar{z}(n, \theta) > 1) \\ & \leqslant \frac{1}{n} \sum h_{M,N}(\tilde{z}_i(n, \theta)) + \frac{1}{n} \sum \max(0, \log \tilde{z}_i(n, \theta) - N). \end{aligned}$$

Obviously, $\sum \tilde{z}_i(n, \theta) \leqslant n$, so Lemma 4 gives that

$$A_n(N, \theta) \xrightarrow{p} 0,$$

uniformly in θ and n. As $\tilde{z}_i(n, \theta) \leqslant z_i(n, \theta)$ and $h_{M,N}(x)$ is a non-decreasing function we get that

$$S_n(\theta) \leqslant H_n(M, N, \theta) + A_n(N, \theta).$$

Lemma 3 gives that $H_n(M, N, \theta) \xrightarrow{p} H(M, N, \theta)$, uniformly in θ as $n \to \infty$. Note that

$$H(M, N, \theta) + A(N, \theta) = T(M, \theta)$$

and let $\hat{\theta}_n$ denote the MSP-estimate. For N large the following holds with probability going to one as n tends to infinity:

$$\begin{aligned} S_n(\hat{\theta}_n) &\leqslant \frac{1}{n}\sum h_{M,N}(z_i(n,\hat{\theta}_n) + \frac{\varepsilon}{4} \\ &\leqslant H(M,N,\hat{\theta}_n) + \frac{\varepsilon}{2} \\ &\leqslant H(M,N,\hat{\theta}_n) + A(N,\hat{\theta}_n) + \frac{\varepsilon}{2} = T(M,\hat{\theta}_n) + \frac{\varepsilon}{2}. \end{aligned}$$

As $S_n(\theta^o) \xrightarrow{p} -\gamma$ we get

$$T(M,\hat{\theta}_n) + \frac{\varepsilon}{2} > S_n(\hat{\theta}_n) > S_n(\theta^o) > T(\theta^o) - \frac{\varepsilon}{2},$$

which implies that

$$T(M,\hat{\theta}_n) > T(\theta^o) - \varepsilon.$$

Now the identifiability condition C2 gives

$$|\hat{\theta}_n - \theta^o| < \delta$$

which completes the proof. □

4. Simulation results

Here we present the results of the simulation study we conducted in order to confirm and support our hypothesis about the consistency and asymptotic normality of the MSP-estimates for multivariate observations. It was also of interest to compare the estimation based on Dirichlet tesselation with that based on nearest-neighbor balls. In all the figures and tables, MSPE stands for MSP-estimate.

4.1. Gaussian density

We simulated bivariate Gaussian random variables with the following mean vector and covariance matrix:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

Then we constructed two spacing functions—one based on nearest neighbor balls and one based on Dirichlet cells. These functions were subsequently maximized by random search algorithm. As the starting point in the parameter space, to speed up the convergence, the true value of the parameter vector was chosen. We have also experimented with different starting points, and always the outcomes (e.g. the maximizers) were identical. Each spacing function was maximized twice—first based on 50, then on 200 observations, and both experiments were repeated 500 times. The mean and the covariance matrix of the outcomes

Table 1
The covariance matrix ($\times n$) of the MSPE of the parameters of Gaussian distribution

n		Nearest neighbor circles					Dirichlet cells				
		μ_1	μ_2	σ_1	σ_2	ρ	μ_1	μ_2	σ_1	σ_2	ρ
50	μ_1	1.20	0.66	−0.07	−0.02	−0.09	1.04	0.56	−0.08	−0.02	−0.07
	μ_2	0.66	1.31	0.05	0.07	0.06	0.56	1.07	0.02	0.01	0.04
	σ_1	−0.07	0.05	0.88	0.26	0.82	−0.08	0.02	0.56	0.17	0.54
	σ_2	−0.02	0.07	0.26	0.79	0.74	−0.02	0.01	0.17	0.54	0.59
	ρ	−0.09	0.06	0.82	0.74	1.88	−0.07	0.04	0.54	0.59	1.51
200	μ_1	1.29	0.65	−0.06	−0.05	−0.13	1.10	0.55	−0.07	−0.04	−0.06
	μ_2	0.65	1.35	−0.05	0.03	−0.08	0.55	1.05	−0.03	0.03	0.00
	σ_1	−0.06	−0.05	0.79	0.33	0.80	−0.07	−0.03	0.53	0.18	0.55
	σ_2	−0.05	0.03	0.33	0.90	0.89	−0.04	0.03	0.18	0.53	0.57
	ρ	−0.13	−0.08	0.80	0.89	2.05	−0.06	0.00	0.55	0.57	1.36

Table 2
The bias of the MSPE of the parameters of Gaussian distribution

	μ_1	μ_2	σ_1	σ_2	ρ
NN circles ($n = 50$)	−0.0060	0.0107	0.0025	0.0140	0.0307
Dirichlet cells ($n = 50$)	−0.0091	0.0064	0.0518	0.0598	0.0116
NN circles ($n = 200$)	−0.0010	−0.0003	0.0032	0.0019	0.0101
Dirichlet cells ($n = 200$)	−0.0041	−0.0004	0.0244	0.0221	0.0034

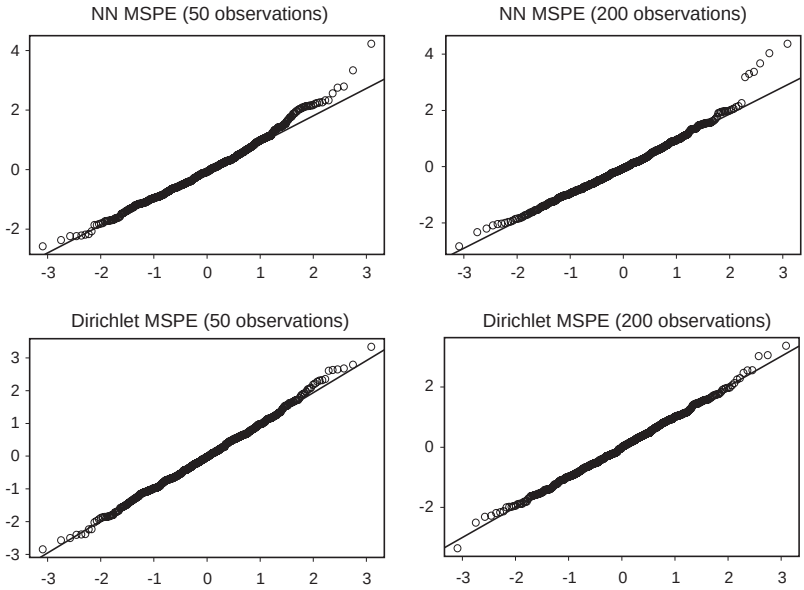

Fig. 1. Quantile–quantile plots of the MSPE for σ_2 (Gaussian density).

were calculated and averaged over 500 repetitions (see Tables 1 and 2). The estimated covariance matrix was compared to the Cramér–Rao bound

$$I^{-1} = \begin{pmatrix} 1 & 0.5 & 0 & 0 & 0 \\ 0.5 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.125 & 0.5 \\ 0 & 0 & 0.125 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 1.25 \end{pmatrix}.$$

The results confirm the consistency of the MSP-estimates. Besides, we note that the MSP-estimate based on Dirichlet cells is much closer to being efficient than the estimate based on the nearest neighbor balls. Fig. 1 displays the normal quantile–quantile plots of the estimates of one of the parameters. We note that the plots support the conjecture of asymptotic normality of the MSP-estimates.

Table 3
The covariance matrix ($\times n$) of the MSPE of the parameters of Gaussian mixture

n		Nearest neighbor circles					Dirichlet cells				
		μ_1	μ_2	σ_1	σ_2	ρ	μ_1	μ_2	σ_1	σ_2	ρ
50	μ_1	2.61	1.11	0.39	0.04	0.49	1.96	0.86	0.25	−0.09	0.24
	μ_2	1.11	1.72	0.16	0.12	0.25	0.86	1.59	0.16	0.16	0.32
	σ_1	0.39	0.16	1.02	0.22	1.02	0.25	0.16	1.04	0.27	0.95
	σ_2	0.04	0.12	0.22	1.14	1.04	−0.09	0.16	0.27	1.10	0.89
	ρ	0.49	0.25	1.02	1.04	2.53	0.24	0.32	0.95	0.89	2.43
200	μ_1	2.50	1.16	0.41	0.06	0.51	1.61	0.88	0.20	0.03	0.25
	μ_2	1.16	1.99	0.14	0.10	0.22	0.88	1.69	0.06	0.11	0.13
	σ_1	0.41	0.14	1.12	0.26	0.84	0.20	0.06	0.91	0.19	0.82
	σ_2	0.06	0.10	0.26	1.12	0.87	0.03	0.11	0.19	0.87	0.80
	ρ	0.51	0.22	0.84	0.87	2.71	0.25	0.13	0.82	0.80	2.10

Table 4
The bias of the MSPE of the parameters of Gaussian mixture

	μ_1	μ_2	σ_1	σ_2	ρ
NN circles ($n=50$)	0.0060	−0.0044	−0.0313	−0.0287	0.0057
Dirichlet cells ($n=50$)	−0.0051	−0.0416	0.0198	0.0391	0.0125
NN circles ($n=200$)	0.0021	−0.0009	0.0191	0.0302	0.0025
Dirichlet cells ($n=200$)	−0.0055	−0.0017	0.0179	0.0124	0.0096

4.2. Mixture of two Gaussian densities

The second simulation was performed for the mixture of two bivariate Gaussian densities. This is a well-known example, where the ML-method breaks down. Particularly, if we set the mean of the first (say) component of the mixture equal to one of the observations, e.g. $\mu_1 = x_1$, $\mu_2 = y_1$ then the ML-function will tend to infinity as σ_1 goes to zero. We simulated the bivariate observations with the following density:

$$g(x, y, \theta) = 0.8 \times f_1(x, y, \mu_1, \mu_2, \sigma_1, \sigma_2, \rho_1) + 0.2 \times f_2(x, y, \mu_3, \mu_4, \sigma_3, \sigma_4, \rho_2),$$

where $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$, $\rho_1 = 0.5$, $\mu_3 = \mu_4 = 2$, $\sigma_3 = \sigma_4 = 2$ and $\rho_2 = 0$. Only the parameters of f_1 were considered unknown. We constructed the spacing functions for 50 and 200 observations. Each experiment was again repeated 500 times as in the previous example. The results can be seen in Tables 3 and 4. The estimated covariance matrix was again compared to the Cramér–Rao bound

$$I^{-1} = \begin{pmatrix} 1.53 & 0.78 & 0.19 & 0.10 & 0.26 \\ 0.78 & 1.53 & 0.10 & 0.19 & 0.26 \\ 0.19 & 0.10 & 0.87 & 0.23 & 0.87 \\ 0.10 & 0.19 & 0.23 & 0.87 & 0.87 \\ 0.26 & 0.26 & 0.87 & 0.87 & 2.15 \end{pmatrix}.$$

Although, as we have noted, this case is more complicated than the previous one, the results of the simulation are very satisfactory. Apparently, the MSP-estimate based on Dirichlet cells seems consistent, with variance, approaching, and sometimes surpassing, the Cramér–Rao bound. It is also asymptotically normal, similarly to the Gaussian case (Fig. 2).

5. Discussion

In the present paper we have proved consistency for multivariate MSP-estimates based on NN-balls. Using results from Jimenez and Yukich (2002) it is possible to prove consistency also for the version based on Dirichlet cells. For the univariate version of the MSP-method the estimators are normally distributed and asymptotically efficient. Results from our simulation study indicate that the estimators based on both versions are asymptotically normally distributed but that the variances for the NN-version are much larger than the Cramér–Rao bound. For both versions proofs of asymptotic normality are still missing. However, results in a recent paper by Baryshnikov and Yukich may be used to prove asymptotic normality

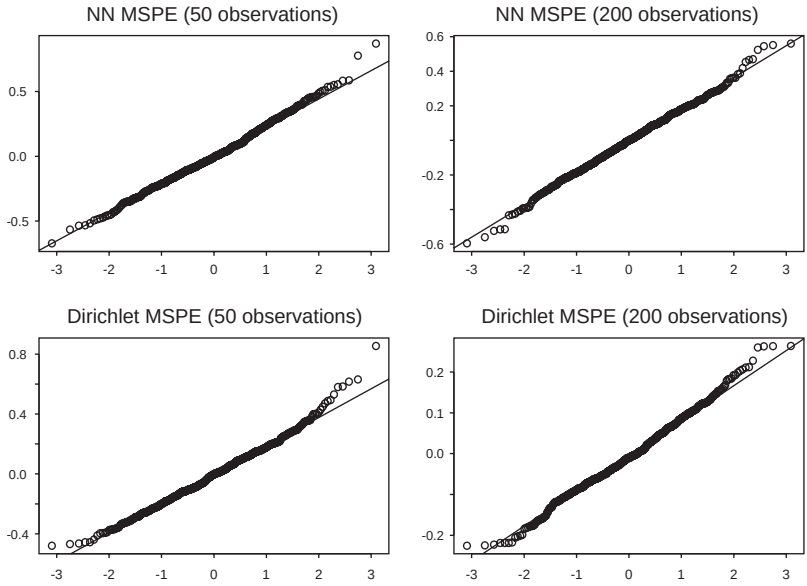

Fig. 2. Quantile–quantile plots of the MSPE for μ_1 (Gaussian mixture).

for the NN-version. As mentioned in Ranneby (1984), for univariate MSP-estimates it is possible to check the validity of the model at the same time as the estimation problem is solved, see also Cheng and Stephens (1989) and Cheng and Traylor (1995). The results from Baryshnikov and Yukich (2003) can be used to give confidence limits for $S_n(\theta_0)$. Thus for multivariate MSP-estimates based on NN-balls it is possible to check the validity of the model. Another advantage with the NN-version is that it is much easier to handle, especially in higher dimensions. The drawback is of course the lack of efficiency. In situations where the maximum likelihood method fails that is usually because of global reasons. Locally the method may still give satisfactory results. To be specific for mixture distributions it should be possible to use the ML method if the starting values for the maximization are in the neighborhood of the true values. Thus in these situations it should be possible to use the MSP-method based on NN-balls to get consistent estimators which can be used as starting values for the ML estimation.

Acknowledgements

This research was conducted using the resources of High Performance Computing Center North (HPC2N) in Sweden. The authors would like to thank Prof. Yuri Belyaev and Dr. Jun Yu for valuable comments and discussion.

References

Baryshnikov, Y., Yukich, J.E., 2003. Generalized spacing statistics and Gaussian fields. Technical Report, Lehigh University, USA.

Basu, D., 1955. An inconsistency of the method of maximum likelihood. Ann. Math. Statist. 26, 144–145.
Bickel, P.J., Breiman, L., 1983. Sum of functions of nearest neighbor distances, moment bounds, limit theorems, and a goodness of fit test. Ann. Probab. 11, 185–214.
Cheng, R.C.H., Amin, N.A.K., 1983. Estimating parameters in continuous univariate distributions with a shifted origin. J. Roy. Statist. Soc. B 45, 394–403.
Cheng, R.C.H., Stephens, M.A., 1989. A goodness-of-fit test using Moran's statistic with estimated parameters. Biometrika 76, 385–392.
Cheng, R.C.H., Traylor, L., 1995. Non-regular maximum likelihood problems (with discussions). J. Roy. Statist. Soc. B 57, 3–44.
Darling, D.A., 1953. On a class of problems related to the random division of an interval. Ann. Math. Statist. 24, 239–253.
Jammalamadaka, S.R., Janson, S., 1986. Limit theorems for a triangular scheme of U-statistics with applications to inter-point distances. Ann. Probab. 14, 1347–1358.
Jammalamadaka, S.R., Zhou, S., 1993. Goodness of fit in multidimensions based on nearest neighbor distances. J. Nonparametric Statist. 2, 271–284.
Jimenez, R., Yukich, J.E., 2002. Asymptotics for statistical distances based on Voronoi tesselation. J. Theoret. Probab. 15, 503–541.
Mattila, P., 1995. Geometry of Sets and Measures in Euclidean Spaces, Fractals and Rectifiability, Cambridge University Press, Cambridge, UK.
Pyke, R., 1965. Spacings (with discussion). J. Roy. Statist. Soc. B 27, 395–449.
Ranneby, B., 1984. The maximum spacing method. An estimation method related to the maximum likelihood method. Scand. J. Statist. 11, 93–112.
Ranneby, B., 1996. Spatial and temporal models in contextual classification. In: Todd Mowrer, H., Czaplewski, R.L., Hamre, R.H. (Eds.), Spatial Accuracy Assessment in Natural Resources and Environmental Sciences: Second International Symposium. Fort Collins, Colorado, pp. 451–458.
Ranneby, B., Ekström, M., 1997. Maximum spacing estimates based on different metrics. Research Report No. 5, Department of Mathematical Statistics, Umeå University.
Rao, C.R., 1973. Linear Statistical Inference and its Applications, 2nd Edition. Wiley, New York.
Reiss, R.D., 1989. Approximate Distributions of Order Statistics, with Applications to Nonparametric Statistics, Springer, New York.
Rényi, A., 1953. On the theory of order statistics. Acta Math. Acad. Sci. Hungar. 4, 191–231.
Titterington, D.M., 1985. Comment on estimating parameters in continuous univariate distribution. J. Roy. Statist. Soc. B 47, 115–116.