Mathematica Balkanica

Mathematical Society of South-Eastern Europe
A quarterly published by
the Bulgarian Academy of Sciences – National Committee for Mathematics

Mathematica
Balkanica

New Series Vol. 15, 2001, Fasc. 3-4

Maximum Likelihood Estimation of k–Dimensional Diffusion–Discrete Model

Dimitrinka I. Vladeva

Presented by V. Kiryakova

In this article we consider a k-dimensional diffusion process with constant drift and diffusion parameters: unknown vector A and positive definite matrix B, respectively. We suppose that the discrete moments of observations are a point process, independent on the considered diffusion process. The maximum likelihood estimates for the unknown parameters A and B are found.

AMS Subj. Classification: 62H12, 62M05

Key Words: diffusion process, k-dimensional Wiener process, maximum likelihood estimation, discrete random sampling

1. Introduction

In this article we consider the diffusion process $X_t = (X_t^1, X_t^2, ..., X_t^k)'$, $t \geq 0$, defined by the stochastic differential equation

$$dX_t = Adt + B^{\frac{1}{2}} dW_t \,, \; t \geq 0 \,, \; X_0 = 0, \tag{1}$$

where $A = (a^1, a^2, ..., a^k)'$ and

$$B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ \cdot & \cdot & \cdots & \cdot \\ b_{k1} & b_{k2} & \cdots & b_{kk} \end{pmatrix}$$

are unknown constant vector and positive definite symmetric matrix, respectively, and $W_t = (W_t^1, W_t^2, ..., W_t^k)'$, is a standard Wiener process with mean 0 and variance the identity matrix I_k.

The solution of the differential equation (1) exists in a strong sense, it is unique and it is the process

$$X_t = At + B^{\frac{1}{2}}W_t\,,\, t \geq 0\,. \tag{2}$$

More theoretical statements can be found in [1].

The maximum likelihood estimation problem for the model (1), when we observe the process X_t, $t \geq 0$ continuously in the interval $[0,T]$, is solved and for one-dimensional case the solution can be seen in [2]. In the case when we have in disposal the non-random discrete observations A. Le Breton has solved the same problem in [3].

In the recent years many authors (see [4] and [5]) consider continuous diffusion processes, when the observations are provided in discrete moments, belonging to the interval $[0,T]$.

At first, the random sampling scheme have been used by J. Beutler in [6]. We use this sampling scheme, which is natural for some practical problems. Let us dispose the observations $x_{t_1},\ldots,x_{t_N}$, where $x_{t_i} = (x^1_{t_i}, x^2_{t_i},\ldots,x^k_{t_i})'$ and $t_1,\ldots,t_N$. The moments $t_1,\ldots,t_N$ are the first N points of an arbitrary point process $\{T_i\}$, $i = 1,\ldots,N$ with independent identically distributed increments. The process $\{T_i\}$, $i = 1,\ldots,N$ is independent of the process X_t, $t \geq 0$ and we compute $E(F(X_{T_i})) = E_{T_i}(E_X(F(X_{T_i} \,/\, T_i = t_i)))$. The problem is to find the maximum likelihood estimates of the unknown constant vector A and the matrix B and to prove their properties. In the one-dimensional case this problem is solved in [7].

Using the maximum likelihood method we prove the following results.

2. Maximum likelihood estimation

We denote $x_i = x_{t_i}$, $\Delta x_i = x_i - x_{i-1}$, $\Delta_i = t_i - t_{i-1}$, $i = 1,\ldots,N$, $B_1 = B^{\frac{1}{2}}$.

Theorem 1. *If $N \geq 2$, the statistics*

$$\hat{A}_N = \frac{x_N}{t_N} \tag{3}$$

is a maximum likelihood estimate for the unknown vector A.

We prove this theorem using the standard maximum likelihood procedure.

P r o o f. The statistical structure for estimation, based on given observations is:

$$\left\{C,\, L\,, P_{A,B}, A \in\, L_1(R^k), B \in\, L_2(R^k)\right\},$$

where:

C is the set of continuous functions from $R^+ \times R^1$ to $R^k \times R^1$;

L is the σ-algebra of subsets of C generated by the family of evaluation functionals on C;

$L_1(R^k)$ is the set of all k - dimentional vectors;

$L_2(R^k)$ is the set of all $k \times k$ positive definite symmetric matrices.

For all $(A, B) \in\, L_1(R^k) \times\, L_2(R^k)$, $P_{A,B}$ denotes the measure on (C, L) induced every process which is a solution of (1). So $P^N_{A,B}$ is the joint distribution of $N.k + N$-dimensional random vector $(X_1, \ldots, X_N, T_1, \ldots, T_N)$, and λ be the Lebesque measure of the same dimension. Here $X_i = X_{T_i} = \left(X^1_{T_i}, \ldots, X^k_{T_i}\right)'$, $i = 1, \ldots, N$. Then having in mind that the process $\{T_i\}^N_{i=1}$ is independent on X_t, $t \geq 0$ and the increments $\Delta X_i, i = 1, \ldots, N$ of the process X_t are independent and conditional Gaussian distributed with mean $A\Delta_i$ and variance $B\Delta_i$, we compute the likelihood function

$$f_{X_1,\ldots,X_N,T_1,\ldots,T_N}(x_1, \ldots, x_N, t_1, \ldots, t_N) = \frac{dP^N_{A,B}}{d\lambda}$$

$$= \prod_{i=1}^{N}((2\pi)^k|B|\Delta_i)^{-\frac{1}{2}}\, g(\Delta_i)\,.\, \exp\left\{-\sum_{i=1}^{N}\frac{(\Delta x_i - A\Delta_i)'B^{-1}(\Delta x_i - A\Delta_i)}{2\Delta_i}\right\},$$

where $g(\Delta_i)$ is the density of $T_i - T_{i-1}$ and $\Delta_i = t_i - t_{i-1}$, $i = 1, \ldots, N$.

We apply the maximum likelihood method and find the maximum of $l(A, B) = \ln f(A, B)$.

Let c_{ij}, $i, j = 1, \ldots, k$ be the elements of the matrix B^{-1}.

We compute:

$$\frac{\partial l(A, B)}{\partial a_m} = -2\frac{\partial}{\partial a_m}\sum_{i=1}^{N}\sum_{j=1, j\neq m}^{k} c_{jm}\frac{(\Delta x^j_i - a_j\Delta_i)(\Delta x^m_i - a_m\Delta_i)}{2\Delta_i}$$

$$-\frac{\partial}{\partial a_m}\sum_{i=1}^{N} c_{mm}\frac{(\Delta x^m_i - a_m\Delta_i)^2}{2\Delta_i}$$

$$= \sum_{i=1}^{N}\sum_{j=1, j\neq m}^{k} c_{jm}(\Delta x^j_i - a_j\Delta_i) + \sum_{i=1}^{N} c_{mm}(\Delta x^m_i - a_m\Delta_i)$$

$$= \sum_{j=1}^{k} c_{jm}(x_N^j - a_j t_N)$$

for all $m = 1, \dots, k$.

For $a_j = \frac{x_N^j}{t_N}$, $j = 1, \dots, k$, the equalities $\frac{\partial l(A,B)}{\partial a_m} = 0$, are true, where $m = 1, \dots, k$ are satisfied.

For the second partial derivatives we find:

$$\frac{\partial^2 l(A,B)}{\partial a_m^2} = -t_N c_{mm} < 0, \quad \frac{\partial^2 l(A,B)}{\partial a_m \partial a_n} = -t_N c_{mn}, \ \forall n \neq m.$$

The matrix of the second partial derivatives is:

$$J = \begin{pmatrix} -t_N c_{11} & -t_N c_{12} & \cdots & -t_N c_{1k} \\ \cdot & \cdot & \cdots & \cdot \\ -t_N c_{k1} & -t_N c_{k2} & \cdots & -t_N c_{kk} \end{pmatrix} = -B^{-1} t_N.$$

The matrix B^{-1} is positive definite hence the matrix J is negative definite.

Thus we establish that the statistics (3) is the maximum likelihood estimate for the unknoun vector A.

Theorem 2. *If $N > k$, the statistics*

$$\hat{B}_N = \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{\Delta x_i \Delta x_i'}{\Delta_i} - \frac{x_N x_N'}{t_N} \right\} \tag{4}$$

is the maximum likelihood estimate for the unknown matrix B, when $A \neq 0$.

For $A = 0$ the maximum likelihood estimate is

$$\hat{B}_N = \frac{1}{N} \sum_{i=1}^{N} \frac{\Delta x_i \Delta x_i'}{\Delta_i}.$$

The approach is different from others used in the proofs of similar propositions. For example, see [8, p.75]. Our proof is based on the following lemma.

Lemma 1. *Let* $y_i = (y_i^1, \dots, y_i^k)'$, $i = 1, \dots, N$, $N > k$, *be* k-*dimensional vectors such that* $B = \sum_{i=1}^{N} y_i y_i'$ *is a non-singular matrix. Then:*

a) the matrix B *is symmetric and positive definite;*

b) $C = \sum_{i=1}^{N} y_i' \left(\sum_{j=1}^{N} y_j y_j' \right)^{-1} y_i = k.$

Proof. a) We establish easily that the matrix B is symmetric.

$$B = \sum_{i=1}^{N} y_i y_i' = \sum_{i=1}^{N} \begin{pmatrix} (y_i^1)^2 & y_i^1 y_i^2 & \cdots & y_i^1 y_i^k \\ \cdot & \cdot & \cdots & \cdot \\ y_i^k y_i^1 & y_i^k y_i^2 & \cdots & (y_i^k)^2 \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^{N} (y_i^1)^2 & \sum_{i=1}^{N} y_i^1 y_i^2 & \cdots & \sum_{i=1}^{N} y_i^1 y_i^k \\ \cdot & \cdot & \cdots & \cdot \\ \sum_{i=1}^{N} y_i^k y_i^1 & \sum_{i=1}^{N} y_i^k y_i^2 & \cdots & \sum_{i=1}^{N} (y_i^k)^2 \end{pmatrix}.$$

Let $Z = (z^1, \dots, z^k)'$ be an arbitrary k-dimensional vector. Then

$$Z'BZ = \sum_{l=1}^{k} \sum_{m=1}^{k} b_{lm} z^l z^m,$$

where $b_{lm} = \sum_{i=1}^{N} y_i^l y_i^m$. Hence

$$Z'BZ = \sum_{i=1}^{N} \sum_{l=1}^{k} \sum_{m=1}^{k} (y_i^l z^l)(y_i^m z^m).$$

We consider the vectors $c_i = (y_i^1 z^1, \dots, y_i^k z^k)$, $i = 1, \dots, N$ and write:

$$Z'BZ = \sum_{i=1}^{N} \sum_{l=1}^{k} \sum_{m=1}^{k} c_i^l c_i^m = \sum_{i=1}^{N} (c_i^1 + \cdots + c_i^k)^2 \geq 0.$$

For an arbitrary vector Z the vectors $y_1, \ldots, y_N, Z$ are linearly dependent. It means that the dot products $y_i.Z$ can not be equal to zero, for all $i = 1,, \ldots, N$. Hence it is imposible to have $c_i^1 + \cdots + c_i^k = 0$, for all $i = 1,, \ldots, N$. So, the matrix B is positive definite.

b) Let B^{-1} be the inverse matrix of B and B_{lm} be the algebraic cofactor of the element b_{lm}, $l, m = 1, \ldots, k$,

$$B^{-1} = \frac{1}{|B|} \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1k} \\ \cdot & \cdot & \cdots & \cdot \\ B_{k1} & B_{k2} & \cdots & B_{kk} \end{pmatrix}.$$

Then

$$C = \sum_{i=1}^{N} y_i' \frac{1}{|B|} \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1k} \\ \cdot & \cdot & \cdots & \cdot \\ B_{k1} & B_{k2} & \cdots & B_{kk} \end{pmatrix} y_i$$

$$= \frac{1}{|B|} \sum_{i=1}^{N} \left(\sum_{j=1}^{k} y_i^j B_{1j} \, , \sum_{j=1}^{k} y_i^j B_{2j} \, , \cdots , \sum_{j=1}^{k} y_i^j B_{kj} \right) (y_i^1, \cdots, y_i^k)'$$

$$= \frac{1}{|B|} \sum_{i=1}^{N} \left(\sum_{j=1}^{k} y_i^j B_{1j} y_i^1 + \sum_{j=1}^{k} y_i^j B_{2j} y_i^2 + \cdots + \sum_{j=1}^{k} y_i^j B_{kj} y_i^k \right)$$

$$= \frac{1}{|B|} \sum_{j=1}^{k} \left(B_{1j} \sum_{i=1}^{N} y_i^j y_i^1 + B_{2j} \sum_{i=1}^{N} y_i^j y_i^2 + \cdots + B_{kj} \sum_{i=1}^{N} y_i^j y_i^k \right)$$

$$= \frac{1}{|B|} \sum_{j=1}^{k} (B_{1j} b_{1j} + B_{2j} b_{2j} + \cdots + B_{kj} b_{kj}) = \frac{1}{|B|} k |B| = k.$$

The result can be written in the form:

$$\sum_{i=1}^{N} y_i' \left(\frac{1}{N} \sum_{j=1}^{N} y_j y_j' \right)^{-1} y_i = Nk. \tag{5}$$

The proof of Lemma 1 can be written shortly using the properties of a trace of a matrix.

P r o o f o f T h e o r e m 2. Let $\lambda_i > 0$, $i = 1, \ldots, k$ be the eigenvalues of the symmetric positive definite matrix B, $\dfrac{1}{\lambda_i} > 0$, $i = 1, \ldots, k$ be the eigenvalues of the inverse matrix B^{-1} and $|B| = \lambda_1.\lambda_2 \ldots \lambda_k$ be the determinant of B.

We will find a symmetric and positive definite matrix $\hat{B}_N$, which maximizes the function $l(A, B) = \ln f(A, B)$. We consider $l(A, B)$ as a function of $\lambda_1, \lambda_2, \ldots, \lambda_k$. Let $E_m, m = 1, \ldots, k$ be the $k \times k$ matrices, whose elements $e_{i,j}$ are equal to zero for all $(i, j) \neq (m, m)$ and $e_{m,m} = 1$. Then there is a orthogonal matrix S, such that $B^{-1} = S \sum_{j=1}^{k} E_j \frac{1}{\lambda_j} S'$.

After substituting $Y_i = \Delta x_i - A\Delta_i$, $i = 1, \ldots, N$, we compute the first partial derivatives of l, considered as a function of $\lambda_j, j = 1, \ldots, k$:

$$\frac{\partial l}{\partial \lambda_j} = \frac{\partial}{\partial \lambda_j} \ln((2\pi)^k \lambda_1.\lambda_2 \ldots \lambda_k)^{-\frac{N}{2}} - \sum_{i=1}^{N} \frac{\partial}{\partial \lambda_j} \left(\frac{Y_i' S \sum_{l=1}^{k} E_l \frac{1}{\lambda_l} S' Y_i}{2\Delta_i} + \ln g(\Delta_i) \right)$$

$$= -\frac{1}{2\lambda_j} \left(N - \frac{1}{\lambda_j} \sum_{i=1}^{N} \frac{Y_i' S E_j S' Y_i}{\Delta_i} \right) = 0, \quad \forall\, j = 1, \ldots, k.$$

After summation of the above k equalities we obtain:

$$kN = \sum_{i=1}^{N} \frac{Y_i' B^{-1} Y_i}{\Delta_i}.$$

Applying Lemma 1 with $y_i = \frac{Y_i}{\sqrt{\Delta_i}}$, $i = 1, \ldots, k$ we see, that $B = \frac{1}{N} \sum_{i=1}^{N} \frac{Y_i Y_i'}{\Delta_i}$ satisfies this equality.

Let us suppose, that a positive definite and symmetric matrix D_N exists, for which the likelihood function atteins it's maximum. Then it is true that

$$kN = \sum_{i=1}^{N} \frac{Y_i' D_N^{-1} Y_i}{\Delta_i}.$$

Let λ_i^D be the eigenvalues of the matrix D_N, and let λ_i^B be the eigenvalues of the matrix $\hat{B}_N$. These eigenvalues are determined from the maximum likelihood equation and hence $\lambda_i^D = \lambda_i^B$. Then for the likelihood function we obtain:

$$f\left(\hat{B}_N\right) = \prod_{i=1}^{N} \frac{1}{\sqrt{\pi|\hat{B}_N|}} \exp\left\{-\frac{Y_i' \hat{B}_N^{-1} Y_i}{\Delta_i}\right\},$$

$$f(D_N) = \prod_{i=1}^{N} \frac{1}{\sqrt{\pi|D_N|}} \exp\left\{-\frac{Y_i' D_N^{-1} Y_i}{\Delta_i}\right\}.$$

So $|D_N| = \lambda_1^D \dots \lambda_k^D = \lambda_1^B \dots \lambda_k^B = |\hat{B}_N|$ and

$$\sum_{i=1}^{N} \frac{Y_i'\, D_N^{-1}\, Y_i}{\Delta_i} = \sum_{i=1}^{N} \frac{Y_i'\, \hat{B}_N^{-1}\, Y_i}{\Delta_i} = kN.$$

Hence $L(D_N) = L\left(\hat{B}_N\right)$.

The matrix $\hat{B}_N$ is determined to within similarity matrix. But there is not a different similar matrix for which the equality b) from Lemma 1 is satisfied for the same vectors y_i.

The obtained extremum is maximum, because

$$\frac{\partial^2 l}{\partial \lambda_j^2} = \frac{1}{2\lambda_j^2}\left(N - \frac{1}{\lambda_j}\sum_{i=1}^{N} \frac{Y_i'\, S\, E_j\, S'\, Y_i}{\Delta_i}\right) - \frac{1}{2\lambda_j^3}\sum_{i=1}^{N} \frac{Y_i'\, S\, E_j\, S'\, Y_i}{\Delta_i}$$

$$= -\frac{1}{2\lambda_j^3}\sum_{i=1}^{N} \frac{Y_i'\, S\, E_j\, S'\, Y_i}{\Delta_i} < 0, \quad \forall\, i = 1, \dots, k.$$

Substituting $Y_i = \Delta x_i - A\Delta_i$ and $\hat{A}_N = \frac{x_N}{t_N}$, we derive for $\hat{B}_N$:

$$\hat{B}_N = \frac{1}{N}\sum_{i=1}^{N} \frac{(\Delta x_i - A\Delta_i)(\Delta x_i - A\Delta_i)'}{\Delta_i}$$

$$= \frac{1}{N}\sum_{i=1}^{N} \frac{\Delta x_i \Delta x_i'}{\Delta_i} - \frac{1}{N}\sum_{i=1}^{N} \frac{\Delta x_i A' \Delta_i}{\Delta_i} - \frac{1}{N}\sum_{i=1}^{N} \frac{A \Delta x_i' \Delta_i}{\Delta_i} + \frac{1}{N}\sum_{i=1}^{N} \frac{A\, A' \Delta_i^2}{\Delta_i}$$

$$= \frac{1}{N}\sum_{i=1}^{N} \frac{\Delta x_i \Delta x_i'}{\Delta_i} - \frac{1}{N}\frac{x_N\, x_N'}{t_N} - \frac{1}{N}\frac{x_N\, x_N'}{t_N} + \frac{1}{N}\frac{x_N\, x_N' t_N}{t_N^2}$$

$$= \frac{1}{N}\left(\sum_{i=1}^{N} \frac{\Delta x_i \Delta x_i'}{\Delta_i} - \frac{x_N\, x_N'}{t_N}\right).$$

Hence $\hat{B}_N$ is maximum likelihood estimate for unknown matrix B.

4. Comments

We would like to highlight the following facts.

1. From the proofs of the Theorem 1 and Theorem 2 it follows that to estimate one of the parameters (A or B), it is not necessary to know the other.

2. The estimator $\hat{A}_N$ depends only on the last observation, the same as a continuous time sampling, given in [2] and in the case when equidistant moments of observation are used. It is interesting to compare the estimations given by different sampling schemes: the point process can be Poisson, geometric and uniform (results of this kind can be seen in [9] and [10]).

3. The used sampling scheme is natural. We add the $(N+1)$-th observation to the first N observations and do not need a new $N+1$ observation. We prove good properties of the estimations without the condition $\max\limits_{1\leq i\leq N} \Delta_i \to 0$ when $N \to \infty$, as in other sampling schemes.

4. The estimation of B given by (4) is not unbiased. The unbiased estimation is

$$\widetilde{B}_N = \frac{1}{N-1}\sum_{i=1}^{N}\left\{\frac{\Delta x_i \Delta x_i'}{\Delta_i} - \frac{x_N x_N'}{t_N}\right\}.$$

We can compute the variance of $\widetilde{B}_N$ and it is equal to $\dfrac{k+1}{(N-1)}B^2$. Obviously, it is independent on the distribution of the random point process $T_1, T_2, \ldots T_N, \ldots$ and tends to zero as $O(N^{-1})$, by $N \to \infty$.

The same result for $k = 1$ is given in [7], i.e. the obtained results generalize the one-dimensional case.

5. The our algebraic approach in proof of Theorem 2 can be used for a new proof concerning the maximum likelihood estimates for a multivariate normal distribution.

References

[1] I. G i c h m a n, A. S k o r o c h o d, *Stochastic Differential Equations*, Naukova Dumka, Kiev (1982).

[2] R. L i p t z e r, A. S h i r y a y e v, *Statistics of Random Processes*, Nauka, Moskow (1974).

[3] A. L e B r e t o n, *Estimation des Parameters d'un Systemme Bilineaire Autonome*, Rapport Rech. **166**, Univ. Grenoble (1979).

[4] G e n o n - C a t a l o t, J. J a c o d, Estimation of the diffusion coefficient for diffusion processes: Random sampling, *Scand. J. of Stat.*, **21**, No 2 (1994), 193-221.

[5] M. K e s s l e r, Estimation of an ergodic diffusion from discrete observations, *Scand. J. of Stat.*, **24**, No 2 (1997), 211-230.

[6] F. B e u t l e r, Alias free randomly time sampling of stochastic processes, *IEEE Trans. Inform. Theory*, **IT - 16** (1970), 147-154.

[7] J. S t o y a n o v, D. V l a d e v a, Estimation of unknown parameters of continuous time stochastic process by observations at random moments, *Compt. Rend. Bulg. Acad. Sci.*, **35**, No 2 (1982), 153-156.

[8] N. G i r i, *Multivariable Statistical Inference*, Acad. Press, Vol. 4 (1977).

[9] D. V l a d e v a, Estimation of unknown parameters in a diffusion - Geometric model, *Math. and Educ. in Math. (Proc. of 13-th Spring Conf. of UBM, Sunny Beach)* (1984), 265-271.

[10] D. V l a d e v a, Parameter estimation of k-dimensional Wiener process by random observations, *Summer School on Appl. Math., Tech. Univ. of Sofia, Sozopol'1999* (1999), 154-156.

Dept. of Mathematics and Informatics
Higher Transport Engineering School "T. Kableshkov"
kv. "Slatina"
Sofia - 1574, BULGARIA

Received: 25.01.2000