KYBERNETIKA — VOLUME 31 (1995), NUMBER 5, PAGES 489–507

EXPONENTIAL RATE OF CONVERGENCE OF MAXIMUM LIKELIHOOD ESTIMATORS FOR INHOMOGENEOUS WIENER PROCESSES

FRIEDRICH LIESE AND ANDREAS WIENKE

The distribution of an inhomogeneous Wiener process is determined by the mean function $m(t) = E\,W(t)$ and the variance function $b(t) = V(W(t))$ which depend on unknown parameter $\vartheta \in \Theta$. Observations are assumed to be in discrete time points where the sample size tends to infinity. Using the general theory of Ibragimow, Hasminskij [2] sufficient conditions for consistency of MLE $\hat{\vartheta}_n$ are established. Exponential bounds for $P(|\hat{\vartheta}_n - \vartheta| > \varepsilon)$ are given and applied to prove strong consistency of $\hat{\vartheta}_n$.

INTRODUCTION

Inhomogeneous Wiener processes with mean function $m(t)$ and variance function $b(t)$ are continuous stochastic processes $W(t)$, $0 \le t \le T$, with independent increments so that $W(t) - W(s)$ is normally distributed with mean $m(t) - m(s)$ and variance $b(t) - b(s)$ and $W(0) = 0$. The functions $b(t)$ and $m(t)$ are continuous and supposed to be known up to the unknown parameter $\vartheta \in \Theta$. $b(t)$ is clearly a nondecreasing function.

Such processes are of interest both of the theoretical and practical point of view. In applications an inhomogeneous Wiener process is often used as a model of the fret of a technical system. It is supposed that the fret in time interval (s,t) is normally distributed and the state of the technical system at time t is the superposition of the independent frets in disjoint time intervals. Let P_ϑ denote the distribution of the process W. P_ϑ is defined on the Borel sets of $C[0,T]$, the space of continuous functions on $[0,T]$. If b not depends on ϑ, then under weak conditions on $m(t,\vartheta)$ it holds $P_\vartheta \sim P_\eta$, $\vartheta, \eta \in \Theta$ and it's possible to evaluate explicitly the density function of P_ϑ with respect to P_η [3]. If $b(t)$ is depending on ϑ then $P_\vartheta \perp P_\eta$ for $\vartheta \neq \eta$. Therefore there is no natural way to develop a reasonable likelihood theory and we suppose that the process is observed at time $t_i \in [0,T]$, where $0 = t_0 \le t_1 \le \ldots \le t_n = T$. Instead of the values $W(t_i)$ we may deal with the increments $W(t_i) - W(t_{i-1})$ which are normally distributed with mean $m(t_i,\vartheta) - m(t_{i-1},\vartheta)$ and variance $b(t_i,\vartheta) - b(t_{i-1},\vartheta)$, respectively. We are interested in asymptotic results. Therefore we suppose that a sequence $0 = t_{0,n} < t_{1,n} < \ldots < t_{k_n,n} = T$ of partitions of $[0, T]$ is given. In difference to Jacod [3] we suppose, that the decompositions $Z_n := \{t_{0,n},\ t_{1,n},\ \dots\ t_{k_n,n}\}$ of the interval $[0, T]$ are nonrandom. Furthermore we suppose that m_n independent replications of the process W are available. The corresponding differences $W_j(t_{i,n}) - W_j(t_{i-1,n})$ then form a double array of random variables. In our asymptotic considerations we do not nessesarily assume, in general, that both k_n and m_n tend to infinity. To establish bounds for the probability $P(\|\tilde{\vartheta}_n - \vartheta\| > \varepsilon)$ where $\tilde{\vartheta}_n$ is the Maximum Likelihood Estimation (MLE), we apply the technique from Ibragimov – Hasminskij [2] which was developed for arbitrary sequences of statistical models. The conditions formulated there imply Hellinger integrals and Hellinger distances of the distributions from the model. To prepare the main results we therefore collect in the first part estimates of product measures whose components are normal distributions. Using these inequalities which include regularity conditions on $m(t, \vartheta)$ and $b(t, \vartheta)$ we are able to apply the general results from [2] to the distribution $\mathcal{L}\big(W_j(t_{i,n}) - W_j(t_{i-1,n})\big)$, $(1 \le i \le k_n,\ j = 1, \dots, m_n)$. Our main results are the Theorems 2 and 3 which include exponential bounds for $P(\|\tilde{\vartheta}_n - \vartheta\| > \varepsilon)$. If the function b depends on ϑ then the upper bound established in these theorems tends to zero with exponentially rate as $k_n \to \infty$. The number of independent replications need not tend to infinity, $m_n = 1$ is enough. This result corresponds to the fact $P_\vartheta \perp P_\eta$ for $\vartheta \neq \eta$ if $b(t, \vartheta)$ depends on ϑ. Conversely, if $b(t, \vartheta)$ is independent of ϑ then we need an increasing number of replications to guarantee consistency of the MLE and to derive exponential bounds for $P(\|\tilde{\vartheta}_n - \vartheta\| > \varepsilon)$. In the general mixed situation both k_n and m_n tend to infinity. Our bounds for the probability $P(\|\tilde{\vartheta}_n - \vartheta\| > \varepsilon)$ include this general situation.

In the last part of the paper we apply the exponential bounds to establish the strong consistency of the MLE $\tilde{\vartheta}_n$.

If the consistency of the MLE is clear, the asymptotical normality of the MLE in the present model can be derived with the methods of [5].

1. HELLINGER INTEGRALS OF NORMAL DISTRIBUTIONS

General conditions for consistency of MLE were established in [2]. The conditions there are formulated in terms of Hellinger integrals. As an auxiliary result we firstly establish inequalities for Hellinger integrals of special k-dimensional normal distributions. For fixed n let U_n be a subset of $\mathbb{R}^k$ and $\mu_{ui}, \sigma_{u,i}$ $(i = 1, \dots, n;\ u \in U_n)$ functions with values in $\mathbb{R}^1$ and $(0, \infty)$, respectively. These functions are used as mean and variance of one dimensional normal distributions. We need the following assumptions:

(V1) $$\Big[\sup_{1 \le i \le n, u \in U_n} \sigma_{ui}^2\Big]\Big[\inf_{1 \le i \le n, u \in U_n} \sigma_{ui}^2\Big]^{-1} \le D_1$$

(V2) $$\sum_{i=1}^{n} \left|\frac{\sigma_{ui}^2}{\sigma_{vi}^2} - 1\right|^2 \le D_2 \|u - v\|^2 \quad \forall\, u, v \in U_n$$

(V3) $$\sum_{i=1}^{n} |\frac{\sigma_{ui}^2}{\sigma_{vi}^2} - 1|^2 \geq D_3 ||u - v||^2 \quad \forall\, u,\, v \in U_n$$

(V4) $$\sum_{i=1}^{n} \frac{(\mu_{ui} - \mu_{vi})^2}{\sigma_{ui}^2 + \sigma_{vi}^2} \leq D_4 ||u - v||^2 \quad \forall\, u,\, v \in U_n$$

(V5) $$\sum_{i=1}^{n} \frac{(\mu_{ui} - \mu_{vi})^2}{\sigma_{ui}^2 + \sigma_{vi}^2} \geq D_5 ||u - v||^2 \quad \forall\, u,\, v \in U_n,$$

where $D_1, \ldots, D_5$ are some nonnegative constants. Let $N(\mu, \sigma^2)$ denote the normal distribution on $\mathbb{R}^1$ with mean μ and variance σ^2. We set for $i = 1, \ldots, n$

$$P_{u,v}^{(i)} = N(\mu_{ui}, \sigma_{vi}^2), \quad u,\, v \in U_n.$$

Definition 1. Let P, Q be probability measures on the measurable space $(\mathcal{X}; \mathcal{A})$ and ν a dominating σ-finite measure. Let p and q denote the densities. Then

$$H_s(P, Q) = \int p^s q^{1-s}\, \mathrm{d}\nu, \quad 0 < s < 1$$

is called the Hellinger integral of order s.

Lemma 1. Let the assumptions (V1), (V2) and (V4) be satisfied. Then there exists a constant $c_1 = c_1(D_1, D_2, D_4)$

$$\sum_{i=1}^{n} \left(1 - H_{\frac{1}{2}}(P_{u,u}^{(i)}, P_{v,v}^{(i)})\right) \leq c_1 ||u - v||^2.$$

Proof. An easy calculation shows

$$H_{\frac{1}{2}}(P_{u,u}^{(i)}, P_{v,v}^{(i)}) = \sqrt{\frac{\pi_i^{1/2}}{\frac{1}{2}(\pi_i + 1)}} \exp\left\{-\frac{1}{4} \frac{(\mu_{ui} - \mu_{vi})^2}{\sigma_{ui}^2 + \sigma_{vi}^2}\right\}$$

with $\pi_i := \sigma_{ui}^2 / \sigma_{vi}^2$. Using the inequality $1 - e^{-a} \leq a$ $(a > 0)$ we get

$$1 - H_{1/2}(P_{u,u}^{(i)}, P_{v,v}^{(i)}) \leq \frac{1}{2} \ln\left(\frac{1}{2}(\pi_i + 1)\right) - \frac{1}{4} \ln \pi_i + \frac{1}{4} \frac{(\mu_{ui} - \mu_{vi})^2}{\sigma_{ui}^2 + \sigma_{vi}^2}.$$

We now investigate the function $g(\pi) := \ln\left(\frac{1}{2}(\pi + 1)\right) - \frac{1}{2} \ln \pi$. An expansion in a Taylor series at point 1 up to the second order term shows

$$g(\pi) \geq \frac{1}{8}(\pi - 1)^2, \qquad 0 < \pi \leq 1, \qquad g(\pi) \leq \frac{1}{8}(\pi - 1)^2, \qquad 1 \leq \pi < \infty.$$

Because of $g(\frac{1}{\pi}) = g(\pi)$ it follows

$$\frac{1}{8}\min\left(1, \frac{1}{\pi^2}\right)(\pi-1)^2 \le g(\pi) \le \frac{1}{8}\max\left(1, \frac{1}{\pi^2}\right)(\pi-1)^2. \tag{1.1}$$

By assumption (V1) we get $\pi_i \ge \frac{1}{D_1}$ and therefore

$$\sum_{i=1}^{n} g(\pi_i) \le \frac{D_1^2}{8}\sum_{i=1}^{n}(\pi_i - 1)^2.$$

The assertion now follows from (V2) and (V4). □

Definition 2. Let P and Q be two probability measures on the measurable space $(\mathcal{X}; \mathcal{A})$ dominated by the σ-finite measure ν. Let p, q be the densities of P and Q. Then

$$\rho_m(P, Q) := \left(\int |p^{\frac{1}{m}} - q^{\frac{1}{m}}|^m \, \mathrm{d}\nu\right)^{\frac{1}{m}}$$

is called the Hellinger distance of the order m of P and Q for all integer $m > 0$.

Remark. The Hellinger distance is a metric for $m \ge 1$ (see [4]).

Lemma 2. Suppose the assumptions (V1) and (V4) are satisfied. For every integer $m > 1$ there exists a constant $c_2 = c_2(m, D_4)$ so that

$$\sum_{i=1}^{n} \rho_{2m}^2(P_{u,w}^{(i)}, P_{v,w}^{(i)}) \le c_2\|u - v\|^2.$$

Proof. To evaluate the expression $\rho_{2m}^{2m}(Q, R)$ we apply the binomial formula:

$$\rho_{2m}^{2m}(P_{u,w}^{(i)}, P_{v,w}^{(i)}) = \sum_{k=0}^{2m}(-1)^k\binom{2m}{k}\exp\left\{\left(\frac{k^2}{8m^2} - \frac{k}{4m}\right)\frac{(\mu_{ui} - \mu_{vi})^2}{\sigma_{wi}^2}\right\}.$$

To simplify the notations we set

$$\begin{aligned}
&\sum_{k=0}^{2m}(-1)^k\binom{2m}{k}\exp\left\{\left(\frac{k^2}{8m^2} - \frac{k}{4m}\right)\frac{(\mu_{ui} - \mu_{vi})^2}{\sigma_{wi}^2}\right\} \\
= \; &\sum_{k=0}^{2m}(-1)^k\binom{2m}{k}\lambda^{k(2m-k)} = P_m(\lambda) \\
&\qquad \text{with } \lambda := \exp\left\{\frac{-1}{8m^2}\frac{(\mu_{ui} - \mu_{vi})^2}{\sigma_{wi}^2}\right\}.
\end{aligned}$$

Similarly as in [2] (p. 274) we expand $P_m(\lambda)$ in a Taylor series at point $\lambda = 1$. It is easy to see, that the derivatives $P_m^{(j)}(1)$ are linear combinations of sums

$$S_i = \sum_{k=0}^{2m} (-1)^k \binom{2m}{k} k^i \quad (i = 1, \dots, 2j).$$

S_i can be written in the following form:

$$\begin{aligned}
S_1 &= \sum_{k=0}^{2m} (-1)^k \binom{2m}{k} k = \sum_{k=0}^{2m} (-1)^k \binom{2m}{k} k z^{k-1} \Big|_{z=1} = \frac{\partial}{\partial z}(z-1)^{2m} \Big|_{z=1} \\
S_2 &= \sum_{k=0}^{2m} (-1)^k \binom{2m}{k} k^2 = \sum_{k=0}^{2m} (-1)^k \binom{2m}{k} k(k-1) z^{k-2} \Big|_{z=1} \\
&\quad + \sum_{k=0}^{2m} (-1)^k \binom{2m}{k} k z^{k-1} \Big|_{z=1} = \frac{\partial^2}{\partial z^2}(z-1)^{2m} \Big|_{z=1} + \frac{\partial}{\partial z}(z-1)^{2m}|_{z=1}.
\end{aligned}$$

Obviously, the term S_i is a linear combination of $\frac{\partial^j}{\partial z^j}(z-1)^{2m}|_{z=1}$ $(j = 1, \dots, i)$ and it holds $S_0 = S_1 = \dots = S_{2m-1} = 0$. Hence $P_m^{(0)} = P_m^{(1)} = \dots P_m^{(m-1)} = 0$ and we get $P_m(\lambda) = (1-\lambda)^m P_m^{(m)}(\psi)$ with some ψ $(0 \le \psi \le 1)$. For the concrete form of the constant $P_m^{(m)}(\psi)$ we refer to [2] (p. 276). Using the inequality $1 - e^{-a} \le a$ $(a > 0)$ we obtain with some absolute constant

$$\rho_{2m}^{2m}(P_{u,w}^{(i)}, P_{v,w}^{(i)}) \le \hat{c}_2^m (1-\lambda)^m \le \hat{c}_2^m \left(\frac{(\mu_{ui} - \mu_{vi})^2}{\sigma_{wi}^2} \right)^m.$$

Assumption (V1) implies

$$\sigma_{wi}^2 \ge \frac{\sigma_{ui}^2 + \sigma_{vi}^2}{2D_1}.$$

The assertion of Lemma 2 now follows from the assumptions (V1) and (V4):

$$\sum_{i=1}^{n} \rho_{2m}^2 (P_{u,w}^{(i)}, P_{v,w}^{(i)}) \le 2\hat{c}_2 D_1 \sum_{i=1}^{n} \frac{(\mu_{ui} - \mu_{vi})^2}{\sigma_{ui}^2 + \sigma_{vi}^2} \le 2\hat{c}_2 D_1 D_4 \|u - v\|^2. \qquad \square$$

Now we estimate Hellinger integrals of normal distributions with different variances. We recall to the definition to the Hellinger integral in Definition 1 and get with $\pi = \sigma_{ui}^2 / \sigma_{vi}^2$:

$$H_{\frac{k}{2m}}(N(\mu_{wi}, \sigma_{ui}^2), N(\mu_{wi}, \sigma_{vi}^2)) = \exp\left\{ \frac{k}{4m} \ln \pi - \frac{1}{2} \ln \left(\frac{k}{2m}(\pi - 1) + 1 \right) \right\}.$$

Lemma 3. Suppose (V1) and (V2) are satisfied. Then for every integer $m > 0$ there exists a constant $c_3 = c_3(m, D_1, D_2)$ so that

$$\sum_{i=1}^{n} \rho_{2m}^2(P_{w,u}^{(i)}, P_{w,v}^{(i)}) \le c_3 \|u - v\|^2.$$

P r o o f. To evaluate the expression $\rho_{2m}^{2m}(P, S)$ we set $\pi = \pi_i = \sigma_{ui}^2/\sigma_{vi}^2$:

$$\begin{aligned}
\rho_{2m}^{2m}(P_{w,u}^{(i)}, P_{w,v}^{(i)}) &= \sum_{k=0}^{2m} (-1)^k \binom{2m}{k} H_{\frac{k}{2m}}(P_{w,u}^{(i)}, P_{w,v}^{(i)}) \\
&= \sum_{k=0}^{2m} (-1)^k \binom{2m}{k} \exp\left\{ \frac{k}{4m} \ln \pi - \frac{1}{2} \ln \left(\frac{k}{2m}(\pi - 1) + 1 \right) \right\}.
\end{aligned}$$

Now let us use the notation $s = \frac{k}{2m}$, $g_k(\pi) = \frac{1}{2} s \ln \pi - \frac{1}{2} \ln(s\pi - s + 1)$, $f_k(\pi) = \exp\{g_k(\pi)\}$ and

$$h(\pi) = \sum_{k=0}^{2m} (-1)^k \binom{2m}{k} f_k(\pi).$$

To expand the function $h(\pi)$ in a Taylor series at point $\pi = 1$ we need the derivatives of the functions $f_k(\pi)$ up to the order $2m$. The jth derivative ($j = 1, \ldots, 2m$) can be expressed by linear combinations of terms

$$\left(\prod_{i=1}^{j} (g_k^{(i)}(\pi))^{\alpha_i} \right) \exp\{g_k(\pi)\},$$

where $\sum_{i=1}^{j} i\alpha_i = j$ for $\alpha_i \in \mathbb{N}$. We consider the function $g_k(\pi)$ and its derivatives. By induction one can prove

$$g_k^{(i)}(\pi) = (-1)^{i+1} \frac{(i-1)!}{2} \left(\frac{s}{\pi^i} - \frac{s^i}{(s\pi - s + 1)} \right).$$

Then the derivative of $g_k(\pi)$ at point $\pi = 1$ is

$$g_k^{(i)}(1) = (-1)^{i+1} \frac{(i-1)!}{2} (s^i - s) \quad (i = 1, \ldots, j).$$

Hence $f_k^{(j)}(1)$ is a linear combination of the products

$$\prod_{i=1}^{j} (s^i - s)^{\alpha_i} \quad \left(\sum_{i=1}^{j} i\alpha_i = j;\ \alpha_i \in \mathbb{N},\ j = 1, \ldots, m \right).$$

By termwise differentiation of function $h(\pi)$ we get

$$h^{(j)}(1) = \sum_{k=0}^{2m} (-1)^k \binom{2m}{k} f_k^{(j)}(1).$$

Because of the special form of $f_k^{(j)}(1)$ the term $h^{(j)}(1)$ is a linear combination of S_i (see proof of Lemma 2). Taking into account $S_i = 0$ $(i = 1, \dots, 2m-1)$ $h(\pi)$ can be estimated by Taylor expansion at point $\pi = 1$ in the following form

$$h(\pi) = \sum_{k=0}^{2m} (-1)^k \binom{2m}{k} f^k(\xi)\,(\pi - 1)^{2m} \leq \hat{c}_3^m (\pi - 1)^{2m},$$

where $\frac{1}{D_1} \leq \pi \leq D_1$ because of (V1) and $0 < \xi < \pi$. Taking again $\pi = \pi_i$ we get

$$\rho_{2m}^{2m}(P_{w,u}^{(i)}, P_{w,v}^{(i)}) = h(\pi_i) \leq \hat{c}_3^m (\pi_i - 1)^{2m}.$$

Under the assumption (V2) we have

$$\sum_{i=1}^{n} \rho_{2m}^2 (P_{w,u}^{(i)}, P_{w,v}^{(i)}) \leq \hat{c}_3 \sum_{i=1}^{n} (\pi_i - 1)^2 \leq \hat{c}_3 D_2 \|u - v\|^2$$

which proves Lemma 3. □

The following inequality was proved in [1], Corollary 3.1.

Lemma 4. Let P_i and Q_i, $i = 1, \dots, n$ be probability measures on the measurable space $(\mathcal{X}, \mathcal{A})$. Denote by $P_1 \times \ldots \times P_n$ and $Q_1 \times \ldots \times Q_n$ the corresponding product measures. Then for each $m > 1$ there exists a constant $c_4 = c_4(m)$ (independent of n) with

$$\rho_{2m}^{2m}(P_1 \times \ldots \times P_n, Q_1 \times \ldots \times Q_n) \leq c_4 \sum_{i=1}^{n} \rho_{2m}^{2m}(P_i, Q_i) + c_4 \Big(\sum_{i=1}^{n} \big(1 - H_{\frac{1}{2}}(P_i, Q_i)\big) \Big)^m.$$

Lemma 5. Assume conditions (V1), (V2) and (V4) are satisfied. For every $m \geq \frac{1}{2}$ there exists a constant $c_5 = c_5(m, D_1, D_2, D_4)$ so that

$$\rho_{2m}^2 \left(\prod_{i=1}^{n} P_{u,u}^{(i)}, \prod_{i=1}^{n} P_{v,v}^{(i)} \right) \leq c_5 \|u - v\|^2.$$

Proof. Because of Lemma 4 and by using the metric property of ρ_{2m} and the elementary inequalities $(a+b)^2 \leq 2(a^2+b^2)$ and $\sum_{i=1}^n (a_i^2+b_i^2)^m \leq (\sum_{i=1}^n a_i^2+b_i^2)^m$ we get

$$\begin{aligned}
\rho_{2m}^{2m} \left(\prod_{i=1}^{n} P_{u,u}^{(i)}, \prod_{i=1}^{n} P_{v,v}^{(i)} \right) &\leq c_4 \sum_{i=1}^{n} \rho_{2m}^{2m}(P_{u,u}^{(i)}, P_{v,v}^{(i)}) + c_4 \Big(\sum_{i=1}^{n} \big(1 - H_{\frac{1}{2}}(P_{u,u}^{(i)}, P_{v,v}^{(i)})\big) \Big)^m \\
\sum_{i=1}^{n} \rho_{2m}^{2m}(P_{u,u}^{(i)}, P_{v,v}^{(i)}) &\leq \sum_{i=1}^{n} \Big(\rho_{2m}(P_{u,u}^{(i)}, P_{u,v}^{(i)}) + \rho_{2m}(P_{u,v}^{(i)}, P_{v,v}^{(i)}) \Big)^{2m} \\
&\leq 2^m \Big(\sum_{i=1}^{n} \rho_{2m}^2(P_{u,u}^{(i)}, P_{u,v}^{(i)}) + \sum_{i=1}^{n} \rho_{2m}^2(P_{u,v}^{(i)}, P_{v,v}^{(i)}) \Big)^m.
\end{aligned}$$

Since Lemma 2 and 3 hold for every w we get

$$\rho_{2m}^{2m}\left(\prod_{i=1}^{n} P_{u,u}^{(i)}, \prod_{i=1}^{n} P_{v,v}^{(i)}\right) \le 2^m \left(c_2\|u-v\|^2 + c_3\|u-v\|^2\right)^m.$$

The statement

$$\sum_{i=1}^{n}\left(1 - H_{\frac{1}{2}}(P_{u,u}^{(i)}, P_{v,v}^{(i)})\right) \le c_1\|u-v\|^2$$

from Lemma 1 now completes the proof. □

Lemma 6. If the assumptions (V1), (V3) and (V5) are satisfied then

$$\prod_{i=1}^{n} H_{\frac{1}{2}}(P_{u,u}^{(i)}, P_{v,v}^{(i)}) \le \exp\left\{-\left(\frac{D_3}{16D_1^2} + \frac{D_5}{4}\right)\|u-v\|^2\right\}.$$

Proof. Using the notations above and $\pi_i = \sigma_{ui}^2/\sigma_{vi}^2$ we get:

$$\prod_{i=1}^{n} H_{\frac{1}{2}}(P_{u,u}^{(i)}, P_{v,v}^{(i)}) = \prod_{i=1}^{n}\sqrt{\frac{\pi_i^{1/2}}{\frac{1}{2}(\pi_i+1)}}\exp\left\{-\frac{1}{4}\frac{(\mu_{ui}-\mu_{vi})^2}{\sigma_{ui}^2+\sigma_{vi}^2}\right\},$$

$$= \exp\left\{\frac{1}{2}\sum_{i=1}^{n}\left(\frac{1}{2}\ln\pi_i - \ln\left(\frac{1}{2}(\pi_i+1)\right) - \frac{1}{2}\frac{(\mu_{ui}-\mu_{vi})^2}{\sigma_{ui}^2+\sigma_{vi}^2}\right)\right\}.$$

Assumption (V1) implies $\frac{1}{D_1} \le D_1$. Using inequality (1.1) for $g(\pi)$ we arrive at

$$\frac{1}{8D_1^2}\sum_{i=1}^{n}(\pi_i-1)^2 \le \sum_{i=1}^{n}\ln\left(\frac{1}{2}(\pi_i+1)\right) - \frac{1}{2}\ln\pi_i.$$

The assertion now follows from the assumptions (V3) and (V5). □

2. GENERAL CRITERIA FOR EXPONENTIAL RATE OF CONVERGENCE OF THE MAXIMUM LIKELIHOOD ESTIMATOR

At first we consider a general sequence of experiments $(\mathcal{X}^n; \mathcal{A}^n; P_\vartheta^n, \vartheta \in \Theta)$, where the parameter space $\Theta \subseteq \mathbb{R}^k$ is open. For all n the family P_ϑ^n $(\vartheta \in \Theta)$ is supposed to be dominated by a σ-finite measure ν^n. For the proof of the exponential convergence rate of maximum likelihood estimators for inhomogeneous Wiener processes we apply results from the monograph by Ibragimov–Hasminskij [2]. Let φ_n be a sequence of nonsingular $k \times k$ matrices. We define $U_n(\vartheta) := \varphi_n^{-1}(\Theta - \vartheta)$ and denote by p_n the density of P_ϑ^n w.r.t. ν^n

$$p_n(x_n, \vartheta) := \frac{\mathrm{d}P_\vartheta^n}{\mathrm{d}\nu^n}(x_n)$$

for $x_n \in \mathcal{X}^n$. We assume, that the experiments are homogeneous, which means that $P_\vartheta^n \sim P_\eta^n$ for all $\vartheta, \eta \in \Theta$ and all n. The likelihood ratio in a localized form is given by

$$Z_{n,\vartheta}(x_n, u) := \frac{p_n(x_n, \vartheta + \varphi_n u)}{p_n(x_n, \vartheta)},$$

where $u \in U_n(\vartheta)$ and $x_n \in \mathcal{X}^n$. Let $\tilde{\vartheta}_n$ be a measurable mapping from $(\mathcal{X}^n, \mathcal{A}^n)$ in $(\Theta^c, \mathcal{B}_\Theta)$ with $\mathcal{B}_\Theta$ as σ-algebra of the Borel sets of Θ^c so that

$$Z_{n,\vartheta}\left(x_n, \varphi_n^{-1}(\tilde{\vartheta}_n - \vartheta)\right) = \sup_{u \in U_n} Z_{n,\vartheta}(x_n, u)$$

holds for all $\vartheta \in \Theta$. Then $\tilde{\vartheta}_n$ is called a maximum likelihood estimator. $Z_{n,\vartheta}(x_n, 0) = 1$ implies

$$P_\vartheta^n\left(|\varphi_n^{-1}(\tilde{\vartheta}_n - \vartheta)| \geq \gamma\right) \leq P_\vartheta^n\left(\sup_{|u| \geq \gamma} Z_{n,\vartheta}(x_n, u) \geq 1\right)$$

for all $\gamma > 0$. Hence to prove consistency it is enough to estimate the term on the right hand side. We denote by $\mathcal{G}$ the set of all sequences of functions $g_n(y)$ with the following properties:

a) For fixed n $g_n(y)$ is nondecreasing on $[0, \infty)$.

b) For all $\alpha > 0$ holds $\lim\limits_{\substack{y \to \infty \\ n \to \infty}} y^{-\alpha} \exp\{-g_n(y)\} = 0$.

The following Theorem 1 (Theorem 5.1 in [2]) plays a basic role to establish the exponential rate of convergence in our proof of consistency. It gives general criteria for consistency and exponential rate of convergence for maximum likelihood estimators. In the sequel we will derive easy manageable conditions on the functions m and b with the help of this theorem.

Theorem 1. Assume that for fixed n and P_ϑ^n-almost all $x_n \in \mathcal{X}^n$ the $Z_{n,\vartheta}(x_n, u)$ is a continuous function of u. Suppose that for every compact $K \subset \Theta$ there exist numbers $m(K)$, $M(K)$ and an $\alpha > k$ ($k = \dim \Theta$) and $l \geq \alpha$, such that for all $\vartheta \in K$

$$\sup_{\|u\|, \|v\| \leq R} |u - v|^{-\alpha} \mathsf{E}_{\vartheta,n} |Z_{n,\vartheta}^{\frac{1}{l}}(u) - Z_{n,\vartheta}^{\frac{1}{l}}(v)|^l \leq M(1 + R^m). \tag{2.1}$$

If there exists a sequence $(g_n) \in \mathcal{G}$ such that for all $u \in U_n, \vartheta \in K$

$$\mathsf{E}_{\vartheta,n} Z_{n,\vartheta}^{\frac{1}{2}}(u) \leq \exp\{-g_n(\|u\|)\} \tag{2.2}$$

then there exists a n_0 so that for all $n > n_0$

$$\sup_{\vartheta \in K} P_\vartheta^n\left(|\varphi_n^{-1}(\tilde{\vartheta}_n - \vartheta)| \geq H\right) \leq B \exp\{-b g_n(H)\}, \tag{2.3}$$

where b and B are nonnegative constants, which only depend on $K \subset \Theta$.

Suppose $\Theta \subseteq \mathbb{R}^k$ and let $W(t)$ $(0 \le t \le T)$ be an inhomogeneous Wiener process with mean function $m(t,\vartheta)$ and variance function $b(t,\vartheta)$, $\vartheta \in \Theta$. On one side the process shall be observed at a grown number of observation points. On the other side we shall assume that independent replications of the process are available. Let $W_1(t), \ldots, W_{m_n}(t)$ $(0 \le t \le T)$ be independent Wiener processes with the same distribution as $W(t)$ $(0 \le t \le T)$. These processes will be observed at time points $0 = t_{0,n} < \ldots < t_{k_n,n} \le T$. Because of $W_j(0) = 0$ for the estimation of the parameter vector ϑ we can deal with the increments $W_j(t_{i,n}) - W_j(t_{i-1,n})$ instead of $W_j(t_{i,n})$ without loss of information. The sequence of associated statistical experiments is $(\mathbb{R}^{k_n m_n}; \mathcal{B}^{k_n m_n}; P_\vartheta^n, \vartheta \in \Theta)$, where

$$P_\vartheta^n = \prod_{i=1}^{k_n} N^{m_n}\big(m(t_{i,n},\vartheta) - m(t_{i-1,n},\vartheta), b(t_{i,n},\vartheta) - b(t_{i-1,n},\vartheta)\big).$$

Here N^{m_n} is the m_n-times product measure. To avoid irregularities let the following conditions hold. Condition (A2) avoid a degeneration of the model. The considered increments of the process have always positive variances. For abbreviation we define $\Delta m(t_{i,n},\vartheta) := m(t_{i,n},\vartheta) - m(t_{i-1,n},\vartheta)$ and $\Delta b(t_{i,n},\vartheta) := b(t_{i,n},\vartheta) - b(t_{i-1,n},\vartheta)$. Assume that the functions m and b behave regularly along the sequence of observation points.

(A0) $\max\limits_{1\le i\le k_n} (t_{i,n} - t_{i-1,n}) \le C \min\limits_{1\le i\le k_n} (t_{i,n} - t_{i-1,n})$.

(A1) Let $\Theta \subseteq \mathbb{R}^k$ be open and Θ^c compact. $m(t,\vartheta)$, $b(t,\vartheta)$ are continuous on $[0,T] \times \Theta^c$ and differentiable w.r.t. $t \in [0,T]$ for all $\vartheta \in \Theta$. Suppose that the derivative $\dot b = \frac{\partial b}{\partial t}$ is a continuous function on $[0,T] \times \Theta^c$.

(A2) Assume that $\sup\limits_{0\le t\le T} \sup\limits_{\vartheta\in\Theta} \dot b(t,\vartheta) < B_1$, $\inf\limits_{0\le t\le T} \inf\limits_{\vartheta\in\Theta} \dot b(t,\vartheta) > B_2$ for some $B_1, B_2 > 0$.

(A3) There exist some nonnegative constants $c_{1,b}$ and $c_{2,b}$ with $c_{1,b}\|\vartheta - \eta\|^2 \le k_n \sum\limits_{i=1}^{k_n} \big(\Delta b(t_{i,n},\vartheta) - \Delta b(t_{i,n},\eta)\big)^2 \le c_{2,b}\|\vartheta - \eta\|^2 \ \ \forall\, \vartheta, \eta \in \Theta$.

(A4) There exist some nonnegative constants $c_{1,m}$ and $c_{2,m}$ with $c_{1,m}\|\vartheta - \eta\|^2 \le k_n \sum\limits_{i=1}^{k_n} \big(\Delta m(t_{i,n},\vartheta) - \Delta m(t_{i,n},\eta)\big)^2 \le c_{2,m}\|\vartheta - \eta\|^2 \ \ \forall\, \vartheta, \eta \in \Theta$.

Theorem 2. If the assumptions (A0), ..., (A4) hold and the sequence N_n satisfies

$$\sup_n \frac{k_n m_n}{N_n} < \infty \tag{2.4}$$

$$\liminf_{n\to\infty} \big(c_{1,m}\frac{m_n}{N_n} + c_{1,b}\frac{k_n m_n}{N_n}\big) > 0 \tag{2.5}$$

then for every compact subset $K \subseteq \Theta$ there are nonnegative constants $B = B(K)$ and $b = b(K)$ depending only on K so that

$$\sup_{\vartheta\in K} P_\vartheta^n\left(\sqrt{N_n}\|\tilde{\vartheta}_n - \vartheta\| > H\right) \le B \exp\left\{-bF\left(c_{1,m}\frac{m_n}{N_n} + c_{1,b}\frac{k_n m_n}{N_n}\right)H^2\right)\right\}$$

for every sufficiently large n and every $H > 0$ where

$$F := \min\left(\frac{B_2^2}{16B_1^4C^6T^2}, \frac{1}{8B_1CT}\right).$$

P r o o f. We denote by $p_n(x_n, \vartheta)$ the density of P_ϑ^n w.r.t. the Lebesgue measure on $\mathbb{R}^{k_n m_n}$ and define $\varphi_n := \frac{1}{\sqrt{N_n}} I_k$, where I_k is the unit matrix. The notations for $U_n(\vartheta) = \varphi_n^{-1}(\Theta - \vartheta)$ and $x_n \in \mathbb{R}^{k_n m_n}$

$$Z_n(x_n, u) := \frac{p_n(x_n, \vartheta + \varphi_n u)}{p_n(x_n, \vartheta)}$$

are the same as in the beginning of this chapter. The continuity of $Z_n(x_n, u)$ with respect to u follows from the structure of the density p_n as a product of onedimensional normal distributions and the continuity of $m(t, \vartheta)$ and $b(t, \vartheta)$ as functions of ϑ. Using the following notation

$$P_{i,u} = N\left((\Delta m)(t_{i,n}, \vartheta + \varphi_n u), (\Delta b)(t_{i,n}, \vartheta + \varphi_n u)\right)$$

$(i = 1, \ldots, k_n)$ we get the relations

$$\mathsf{E}_{\vartheta,n}|Z_{n,\vartheta}^{\frac{1}{l}}(u) - Z_{n,\vartheta}^{\frac{1}{l}}(v)|^l = \rho_l^l\left(\prod_{j=1}^{m_n}\prod_{i=1}^{k_n} P_{i,u}, \prod_{j=1}^{m_n}\prod_{i=1}^{k_n} P_{i,v}\right) \tag{2.6}$$

$$\mathsf{E}_{\vartheta,n} Z_{n,\vartheta}^{\frac{1}{2}}(u) = H_{\frac{1}{2}}\left(\prod_{j=1}^{m_n}\prod_{i=1}^{k_n} P_{i,u}, \prod_{j=1}^{m_n}\prod_{i=1}^{k_n} P_{i,0}\right). \tag{2.7}$$

We set $\mu_{ui} = \Delta m(t_{i,n}, \vartheta + \frac{1}{\sqrt{N_n}}u)$ and $\sigma_{ui}^2 = \Delta b(t_{i,n}, \vartheta + \frac{1}{\sqrt{N_n}}u)$. We now verify the conditions (V1),...,(V5). The condition (V1) follows from (A0) and (A2). To establish (V2), (V3) and (V4) we remark that in view of (A0)

$$\frac{T}{Ck_n} \le t_{i,n} - t_{i-1,n} \le \frac{CT}{k_n}. \tag{2.8}$$

Hence by (A2)

$$\frac{B_2 T}{Ck_n} \le \sigma_{ui}^2 \le \frac{B_1 CT}{k_n}. \tag{2.9}$$

Hence (A3) and (A4) imply:

$$\frac{c_{1,b}k_n}{(B_1CT)^2 N_n}\|u - v\|^2 \le \sum_{i=1}^{k_n} \frac{(\sigma_{ui}^2 - \sigma_{vi}^2)^2}{\sigma_{v,i}^4} \le \frac{c_{2,b}C^2 k_n}{(B_2T)^2 N_n}\|u - v\|^2$$

$$\frac{c_{1,m}}{2B_1CTN_n}||u-v||^2 \leq \sum_{i=1}^{k_n}\frac{(\mu_{ui}-\mu_{vi})^2}{\sigma_{ui}^2+\sigma_{vi}^2} \leq \frac{c_{2,m}C}{2B_2TN_n}||u-v||^2.$$

Now we apply Lemma 5 with m_nk_n instead of n. Because of condition (2.4)

$$D_2 = \frac{c_{2,b}C^2}{(B_2T)^2}\left(\sup_n \frac{m_nk_n}{N_n}\right) \quad \text{and} \quad D_4 = \frac{c_{2,m}C}{(2B_2T)^2}\left(\sup_n \frac{m_n}{N_n}\right)$$

satisfy the conditions (V2) and (V4). We set $\alpha = 1 = 2k$ in Theorem 1 and $m = k$ in Lemma 5 and get from (2.6)

$$\mathsf{E}_{\vartheta,n}\left|Z_{n,\vartheta}^{\frac{1}{2k}}(u) - Z_{n,\vartheta}^{\frac{1}{2k}}(v)\right|^{2k} = \rho_{2k}^{2k}\left(\prod_{j=1}^{m_n}\prod_{i=1}^{k_n}P_{i,u}, \prod_{j=1}^{m_n}\prod_{i=1}^{k_n}P_{i,v}\right) \leq c_5^k||u-v||^{2k}$$

for every u, $v \in U_n(\vartheta)$. Hence condition (2.1) in Theorem 1 is fulfilled. We now turn to condition (2.2). Inequality (2.9) allows us to put $D_1 = \frac{B_1}{B_2}K^2$ in (V1). Set for fixed n

$$D_3 = \frac{c_{1,b}m_nk_n}{(B_1CT)^2N_n} \quad \text{and} \quad D_5 = \frac{c_{1,m}m_n}{2B_1CTN_n}$$

then Lemma 6 implies

$$\begin{aligned}
\mathsf{E}_{\vartheta,n}Z_{n,\vartheta}^{\frac{1}{2}}(u) &\leq \exp\left\{-(\frac{D_3}{16D_1^2}+\frac{D_5}{4})||u||^2\right\} \\
&= \exp\left\{-(\frac{c_{1,b}B_2^2m_nk_n}{16B_1^4C^6T^2N_n}+\frac{c_{1,m}m_n}{8B_1CTN_n})||u||^2\right\} \\
&\leq \exp\left\{-F(c_{1,m}\frac{m_n}{N_n}+c_{1,b}\frac{m_nk_n}{N_n})||u||^2\right\},
\end{aligned}$$

where

$$F = \min\left(\frac{B_2^2}{16B_1^4C^6T^2}; \frac{1}{8B_1CT}\right).$$

Now we set

$$g_n(x) = F\left(c_{1,m}\frac{m_n}{N_n}+c_{1,b}\frac{m_nk_n}{N_n}\right)x^2.$$

Obviously $g_n(x)$ is nondecreasing in $[0,\infty)$. The relation $\lim_{\substack{y\to\infty\\ n\to\infty}} y^{-\alpha}\exp\{-g_n(y)\} = 0$ is a consequence of assumption (2.5). The assertion now follows directly from Theorem 1. □

We reformulate the result of the last theorem by setting $H = \sqrt{N_n}\varepsilon$, $\varepsilon > 0$, $N_n = k_nm_n$.

Proposition 1. Under the assumptions of Theorem 2 it holds

$$\sup_{\vartheta \in K} P_\vartheta^n\left(\|\tilde{\vartheta}_n - \vartheta\| > \varepsilon\right) \le B \exp\left\{-bF\left(c_{1,m}m_n + c_{1,b}k_n m_n\right)\varepsilon^2\right)\right\}.$$

Let us discuss the last proposition in more details. If the variance function $b(t, \vartheta)$ of the Wiener process under consideration really depends on ϑ then one may expect that $c_{1,b} > 0$. Then if $k_n \to \infty$ as $n \to \infty$ one sample ($m_n = 1$) is enough to come to a consistent estimator. This corresponds to the fact that the distribution of Wiener processes with different parameter values are mutually singular. If more replications are available then the rate of convergence increases. If $b(t, \vartheta)$ is independent of ϑ then $c_{1,b} = 0$. Then we have to assume $c_{1,m} > 0$ and $m_n \to \infty$ to guarantee consistency. Then condition $c_{1,m} > 0$ means that the mean value function $m(t, \vartheta)$ differs essentially from $m(t, \eta)$, $\vartheta \neq \eta$. The condition $m_n \to \infty$ says that the number of independent replications of the Wiener process tends to infinity. It is clear that for $b(t, \vartheta)$ being independent of ϑ this condition is nessesary to arrive to a consistent MLE since the distribution of the Wiener process for different parameter values are measure theoretically equivalent under weak conditions to the mean value function $m(t, \vartheta)$.

In the following we need a simple technical lemma which is a direct consequence of the definition of the Riemann integral.

Lemma 7. Let $h_\lambda(t)$ $(0 \le t \le T,\ \lambda \in \Lambda)$ be a family which is equicontinuous and uniformly bounded. Suppose $0 = t_{0,n} \le \ldots \le t_{k_n,n} = T$ is a sequence of partitions of $[0, T]$ with

$$\max_{1 \le i \le k_n} (t_{i,n} - t_{i-1,n}) \underset{n \to \infty}{\longrightarrow} 0. \tag{2.10}$$

Then

$$\sup_{\lambda \in \Lambda} \left| \sum_{i=1}^{k_n} h_\Lambda(t_{i,n})(t_{i,n} - t_{i-1,n}) - \int_0^T h_\lambda(t)\,\mathrm{d}t \right| \underset{n \to \infty}{\longrightarrow} 0.$$

We now ask for better manageable conditions which imply (A3) and (A4), respectively. Let $\Theta \subseteq \mathbb{R}^k$ be open. For $\varepsilon > 0$ we set

$$\Theta^\varepsilon = \{\eta : \eta = \vartheta + \xi,\ \vartheta \in \Theta,\ \|\xi\| \le \varepsilon\}.$$

The following regularity condition on a function $f(t, \vartheta)$ is useful to derive sufficient conditions for (A3) and (A4). Set $\dot{f}(t, \vartheta) = \frac{\partial}{\partial t} f(t, \vartheta)$ and $D\dot{f}(t, \vartheta) = \left(\frac{\partial}{\partial \vartheta_1}\dot{f}(t, \vartheta), \ldots, \frac{\partial}{\partial \vartheta_k}\dot{f}(t, \vartheta)\right)$.

(A) There exists $\varepsilon > 0$ so that f is defined on $[0, T + \varepsilon) \times \Theta^\varepsilon$. For fixed $\vartheta \in \Theta^\varepsilon$ the function $t \longmapsto f(t, \vartheta)$ is differentiable on $[0, T + \varepsilon)$ and for fixed $0 \le t < T + \varepsilon$ the function $\vartheta \longmapsto \dot{f}(t, \vartheta)$ is differentiable w.r.t. the components of ϑ. $(t, \vartheta) \longmapsto D\dot{f}(t, \vartheta)$ is continuous on $[0, T + \varepsilon) \times \Theta^\varepsilon$.

We introduce a function F by

$$F(h,s,e,t,\eta)=[f(t+h,\eta+se)-f(t,\eta+se)-\big(f(t+h,\eta)-f(t,\eta)\big)]h^{-1}s^{-1}$$

if $0<h\le\varepsilon,\ -\varepsilon\le s\le\varepsilon,\ \|e\|=1,\ s\ne 0$. Put

$$F(0,s,e,t,\eta)=[\dot f(t,\eta+se)-\dot f(t,\eta)]s^{-1}$$

if $-\varepsilon\le s\le\varepsilon,\ \|e\|=1,\ s\ne 0$. Set

$$F(h,0,e,t,\eta)=\langle(D\dot f)(t+h,\eta),e\rangle-\langle(D\dot f)(t,\eta),e\rangle$$

$0<h\le\varepsilon,\ \|e\|=1$, where $\langle\cdot,\cdot\rangle$ is the scalar product. Put finally

$$F(0,0,e,t,\eta)=\langle(D\dot f)(t,\eta),e\rangle.$$

Let Θ^c be denote the closure of Θ which is assumed to be compact. S_1 is the closed unit sphere around the origin. The representation

$$f(t+h,\eta)-f(t,\eta)-f(t+h,\eta+se)+f(t,\eta+se)=\int_t^{t+h}\int_0^s\langle(D\dot f)(x,\vartheta+\xi e),e\rangle\,\mathrm{d}x\,\mathrm{d}\xi$$

shows that $(h,s,e,t,\eta)\longmapsto F(h,s,e,t,\eta)$ is a continuous function on $[0,\frac{\varepsilon}{2}]\times[-\frac{\varepsilon}{2},\frac{\varepsilon}{2}]\times S_1\times[0,T]\times\Theta^c$ provided f fulfils (A).

Lemma 8. If f fulfils the condition (A) and $\{t_{i,n}\}$ satisfies (A0) then

$$\sup_{\substack{-\varepsilon\le 2s\le\varepsilon,s\ne0\\ e\in S_1,\eta\in\Theta^c}}\Bigg|\sum_{i=1}^{k_n}\Big(\frac{(\Delta f)(t_{i,n},\eta+se)-(\Delta f)(t_{i,n},\eta)}{(t_{i,n}-t_{i-1,n})s}\Big)^2(t_{i,n}-t_{i-1,n})-\int_0^T\Big(\frac{\dot f(t,\eta+se)-\dot f(t,\eta)}{s}\Big)^2\mathrm{d}t\Bigg|\underset{n\to\infty}{\longrightarrow}0.$$

P r o o f. Because of the continuity of F on the compact set $[0,\frac{\varepsilon}{2}]\times[-\frac{\varepsilon}{2},\frac{\varepsilon}{2}]\times S_1\times[0,T]\times\Theta^c$ we have

$$\Delta_n:=\sup_{\substack{e\in S_1,\eta\in\Theta^c\\ -\varepsilon\le2s\le\varepsilon,s\ne0}}\max_i\big|F^2(t_{i,n}-t_{i-1,n},s,e,t_{i,n},\eta)-F^2(0,s,e,t_{i,n},\eta)\big|\underset{n\to\infty}{\longrightarrow}0$$

and by the definition of F:

$$\sup_{\substack{e\in S_1,\eta\in\Theta^c\\ -\varepsilon\le2s\le\varepsilon,s\ne0}}\Bigg|\sum_{i=1}^{k_n}\Big(\frac{(\Delta f)(t_{i,n},\eta+se)-(\Delta f)(t_{i,n},\eta)}{(t_{i,n}-t_{i-1,n})s}\Big)^2(t_{i,n}-t_{i-1,n})-\sum_{i=1}^{k_n}\Big(\frac{\dot f(t_{i,n},\eta+se)-\dot f(t_{i,n},\eta)}{s}\Big)^2(t_{i,n}-t_{i-1,n})\Bigg|$$
$$\le\ \Delta_n\sum_{i=1}^{k_n}(t_{i,n}-t_{i-1,n})=\Delta_nT\underset{n\to\infty}{\longrightarrow}0.$$

We set $\lambda = \eta + se$ and apply Lemma 7 to $\dot{f}(t, \eta + se)$ to get the assertion. □

Now we introduce the function

$$G(t,h,\vartheta,\eta) = \frac{f(t+h,\vartheta) - f(t,\vartheta) - \big(f(t+h,\eta) - f(t,\eta)\big)}{h\|\eta - \vartheta\|}$$

on $0 < t \leq T$, $0 < h \leq \varepsilon$, $\vartheta, \eta \in \Theta^c$, $\|\vartheta - \eta\| \geq \delta > 0$ and set

$$G(t,0,\vartheta,\eta) = \frac{\dot{f}(t,\vartheta) - \dot{f}(t,\eta)}{\|\eta - \vartheta\|}$$

$G(t,h,\vartheta,\eta)$ is again continuous. The following lemma may be proved quite analogously to Lemma 8.

Lemma 9. Suppose f fulfils the condition (A) and $\{t_{i,n}\}$ satisfies (A0). Then

$$\sup_{\substack{\|\vartheta-\eta\|>\delta \\ \vartheta,\eta\in\Theta^c}} \Bigg| \sum_{i=1}^{k_n} \left(\frac{\Delta f(t_{i,n},\vartheta) - \Delta f(t_{i,n},\eta)}{(t_{i,n} - t_{i-1,n})\|\vartheta - \eta\|} \right)^2 (t_{i,n} - t_{i-1,n}) - \int_0^T \left(\frac{\dot{f}(t,\vartheta) - \dot{f}(t,\eta)}{\|\vartheta - \eta\|} \right)^2 \mathrm{d}t \Bigg| \underset{n\to\infty}{\longrightarrow} 0.$$

Proposition 2. Suppose $\Theta \subseteq \mathbb{R}^k$ is open and bounded. Assume f fulfils condition (A) and there exist a constant $c > 0$ so that

$$\int_0^T \big(\dot{f}(t,\vartheta) - \dot{f}(t,\eta)\big)^2 \,\mathrm{d}t \geq c\|\vartheta - \eta\|^2 \tag{2.11}$$

for every $\vartheta, \eta \in \Theta^c$. If $\{t_{i,n}\}$ is a sequence which fulfils (A0) then there are constants $c_{1,f}, c_{2,f}$ with

$$c_{1,f}\|\vartheta - \eta\|^2 \leq k_n \sum_{i=1}^{k_n} \big(\Delta f(t_{i,n},\vartheta) - \Delta f(t_{i,n},\eta)\big)^2 \leq c_{2,f}\|\vartheta - \eta\|^2 \quad \forall\, \vartheta, \eta \in \Theta^c$$

and every sufficiently large n.

P r o o f. Set

$$\begin{aligned}\Phi(u,v) &= f(t_{i-1,n} + u(t_{i,n} - t_{i-1,n}), \eta + v(\vartheta - \eta)) - f(t_{i-1,n}, \eta + v(\vartheta - \eta)) \\ &\quad - f(t_{i-1,n} + u(t_{i,n} - t_{i-1,n}), \eta) + f(t_{i-1,n}, \eta).\end{aligned}$$

Using the relation

$$f(t+h,\eta) - f(t,\eta) - f(t+h,\eta+se) + f(t,\eta+se) = \int_t^{t+h} \int_0^s \langle (D\dot{f})(x,\vartheta+\xi e), e\rangle \,\mathrm{d}x\,\mathrm{d}\xi$$

we see that for $0 \le u, v \le 1$

$$|\Phi(u,v)| \le \sup_{0\le t\le T, \vartheta\in\Theta} \|(D\dot{f})(t,\vartheta)\| \, |t_{i,n}-t_{i-1,n}| \, \|\vartheta-\eta\| = \Gamma_1 |t_{i,n}-t_{i-1,n}| \, \|\vartheta-\eta\|.$$

Condition (A0) implies $\Gamma_2 = \sup_n k_n \max_{1\le i\le k_n} |t_{i,n}-t_{i-1,n}| < \infty$. Hence

$$k_n \sum_{i=1}^{k_n} |\Phi_i(u,v)|^2 \le k_n \Gamma_1^2 \sum_{i=1}^{k_n} |(t_{i,n}-t_{i-1,n})|^2 \, \|\vartheta-\eta\|^2 \le \Gamma_1^2\Gamma_2 T \|\vartheta-\eta\|^2.$$

Putting $c_{2,f} = \Gamma_1^2\Gamma_2 T$ we get the right hand inequality in the statement. To prove the other inequality we firstly consider the case $\|\vartheta-\eta\| < \frac{\varepsilon}{2}$. Then $\vartheta = \eta + se$, $e \in S_1$, $0 < s < \frac{\varepsilon}{2}$ and $s = \|\vartheta-\eta\|$. The assumption on f says

$$\int_0^T \left(\frac{\dot{f}(t,\eta+se)-\dot{f}(t,\eta)}{s}\right)^2 \mathrm{d}t \ge c.$$

Condition (A0) now implies the existence of a constant $\Gamma_3 > 0$ with

$$\frac{\Gamma_3}{\min\limits_{1\le i\le k_n} |t_{i,n}-t_{i,n}|} \le k_n.$$

By Lemma 8 we have for all sufficiently large n and all $e \in S_1$, $-\frac{\varepsilon}{2} < s < \frac{\varepsilon}{2}$, $s \neq 0$:

$$\begin{aligned} & k_n \sum_{i=1}^{k_n} \left(\frac{\Delta f(t_{i,n},\vartheta)-\Delta f(t_{i,n},\eta)}{\|\vartheta-\eta\|}\right)^2 \\ \ge \; & \Gamma_3 \left(\frac{\Delta f(t_{i,n},\vartheta)-\Delta f(t_{i,n},\eta)}{(t_{i,n}-t_{i-1,n})\|\vartheta-\eta\|}\right)^2 (t_{i,n}-t_{i-1,n}) \ge \frac{c}{2}. \end{aligned}$$

This yields the left hand inequality of the statement for $\|\vartheta-\eta\| < \frac{\varepsilon}{2}$. The case $\|\vartheta-\eta\| \ge \frac{\varepsilon}{2}$ may be treated analogously with the help of Lemma 9. □

Now we derive from Theorem 2 easier manageable conditions on the functions $m(t,\vartheta)$ and $b(t,\vartheta)$ which imply an exponential rate of convergence of maximum likelihood estimators.

Theorem 3. Suppose $\{t_{i,n}\}$ fulfils (A0) and $\Theta \subseteq \mathbb{R}^k$ is open and Θ^c compact. Assume that both $m(t,\vartheta)$, $b(t,\vartheta)$ fulfil the condition (A). Suppose b fulfils (A2). Put

$$c_b = \inf_{\substack{\vartheta,\eta\in\Theta^c \\ \vartheta\neq\eta}} \frac{1}{\|\vartheta-\eta\|^2} \int_0^T \left(\dot{b}(t,\vartheta)-\dot{b}(t,\eta)\right)^2 \mathrm{d}t \tag{2.12}$$

$$c_m = \inf_{\substack{\vartheta,\eta\in\Theta^c \\ \vartheta\neq\eta}} \frac{1}{\|\vartheta-\eta\|^2} \int_0^T \left(\dot{m}(t,\vartheta)-\dot{m}(t,\eta)\right)^2 \mathrm{d}t. \tag{2.13}$$

Then for every sequence $\tilde{\vartheta}_n$ of maximum likelihood estimators of ϑ by observation of m_n independent Wiener processes at the points $\{t_{i,n}\}$ $(1 \le i \le k_n)$, and every compact set $K \subseteq \Theta$ there are two nonnegative constants $b = b(K)$, $B = B(K)$ so that with F from Theorem 2

$$\sup_{\vartheta \in K} P_\vartheta^n\left(\sqrt{k_n m_n}\|\tilde{\vartheta}_n - \vartheta\| > H\right) \leq B \exp\left\{-bF\left(\frac{c_m}{k_n} + c_b\right)H^2\right\} \tag{2.14}$$

provided $c_b > 0$ and $k_n \underset{n\to\infty}{\longrightarrow} \infty$. It holds

$$\sup_{\vartheta \in K} P_\vartheta^n\left(\sqrt{m_n}\|\tilde{\vartheta}_n - \vartheta\| > H\right) \leq B \exp\left\{-bF(c_m + c_b k_n)H^2\right\} \tag{2.15}$$

provided $c_m > 0$ and $m_n \underset{n\to\infty}{\longrightarrow} \infty$.

The proof of Theorem 3 is a direct consequence of Theorem 2 and Proposition 2. □

The statement (2.14) says that in case the variance function actually depends on the parameter formulated as condition $c_b > 0$ then a sequence of observations at an increasing number of observation points of one continuous realization is enough to generate consistency of the MLE. The m_n replications increase the rate of convergence. If we only know that $c_m > 0$ then the number of replications m_n must tend to infinity to generate consistency. The increasing number of observation points $t_{i,n}$ $(i = 1, \ldots, k_n)$ is only used to realize the left hand inequality in (A4) with help of the assumption $c_m > 0$.

An analysis of the proof of Theorem 2 shows that the statement (2.15) continues to hold if k_n is fixed $k_n = k_{n_0}$, but n_0 is large enough in the sense that the left hand inequality in (A4) holds with $c_m \geq 0$. If $k_n = k_{n_0}$ is fixed and $m_n \to \infty$ then we arrange the increments of the process $W_j(t)$ in vectors

$$Z_j = \left(W_j(t_{1,n_0}), W_j(t_{2,n_0}) - W_j(t_{1,n_0}), \ldots, W_j(t_{k_{n_0},n_0}) - W_j(t_{k_{n_0}-1,n_0})\right)$$

and see that $Z_1, \ldots, Z_{m_n}$ are i.i.d. random vectors. The exponential rate of convergence of MLE is then a well known fact, see [2] and [6].

We ask now for sufficent conditions of the basic assumptions (2.12) and (2.13) in the special case of $k_n := n$, $t_{i,n} := \frac{iT}{n}$, $i = 0, \ldots, n$. Consider the expression

$$\begin{aligned}
&\frac{1}{\|\vartheta - \eta\|^2} \int_0^T \left(\dot{b}(t,\vartheta) - \dot{b}(t,\eta)\right)^2 \mathrm{d}t \\
= \;& \frac{1}{\|\vartheta - \eta\|^2} \int_0^T \int_0^1 \Big(\sum_{l=1}^k \frac{\partial \dot{b}}{\partial \vartheta_l}(t, \vartheta + s(\eta - \vartheta))(\vartheta_l - \eta_l)\Big)^2 \mathrm{d}s\,\mathrm{d}t \\
= \;& \frac{1}{\|\vartheta - \eta\|^2} \int_0^1 \int_0^T \sum_{l,m=1}^k \frac{\partial \dot{b}}{\partial \vartheta_l}(t, \vartheta + s(\eta - \vartheta)) \\
&\qquad \frac{\partial \dot{b}}{\partial \vartheta_m}(t, \vartheta + s(\eta - \vartheta))(\vartheta_l - \eta_l)(\vartheta_m - \eta_m)\,\mathrm{d}t\mathrm{d}s \\
\geq \;& \frac{1}{\|\vartheta - \eta\|^2} \int_0^1 \lambda_{1,b}(\vartheta + s(\eta - \vartheta))\|\vartheta - \eta\|^2\,\mathrm{d}s \geq \inf_{\vartheta \in \Theta} \lambda_{1,b}(\vartheta),
\end{aligned}$$

where the second last inequality holds because the sum under the integral is a quadratic form, which can estimated with help of the smallest eigenvalue $\inf_{\vartheta\in\Theta}\lambda_{1,b}(\vartheta)$ of the associated matrix

$$A_b(\vartheta) := \Big(\int_0^T \frac{\partial \dot{b}}{\partial\vartheta_l}(t,\vartheta)\frac{\partial \dot{b}}{\partial\vartheta_m}(t,\vartheta)\,\mathrm{d}t\Big)_{1\le l,m\le k}.$$

Obvious sufficient conditions for (2.12) and (2.13) are the uniform positivity of the smallest eigenvalues $\lambda_{1,b}(\vartheta)$ and $\lambda_{1,m}(\vartheta)$ of the matrices $A_b(\vartheta)$ and $A_m(\vartheta)$, respectively.

Proposition 3. Suppose $\{t_{i,n}\}$ fulfils (A0) and $\Theta \subseteq \mathbb{R}^k$ is open and convex and Θ^c is compact. Assume that both $m(t,\vartheta)$ and $b(t,\vartheta)$ fulfil the condition (A). If $\inf_{\vartheta\in\Theta^c}\lambda_{1,b}(\vartheta) > 0$ and $\inf_{\vartheta\in\Theta^c}\lambda_{1,m}(\vartheta) > 0$ then (2.15) in Theorem 3 holds.

Example. Suppose both $m(t,\vartheta)$ and $b(t,\vartheta)$ are linear in the parameters, i.e. we have a quasilinear model

$$b(t,\vartheta) = \sum_{j=1}^k \vartheta_j\psi_j(t), \quad m(t,\vartheta) = \sum_{j=1}^k \vartheta_j\varphi_j(t).$$

Suppose φ_j, ψ_j are continuously differentiable in $[0, T+\varepsilon)$. Then

$$A_b(\vartheta) = \Big(\int_0^T \dot{\psi}_l(t)\dot{\psi}_m(t)\,\mathrm{d}t\Big)_{1\le l,m\le k}.$$

It is easy to see that this matrix is nonsingular iff the functions $\dot{\psi}_j(t)$ are linearly independent, i.e.

$$\sum_{j=1}^k \vartheta_j\dot{\psi}_j(t) = 0 \quad \text{a.s.}$$

w.r.t. the Lebesgue measure holds iff $c_1 = \ldots = c_k = 0$. Consequently the only conditions on m, b to guarantee consistency in the quasilinear model are the continuous differentiability of the φ_j, ψ_j and the condition that the model is not overparametrized, i.e. the $\{\varphi_j\}$ and $\{\psi_j\}$, respectively, systems are linearly independent.

3. STRONG CONSISTENCY OF MLE

In this chapter we apply the exponential bounds to establish strong consistency of MLE and to formulate results on the rate of the a.s. convergence.

Suppose $(\mathcal{X};\mathcal{A};P_\vartheta,\vartheta\in\Theta)$ is a probability space on which the i.i.d. Wiener processes $W_1, W_2,\ldots$ are defined. A special class of sequences of real numbers becomes importance in the sequel. Let N_n be a given sequence of naturals. Denote by $\mathcal{U}(\{N_n\})$ the class of all sequences $\{a_n\}$, $a_n\in\mathbb{R}^1$ so that

$$\sum_{n=1}^\infty q^{\frac{N_n}{a_n^2}} \le \infty$$

for every $0<q<1$. Obvious examples for $N_n = n$ are $a_n = n^\alpha$, $0<\alpha<1/2$.

Theorem 4. Suppose the assumptions in Theorem 3 are fulfilled. The conditions $c_b > 0$ and $\{a_n\} \in \mathcal{U}(\{k_n m_n\})$ imply

$$P_\vartheta\left(\lim_{n\to\infty} a_n \|\tilde{\vartheta}_n - \vartheta\| = 0\right) = 1.$$

The conditions $c_m > 0$ and $\{b_n\} \in \mathcal{U}(\{m_n\})$ yield

$$P_\vartheta\left(\lim_{n\to\infty} b_n \|\tilde{\vartheta}_n - \vartheta\| = 0\right) = 1.$$

Proof. We use the well-known fact (see [7]) that for a sequence of random variables

$$P_\vartheta\left(\lim_{n\to\infty} X_n = 0\right) = 1 \quad \text{iff} \quad P_\vartheta\left(\left|\sup_{m\geq n} X_m\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0$$

for every $\varepsilon > 0$. Using the inequality $P_\vartheta(|\sup_{m\geq n} X_m| > \varepsilon) \leq \sum_{k=n}^{\infty} P_\vartheta(|X_k| > \varepsilon)$ the two statements of Theorem 4 follow directly from (2.14) and (2.15) and the definition of $\mathcal{U}(\{k_n m_n\})$ and $\mathcal{U}(\{m_n\})$, respectively. □

(Received August 20, 1993.)

REFERENCES

[1] K. Dzhaparidze and E. Valkeila: On the Hellinger type distances for filtered experiments. Probab. Theory Related Fields *85* (1990), 105–117.
[2] I. A. Ibragimov and R. Z. Hasminskij: Asymptotic Estimation Theory (in Russian). Nauka, Moscow 1979.
[3] J. Jacod: Random Sampling in Estimation Problems for Continuous Gaussian Processes w: h Independent Increments. Preprint No. 57 Laboratoire de Probabiltès de l' Universitè Paris V.
[4] F. Liese and I. Vajda: Convex Statistical Distances. Teubner–Verlag, Leipzig 1987.
[5] F. Liese and A. Wienke: Asymptotische Normalität von Maximum–Likelihood–Schätzern in inhomogenen Wienerprozessen. Rostock. Math. Kolloq. *47* (1994), 31–48.
[6] T. Pfaff: Quick consistency of quasi maximum likelihood estimators. Ann. Math. Statist. *10* (1982), 3, 900–1005.
[7] A. Rényi: Wahrscheinlichkeitsrechnung mit einem Anhang über Informationstheorie. VEB Deutscher Verlag der Wissenschaften, Berlin 1977.

Prof. Dr. Friedrich Liese, Universität Rostock, Fachbereich Mathematik, Universitätsplatz 1, 18051 Rostock. Federal Republic of Germany.

Andreas Wienke, Universität Rostock, Fachbereich Mathematik, Universitätsplatz 1, 18051 Rostock. Federal Republic of Germany.