KYBERNETIKA — VOLUME 39 (2003), NUMBER 6, PAGES 703- 717

BAYESIAN MCMC ESTIMATION OF THE ROSE OF DIRECTIONS

MICHAELA PROKEŠOVÁ

The paper concerns estimation of the rose of directions of a stationary fibre process in $\mathbb{R}^3$ from the intersection counts of the process with test planes. A new approach is suggested based on Bayesian statistical techniques. The method is derived from the special case of a Poisson line process however the estimator is shown to be consistent generally. Markov chain Monte Carlo (MCMC) algorithms are used for the approximation of the posterior distribution. Uniform ergodicity of the algorithms used is shown. Properties of the estimation method are studied both theoretically and by simulation.

Keywords: rose of directions, planar section, fibre process, Bayesian statistics, MCMC algorithm

AMS Subject Classification: 62M30, 62F15, 65B05

1. INTRODUCTION

The analysis of the anisotropy of random fibre systems is a frequent problem in many applied sciences like biology and metallography. If we model these systems as stationary fibre processes in $\mathbb{R}^3$ then the anisotropy can be quantitatively characterized by the directional distribution $\mathcal{P}$. Throughout the paper the non-oriented directional distribution, called also the rose of directions, is studied, while the oriented one is typical for surface normal orientations of particle systems. The fibre systems are usually examined using section probes of lower dimension. In $\mathbb{R}^3$ the data which are used for the estimation procedure are the intersection counts (observed in finitely many test windows) of the fibre process with test planes. Several methods have been suggested for the estimation of $\mathcal{P}$ including parametric models, interpolation with smooth functions or the use of the associated zonoid. See e.g. Cruz-Orive et al [3], Hilliard [4], Mecke and Nagel [11], Campi, Haas, Weil [2], Kiderlen [7].

In this paper we present a new approach, based on Bayesian statistical techniques. In Section 2 we recall definitions and notation that will be needed in the sequel. In Section 3 we formulate our assumptions and review the two estimators of Kiderlen [7]. In Section 4 we present the new estimator of $\mathcal{P}$. We consider a discretized parametric version $\mathcal{R}$ of the directional distribution $\mathcal{P}$. The support $\operatorname{supp}\mathcal{R}$ of the measure $\mathcal{R}$ is a finite set of vectors which depends only on the intersecting test planes, not on the data. Using the relation between the rose of directions and the rose of

intersections we determine the likelihood of the observed rose of intersections and also the posterior distribution on possible values of $\mathcal{R}$. The posterior mean is then our estimator of $\mathcal{P}$. Since we do not make any assumptions about the functional form of the distribution $\mathcal{P}$ nor we use smooth functions for the approximation of $\mathcal{P}$ the resulting estimator we get using the 'discretization' procedure is able to detect also sharp anisotropies and multimodal directional distributions. Consistency of the estimator is shown for general stationary fibre processes.

Since it is not possible to compute the estimator for concrete data analytically we need to use one of the Markov Chain Monte Carlo methods to get a numerical solution. In Section 5 we describe the algorithm that we use, and discuss its properties including the uniform ergodicity. Finally in Section 6 we show simulation results – we apply the estimation procedure to several examples and compare empirically the new estimator with the EM and LP estimators of Kiderlen [7] by measuring the Prohorov distance between the original rose of directions and its estimate.

2. BACKGROUND

As a standard notation we use the scalar product $|\langle .,. \rangle|$ in $\mathbb{R}^3$, $\mathcal{S}^2$ the unit sphere in $\mathbb{R}^3$, H^k, $k \in \mathbb{N}$, the k-dimensional Hausdorff measure, μ^k_{Leb} the Lebesgue measure in $\mathbb{R}^k$, $\mathcal{B}(A)$ the Borel σ-algebra on the set A. We shall call a measure on $\mathcal{S}^2$ even, if it assigns the same mass to the set $B \in \mathcal{B}(\mathcal{S}^2)$ and to its reflection at the origin $-B$.

By a line process Φ in $\mathbb{R}^3$ we mean a random variable with values in the set L of locally finite collections of lines in $\mathbb{R}^3$ equipped with the σ-algebra $\mathcal{F} = \mathcal{A}|_L$ – the restriction of the hitting σ-algebra $\mathcal{A}$ to the set F where $\mathcal{A} = \sigma\{F_K, K \subseteq \mathbb{R}^3, K \text{ is compact}\}$ is the smallest σ-algebra generated by the sets F_K and $F_K = \{F \subseteq \mathbb{R}^3, F \text{ closed} : F \cap K \neq \emptyset\}$ for K a compact set (Matheron [9]). If the number of lines hitting an arbitrary compact test set is Poisson distributed then Φ is a Poisson process.

We consider a stationary line process Φ in $\mathbb{R}^3$. Its distribution is translation invariant and can be characterized by the length intensity and the directional distribution. For the definitions let A be a Borel set of unit volume. The length intensity Λ is the mean length of the union of all line segments from the intersection of the process with the set A, $\Lambda = \mathrm{E}\, H^1(\Phi \cap A)$. We suppose throughout the following that $\Lambda > 0$. The directional distribution $\mathcal{P}$ of the process is an even measure on $\mathcal{S}^2$ defined by $\mathcal{P}(B) = \frac{\eta(B)}{\Lambda}$, where $B \in \mathcal{B}(\mathcal{S}^2)$ is a centrally symmetric set and $\eta(B)$ is the mean length of the union of all line segments from the intersection of A with the lines of the process which have directions in the set B. Since $\eta(\mathcal{S}^2) = \Lambda$, we see that $\mathcal{P}$ is really a probability measure. Both Λ and $\mathcal{P}$ are uniquelly determined and do not depend on the set A.

Let $h \in \mathcal{S}^2$ be a unit normal vector of a plane $h^\perp$. Then the intersection $\Phi \cap h^\perp$ is almost surely a point process. It is a stationary process in $h^\perp$ and its intensity (the mean number of points per unit area in $h^\perp$) as a function of h is an even continuous function on $\mathcal{S}^2$ called the rose of intersections. (Since the process Φ is stationary the distribution of the intersection process does not depend on the location of the plane $h^\perp$ but only on its normal vector orientation).

Let us denote by $\mathcal{F}_{\mathcal{P}}$ the cosine transform of the directional distribution $\mathcal{P}$

$$\mathcal{F}_{\mathcal{P}}(h) = \int_{\mathcal{S}^2} |\langle h, u\rangle| \mathcal{P}(\mathrm{d}u), \quad h \in \mathcal{S}^2. \tag{1}$$

Then the function $V(h) = \Lambda \mathcal{F}_{\mathcal{P}}(h)$, $h \in \mathcal{S}^2$ is the cosine transform of the directional measure $\eta = \Lambda \mathcal{P}$ of the line process Φ, hence it determines $\Lambda \mathcal{P}$ uniquely (Schneider [14]). Moreover, $V(h)$ is exactly the rose of intersections (Stoyan, Kendall, Mecke [15], Chapter 9.4), which is the fact that makes the estimation of η possible.

3. MODEL

In this section, we shall present a new estimator for the directional distribution of a stationary Poisson line process Φ. Let $h_1^\perp, \dots h_n^\perp$ be n test planes with corresponding normal unit vectors $h_1, \dots, h_n$ (due to the stationarity of Φ we restrict to planes containing the origin). Set $\mathcal{H} = \{h_1, \dots, h_n\} \subseteq \mathcal{S}^2$. Let V_i, $i \in \{1, \dots, n\}$ be the random number of points in the intersection $\Phi \cap h_i^\perp$ which lie in a given Borel test set $A_i \subseteq h_i^\perp$. Without loss of generality we may assume $H^2(A_i) = 1$, $i = 1, \dots, n$. According to the Poisson assumption and equation (1), the random variable V_i is Poisson distributed with the mean value $V(h_i)$. A sample of the random vector $(V_1, \dots, V_n)$ denoted by $\vec{v} = (v_1, \dots, v_n)$ presents the measured data.

We shall make three assumptions for the derivations as follows.

Assumptions. (i) Test planes do not contain a common line, $\mathrm{lin}\{h_1, \dots, h_n\} = \mathbb{R}^3$. Here lin D denotes the linear hull of a set $D \subseteq \mathbb{R}^3$.
(ii) Random variables $V_1, \dots, V_n$ are stochastically independent. (This can be achieved using independent copies of the process Φ to obtain the intersections with the test planes $h_i^\perp$.)
(iii) It holds $\vec{v} \neq 0$. (Otherwise the zero measure would be a reasonable estimator for η.)

During the estimation procedure we first define a suitable discrete approximation $\mathcal{R}$ (a probability measure on $\mathcal{S}^2$ with finite support and depending on a finite number of parameters) which we use as a model for $\mathcal{P}$. Then by the use of Bayesian techniques we estimate the parameters of this model $\mathcal{R}$ from the measured data. Substituting the estimated parameters into the formula for $\mathcal{R}$ we get the resulting estimate $\widehat{\mathcal{R}}$ of the rose of directions $\mathcal{P}$.

Let δ_{t_j} be an even probability measure with the support formed by the vectors $\pm t_j \in \mathcal{S}^2$. Then $\mathcal{R}$ can be written as a finite mixture of such measures:

$$\mathcal{R} = \sum_{j=1}^{k} a_j \delta_{t_j}, \quad a_j \geq 0, \quad \sum_{j=1}^{k} a_j = 1, \quad t_j \in T. \tag{2}$$

The choice of the set $T \subseteq \mathcal{S}^2$ is essential for the quality of the approximation. We define

$$T = \{h_r \times h_l \mid h_r \neq h_l \in \mathcal{H}\} \tag{3}$$

as the set of all (different) vector products of the vectors from $\mathcal{H}$. Optimality of this choice is proved in Kiderlen [7]. Let us briefly review the problem and its solution here.

Kiderlen [7] considered the problem of finding a measure-valued maximum likelihood estimator of the directional measure η of a Poisson line process Φ in $\mathbb{R}^d$, $d \geq 2$. We refer to the case $d = 3$. Let us denote by $\mathcal{M}$ the set of all centrally symmetric measures on $\mathcal{S}^2$ and compute the log-likelihood function $\mu \to \log P(H^0(\Phi \cap A \cap h_i^\perp) = v_i)$, $i = 1, \ldots, n$ (where A is some unit volume test window). The following equality holds up to an additive constant.

$$L(\mu) = \sum_{i=1}^{n} (v_i \log(\mathcal{F}_\mu(h_i)) - \mathcal{F}_\mu(h_i)). \tag{4}$$

The original problem can now be reformulated as a convex optimization problem

$$\begin{array}{ll} \text{minimize:} & -L(\mu) \\ \text{subject to:} & \mu \in \mathcal{M}\,. \end{array} \tag{5}$$

A solution of problem (5) always exists and any two different solutions μ^*, σ^* are tomographically equivalent, i.e. $\mathcal{F}_{\mu^*}(h_i) = \mathcal{F}_{\sigma^*}(h_i)$, $i = 1, \ldots, n$ (for proofs see Mair [8]). However, (5) cannot, in general, be solved in a closed form and a numerical optimization has to be used. Since $\mathcal{M}$ is infinite dimensional it has to be replaced by a finite dimensional subset $\mathcal{M}(T) = \{\sigma \in \mathcal{M} : \operatorname{supp} \sigma \subseteq T\}$ where $T = \{t_j\}_{j=1}^k$ is a finite subset of $\mathcal{S}^2$. So we get a modified problem

$$\begin{array}{ll} \text{minimize:} & -L(\mu) \\ \text{subject to:} & \mu \in \mathcal{M}(T). \end{array} \tag{6}$$

Again according to Mair [8], this problem is solvable and a measure $\mu^* \in \mathcal{M}(T)$ is a solution of (6) if and only if

$$\sum_{i=1, v_i \neq 0}^{n} \frac{v_i}{\mathcal{F}_{\mu^*}(h_i)} |\langle h_i, t_j \rangle| \;\leq\; \sum_{i=1}^{n} |\langle h_i, t_j \rangle| \tag{7}$$

for $j = 1, \ldots, k$ with equality for all t_j in $\operatorname{supp} \mu^*$. We get only finitely many inequalities here and the measure μ^* solving the problem (7) can be computed numerically using the iterative EM algorithm (the algorithm is described in detail in McLachlan [10]).

We get the following proposition as a special case of Theorem 1 (Kiderlen [7])

Proposition 1. Let the Assumptions (i)–(iii) be fulfilled. Then (5) has a solution. If T is chosen according to (3) then any solution of the discretized problem (6) is a solution of the original problem (5).

Thus the directional measure and the rose of directions as its $1/\Lambda$ multiple can be effectively approximated by finite-support measures with support in T. Kiderlen [7] suggests two estimators.

The first estimator of η is the solution of (7). It is found using the iterative EM algorithm. The concrete iteration procedure for the problem (7) is the following. We start with $(\vec{a})^{(0)} = (1, 1, \dots, 1) \in \mathbb{R}^k$. Then, given $(\vec{a})^{(m)} = (a_1^{(m)}, \dots a_k^{(m)})$, we define $(\vec{a})^{(m+1)} = (a_1^{(m+1)}, \dots, a_k^{(m+1)})$ by

$$a_j^{(m+1)} = \frac{a_j^{(m)}}{\rho_j} \left(\sum_{i=1, v_i \neq 0}^{n} \frac{v_i}{\gamma_i^{(m)}} |\langle h_i, t_j \rangle| \right), \qquad j = 1, \dots, k,$$

where

$$\gamma_i^{(m)} = \sum_{s=1}^{k} |\langle h_i, t_s \rangle| \, a_s^{(m)}, \qquad i = 1, \dots, n,$$

and

$$\rho_j = \sum_{r=1}^{n} |\langle h_r, t_j \rangle|, \qquad j = 1, \dots, k.$$

If we denote $(\vec{a})^* = \lim_{n \to \infty} (\vec{a})^{(m)}$ then $\sum_{j=1}^{k} a_j^* \delta_{t_j}$ is the solution of (7) and setting $(\vec{a})^{EM} = \frac{(\vec{a})^{(m)}}{|(\vec{a})^{(m)}|}$ for an $m \in \mathbb{N}$ big enough the first estimator of $\mathcal{P}$ is $\sum_{j=1}^{k} a_j^{EM} \delta_{t_j}$. Denote it by EM.

The second estimator of η is defined by $\sum_{j=1}^{k} a_j^{**} \delta_{t_j}$ where $(\vec{a})^{**}$ is the solution of the linear program

$$\begin{array}{ll} \text{minimize:} & \sum_{i=1}^{n} \left(v_i - \sum_{j=1}^{k} a_j |\langle t_j, h_i \rangle| \right), \\ \text{subject to:} & \sum_{j=1}^{k} a_j |\langle t_j, h_i \rangle| \leq v_i, \qquad i = 1, \dots, n, \\ & a_j \geq 0, \qquad j = 1, \dots, k. \end{array} \tag{8}$$

The solution of (8) is found using the simplex algorithm. Thus defining again $(\vec{a})^{LP} = \frac{(\vec{a})^{**}}{|(\vec{a})^{**}|}$ we get the second estimator (denoted by LP) $\sum_{j=1}^{k} a_j^{LP} \delta_{t_j}$ of $\mathcal{P}$.

4. METHOD

Our estimation procedure is different from those of Kiderlen [7] but since it makes use of the likelihood we have chosen the set T in (3) as the support of the discrete approximation $\mathcal{R}$ of $\mathcal{P}$ again. Note the important fact that T is completely determined by the set $\mathcal{H}$ and depends neither on the sample $\vec{v}$ nor on the distribution $\mathcal{P}$.

Now we proceed with the Bayesian estimation of the parameters $(a_1, \dots, a_k) = \vec{a}$, and Λ. Under our model the rose of intersections satisfies

$$V(h_i) = \Lambda \mathcal{F}_{\mathcal{R}}(h_i) = \Lambda \sum_{j=1}^{k} a_j |\langle h_i, t_j \rangle|, \quad i = 1, \dots, n, \quad \Lambda \in \mathbb{R}^+.$$

Under the assumption (ii) the components of the random vector $(V_1, \dots, V_n)$ are independently Poisson distributed with mean values $(\Lambda \mathcal{F}_{\mathcal{R}}(h_1), \dots, \Lambda \mathcal{F}_{\mathcal{R}}(h_n))$, hence the likelihood under our model is

$$L(\Lambda, a_1, \dots, a_k | \vec{v}) \propto \Lambda^{\sum_{i=1}^{n} v_i} \exp\{-\Lambda \sum_{i=1}^{n} \mathcal{F}_{\mathcal{R}}(h_i)\} \prod_{i=1}^{n} (\mathcal{F}_{\mathcal{R}}(h_i))^{v_i}. \tag{9}$$

We assume that there exists an upper bound $\Lambda_{\max}$ of possible values of the length intensity Λ (which is not a restrictive assumption in practical applications). We use independent noninformative priors for Λ and the vector $\vec{a}$. Let the distribution $\mathcal{L}(\Lambda)$ be uniform on $[0, \Lambda_{\max}]$ and let $\mathcal{L}(\vec{a})$ be uniform on the simplex

$$N = \{(a_1, \dots, a_k) \in \mathbb{R}^k : \sum_{j=1}^{k} a_j = 1, a_j \geq 0,\ j \in \{1, \dots, k\}\}.$$

Then the posterior distribution Π of $(\Lambda, a_1, \dots, a_k)$ given $\vec{v}$ has the density

$$\pi(\Lambda, a_1, \dots, a_k | \vec{v}) \propto \Lambda^{\sum_{i=1}^{n} v_i} \exp\{-\Lambda \sum_{i=1}^{n} \sum_{j=1}^{k} a_j |\langle h_i, t_j \rangle|\} \cdot \prod_{i=1}^{n} \left(\sum_{j=1}^{k} a_j |\langle h_i, t_j \rangle| \right)^{v_i} \tag{10}$$

on $[0, \Lambda_{\max}] \times N \equiv M$, with respect to $\mu \times \tilde{\mu}$, where μ is the Lebesgue measure on $[0, \Lambda_{\max}]$ and $\tilde{\mu}$ is the uniform measure on N, $\tilde{\mu}(N) = 1$. It holds $\pi(\Lambda, a_1, \dots, a_k | \vec{v}) = 0$ outside M.

We define the estimators for the parameters $\Lambda, a_1, \dots, a_k$ as the marginal posterior means and denote them by $\widehat{\Lambda}, \widehat{a_1}, \dots, \widehat{a_k}$. By substituting these values in the general formula (2) for $\mathcal{R}$ we get the desired estimator $\widehat{\mathcal{R}}$ for the directional distribution $\mathcal{P}$ of the line process Φ

$$\widehat{\mathcal{R}} = \sum_{j=1}^{k} \widehat{a_j} \delta_{t_j}. \tag{11}$$

Because of the complicated form of the density π, the mean posterior values are not accessible directly and have to be computed numerically. In the next section we give the detailed description of this procedure.

Even though the presented estimator was derived from the Poisson assumption it is applicable to the general stationary spatial processes. In this case consistency of the ML estimator was proved under mild assumptions in Kiderlen [7], Theorem 3. It is easy to see that for a compact parametric space (as is our M) from the consistency of the ML estimator the consistency of the Bayes estimator based on the posterior mean follows.

For the comparison of the different estimators we used the empirical mean Prohorov distance of the precise directional distribution and the estimator. Since the

Prohorov distance corresponds to the weak convergence on the set of measures $\mathcal{M}$ (Matheron [9]) it is a convenient metric for comparing the precision of the estimators. Another quantitative property is the variability of the estimators which is evaluated in various ways from the empirical covariance matrix of the estimated $\vec{a}$.

5. ALGORITHM

For the approximation of the posterior means of the parameters $\Lambda, a_1, \ldots, a_k$, we use the method of Markov chain Monte Carlo (MCMC), specifically the random walk Metropolis algorithm (Metropolis et al [12]). This method is generally used for the approximation of such mean values of various functions of a probability distribution Π on a state space $\mathbb{S}$ that cannot be computed analytically. The method proceeds as follows. A homogeneous Markov chain $X = \{X(t)\}_{t=1}^{T}$, $T \in \mathbb{N}$ on $\mathbb{S}$ with limiting distribution $\Pi(x)$ is constructed and the desired mean values $\mathrm{E}_\Pi f = \int f(x)\ \Pi(\mathrm{d}x)$ are then approximated by the ergodic averages $\bar{f} = \frac{1}{T}\sum_{t=1}^{T} f(X(t))$ of one realization of the chain. The transition probability kernel P of the chain X is constructed from an arbitrary transition probability kernel Q on $\mathbb{S}$. We suppose that Q has a density $q(x,y)$ with respect to some reference measure ν on the state space $\mathbb{S}$. Defining the acceptance probability

$$\alpha(x,y) = \begin{cases} \min\left\{\frac{\pi(y)}{\pi(x)}\frac{q(y,x)}{q(x,y)}, 1\right\}, & \pi(x)q(x,y) > 0, \quad x,y \in \mathbb{S}, \\ 1, & \pi(x)q(x,y) = 0, \quad x,y \in \mathbb{S}, \end{cases}$$

where $\pi(x)$ is the density of the limiting distribution Π with respect to ν, and

$$\begin{aligned} p(x,y) &= \alpha(x,y)q(x,y) \quad \text{for} \quad x \neq y, \quad x,y \in \mathbb{S}, \\ p(x,x) &= 0, \\ r(x) &= \int_{\mathbb{S}} (1-\alpha(x,y))q(x,y)\nu(\mathrm{d}y), \end{aligned}$$

and putting,

$$P(x,\mathrm{d}y) = p(x,y)\nu(\mathrm{d}y) + r(x)\delta_x(\mathrm{d}y) \qquad x,y \in \mathbb{S},$$

where δ_x is the unit mass at x; the resulting probability transition kernel P has the stationary distribution Π.

If the proposal density $q(x,y)$ depends only on the difference $(x-y)$, then the MCMC algorithm is called random walk Metropolis. For a general introduction into MCMC algorithms the reader is referred to Geyer [5].

The limiting distribution is the posterior distribution $\Pi(\Lambda, a_1, \ldots, a_k \,|\vec{v})$ and the state space $\mathbb{S} = \mathbb{R} \times \{\vec{x} \in \mathbb{R}^k : \sum_{i=1}^{k} x_i = 1\}$ in our case. For simplicity of notation, we write $\pi(\Lambda, \vec{a})$ instead of $\pi(\Lambda, a_1, \ldots, a_k \,|\vec{v})$ in (10).

Let E be the canonical base of $\mathbb{R}^{k-1}$, B an orthonormal base of the hyperplane $L = \{\vec{x} \in \mathbb{R}^k : \sum_{i=1}^{k} x_i = 1\}$ and $\beta : \mathbb{R}^{k-1} \to L$ a linear mapping satisfying $\beta(E) = B$. Define the reference measure ν by the k-dimensional Lebesgue measure on $\mathbb{R} \times L$. Let $\mathcal{L}_1$ be the one-dimensional centred normal distribution with the variance $\sigma_\Lambda^2 > 0$,

and $\mathcal{L}_2$ the $(k-1)$-dimensional centred normal distribution with the covariance matrix $\sigma_{\vec{a}}^2 I$, $\sigma_{\vec{a}}^2 > 0$ and denote by $f(z)$, $z \in \mathbb{S}$ the density with respect to ν of the probability distribution $\mathcal{L}_1 \times \mathcal{L}_2\ \beta^{-1}$ on $\mathbb{R} \times L$. The proposal density is defined by $q(x,y) = f(x-y), x, y \in \mathbb{S}$.

With the notation

$$C = \{C_{i,j} = |\langle h_i, t_j \rangle|\ \}_{i=1,j=1}^{n,k}, \qquad w = C\vec{a},$$

the acceptance probability is

$$\begin{aligned} \alpha((\Lambda,\vec{a}),(\Lambda',\vec{a}')) &= 1, && (\Lambda,\vec{a}) \notin M, \quad (12)\\ &= 0, && (\Lambda,\vec{a}) \in M,\ (\Lambda',\vec{a}') \notin M,\\ &= \min\left\{1, \left(\frac{\Lambda'}{\Lambda}\right)^{\sum v_i} \exp\{\Lambda \sum_{i=1}^{n} w_i - \Lambda' \sum_{i=1}^{n} w_i'\} \frac{\prod_{i=1}^{n} (w_i')^{v_i}}{\prod_{i=1}^{n} w_i^{v_i}}\right\}, \\ & && (\Lambda,\vec{a}),(\Lambda',\vec{a}') \in M. \end{aligned}$$

Since the set M is absorbing, it is sufficient for the simulation to consider the resulting Markov kernel P restricted only to the set $M \times \mathcal{B}(M)$ ($\mathcal{B}(M)$ denotes the Borel σ-algebra); let us denote it by $\tilde{P} = P|_{M \times \mathcal{B}(M)}$.

Note that if we had restricted to the set M already the proposal distribution Q then we would need to normalize the density $q(x,y) = \frac{f(x-y)}{\int_M f(x-z)\mathrm{d}z}$ which could be done again only numerically and, moreover, the density $q(x,y)$ would not be symmetric and we would get into further problems when computing the acceptance probability α. None of these problems occur with our choice, the algorithm is easy to implement and it has good properties as will be shown in the sequel.

Lemma 1. Let the Assumptions (i) and (iii) be fulfilled. Then the Markov kernel $\tilde{P}$ is aperiodic and irreducible.

Proof. The matrix C has no row with all zeros according to the assumption (i). It follows that $p(x,y) = q(x,y).\alpha(x,y) > 0$ for all $x, y \in \text{relint } M$ (here relint M means the interior of M in the space $\mathbb{S}$). If we denote by ψ the k-dimensional Hausdorff measure on M, $\tilde{P}(x,A) > 0$ will hold for all $A \in \mathcal{B}(M)$, $\psi(A) > 0$, $x \in M$, hence the kernel $\tilde{P}$ is ψ-irreducible and aperiodic. □

Theorem 1. Under the Assumptions (i) and (iii) the Markov kernel $\tilde{P}$ is uniformly ergodic.

Proof. According to the Theorem 16.2.2 in Meyn and Tweedie [13], it is enough to show that the whole set M is small. We need to find a probability measure ρ on $\mathcal{B}(M)$ and $\delta > 0$ such that

$$\tilde{P}(x,A) \geq \delta\rho(A), \qquad \forall\, x \in M,\ \forall A \in \mathcal{B}(M).$$

Trivially a sufficient condition for this is the existence of $l > 0$, $M_1 \subseteq M$, $\nu(M_1) > 0$ such that $p(x,y) \geq l$, $\forall\, y \in M_1$, $x \in M$.
Let us define

$$M_1 = [\Lambda_1, \Lambda_{\max}] \times \{a \in \mathbb{R}^k : \sum_{j=1}^{k} a_j = 1,\ a_j \geq d_{m_1}\ \forall j\},$$

with some fixed $0 < \Lambda_1 < \Lambda_{\max}$, $d_{m_1} > 0$ such that $\nu(M_1) > 0$. Then M_1 and M are compact sets and the proposal densities $\{q(x,y),\ x \in M\}$ are positive and continuous on M_1 which implies the existence of a lower bound $q_m > 0$ such that $q(x,y) \geq q_m$, $\forall\, x \in M$, $y \in M_1$.
The last step now is to bound from below the acceptance probability $\alpha(x,y)$. This can be done easily, as the following inequalities show

$$\pi(x) \leq (\Lambda_{\max})^{\sum_{i=1}^{n} v_i} . 1 . \prod_{i=1}^{n} \left(\sum_{j=1}^{k} C_{i,j} \right)^{v_i} = K, \qquad \forall\, x \in M,$$

$$\pi(y) \geq (\Lambda_1)^{\sum_{i=1}^{n} v_i} . \exp\left\{ -\Lambda_{\max} \sum_{i=1}^{n} \sum_{j=1}^{k} C_{i,j} \right\} . \prod_{i=1}^{n} d_{m_1} \left(\sum_{j=1}^{k} C_{i,j} \right)^{v_i} = \kappa > 0,$$

$$\forall\, y \in M_1.$$

The choice of $l = \frac{\kappa\, q_m}{K}$ completes the proof. □

From the uniform ergodicity of $\tilde{\mathcal{P}}$, the central limit theorem holds for ergodic averages of any square integrable function of the limiting distribution Π and any initial condition (any starting value of the realization of the chain X) (see Tierney [16]). Thus the ergodic averages converge to the estimated posterior means as $O(1/\sqrt{n})$.

6. SIMULATIONS

We present some simulation results in this section. The aim is to compare our estimator (11) denoted by MH with two estimators of Kiderlen [7]. First, let us present one particular example before a more detailed comparison will be made.

Let us recall here the definition of the Fisher distribution on $\mathcal{S}^2$. This distribution is determined by two parameters – the axis of symmetry $u \in \mathcal{S}^2$ and the concentration parameter $\kappa \in \mathbb{R}^+$. The density of the distribution is then defined as

$$f(x) = \frac{e^{\kappa |\langle x,u \rangle|}}{\int_{\mathcal{S}^2} e^{\kappa |\langle x,u \rangle|}\, \omega_2(\mathrm{d}x)}, \qquad x \in \mathcal{S}^2,$$

where ω_2 is the spherical Lebesgue measure on $\mathcal{S}^2$. We denote by $\mathcal{P}_1$ the Fisher distribution with concentration parameter $\kappa = 10$ and the axis $u_0 = [0.572, 0.572, 0.588]$.

In the example, we considered a stationary Poisson line process with length intensity $\Lambda = 100$ and the directional distribution $\mathcal{P}_1$. The set $\mathcal{H}$ were normals to the

faces of a regular icosahedron in the standard orientation (the intersection with the $x^{\perp}$ plane being the regular hexagon with one vertice on the z axis). Thus $n = 10$ and 10 independent realizations of intersection processes were evaluated in order to get the input data vector $\vec{v}$. The set T chosen according to (3) consists of $k = 45$ (pairs of) vectors. We determined the parameters $\widehat{\Lambda}, \widehat{a_1}, \ldots, \widehat{a_k}$ according to the MCMC algorithm described in the previous section. For the algorithm we set the starting iteration

$$a_i^{(0)} = \frac{1}{k}, \quad \text{for all } i \in \{1, \ldots, k\}, \tag{13}$$

$$\Lambda^{(0)} = 2\frac{\sum v_i}{n}. \tag{14}$$

$\Lambda^{(0)}$ is an unbiased estimator of the length intensity in the isotropic case due to the assumption
$H^2(A_i) = 1$, $i = 1, \ldots, n$, and $(\vec{a})^{(0)}$ corresponds to the uniform distribution on T. The variances of the proposal distributions were $\sigma_\Lambda^2 = 1$ and $\sigma_{\vec{a}}^2 = 0.001$.

The rate of the convergence of the algorithm depends on the variances of the proposal distributions radically. It is necessary to balance the mixing of the chain (the proposal steps must not be too small) and the average acceptance rate $\bar{\alpha}$ (the proportion of the proposed steps in the chain which were accepted according to (12) – bigger steps are more likely to be rejected). However, the problem of optimal tuning of the parameters exceeds the scope of this paper and the parameters were chosen as fairly reasonable even if not optimal (for references on optimal tuning see for example Tierney [16]) as shows also the value of $\bar{\alpha} = 0.322$ of our simulation.

We took 5000 iterations from the simulated Markov chain as the burn-in and then took 10 000 values with the step of 100 iterations

$$\widehat{a}_i = \frac{1}{10000} \sum_{t=1}^{10000} a_i^{(100t+5000)}, \qquad i = 1, \ldots, k. \tag{15}$$

Figure 1 shows the resulting MH estimator. In each point $\pm t_j$, $t_j \in T \subseteq \mathcal{S}^2$, a sphere is drawn with radius proportional to $\widehat{a_j}$. The axis of the estimated directional distribution is denoted by the triangle. We can see that the estimator detects the anisotropy of $\mathcal{P}_1$ well and it also reflects its symmetry.

We proceed with the comparison of the Kiderlen's estimators. We applied the estimators to two stationary Poisson line processes with the same length intensity $\Lambda = 100$ but different directional distributions. The first distribution was the distribution $\mathcal{P}_1$ already defined above and the second distribution $\mathcal{P}_2$ was a mixture of three Fisher distributions with the same concentration parameter $\kappa = 10$ and different axes

$u_1 = [0.572, 0.572, 0.588]$,
$u_2 = [-0.572, -0.572, 0.588]$,
$u_3 = [-0.588,\ 0, 0.801]$.

We considered three different types of input data corresponding to three different sets of $\mathcal{H}$: normals to the faces of a cube, a regular octahedron and a regular dodecahedron.

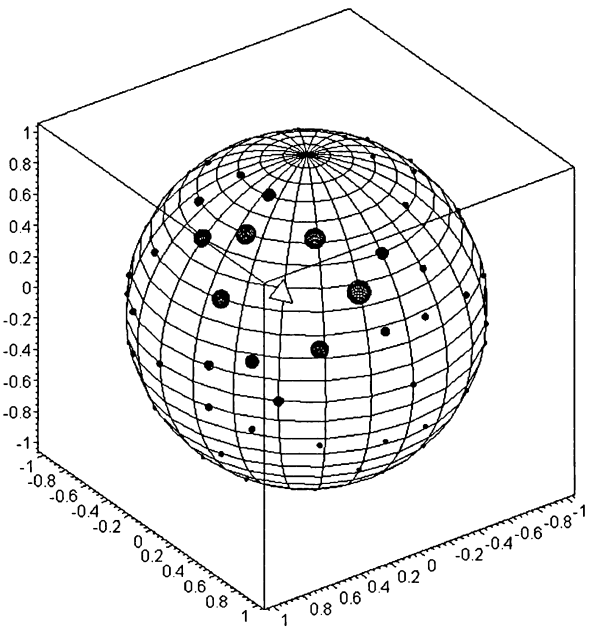

Fig. 1. The MH estimate of the rose of directions for the original directional distribution $\mathcal{P}_1$ being the Fisher distribution. The intersection planes were parallel to the faces of the regular icosahedron. The direction of the axis of the Fisher distribution is marked by the triangle.

The parameters $\widehat{a_1}, \ldots \widehat{a_k}$ of the MH estimate (11) were computed in the same way as in the previous example using the formula (16) with the starting iteration given by (14) and (15). The values of the proposal variances σ^2_Λ, $\sigma^2_{\vec{a}}$ used for the different sets $\mathcal{H}$ (the same for both $\mathcal{P}_1$ and $\mathcal{P}_2$) are given in Table 1 as well as the average acceptance rates $\bar{\alpha}$. The explanation concerning the choice of the parameters from the preceding example applies here as well.

For the comparison of the estimators we used the empirical mean Prohorov distances of the precise directional distribution and the estimators.

The Prohorov distance PD of two measures μ, ν on $\mathcal{B}(\mathcal{S}^2)$ is defined by

$$PD(\mu,\nu) \equiv \inf\{\epsilon > 0 \mid \mu(A) \leq \nu(A^\epsilon) + \epsilon, \nu(A) \leq \mu(A^\epsilon) + \epsilon, \text{ for all closed subsets } A \text{ of} \mathcal{S}^2\},$$

where A^ϵ is the set of all unit vectors whose spherical distance d from A is strictly less then ϵ.

Table 1. Parameter values for the MH estimator for different types of input data. The input data are the intersection counts with planes determined by the set $\mathcal{H}$ of their normal vector orientations. Here $\mathcal{H}$ are unit normal vectors to the faces of a cube, a regular octahedron and a regular dodecahedron, respectively. The parameters are: n – the number of different intersection planes (equivalently the number of elements in $\mathcal{H}$), k – number of different vectors in T (in the support of the MH estimator), σ^2_Λ, $\sigma^2_{\vec{a}}$ – variances of the proposal distributions, $\bar{\alpha}(\mathcal{P}_1)$, $\bar{\alpha}(\mathcal{P}_2)$ – the average acceptance rates when estimating the directional distribution $\mathcal{P}_1$, $\mathcal{P}_2$, respectively.

	σ^2_Λ	$\sigma^2_{\vec{a}}$	n	k	$\bar{\alpha}(\mathcal{P}_1)$	$\bar{\alpha}(\mathcal{P}_2)$
cube	1	0.05	3	3	0.50	0.52
octahedron	1	0.02	4	6	0.50	0.63
dodecahedron	1	0.01	6	15	0.20	0.34

For the evaluation of PD we modified the algorithm of Beneš and Gokhale [1]. The original algorithm was defined for measures on the set $Z = [0, \pi)$. However we can reformulate their Lemma 2 also for measures on $\mathcal{S}^2$, the proof being the same. Thus we have

Proposition 2. Let $\mathcal{R}_n$ be a discrete measure on $\mathcal{S}^2$ with a finite support $\operatorname{supp}\mathcal{R}_n = \{z_1, \ldots, z_n\}$ and $\mathcal{P}$ a measure on $\mathcal{S}^2$. Then it holds

$$PD(\mathcal{R}_n, \mathcal{P}) = \inf\{\epsilon > 0 \mid \mathcal{R}_n(A) \leq \mathcal{P}(A^\epsilon) + \epsilon \text{ for all } A \in \operatorname{supp}\mathcal{R}_n\}. \tag{16}$$

Thus we get a reduction to finitely many conditions, however to test if $PD < \epsilon$ for some $\epsilon > 0$ is still a problem of exponential complexity and efficient heuristics have to be applied to avoid testing (and computing $\int_{A^\epsilon} \mathcal{P}(\mathrm{d}x)$) for all the $(2^n - 1)$ subsets of $\operatorname{supp}\mathcal{R}_n$.

Table 2. The mean empirical Prohorov distances (computed from 200 samples) of the exact distributions $\mathcal{P}_1$ (a Fisher distribution) and $\mathcal{P}_2$ (a mixture of three Fisher distributions) and its MH, EM and LP estimators. The values are computed for three different sets of intersection planes (that means three different types of input data) – planes parallel with the faces of a cube, a regular octahedron and a regular dodecahedron.

	$\mathcal{P}_1$			$\mathcal{P}_2$		
	cube	octahedron	dodecahedron	cube	octahedron	dodecahedron
MH	0.5475	0.3977	0.3032	0.4975	0.4010	0.2992
EM	0.5475	0.3975	0.2850	0.4975	0.3997	0.3027
LP	0.5475	0.3975	0.2897	0.4975	0.3995	0.3030

Two hundred samples of the input data $\vec{v}$ were generated in our simulation for the roses of intersections $\mathcal{F}_{\mathcal{P}_1}$ and $\mathcal{F}_{\mathcal{P}_2}$ for all three choices of $\mathcal{H}$ and three estimators

were computed. The empirical Prohorov mean distances are given in Table 2. The variability of the estimators was quantified by means of traces, determinants and maximal eigenvalues of the empirical covariance matrices of the vector $\vec{a}$, se e Tables 3, 4 and 5 respectively.

Table 3. The traces of the empirical covariance matrix (computed from 200 samples) of the vector $\vec{a}$ of the estimated parameters of the directional distribution. The values are computed for all three examined estimators, three different sets of intersection planes $\mathcal{H}$ (faces of a cube a regular octahedron and a regular dodecahedron) and two different original roses of directions $\mathcal{P}_1$ (the Fisher distribution) and $\mathcal{P}_2$ (mixture of three different Fisher distributions).

	$\mathcal{P}_1$			$\mathcal{P}_2$		
	cube	octahedron	dodecahedron	cube	octahedron	dodecahedron
MH	0.0094	0.0363	0.0745	0.0115	0.0404	0.0663
EM	0.0102	0.0271	0.0952	0.0121	0.0372	0.1203
LP	0.0102	0.0225	0.1034	0.0121	0.0345	0.1119

Table 4. Decadic logarithms of determinants of the empirical covariance matrix (computed from 200 samples) of the vector $\vec{a}$ of the estimated parameters of the directional distribution. The values are computed for all three examined estimators, three different sets of intersection planes $\mathcal{H}$ (faces of a cube a regular octahedron and a regular dodecahedron) and two different original roses of directions $\mathcal{P}_1$ (the Fisher distribution) and $\mathcal{P}_2$ (mixture of three different Fisher distributions).

	$\mathcal{P}_1$			$\mathcal{P}_2$		
	cube	octahedron	dodecahedron	cube	octahedron	dodecahedron
MH	–25	–22	–58	–10	–17	–48
EM	–10	–19	–43	–9	–17	–43
LP	–10	–17	–40	–9	–33	–41

Many plots of the type shown in Figure 1 and the corresponding Prohorov distances are presented in Hlawiczkova et al [6], where fibre processes of tessellation edges in $\mathbb{R}^3$ are investigated for Voronoi tessellations generated by various point processes. Only the EM estimator is used for the rose of directions estimation in this paper.

7. CONCLUSIONS

It makes sense to compare estimators EM, LP and MH. Theoretically, maximum likelihood and Bayesian estimators are asymptotically equivalent but their small sample properties may differ. Bayesian estimators are always admissible (Wald [17]) which need not hold for the maximum likelihood estimators. The consistency of all estimators valid for the general stationary fibre processes is an important property for practical applications.

Table 5. Maximal eigenvalues of the empirical covariance matrix (computed from 200 samples) of the vector $\vec{a}$ of the estimated parameters of the directional distribution. The values are computed for all three examined estimators, three different sets of intersection planes $\mathcal{H}$ (faces of a cube a regular octahedron and a regular dodecahedron) and two different original roses of directions $\mathcal{P}_1$ (the Fisher distribution) and $\mathcal{P}_2$ (mixture of three different Fisher distributions).

	$\mathcal{P}_1$			$\mathcal{P}_2$		
	cube	octahedron	dodecahedron	cube	octahedron	dodecahedron
MH	0.0021	0.0173	0.0371	0.0062	0.0180	0.0181
EM	0.0060	0.0115	0.0304	0.0066	0.0163	0.0327
LP	0.0060	0.0830	0.0238	0.0066	0.0160	0.0293

Only minor differences between the estimators EM, LP and MH were revealed by our simulations. According to the Prohorov distances the estimators perform almost equally. However, the variability is smaller for MH estimator in the most interesting dodecahedral case. This effect corresponds to the general knowledge that Bayesian estimators may yield smaller variances.

ACKNOWLEDGEMENT

The research was supported by Grant A1075201 of the Grant Agency of the Academy of Sciences of the Czech Republic and by Grant 201/03/0946 of the Grant Agency of the Czech Republic. The author wishes to thank Prof. L. Klebanov (Charles University) for his helpful comments.

(Received January 24, 2003.)

REFERENCES

[1] V. Beneš and A. M. Gokhale: Planar anisotropy revisited. Kybernetika *36* (2000), 149–164.

[2] S. Campi, D. Haas, and W. Weil: Approximaton of zonoids by zonotopes in fixed directions. Discrete Comput. Geom. *11* (1994), 419–431.

[3] L. M. Cruz-Orive, H. Hoppeler, O. Mathieu, and E. R. Weibel: Stereological analysis of anisotropic structures using directional statistics. Appl. Statist. *34* (1985), 14–32.

[4] J. E. Hilliard: Specification and measurement of microstructural anisotropy. Trans. Metall. Soc. AIME *224* (1962), 1201–1211.

[5] C. J. Geyer: Practical Markov chain Monte Carlo (with discussion). Statist. Sci. *7* (1992), 473–511.

[6] M. Hlawiczková, P. Ponížil, and I. Saxl: Estimating 3D fibre process anisotropy. In: Topics in Applied and Theoretical Mathematics and Computer Science (V. V. Kluev and N. E. Mastorakis), WSEAS Press, WSEAS, 2001, 214–219.

[7] M. Kiderlen: Non-parametric estimation of the directional distribution of stationary line and fibre processes. Adv. Appl. Prob. *33* (2001), 6–24.

[8] B. A. Mair, M. Rao, and J. M. M. Anderson: Positron emission tomography, Borel measures and weak convergence. Inverse Problems *12* (1996), 965–976.

[9] G. Matheron: Random Sets and Integral Geometry. Wiley, New York 1975.
[10] G. J. McLachlan and T. Krishnan: The EM Algorithm and Extensions. Wiley, New York 1997.
[11] J. Mecke and W. Nagel: Stationäre räumliche Faserprozesse und ihre Schnittzahlrosen. Elektron. Informationsverarb. Kybernet. *16* (1980), 475–483.
[12] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller: Equation of state calculations by fast computing machines. J. Chem. Phys. *21* (1953), 1087–1091.
[13] S. Meyn and R. Tweedie: Markov Chains and Stochastic Stability. Springer, London 1993.
[14] R. Schneider: Convex Bodies: The Brunn-Minkowski Theory. Encyclopedia Math. Appl. *44* (1993).
[15] D. Stoyan, W. S. Kendall, and J. Mecke: Stochastic Geometry and its Applications. Second edition. Wiley, New York 1995.
[16] L. Tierney: Markov chains for exploring posterior distributions. Ann. Statist. *22* (1994), 1701–1762.
[17] A. Wald: Statistical Decision Functions. Wiley, New York 1950.

Mgr. Michaela Prokešová, Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 186 75 Praha 8. Czech Republic.
e-mail: prokesov@karlin.mff.cuni.cz