Ann Inst Stat Math (2011) 63:431–479
DOI 10.1007/s10463-009-0263-z

Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations

Nakahiro Yoshida

Received: 19 December 2005 / Revised: 15 January 2007 / Published online: 21 May 2010

Abstract The estimate of the probability of the large deviation or the statistical random field is the key to ensure the convergence of moments of the associated estimator, and it also plays an essential role to prove mathematical validity of the asymptotic expansion of the estimator. For non-linear stochastic processes, it involves technical difficulties to show a standard exponential type estimate; besides, it is not necessary for these purposes. In this paper, we propose a polynomial-type large deviation inequality which is easily verified by the L^p-boundedness of certain functionals; usually they are simple additive functionals. We treat a statistical random field with multi-grades and discuss M and Bayesian type estimators. As an application, we show the behavior of those estimators, including convergence of moments, for the statistical random field in the quasi-likelihood analysis of the stochastic differential equation that is possibly multi-dimensional and non-linear. The results are new even for stochastic differential equations, while they obviously apply to other various statistical models.

This work was in part supported by JSPS Grants-in-Aid for Scientific Research No. 19340021, JST Basic Research Programs PRESTO, Cooperative Research Program of the Institute of Statistical Mathematics, and by 21st Century COE Program "Base for New Development of Mathematics to Science and Technology" and the Global COE Program "The research and training center for new development in mathematics", Graduate School of Mathematical Sciences, University of Tokyo. The author thanks the referees for their valuable comments. Also he wishes to acknowledge the useful correspondence with Prof. Yury Kutoyants, Prof. Masayuki Uchida, Mr. T. Ogihara and Mr. H. Inatsugu.

N. Yoshida (✉)
Graduate School of Mathematical Sciences, The University of Tokyo,
3-8-1 Komaba, Meguro, Tokyo 153-8914, Japan
e-mail: nakahiro@ms.u-tokyo.ac.jp

N. Yoshida
Japan Science and Technology Agency, Saitama, Japan

Keywords Large deviation · Quasi-likelihood analysis · Random field · Diffusion process

1 Introduction

Analysis of the likelihood ratio is the first key step in order to investigate the performance of statistics. After the initiation of the local asymptotic normality by Le Cam and Hájek, a new paradigm of analysis was established by Ibragimov and Has'minskii (1972a,b, 1981). That is, all asymptotic properties of statistics in likelihood analysis could be reduced in the convergence of the random field formed by the likelihood ratios. This program was successfully implemented mainly for i.i.d. settings and white Gaussian noise models.

We shall recall Ibragimov–Has'minskii's result briefly. Let $\mathcal{E}^\epsilon = \{\mathcal{X}^\epsilon, \mathcal{A}^\epsilon, P^\epsilon_\theta, \theta \in \Theta\}$ be a sequence of statistical experiments with $\epsilon \in (0, 1]$. Θ denotes a parameter space in $\mathbb{R}^m$. $\varphi(\epsilon)$ is a positive normalizing factor tending to zero as $\epsilon \downarrow 0$. For a $\theta_0 \in \Theta$, define a random field Z_ϵ by

$$Z_\epsilon(u) = \frac{\mathrm{d}P^\epsilon_{\theta_0+\varphi(\epsilon)u}}{\mathrm{d}P^\epsilon_{\theta_0}}(X^\epsilon)$$

for $u \in \mathbb{R}^m$. The following is a simplified version of their result.[1]

Theorem 0 (Ibragimov and Has'minskii 1972a,b, 1981) *Suppose that Z_ϵ satisfies the following conditions:*

(i) *There exist $\alpha > m$ and $k \geq \alpha$ such that for some constant C,*

$$P^\epsilon\left[|Z_\epsilon(u_2)^{1/k} - Z_\epsilon(u_1)^{1/k}|^k\right] \leq C|u_2 - u_1|^\alpha \quad (\forall u_1, u_2, \epsilon). \tag{1}$$

(ii) *For some $\gamma > 0$ and $c > 0$,*

$$P^\epsilon\left[Z_\epsilon(u) \geq e^{-c|u|^\gamma}\right] \leq e^{-c|u|^\gamma}. \tag{2}$$

(iii) *Finite-dimensional convergence: $Z_\epsilon \to^{d_f} Z$, where Z is a $\hat{C}(\mathbb{R}^m)$-valued random variable.*[2]

Then $(P^\epsilon)^{Z_\epsilon} \to \mathcal{L}\{Z\}$. Moreover,

$$P^\epsilon\left[\sup_{u:|u|\geq r} Z_\epsilon(u) > e^{-c_1 r^\gamma}\right] \leq e^{-c_1 r^\gamma}.$$

[1] P^ϵ denotes $P^\epsilon_{\theta_0}$. By convention, P^ϵ functions as the expectation for a random variable.

[2] $\hat{C}(\mathbb{R}^m)$ is the space of continuous functions on $\mathbb{R}^m$ that tends to zero at the infinity.

If $\hat{u}$ uniquely attains the maximum of $Z(u)$, then for any sequence of the maximum likelihood estimator $\hat{\theta}_\epsilon$ for θ, $\hat{u}_\epsilon := \varphi(\epsilon)^{-1}(\hat{\theta}_\epsilon - \theta_0) \to^d \hat{u}$ and moreover,

$$P^\epsilon\left[f(\hat{u}_\epsilon)\right] \to P\left[f(\hat{u})\right]$$

for every $f \in C_\uparrow(\mathbb{R}^m)$.[3]

Due to the convergence $Z_\epsilon \to^d Z$ in $\hat{C}(\mathbb{R}^m)$, one has $\sup_{u\in A} Z_\epsilon \to^d \sup_{u\in A} Z$ for every measurable set A; as a consequence, the convergence of Z_ϵ validates the convergence $\hat{u}_\epsilon = \mathrm{argmax}_u Z_\epsilon \to^d \hat{u} = \mathrm{argmax}_u Z$. In the LAN case, $Z(u)$ takes the form of $Z(u) = \exp\left(\Delta(\theta_0)[u] - 2^{-1}I(\theta_0)[u,u]\right)$, where $\Delta(\theta_0) \sim N_m(0, I(\theta_0))$ and $I(\theta_0)$ is the Fisher information matrix. In this case, the asymptotic normality of the maximum likelihood estimator is obvious from the convergence $\hat{u}_\epsilon \to^d \hat{u}$ since $\hat{u} = I(\theta_0)^{-1}\Delta(\theta_0)$. This is the first advantage of Ibragimov-Has'minskii's approach, namely the limit distribution of the maximum likelihood estimator can be obtained in a quite natural manner. Symbolically, once the convergence of the likelihood ratio random field is established, all asymptotic properties of the estimators appearing in likelihood analysis follow. However, in order to obtain the convergence $Z_\epsilon \to^d Z$ in $\hat{C}(\mathbb{R}^m)$, it is not necessary to assume such a strong condition of separation as (2) (Yoshida 1990). In this sense, it may be said that the most essential part of their approach is in the use of the large deviation estimate (2) and strong properties derived from it such as the convergence of the moments of the estimator.

Convergence of moments or the estimate of the error probability of an estimator plays an essential role in key steps of theoretical statistics. For mean bias correction to an estimator, we need the existence of the mean of it. The expected value of the plug-in functional

$$P^\epsilon\left[\log\ \text{likelihood}\ (\hat{u}_\epsilon)\right]$$

is necessary in prediction theory and in construction of information criteria. For example, the number of parameters appearing in AIC as the correction term is nothing but the mean square of $\hat{u}_\epsilon$[4] It is impossible to develop the higher-order asymptotic theory without the following type estimate:

$$\mathbf{P}\left[|\hat{\theta}_n - \theta_0| > n^{-\beta}\right] \le n^{-N}$$

for $\beta < 1/2$. Also, for the same estimates for Bayes estimators, we need an estimate of the large deviation probability for the likelihood ratio random field. The necessity of the polynomial-type large deviation inequalities in the theory of asymptotic expansion of the estimators for stochastic processes motivates this article (Yoshida 1995; Sakamoto and Yoshida 1999, 2004).

[3] The set of continuous functions on $\mathbb{R}^m$ of at most polynomial growth.

[4] In the literature of asymptotic statistics, unfortunately, such an expectation has been very often assumed to exist without any mathematical backing. Rigorous treatment of this problem under a certain integrability assumption can be seen, e.g., in Uchida and Yoshida (2001).

Kutoyants found that Ibragimov–Has'minskii's scheme could work for stochastic processes including diffusion type processes and point processes. Among his many results, Kutoyants established a complete theory for processes with small diffusions (Kutoyants 1984). Many applications of this methodology were also presented in Kutoyants (1994). We should note that the Ibragimov–Has'minskii–Kutoyants' scheme was effectively used in derivation of asymptotic expansions to ensure L^p-boundedness of the estimators for a stochastic process with small diffusions (Yoshida 1992a, 1993).

We can use (1) or another simpler continuity inequality in most cases in practice. Contrarily, as the core of the theory, the large deviation inequality (2) is not easy to obtain for most of stochastic processes, even for nonlinear ergodic diffusions; asymptotic theory for small perturbed systems is special in this sense. As seen in Kutoyants (1984), for ergodic diffusions, is necessary an explicit expression of the moment generating function of a kind of deviation in the statistical model, although it implies very strong results such as the convergence of moments once it is obtained; strong assumptions possibly produce strong results! Without such a strong assumption, the convergence of statistical random fields was proved in Yoshida (1990), however, without moment convergence.

It is an important observation, as some of statisticians have been aware and really it was implicitly written in Ibragimov and Has'minskii's papers, that the exponential type large deviation inequality like (2) is much stronger than our use. The polynomial-type large deviation inequality is sufficient to develop a theory. Here the polynomial-type large deviation inequality means

$$P^{\epsilon}\left[\sup_{u:|u|\geq r} Z_{\epsilon}(u) \geq r^{-N}\right] \leq \frac{C_N}{r^N} \quad (r > 0)$$

because the rate of the convergence of the probability is of or faster than a polynomial order. Some exponential in place of r^{-N} on the left-hand side is very often possible.

Kutoyants (2004) presented a polynomial-type large deviation inequality for one-dimensional diffusions by means of the local time. The aim of the present article is to provide a polynomial-type large deviation inequality in a more abstract setting of the partially locally asymptotically quadratic (PLAQ) sequence of experiments (without any special properties belonging to diffusion processes). From our results, for stochastic processes including nonlinear non-Gaussian time series models and semimartingales as well as multi-dimensional diffusion type processes, it is possible to obtain new convergence results, e.g., the convergence of moments of the M-estimator, and the asymptotic normality of the Bayes estimator, and its convergence of moments. Our results also provide the same consequences even for estimators based on sampled data from diffusions with/without jumps.[5] The grading of the parameters appearing later is necessary for this reason. The author hopes that our approach enables us to connect the Ibragimov–Has'minskii–Kutoyants' program in the asymptotic decision

[5] The asymptotic normality of the M-estimator is already known for jump-diffusions; see Shimizu and Yoshida (2003).

theory for stochastic processes to the *quasi*-likelihood analysis that is inevitable in statistics for sampled continuous-time processes.

Our setting includes the so-called non-ergodic experiments. There, the limit $\hat{u}$ of the estimator does not necessarily admit moments; for example, the limit distribution of the maximum likelihood estimator has the Cauchy distribution in a simple example in the non-ergodic statistics. So no theory that ensures all moments should exist. We shall deal with the order of moment more carefully than in the case of exponential type large deviation inequality.

Exponential- or polynomial-type estimates for Markov chains have been studied. Basawa and Prakasa Rao (1980) presented an exponential-type estimate. Varakin and Veretennikov (2002) used stopping times related to the Markov chain to give a polynomial-type estimate under a kind of *locally* uniform condition for a Hellinger distance. On the other hand, our method does not take advantage of the Markovian property, so it can apply to processes with functional feedback depending on the path.[6]

2 Polynomial type large deviation inequality

Let Θ be a bounded open set in $\mathbb{R}^m$ with diameter C. The unknown parameter space is the product $\Xi = \Theta \times \mathcal{T}$, where $\mathcal{T}$ is an arbitrary set. The statistical inference for the unknown parameter θ will be done based on the random fields $\mathbb{H}_T : \Omega \times \Xi \to \mathbb{R}$.[7] Later, we will consider an estimator maximizing $\mathbb{H}_T$.

We denote by $\xi_0 = (\theta_0, \tau_0)$ the true value of the unknown parameter. In order to study the asymptotic properties of the estimator for θ, we shall replace θ by another local parameter u to focus on the vicinity of $\theta_0 \in \Theta$. Let $\mathbb{U}_T(\xi_0) = \{u \in \mathbb{R}^m;\ \theta_0 + a_T(\xi_0)u \in \Theta\}$ and let $a_T(\xi_0) \in \mathrm{GL}(m)$ be a deterministic sequence in $\mathrm{GL}(m)$[8] satisfying $|a_T(\xi_0)| \to 0$ as $T \to \infty$. We mimic the local asymptotic theory to define the random field $\mathbb{Z}_T : \mathbb{U}_T(\xi_0) \times \mathcal{T} \to (0, \infty)$ by

$$\mathbb{Z}_T(u, \tau; \theta_0) = \exp\left\{\mathbb{H}_T(\theta_0 + a_T(\xi_0)u, \tau) - \mathbb{H}_T(\theta_0, \tau)\right\}.$$

In Sect. 6, we will treat an experiment generated by a sampled diffusion process, where diffusion parameter and drift parameter have different convergence rates. In estimation of the diffusion parameter, the drift parameter is in a sense redundant. However, we then need to control variables uniformly in the drift parameter. In our approach, we climb up the ladder of parameters with multi-scaling , and at each step, we apply its particular law of large numbers (plus moment estimates). By this reason, we introduced an additional parameter τ as well as θ.

---

[6] The author thanks to the referee for information about Varakin and Veretennikov (2002).

[7] Even when a sequence of probability spaces is given, we may consider a single probability space Ω without loss of generality.

[8] The general linear group over $\mathbb{R}$ of degree m.

We assume that the random field $\mathbb{Z}_T$ is partially locally asymptotically quadratic at $\theta_0 \in \Theta$ under P_{ξ_0}:

$$\mathbb{Z}_T(u, \tau; \theta_0) = \exp\left(\Delta_T(\tau; \xi_0)[u] - \frac{1}{2}\Gamma(\tau; \xi_0)[u, u] + r_T(u, \tau; \xi_0)\right) \tag{3}$$

for $(u, \tau) \in \mathbb{U}_T(\theta_0) \times \mathcal{T}$. Here $\Delta_T(\tau; \xi_0)$ is a random linear form on $\mathbb{R}^m$ and $\Gamma(\tau; \xi_0)$ a random bilinear form on $\mathbb{R}^m \times \mathbb{R}^m$, and both are defined on the same probability space as $\mathbb{Z}_T$. The residual term $r_T(u, \tau; \xi_0)$ is expected, under P_{ξ_0}, to be small when $T \to \infty$, but we shall specify its property more precisely later. Here, we wrote ξ_0 in "Δ_T", "Γ" and "r_T" because in the decomposition on the right-hand side, each variable's behavior depends on the true parameter value ξ_0.

In what follows, we fix a set $K \subset \Xi$ and also fix a positive number L. For $\xi_0 \in K$, set $b_T(\xi_0) = \left\{\lambda_{\min}(a_T(\xi_0)' a_T(\xi_0))\right\}^{-1} \uparrow \infty$, and assume that $b_T(\xi_0)^{-1} \leq \lambda_{\max}(a_T(\xi_0)' a_T(\xi_0)) \leq C_1 b_T(\xi_0)^{-1}$ for some constant $C_1 \in [1, \infty)$. Here $\lambda_{\min}$ and $\lambda_{\max}$ denote the minimum eigenvalue and the maximum eigenvalue, respectively.

Let $\alpha \in (0, 1)$, $\delta_T(\xi_0) = b_T(\xi_0)^{-\alpha/2}$, and $U_T(r, \xi_0) = \{u \in \mathbb{U}_T(\xi_0);\ r \leq |u| \leq \delta_T(\xi_0)\, b_T(\xi_0)^{1/2}\}$. Let $\epsilon_1(r) = r^{-\rho_1}$ for some ρ_1, and let $S_T(r; \xi_0) = \{\omega;\ \sup_{(u,\tau)\in U_T(r,\xi_0)\times\mathcal{T}} (1 + |u|^2)^{-1} |r_T(u, \tau; \xi_0)| < \epsilon_1(r)\}$. We will assume[9]

[A1] For some constant C_L,

$$\sup_{\substack{T > 0 \\ \xi_0 \in K}} P_{\xi_0}\left[S_T(r; \xi_0)^c\right] \leq \frac{C_L}{r^L} \quad (r > 0).$$

This is a bit technical condition to estimate the residual term; however, we will replace Condition [A1] later by more handy conditions.

As seen in a basic example of the non-ergodic statistics, the nondegeneracy of the observed information is closely related to the existence of moments of the estimator. Let

$$A(\epsilon, \xi_0) = \left\{\omega;\ \epsilon |u|^2 \leq \frac{1}{4}\Gamma(\tau; \xi_0)[u, u] \text{ for all } (u, \tau) \in \mathbb{R}^m \times \mathcal{T}\right\}$$

and assume

[A2] For some constant C_L,

$$\sup_{\xi_0 \in K} P_{\xi_0}[A(\epsilon_1(r), \xi_0)^c] \leq \frac{C_L}{r^L} \quad (r > 0).$$

[9] We implicitly assume the measurability of the event $S_T(r; \xi_0)$. In most cases, the separability of the random field $\mathbb{H}_T$ ensures the measurability. Because one gains little in practice in removing such a regularity condition, by introducing the outer measure for example, we do not pursue it here. In what follows also, we will put measurability assumptions implicitly.

In our argument, a global identifiability (or distinguishability) condition will play an essential role. Let $\mathbb{Y} : \Theta \times \mathcal{T} \times K \to \mathbb{R}$ be a random field. It is expected to be the limit of $\mathbb{Y}_T$ defined by

$$\mathbb{Y}_T(\theta, \tau; \xi_0) = \frac{1}{b_T(\xi_0)} \left(\mathbb{H}_T(\theta, \tau) - \mathbb{H}_T(\theta_0, \tau)\right).$$

We assume

[A3] For each $\xi_0 \in K$, there are a positive random variable $\chi(\xi_0)$ and a positive (deterministic) constant $\rho = \rho(\xi_0)$ such that

$$\mathbb{Y}(\theta_1, \tau; \xi_0) = \mathbb{Y}(\theta_1, \tau; \xi_0) - \mathbb{Y}(\theta_0, \tau; \xi_0) \le -\chi(\xi_0)|\theta_1 - \theta_0|^\rho$$

for all $\theta_1 \in \Theta$ and all $\tau \in \mathcal{T}$, under P_{ξ_0}.

The parameter $\chi(\xi_0)$ denotes the degree of separation of models, depending on the randomness that remains in the limit. On the other hand, the constant ρ may be regarded as reflecting the smoothness of the parametric model. We often observe that $\rho = 2$.

[A4] Parameters ρ_1, ρ_2 and β_2 satisfy the following inequalities: $0 < \rho_1 < 1, \alpha\rho < \rho_2$, $\beta_2 \ge 0$ and $1 - 2\beta_2 - \rho_2 > 0$.

Roughly speaking, we take small α, and small ρ_2 and β_2 satisfying $\alpha\rho < \rho_2$ and $1 - 2\beta_2 - \rho_2 > 0$.

[A5] For some constant C_L,

$$\sup_{\xi_0 \in K} P_{\xi_0}\left[\chi(\xi_0) \le r^{-(\rho_2 - \alpha\rho)}\right] \le \frac{C_L}{r^L} \quad (r > 0).$$

We will assume a few moment conditions.

[A6] For $M_1 = L(1 - \rho_1)^{-1}$,

$$\sup_{\xi_0 \in K} \sup_{T>0} P_{\xi_0}\left[\left(\sup_{\tau \in \mathcal{T}} |\Delta_T(\tau; \xi_0)|\right)^{M_1}\right] < \infty.$$

For $M_2 = L(1 - 2\beta_2 - \rho_2)^{-1}$,

$$\sup_{\xi_0 \in K} \sup_{T>0} P_{\xi_0}\left[\left(\sup_{\substack{h : \delta_T(\xi_0) \le |h| \\ \tau \in \mathcal{T}}} b_T(\xi_0)^{\frac{1}{2} - \beta_2} |\mathbb{Y}_T(\theta_0 + h, \tau; \xi_0) - \mathbb{Y}(\theta_0 + h, \tau; \xi_0)|\right)^{M_2}\right] < \infty.$$

Remark 1 $\Gamma(\tau, \xi_0)$ and $\chi(\xi_0)$ are in general random. An explosive Ornstein–Uhlenbeck process is the case; really, random limiting information is common in the so-called *non-ergodic* statistics. Condition [A2] restricts the probability of $\Gamma(\tau, \xi_0)$

near zero. Condition [A5] for $\chi(\xi_0)$ is similar. Those conditions would seem more real if we would consider a multi-dimensional explosive Ornstein–Uhlenbeck process, where the limit of the conditional Fisher information after normalization has a chi-square distribution $\chi^2(\nu)$ with $\nu \geq 1$. See Yoshida (1988) for details. If $\Gamma(\tau, \xi_0)$ and $\chi(\xi_0)$ are constants (*ergodic* case), we can remove Conditions [A2] and [A5].

Let $V_T(r, \xi_0) = \{u \in \mathbb{U}_T(\xi_0);\ r \leq |u|\}$. We denote by C_L generic constants depending on L. For functions $f(x)$ and $g(x)$, the symbol $f(x) \lesssim g(x)$ stands for the relation $f(x) \leq cg(x)$ for some constant c independent of x. The following theorem gives a polynomial type large deviation inequality for the statistical random field:

Theorem 1 *Given* $L > 0$*, assume Conditions* [A1]–[A6]. *Then there exists a constant* C_L *such that*

$$\sup_{\xi_0 \in K} P_{\xi_0}\left[\sup_{(u,\tau) \in V_T(r,\xi_0) \times \mathcal{T}} \mathbb{Z}_T(u, \tau; \theta_0) \geq e^{-r^{2-(\rho_1 \vee \rho_2)}/2}\right] \leq \frac{C_L}{r^L} \tag{4}$$

for all $T > 0$ *and* $r > 0$. *Here the supremum of the empty set should read* $-\infty$ *by convention.*

Proof Set $\epsilon_2(r) = r^{-\rho_2}$ and $\epsilon_3(r) = r^{-\beta_2}$. In order to prove Theorem 1, it suffices to estimate the following two probabilities:

$$P_{\xi_0}\left[\sup_{(u,\tau) \in U_T(r,\xi_0) \times \mathcal{T}} \mathbb{Z}_T(u, \tau; \theta_0) \geq e^{-\epsilon_1(r) r^2/2}\right] \tag{5}$$

and

$$P_{\xi_0}\left[\sup_{(u,\tau) \in (V_T(r,\xi_0) \setminus U_T(r,\xi_0)) \times \mathcal{T}} \mathbb{Z}_T(u, \tau; \theta_0) \geq e^{-\epsilon_2(r) r^2/2}\right]. \tag{6}$$

(a) Estimate of the probability (5). By assumption,

$$P_{\xi_0}\left[S_T(r; \xi_0)^c\right] \leq \frac{C_L}{r^L}$$

and

$$P_{\xi_0}\left[A(\epsilon_1(r), \xi_0)^c\right] \leq \frac{C_L}{r^L}.$$

From these inequalities, we obtain

$$
\begin{aligned}
&P_{\xi_0}\left[\sup_{(u,\tau)\in U_T(r,\xi_0)\times\mathcal{T}} \mathbb{Z}_T(u,\tau;\theta_0) \geq e^{-\epsilon_1(r)r^2/2}\right] \\
&\leq \frac{C_L}{r^L} + P_{\xi_0}\left[\sup_{(u,\tau)\in U_T(r,\xi_0)\times\mathcal{T}} \left(|u||\Delta_T(\tau;\xi_0)| - \epsilon_1(r)|u|^2\right)\right. \\
&\quad \left. +\epsilon_1(r) \geq -\frac{1}{2}\epsilon_1(r)r^2\right] \\
&\leq \frac{C_L}{r^L} + P_{\xi_0}\left[\sup_{\tau\in\mathcal{T}} |\Delta_T(\tau;\xi_0)| > 2\epsilon_1(r)r\right] \\
&\quad + P_{\xi_0}\left[r\sup_{\tau\in\mathcal{T}} |\Delta_T(\tau;\xi_0)| - \epsilon_1(r)r^2 + \epsilon_1(r) \geq -\epsilon_1(r)r^2/2\right] \\
&\leq \frac{C_L}{r^L} + 2P_{\xi_0}\left[\sup_{\tau\in\mathcal{T}} |\Delta_T(\tau;\xi_0)| > \epsilon_1(r)r/2 - \epsilon_1(r)/r\right] \\
&\leq \frac{C_L}{r^L} + 2\left(\epsilon_1(r)r/2 - \epsilon_1(r)/r\right)^{-M_1} P_{\xi_0}\left[\left(\sup_{\tau\in\mathcal{T}} |\Delta_T(\tau;\xi_0)|\right)^{M_1}\right] \\
&\lesssim \frac{1}{r^L} + \left(r^{1-\rho_1}\right)^{-M_1} P_{\xi_0}\left[\left(\sup_{\tau\in\mathcal{T}} |\Delta_T(\tau;\xi_0)|\right)^{M_1}\right] \\
&\lesssim \frac{1}{r^L}.
\end{aligned}
$$

(b) Estimate of the probability (6). We know that $r \leq Cb_T(\xi_0)^{1/2}$ essentially, C being the diameter of Θ, because for $h = a_T(\xi_0)u$,

$$
\begin{aligned}
r &\leq |u| \leq |a_T(\xi_0)^{-1}h| \leq \{\lambda_{\min}(a_T(\xi_0)a_T(\xi_0)')\}^{-1/2}|h| \\
&= \{\lambda_{\min}(a_T(\xi_0)'a_T(\xi_0))\}^{-1/2}|h| \leq b_T(\xi_0)^{1/2}C.
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
&P_{\xi_0}\left[\sup_{\substack{h:\,\delta_T(\xi_0)\leq|h| \\ \tau\in\mathcal{T}}} \mathbb{Y}(\theta_0+h,\tau;\xi_0) \geq -C^2\epsilon_2\left(Cb_T(\xi_0)^{1/2}\right)\right] \\
&\leq P_{\xi_0}\left[\chi(\xi_0) \leq C^{2-\rho_2} b_T(\xi_0)^{-(\rho_2-\alpha\rho)/2}\right] \\
&\lesssim b_T(\xi_0)^{-L/2} \\
&\lesssim r^{-L}.
\end{aligned}
$$

With this estimate, we have

$$
\begin{aligned}
&P_{\xi_0}\left[\sup_{(u,\tau)\in(V_T(r,\xi_0)\setminus U_T(r,\xi_0))\times\mathcal{T}} \mathbb{Z}_T(u,\tau;\theta_0) \geq e^{-\epsilon_2(r)r^2/2}\right] \\
&\quad \leq P_{\xi_0}\left[\sup_{\substack{h\,:\,\delta_T(\xi_0)\leq|h| \\ \tau\in\mathcal{T}}} \mathbb{Y}_T(\theta_0+h,\tau;\xi_0) \geq -\frac{C^2}{2}\epsilon_2\left(Cb_T(\xi_0)^{1/2}\right)\right] \\
&\quad \leq P_{\xi_0}\left[\sup_{\substack{h\,:\,\delta_T(\xi_0)\leq|h| \\ \tau\in\mathcal{T}}} \mathbb{Y}(\theta_0+h,\tau;\xi_0) \geq -C^2\epsilon_2\left(Cb_T(\xi_0)^{1/2}\right)\right] \\
&\qquad +P_{\xi_0}\left[\sup_{\substack{h\,:\,\delta_T(\xi_0)\leq|h| \\ \tau\in\mathcal{T}}} |\mathbb{Y}_T(\theta_0+h,\tau;\xi_0)-\mathbb{Y}(\theta_0+h,\tau;\xi_0)| \geq \frac{C^2}{2}\epsilon_2\left(Cb_T(\xi_0)^{1/2}\right)\right] \\
&\quad \leq O(r^{-L}) + \left(\frac{C^2}{2}b_T(\xi_0)^{1/2}\epsilon_3(b_T(\xi_0))\epsilon_2\left(Cb_T(\xi_0)^{1/2}\right)\right)^{-M_2} \\
&\qquad \times P_{\xi_0}\left[\left(\sup_{\substack{h\,:\,\delta_T(\xi_0)\leq|h| \\ \tau\in\mathcal{T}}} b_T(\xi_0)^{1/2}\epsilon_3(b_T(\xi_0))\,|\mathbb{Y}_T(\theta_0+h,\tau;\xi_0)-\mathbb{Y}(\theta_0+h,\tau;\xi_0)|\right)^{M_2}\right] \\
&\quad \lesssim r^{-L} + b_T(\xi_0)^{-M_2(\frac{1}{2}-\beta_2-\frac{1}{2}\rho_2)} \\
&\quad \lesssim r^{-L} + r^{-M_2(1-2\beta_2-\rho_2)} \\
&\quad \lesssim \frac{1}{r^L}.
\end{aligned}
$$

□

Remark 2 If the inequalities in the assumptions are supposed to hold for exponential bounds, then we would have usual exponential-type large deviation inequality as a consequence.

3 Simple sufficient conditions

In this section, a set of simple sufficient conditions for [A1] will be provided. Though it is possible to relax those conditions, we will give simplicity priority over weakness.

3.1 Contrast function of class C^2

Suppose that $\mathbb{H}_T(\theta,\tau)$ is of class C^2 in θ for every $\omega\in\Omega$. We define $\Delta_T(\tau;\xi_0)$ and a random matrix $\Gamma_T(\theta,\tau;\xi_0)$ by

$$\Delta_T(\tau;\xi_0)[u] = \partial_\theta\mathbb{H}_T(\theta_0,\tau)[a_T(\xi_0)u]$$

and

$$\Gamma_T(\theta,\tau;\xi_0)[u,u] = -\partial_\theta^2 \mathbb{H}_T(\theta,\tau)[a_T(\xi_0)u, a_T(\xi_0)u]$$

for $u \in \mathbb{R}^m$. Then for $\Gamma(\tau;\xi_0) = \Gamma(\theta_0,\tau;\xi_0)$ and

$$r_T(u,\tau;\xi_0) = \int_0^1 (1-s)\,\{\Gamma(\tau;\xi_0)[u,u] - \Gamma_T(\theta_0 + s a_T(\xi_0)u,\tau;\xi_0)[u,u]\}\,\mathrm{d}s,$$

we have the PLAQ representation (3) for $\mathbb{Z}_T(u,\tau;\theta_0)$.

Let

$$S'_T(r;\xi_0) = \left\{ \omega;\ \sup_{\substack{(h,\tau)\,\in\,(\Theta-\theta_0)\times\mathcal{T}:\\ b_T(\xi_0)^{-1/2} r \le |h| \le C_1^{1/2}\delta_T(\xi_0)}} |\Gamma_T(\theta_0+h,\tau;\xi_0) - \Gamma(\tau;\xi_0)| < \epsilon_1(r) \right\}.$$

The following is a sufficient condition for [A1]:

[A1′] For some constant C_L,

$$\sup_{\substack{T>0\\ \xi_0\in K}} P_{\xi_0}\left[S'_T(r;\xi_0)^c\right] \le \frac{C_L}{r^L} \quad (r>0).$$

3.2 Contrast function of class C^3

Condition [A1′] is a uniform law of large numbers with convergence rate. It is often verified by applying a uniform central limit theorem. Here, we assume that Θ admits Sobolev's inequality or GRR-inequality (see Sects. 6 and 7) as well as being bounded open, and that the random field $\mathbb{H}_T : \Omega\times\Xi \to \mathbb{R}$ is of class C^3 in θ, and remove the requirement of the uniform (in θ) convergence from the condition.

Set $\beta = \alpha/(1-\alpha)$. We replace Condition [A4] by

[A4′] Parameters β_1, ρ_1, ρ_2 and β_2 satisfy the following inequalities: $0<\beta_1<1/2$, $0<\rho_1<\min\{1,\beta,2\beta_1/(1-\alpha)\}$, $\alpha\rho<\rho_2$, $\beta_2\ge 0$ and $1-2\beta_2-\rho_2>0$.

Furthermore, we assume the following additional conditions:

[A1″] For $M_3 = L(\beta-\rho_1)^{-1}$,

$$\sup_{\substack{T>0\\ \xi_0\in K}} P_{\xi_0}\left[\left(b_T(\xi_0)^{-1} \sup_{(\theta,\tau)\in\Theta\times\mathcal{T}} \left|\partial_\theta^3 \mathbb{H}_T(\theta,\tau)\right|\right)^{M_3}\right] < \infty.$$

Moreover, for $M_4 = L(\frac{2\beta_1}{1-\alpha} - \rho_1)^{-1}$,

$$\sup_{\substack{T>0\\ \xi_0\in K}} P_{\xi_0}\left[\sup_{\tau\in\mathcal{T}} \left(b_T(\xi_0)^{\beta_1}\,|\Gamma_T(\theta_0,\tau;\xi_0) - \Gamma(\tau;\xi_0)|\right)^{M_4}\right] < \infty.$$

Lemma 1 *Condition* [A1′] *holds under* [A4′] *and* [A1″].

Proof We may assume that $r \le \sqrt{C_1}\delta_T(\xi_0)b_T(\xi_0)^{1/2}$. Otherwise, the probability $P_{\xi_0}\left[S'_T(r;\xi_0)^c\right]$ is equal to zero. When $b_T(\xi_0)^{-1/2}r \le |h| \le \sqrt{C_1}\delta_T(\xi_0)$, since

$$b_T(\xi_0)^{-1} \le \left(\frac{\sqrt{C_1}}{r}\right)^{2/(1-\alpha)},$$

we have

$$|h| \le \sqrt{C_1}b_T(\xi_0)^{-\alpha/2} \le C_1^{(1+\beta)/2}r^{-\beta}.$$

Therefore, for some positive constant $C(m)$,

$$\begin{aligned}
&P_{\xi_0}\left[S'_T(r;\xi_0)^c\right] \\
&\le P_{\xi_0}\left[C(m)r^{-\beta}b_T(\xi_0)^{-1}\sup_{(\theta,\tau)\in\Theta\times\mathcal{T}}\left|\partial_\theta^3\mathbb{H}_T(\theta,\tau)\right| \ge \epsilon_1(r)/2\right] \\
&\quad + P_{\xi_0}\left[\sup_{\tau\in\mathcal{T}}|\Gamma_T(\theta_0,\tau;\xi_0)-\Gamma(\tau;\xi_0)| \ge \epsilon_1(r)/2\right] \\
&\lesssim (r^\beta\epsilon_1(r))^{-M_3}P_{\xi_0}\left[\left(b_T(\xi_0)^{-1}\sup_{(\theta,\tau)\in\Theta\times\mathcal{T}}\left|\partial_\theta^3\mathbb{H}_T(\theta,\tau)\right|\right)^{M_3}\right] \\
&\quad + (b_T(\xi_0)^{\beta_1}\epsilon_1(r))^{-M_4}P_{\xi_0}\left[\sup_{\tau\in\mathcal{T}}\left(b_T(\xi_0)^{\beta_1}|\Gamma_T(\theta_0,\tau;\xi_0)-\Gamma(\tau;\xi_0)|\right)^{M_4}\right] \\
&\lesssim (r^\beta\epsilon_1(r))^{-M_3} + (r^{\frac{2\beta_1}{1-\alpha}}\epsilon_1(r))^{-M_4} \\
&\lesssim r^{-M_3(\beta-\rho_1)} + r^{-M_4(\frac{2\beta_1}{1-\alpha}-\rho_1)} \\
&\lesssim \frac{1}{r^L}.
\end{aligned}$$

□

Obviously we obtain

Theorem 2 *Let* $L > 0$. *Suppose that Conditions* [A1″], [A2],[A3], [A4′],[A5] *and* [A6] *are satisfied. Then there exists a constant* C_L *such that Inequality (4) holds for all* $T > 0$ *and* $r > 0$.

3.3 Deterministic limit information

Let us consider the situation where $\Gamma(\tau;\xi_0)$ and $\chi(\xi_0)$ are deterministic. It is the case for the locally asymptotically normal family of statistical experiments. Some of conditions in Sect. 2 can be removed or simplified.

L is a given positive number and $K \subset \Xi$ as before.

[B1] The matrix $\Gamma(\tau;\xi_0)$ is deterministic and positive definite uniformly in $\tau \in \mathcal{T}$ and $\xi_0 \in K$.

[B2] For K, there are a deterministic positive number $\chi = \chi(K)$ and a positive deterministic constant $\rho = \rho(K)$ such that

$$\mathbb{Y}(\theta_1, \tau; \xi_0) = \mathbb{Y}(\theta_1, \tau; \xi_0) - \mathbb{Y}(\theta_0, \tau; \xi_0) \leq -\chi|\theta_1 - \theta_0|^\rho$$

for all $\theta_1 \in \Theta$ and all $\tau \in \mathcal{T}$, under P_{ξ_0}.

Theorem 3 *For some constant C_L, Inequality (4) is valid for all $T > 0$ and $r > 0$ under each one of the following sets of conditions:*

(a) *Conditions* [A1], [A4], [A6], [B1] *and* [B2] *are satisfied,*
(b) *Conditions* [A1′], [A4], [A6], [B1] *and* [B2] *are satisfied and the random field $\mathbb{H}_T$ is of class C^2,*
(c) *Conditions* [A1″], [A4′], [A6], [B1] *and* [B2] *are satisfied and the random field $\mathbb{H}_T$ is of class C^3.*

4 Estimate of the modulus of continuity and convergence of the estimator

Assume that $\mathcal{T}$ is a bounded open set in $\mathbb{R}^{m'}$.[10] In this and the following sections, we assume that the random field $\mathbb{H}_T$ on Ξ is continuous and can be continuously extended to the boundary $\partial\Xi$.

The **M-estimator** $(\hat{\theta}_T, \hat{\tau}_T)$ for (θ, τ) is a measurable mapping that satisfies

$$\mathbb{H}_T(\hat{\theta}_T, \hat{\tau}_T) = \sup_{(\theta,\tau)\in\bar{\Xi}} \mathbb{H}_T(\theta, \tau).$$

That is, the pair $(\hat{\theta}_T, \hat{\tau}_T)$ that maximizes the utility function $\mathbb{H}_T$ is to be called the M-estimator. Let $\hat{u}_T = a_T(\xi_0)^{-1}(\hat{\theta}_T - \theta_0)$. Letter p will denote a positive number.

Before going to limit theorems for the M-estimator, we should here notice that the L^p-estimate of $\hat{u}_T$ follows from the polynomial type large deviation inequality without any specification of the limit of $\mathbb{Z}_T(\cdot, \cdot; \theta_0)$.

Proposition 1 *Let $L > p > 0$. Suppose that there exists a constant C_L such that*

$$\sup_{\xi_0\in K} P_{\xi_0}\left[\sup_{(u,\tau)\in V_T(r,\xi_0)\times\mathcal{T}} \mathbb{Z}_T(u, \tau; \theta_0) \geq 1\right] \leq \frac{C_L}{r^L}$$

for all $T > 0$ and $r > 0$. Then for any sequence of M-estimators $(\hat{\theta}_T, \hat{\tau}_T)$, it holds that

$$\sup_{\xi_0\in K}\sup_{T>0} P_{\xi_0}\left[|\hat{u}_T|^p\right] < \infty. \tag{7}$$

[10] It is just for simplifying descriptions. Finite dimensionality of $\mathcal{T}$ is not essential.

In particular, Inequality (7) holds under Conditions [A1]-[A6].

For $c > 0$, let

$$B_{c,T} = \{u \in \mathbb{R}^m;\ |u| < c,\ \theta_0 + a_T(\xi_0)u \in \Theta\},$$

and for $c > 0$ and $\delta > 0$, let

$$w_T(\delta, c) = \sup_{\substack{u_1, u_2 \in B_{c,T}, |u_2 - u_1| \le \delta \\ \tau \in \mathcal{T}}} \left|\log \mathbb{Z}_T(u_2, \tau; \theta_0) - \log \mathbb{Z}_T(u_1, \tau; \theta_0)\right|.$$

[C1] For every $\epsilon > 0$ and $c > 0$, $\lim_{\delta \downarrow 0} \sup_T P_{\xi_0}[w_T(\delta, c) > \epsilon] = 0$.

Remark 3 Under regularity conditions, it is easy to verify Condition [C1] in a quite similar way as Lemma 4.1 of Yoshida (1990). See Sect. 7.

For $c > 0$ and $\delta > 0$, define $w_T^\dagger(\delta, c)$ by

$$w_T^\dagger(\delta, c) = \sup_{\substack{u \in B_{c,T} \\ \tau_1, \tau_2 \in \mathcal{T}, |\tau_2 - \tau_1| \le \delta}} \left|\log \mathbb{Z}_T(u, \tau_2; \theta_0) - \log \mathbb{Z}_T(u, \tau_1; \theta_0)\right|.$$

We assume the condition

[C2] $\lim_{\delta \downarrow 0} \sup_T P_{\xi_0}\left[w_T^\dagger(\delta, c) > \epsilon\right] = 0$ for all $\epsilon > 0$ and $c > 0$.

Let $\hat{C}(\mathbb{R}^m \times \mathcal{T})$ be the space of continuous functions on $\mathbb{R}^m \times \mathcal{T}$ that have limits as (u, τ) tends to the boundary of $\mathbb{R}^m \times \mathcal{T}$, and tend to zero uniformly in τ as $|u| \to \infty$. Equip $\hat{C}(\mathbb{R}^m \times \mathcal{T})$ with the supremum norm.

We extend $\mathbb{Z}_T(\cdot, \cdot; \theta_0)$ to a function in $\hat{C}(\mathbb{R}^m \times \mathcal{T})$ that is not greater than $\mathbb{Z}_T(\cdot, \cdot; \theta_0)$, and denote it by the same symbol.

Let $\mathbb{Z}(\cdot, \cdot; \theta_0)$ be a $\hat{C}(\mathbb{R}^m \times \mathcal{T})$-valued random variable.

[C3] $\mathbb{Z}_T(\cdot, \cdot; \theta_0) \to^{d_f} \mathbb{Z}(\cdot, \cdot; \theta_0)$ (finite-dimensional convergence).

Here are some applications of the estimates of the statistical random field $\mathbb{Z}_T$.

Theorem 4 *Assume Conditions* [C1], [C2] *and* [C3].

(a) *Extended $\mathbb{Z}_T(\cdot, \cdot; \theta_0) \to^d \mathbb{Z}(\cdot, \cdot; \theta_0)$ in $\hat{C}(\mathbb{R}^m \times \mathcal{T})$ as $T \to \infty$ if for some sequence of positive numbers $\epsilon(r)$ tending to zero as $r \to \infty$,*

$$P_{\xi_0}\left[\sup_{(u,\tau) \in V_T(r,\xi_0) \times \mathcal{T}} \mathbb{Z}_T(u, \tau; \theta_0) \ge \epsilon(r)\right] \le \epsilon(r)$$

for all $T > 0$ and $r > 0$.

(b) *Assume that the limit random field* $\mathbb{Z}$ *of* $\mathbb{Z}_T(\cdot,\cdot;\theta_0)$ *(as in* (a)*) does not depend on* $\tau \in \mathcal{T}$ *and that there exists a random variable* $\hat{u}$ *such that*[11]

$$\mathbb{Z}(\hat{u};\theta_0) = \sup_{u\in\mathbb{R}^m} \mathbb{Z}(u;\theta_0)$$

and $\hat{u}$ *is a (random) unique maximum point a.s. Then*

$$\hat{u}_T \to^d \hat{u}$$

as $T \to \infty$.

(c) *Assume that* $\hat{u}_T \to^d \hat{u}$. *If there exists a number* $L > p$ *and a constant* C_L *such that*

$$P_{\xi_0}\left[\sup_{(u,\tau)\in V_T(r,\xi_0)\times\mathcal{T}} \mathbb{Z}_T(u,\tau;\theta_0) \geq 1\right] \leq \frac{C_L}{r^L}$$

for all $T > 0$ *and* $r > 0$, *then* $P_{\xi_0}\left[f(\hat{u}_T)\right] \to \mathbb{P}\left[f(\hat{u})\right]$ *as* $T \to \infty$ *for all continuous functions* f *satisfying* $\limsup_{|u|\to\infty} |f(u)||u|^{-p} < \infty$. *(*$\mathbb{P}$ *denotes the probability measure on the probability space on which the limit variables are realized.)*

Proof The weak convergence in (a) is a consequence of the tightness of $\{\mathbb{Z}_T(\cdot,\cdot;\theta_0)\}_{T\geq T_0}$. Showing (b) is also easy: for any closed set F in $\mathbb{R}^m$, we have

$$\begin{aligned}
&\limsup_{T\to\infty} P_{\xi_0}[\hat{u}_T \in F] \\
&\quad\leq \limsup_{T\to\infty} P_{\xi_0}\left[\sup_{u\in F} \mathbb{H}_T(\theta_0 + a_T(\xi_0)u, \hat{\tau}_T) - \mathbb{H}_T(\theta_0, \hat{\tau}_T)\right. \\
&\quad\left.\geq \sup_{u\in F^c} \mathbb{H}_T(\theta_0 + a_T(\xi_0)u, \hat{\tau}_T) - \mathbb{H}_T(\theta_0, \hat{\tau}_T)\right] \\
&\quad\leq \limsup_{T\to\infty} P_{\xi_0}\left[\sup_{u\in F}\sup_{\tau\in\mathcal{T}} \mathbb{Z}_T(u,\tau;\theta_0) - \sup_{u\in F^c}\inf_{\tau\in\mathcal{T}} \mathbb{Z}_T(u,\tau;\theta_0) \in \mathbb{R}_+\right] \\
&\quad\leq \mathbb{P}\left[\sup_{u\in F}\mathbb{Z}(u;\theta_0) - \sup_{u\in F^c}\mathbb{Z}(u;\theta_0) \in \mathbb{R}_+\right] \\
&\quad\leq \mathbb{P}\left[\hat{u} \in F\right].
\end{aligned}$$

Therefore, $\hat{u}_T \to^d \hat{u}$. And now (c) is obvious: we can change p in Proposition 1 for uniform integrability. □

Remark 4 If we assume that the conditions are satisfied uniformly in $\xi_0 \in \mathcal{K}$ for a set $\mathcal{K}$ in Ξ and that the convergence of $a_T(\xi_0)$ is uniform as well, then the convergences in

[11] We denote by $\mathbb{Z}(u;\theta_0)$ the random field $\mathbb{Z}(u,\tau;\theta_0)$ independent of τ.

(a) and (c) hold uniformly in $\mathcal{K}$. Additionally, if the uniform regularity condition [UR] below for $Z(u) = \mathbb{Z}(u; \theta_0)$ is satisfied, then the consequence in (b) is valid uniformly in $\mathcal{K}$.

[UR] For every $t \in \mathbb{R}^m$, the distribution function of $\sup_{u \leq t} Z(u) - \sup_{u \not\leq t} Z(u)$ is continuous at 0 uniformly in $\mathcal{K}$.

Here $\{u; u \leq t\}$ denotes the rectangle at the lower left of t.

Remark 5 Obviously, each convergence in distribution in the results stated above can be strengthen to the stable convergence if each assumption of convergence in distribution is replaced by the stable convergence. It is also the case for the results in the following sections.

5 Joint convergence

For each $k \in \{1, \ldots, K\}$, let Θ_k be a bounded open set[12] in $\mathbb{R}^{d_k}$. $\theta = (\theta_1, \ldots, \theta_K)$ is a blocked vector in $\Theta_1 \times \cdots \times \Theta_K$. The true value of θ is denoted by $\theta^* = (\theta_1^*, \ldots, \theta_K^*)$. We assume that $\mathcal{T}$ is a one-point set and omit τ from notation. $a_T^k(\theta_k^*)$ is a deterministic invertible $d_k \times d_k$-matrix and set $a_T(\theta^*) = \mathrm{diag}(a_T^1(\theta_1^*), \ldots, a_T^K(\theta_K^*))$. We write $\theta_k^\dagger$ for $\theta_k^* + a_T^k(\theta_k^*) u_k$, $u_k \in \mathbb{R}^{d_k}$. Also write $\underline{\theta}_k = (\theta_1, \ldots, \theta_k)$ and $\overline{\theta}_k = (\theta_k, \ldots, \theta_K)$. Parameters $\underline{\theta}_k^*$ and $\overline{\theta}_k^*$ are defined in a similar manner. Write $u = (u_1, \ldots, u_K)$, $\theta^\dagger = (\theta_1^\dagger, \ldots, \theta_K^\dagger)$ and $d = \sum_{k=1}^K d_k$.

Let $\Theta = \prod_{k=1}^K \Theta_k$. We consider a sequence of statistical random fields $\mathbb{H}_T : \Omega \times \Theta \to \mathbb{R}$. The exponential statistical random field is now given by

$$\mathbb{Z}_T(u; \theta^*) = \exp\left(\mathbb{H}_T(\theta^\dagger) - \mathbb{H}_T(\theta^*)\right).$$

The kth random field $Z_T^k(u_k; \underline{\theta}_{k-1}, \theta_k^*, \overline{\theta}_{k+1})$ is defined by

$$\mathbb{Z}_T^k(u_k; \underline{\theta}_{k-1}, \theta_k^*, \overline{\theta}_{k+1}) = \exp\left(\mathbb{H}_T(\underline{\theta}_{k-1}, \theta_k^\dagger, \overline{\theta}_{k+1}) - \mathbb{H}_T(\underline{\theta}_{k-1}, \theta_k^*, \overline{\theta}_{k+1})\right). \quad (8)$$

By convention, we neglect symbols with index $K+1$ like $\overline{\theta}_{K+1}$ and ones with index 0 like $\underline{\theta}_0$.

Let

$$V_T^k(r, \theta_k^*) = \{u_k \in \mathbb{U}_T^k(\theta_k^*);\ r \leq |u_k|\},$$

where

$$\mathbb{U}_T^k(\theta_k^*) = \{u_k \in \mathbb{R}^{d_k};\ \theta_k^* + a_T^k(\theta_k^*) u_k \in \Theta_k\}.$$

Let L be a given positive number.

[12] A bounded closed domain is also available if the reader would like to insure for the existence of an exact M-estimator though an asymptotic M-estimator is sufficient to consider in asymptotic theory.

5.1 Convergence of the moments of a multi-scaling M-estimator

Suppose that $\mathbb{H}_T$ can be extended continuously to $\overline{\Theta}$ a.s. We obtain a polynomial type estimate for the M-estimator if we apply Theorem 1 stepwise. More precisely, we have

Proposition 2 *Let $\hat{\theta} \equiv \hat{\theta}_T = (\hat{\theta}_1, \ldots, \hat{\theta}_K)$ be any sequence of M-estimators maximizing $\mathbb{H}_T$. Given $L > 0$ and $\mathcal{K} \subset \Theta$, suppose that the following polynomial type estimates are valid:*

$$\sup_{\theta^* \in \mathcal{K}} P_{\theta^*}\left[\sup_{(u_k, \overline{\theta}_{k+1}) \in V_T^k(r, \theta_k^*) \times \prod_{j=k+1}^K \Theta_j} \mathbb{Z}_T^k(u_k; \underline{\hat{\theta}}_{k-1}, \theta_k^*, \overline{\theta}_{k+1}) \geq 1 \right] \leq \frac{C_L}{r^L}$$

for all $T > 0$, $r > 0$ and $k = 1, \ldots, K$. Then the normalized M-estimator $\hat{u}_T = a_T(\theta^*)^{-1}(\hat{\theta}_T - \theta^*)$ satisfies the polynomial type large deviation inequality

$$\sup_{\theta^* \in \mathcal{K}} P_{\theta^*}\left[|\hat{u}_T| > r \right] \leq \frac{C_L}{r^L} \tag{9}$$

for all $T > 0$ and $r > 0$.

Proof We only observe that by definition of the M-estimator,

$$\begin{aligned} 0 &\leq \mathbb{H}_T(\underline{\hat{\theta}}_{k-1}, \underline{\hat{\theta}}_k, \overline{\hat{\theta}}_{k+1}) - \mathbb{H}_T(\underline{\hat{\theta}}_{k-1}, \theta_k^*, \overline{\hat{\theta}}_{k+1}) \\ &\leq \sup_{(u_k, \overline{\theta}_{k+1}) \in V_T^k(r, \theta_k^*) \times \prod_{j=k+1}^K \Theta_j} \log \mathbb{Z}_T^k(u_k; \underline{\hat{\theta}}_{k-1}, \theta_k^*, \overline{\theta}_{k+1}) \end{aligned}$$

if $|\hat{u}_T^k| > r$. □

In particular, this proposition ensures the tightness of $\{\hat{u}_T\}_{T>T_0}$.

In the previous section we focused our attention to a marginal (elementwise) convergence, but here we are interested in a joint convergence. For this purpose, we need only weak convergence of the statistical random field on compact sets because we already have tightness (or more strongly estimate of tail probability) of the M-estimator.

Write $B(R) = \{u \in \mathbb{R}^d;\ |u| \leq R\}$.

Theorem 5 *Let $\mathcal{K} \subset \Theta$. Suppose that the following conditions are fulfilled:*

- (i) *For every $R > 0$, $\mathbb{Z}_T(\cdot; \theta^*) \to^d \mathbb{Z}(\cdot; \theta^*)$ in $C(B(R))$ as $T \to \infty$, uniformly in $\mathcal{K}$, and $[UR]$ holds for $Z(u) = \mathbb{Z}(u; \theta^*)$.*
- (ii) *There exists a measurable mapping $\hat{u}$ that is a unique maximum point of $Z(\cdot; \theta^*)$ for each $\theta^* \in \mathcal{K}$, and $\{Z(\cdot; \theta^*)\}_{\theta^* \in \mathcal{K}}$ is tight.*

Then

- (a) *$\hat{u}_T \to^d \hat{u}$ as $T \to \infty$ uniformly in $\mathcal{K}$ if $\{\hat{u}_T\}_{T>T_0}$ is tight for some $T_0 \geq 0$ uniformly in $\mathcal{K}$.*

(b) $P_{\theta^*}\left[f(\hat{u}_T)\right] \to \mathbb{P}\left[f(\hat{u})\right]$ *as* $T \to \infty$ *uniformly in* $\mathcal{K}$ *for all continuous functions* f *satisfying* $\lim_{|u|\to\infty} |f(u)||u|^{-p} = 0$ *if* $\limsup_{T\to\infty} \sup_{\theta^*\in\mathcal{K}} \|\hat{u}_T\|_p < \infty$*, in particular, if (9) holds for some* $L > p$*.*

5.2 Convergence of the integration type functionals and Bayes estimators

Theorem 6 *Let* $\mathcal{K} \subset \Theta$*. Suppose that the following conditions are satisfied:*

(i) *For every* $k = 1, \ldots, K$ *and* $R > 0$*,* $\mathbb{Z}_T^k(\cdot; \underline{\theta}_k^*, \cdot) \to^{d(P_{\theta^*})} \mathbb{Z}^k(\cdot; \underline{\theta}_k^*, \cdot)$ *in* $C(\{u_k; |u_k| \le R\} \times \prod_{i=k+1}^K \Theta_i)$ *uniformly in* $\theta^* \in \mathcal{K}$ *as* $T \to \infty$*.*

(ii) *For some* $L > 1$ *and some constants* $D_k > d_k + p$*,*

$$\sup_{\theta^*\in\mathcal{K}} P_{\theta^*}\left[\sup_{(u_k,\overline{\theta}_{k+1})\in V_T^k(r,\theta_k^*)\times\prod_{j=k+1}^K \Theta_j} \mathbb{Z}_T^k(u_k; \underline{\theta}_k^*, \overline{\theta}_{k+1}) \ge \frac{C_1}{r^{D_k}}\right] \le \frac{C_2}{r^L}$$

for all $T > 0$*,* $r > 0$ *and* $k = 1, \ldots, K$*.*

(iii) *For every* $R > 0$*,* $\mathbb{Z}_T(u; \theta^*) \to^{d(P_{\theta^*})} \mathbb{Z}(u; \theta^*)$ *in* $C(\{u; |u| \le R\})$ *uniformly in* $\theta^* \in \mathcal{K}$ *as* $T \to \infty$*.*

Then for any $f \in C(\mathbb{R}^d; \mathbb{R}^s)$ *satisfying* $\sup_{u\in\mathbb{R}^d}(1 + |u|^p)^{-1}|f(u)| < \infty$*,*

$$\int f(u)\mathbb{Z}_T(u; \theta^*)\mathrm{d}u \to^{d(P_{\theta^*})} \int f(u)\mathbb{Z}(u; \theta^*)\mathrm{d}u$$

uniformly in $\theta^* \in \mathcal{K}$ *as* $T \to \infty$*.*

Remark 6 When $K = 1$, i.e., the case of a single grade, we can remove Condition (i) since it is the same as (iii).

Proof of Theorem 6. From (i) and (ii), it follows that uniformly

$$\mathbb{Z}_T^k(\cdot; \underline{\theta}_k^*, \cdot) \to^{d(P_{\theta^*})} \mathbb{Z}^k(\cdot; \underline{\theta}_k^*, \cdot)$$

in $\hat{C}(\mathbb{R}^{d_k} \times \prod_{i=k+1}^K \Theta_i)$, and also that

$$\int_A (1 + |u_k|^p) \sup_{\overline{\theta}_{k+1}\in\prod_{i=k+1}^K \Theta_i} \mathbb{Z}_T^k(u_k; \underline{\theta}_k^*, \overline{\theta}_{k+1})\mathrm{d}u_k$$
$$\to^{d(P_{\theta^*})} \int_A (1 + |u_k|^p) \sup_{\overline{\theta}_{k+1}\in\prod_{i=k+1}^K \Theta_i} \mathbb{Z}^k(u_k; \underline{\theta}_k^*, \overline{\theta}_{k+1})\mathrm{d}u_k$$

for all $A \in \mathbb{B}_{d_k}$. To obtain the last convergence, we note that

$$P_{\theta^*}\left[\int_{|u_j|\geq r}(1+|u_j|^p)\sup_{\overline{\theta}_{j+1}\in\prod_{i=j+1}^K\Theta_i}\mathbb{Z}_T^j(u_j;\underline{\theta}_j^*,\overline{\theta}_{j+1})\mathrm{d}u_j\right.$$
$$\left. > C_1{}'\left(\sum_{l=0}^{\infty}(r+l)^{p+d_j-1-D_j}\right)\right] \leq C_2\sum_{l=0}^{\infty}(r+l)^{-L}$$

for tightness, where C_1' is a positive constant.

Obviously,

$$\int_{\sum_{j=1}^K|u_j|\geq Kr}\left(1+\sum_{k=1}^K|u_k|^p\right)\mathbb{Z}_T(u;\theta^*)\mathrm{d}u$$
$$\leq \sum_{j=1}^K\int_{|u_j|\geq r}\left(1+\sum_{k=1}^K|u_k|^p\right)\mathbb{Z}_T(u;\theta^*)\mathrm{d}u$$

and

$$\int_{|u_j|\geq r}\left(1+\sum_{k=1}^K|u_k|^p\right)\mathbb{Z}_T(u;\theta^*)\mathrm{d}u$$
$$=\int_{|u_j|\geq r}\left(1+\sum_{k=1}^K|u_k|^p\right)\prod_{k=1}^K\mathbb{Z}_T^k(u_k;\underline{\theta}_k^*,\overline{\theta}_{k+1}^\dagger)\mathrm{d}u$$
$$\leq\int_{|u_j|\geq r}(1+|u_j|^p)\sup_{\overline{\theta}_{j+1}\in\prod_{i=j+1}^K\Theta_i}\mathbb{Z}_T^j(u_j;\underline{\theta}_j^*,\overline{\theta}_{j+1})\mathrm{d}u_j$$
$$\times\prod_{k\neq j}\int(1+|u_k|^p)\sup_{\overline{\theta}_{k+1}\in\prod_{i=k+1}^K\Theta_i}\mathbb{Z}_T^k(u_k;\underline{\theta}_k^*,\overline{\theta}_{k+1})\mathrm{d}u_k.$$

Therefore, for every $\epsilon > 0$, we can choose a sufficiently large r such that

$$P_{\theta^*}\left[\int_{\sum_{j=1}^K|u_j|\geq Kr}\left(1+\sum_{k=1}^K|u_k|^p\right)\mathbb{Z}_T(u;\theta^*)\mathrm{d}u > \epsilon\right] < \epsilon$$

uniformly in θ^* and $T \geq T_0$, for some T_0. [In a similar way as above, we see that $\{\mathbb{Z}_T(\cdot;\theta^*)\}_{T\geq T_0}$ is tight in $\hat{C}(\mathbb{R}^d)$ for some T_0.] □

We shall apply our large deviation results to Bayesian estimators in three types of formulations.

We denote by $\tilde{\theta}_T^+$ the **simultaneous Bayes type estimator** for a prior density $\pi : \Theta \to \mathbb{R}_+$:

$$\tilde{\theta}_T^+ = \left(\int_\Theta \exp(\mathbb{H}_T(\theta))\, \pi(\theta)\, \mathrm{d}\theta\right)^{-1} \int_\Theta \theta \exp(\mathbb{H}_T(\theta))\, \pi(\theta)\, \mathrm{d}\theta.$$

Assume that π is continuous and $0 < \inf_{\theta\in\Theta} \pi(\theta) \le \sup_{\theta\in\Theta} \pi(\theta) < \infty$. The scaled parameter space $\mathbb{U}_T(\theta^*) = \{u; \theta^* + a_T(\theta^*)u \in \Theta\}$. Let $\tilde{u}_T^+ = a_T(\theta^*)^{-1}(\tilde{\theta}_T^+ - \theta^*)$, then

$$\tilde{u}_T^+ = \left(\int_{\mathbb{U}_T(\theta^*)} \mathbb{Z}_T(u;\theta^*)\pi(\theta^* + a_T(\theta^*)u)\, \mathrm{d}u\right)^{-1} \times \int_{\mathbb{U}_T(\theta^*)} u\, \mathbb{Z}_T(u;\theta^*)\pi(\theta^* + a_T(\theta^*)u)\, \mathrm{d}u.$$

It will be seen that the right-hand side is asymptotically equivalent to

$$\check{u}_T = \left(\int_{\mathbb{R}^d} \mathbb{Z}_T(u;\theta^*)\, \mathrm{d}u\right)^{-1} \int_{\mathbb{R}^d} u\, \mathbb{Z}_T(u;\theta^*)\, \mathrm{d}u$$

for zero-extended random fields $\mathbb{Z}_T$ under the assumptions we will make. So we may identify these random vectors in asymptotic theory. Let

$$\tilde{u}^+ = \left(\int_{\mathbb{R}^d} \mathbb{Z}(u;\theta^*)\, \mathrm{d}u\right)^{-1} \int_{\mathbb{R}^d} u\, \mathbb{Z}(u;\theta^*)\, \mathrm{d}u.$$

Theorem 7 *Assume the conditions in Theorem 6 for $p = 1$. If the family $\mathcal{L}\{(\int_{\mathbb{R}^d} \mathbb{Z}(u;\theta^*)\mathrm{d}u)^{-1}|P_{\theta^*}\}$ $(\theta^* \in \mathcal{K})$ is uniformly tight, then for the simultaneous Bayes type estimator, $\tilde{u}_T^+ \to^{d(P_{\theta^*})} \tilde{u}^+$ uniformly in $\theta^* \in \mathcal{K}$ as $T \to \infty$.*

Proof In what follows, we may remove π from the integrands and also may only consider the integrations of the random fields $\mathbb{Z}_T$ over the whole space $\mathbb{R}^d$ by asymptotic equivalence and by way of extending of them. For arbitrary $\epsilon > 0$, there exist positive numbers A and T_0 such that

$$\begin{aligned}&\sup_{T\ge T_0} \sup_{\theta^*\in\mathcal{K}} P_{\theta^*}\left[(\int_{\mathbb{R}^d} \mathbb{Z}_T(u;\theta^*)\mathrm{d}u)^{-1} \ge A\right]\\ &\le \sup_{\theta^*\in\mathcal{K}} P_{\theta^*}\left[(\int_{\mathbb{R}^d} \mathbb{Z}(u;\theta^*)\mathrm{d}u)^{-1} \ge \frac{A}{2}\right] + \frac{\epsilon}{2}\\ &< \epsilon\end{aligned}$$

Therefore, the assertion can be proved by using the uniform joint weak convergence

$$\left(\int \mathbb{Z}_T(u;\theta^*)\mathrm{d}u, \int u\, \mathbb{Z}_T(u;\theta^*)\mathrm{d}u\right) \to^{d(P_{\theta^*})} \left(\int \mathbb{Z}(u;\theta^*)\mathrm{d}u, \int u\, \mathbb{Z}(u;\theta^*)\mathrm{d}u\right)$$

given in Theorem 6. □

Remark 7 If $\mathcal{K}$ is a set of a single element, then the additional tightness condition is not necessary because $\mathbb{Z}(\cdot;\theta^*)$ is in $C(\mathbb{R}^d)$ and $\mathbb{Z}(0;\theta^*)=1$ so that $\int_{\mathbb{R}^d}\mathbb{Z}(u;\theta^*)\mathrm{d}u>0$ a.s.

Theorem 8 *Let $\epsilon>0$, $p\geq 1$ and $p-(K-1)\epsilon\geq p_1>0$. Assume the conditions in Theorem 6 for p and some $L>p_1+1$. Suppose*

(i) *For some $T_0>0$,*

$$\sup_{\substack{\theta^*\in\mathcal{K}\\ T\geq T_0}} P_{\theta^*}\left[\left(\int_{\mathbb{U}_T(\theta^*)}\mathbb{Z}_T(u;\theta^*)\mathrm{d}u\right)^{-1}\right]<\infty. \tag{10}$$

(ii) *For some $T_0>0$ and C_3,*

$$\sup_{\substack{\theta^*\in\mathcal{K}\\ T\geq T_0}} P_{\theta^*}\left[\int_{\mathbb{U}_T^k(\theta_k^*)}\sup_{\overline{\theta}_{k+1}\in\prod_{i=k+1}^K\Theta_i}\mathbb{Z}_T^k(u_k;\underline{\theta}_k^*,\overline{\theta}_{k+1})\mathrm{d}u_k\geq r^{\epsilon}\right]\leq\frac{C_3}{r^L} \tag{11}$$

for all $k=1,\dots,K$ and $r>0$.

Then $\tilde{u}_T^+\to^{d(P_{\theta^*})}\tilde{u}^+$ and

$$P_{\theta^*}\left[f(\tilde{u}_T^+)\right]\to\mathbb{P}\left[f(\tilde{u}^+)\right]$$

uniformly in $\theta^*\in\mathcal{K}$ as $T\to\infty$, for any continuous function f satisfying $\limsup_{|u|\to\infty}|u|^{-p_1}|f(u)|<\infty$.

Proof For the jth element $\tilde{u}_{j,T}^+$ of $\tilde{u}_T^+$, we have

$$\begin{aligned}
&P_{\theta^*}\left[|\tilde{u}_{j,T}^+|^{p_1}\right]\\
&\leq P_{\theta^*}\left[\left(\int_{\mathbb{U}_T(\theta^*)}\mathbb{Z}_T(u;\theta^*)\pi(\theta^*+a_T(\theta^*)u)\,du\right)^{-1}\right.\\
&\qquad\left.\times\int_{\mathbb{U}_T(\theta^*)}|u_j|^{p_1}\,\mathbb{Z}_T(u;\theta^*)\pi(\theta^*+a_T(\theta^*)u)\,\mathrm{d}u\right]\\
&\leq C(\pi)\sum_{r=0}^{\infty}(r+1)^{p_1}P_{\theta^*}\left[\left(\int_{\mathbb{U}_T(\theta^*)}\mathbb{Z}_T(u;\theta^*)\,\mathrm{d}u\right)^{-1}\right.\\
&\qquad\left.\times\int_{\{u;r<|u_j|\leq r+1\}\cap\mathbb{U}_T(\theta^*)}\mathbb{Z}_T(u;\theta^*)\,\mathrm{d}u\right]
\end{aligned} \tag{12}$$

$$\le C(\pi)\sum_{r=1}^{\infty}(r+1)^{p_1}\Bigg\{P_{\theta^*}\Bigg[\int_{\{u_j;r<|u_j|\le r+1\}\cap\mathbb{U}_T^j(\theta_j^*)}\sup_{\overline{\theta}_{j+1}\in\prod_{i=j+1}^K\Theta_i}\mathbb{Z}_T^j(u_j;\underline{\theta}_j^*,\overline{\theta}_{j+1})\mathrm{d}u_j$$

$$\times\prod_{k\ne j}\int_{\mathbb{U}_T^k(\theta_k^*)}\sup_{\overline{\theta}_{k+1}\in\prod_{i=k+1}^K\Theta_i}\mathbb{Z}_T^k(u_k;\underline{\theta}_k^*,\overline{\theta}_{k+1})\mathrm{d}u_k>\frac{C_1}{r^{D_j-(K-1)\epsilon-d_j+1}}\Bigg]$$

$$+\frac{C_1}{r^{D_j-(K-1)\epsilon-d_j+1}}P_{\theta^*}\left[\left(\int_{\mathbb{U}_T(\theta^*)}\mathbb{Z}_T(u;\theta^*)\,\mathrm{d}u\right)^{-1}\right]\Bigg\}+C(\pi) \tag{13}$$

$$\le C(\pi)\sum_{r=0}^{\infty}(r+1)^{p_1}\Bigg\{P_{\theta^*}\Bigg[\sup_{(u_j,\overline{\theta}_{j+1})\in V_T^j(r,\theta_j^*)\times\prod_{i=j+1}^K\Theta_i}\mathbb{Z}_T^j(u_j;\underline{\theta}_j^*,\overline{\theta}_{j+1})>\frac{C_1'}{r^{D_j}}\Bigg]$$

$$+(K-1)\frac{C_3}{r^L}+\frac{C_1''}{r^{D_j-(K-1)\epsilon-d_j+1}}P_{\theta^*}\left[\left(\int_{\mathbb{U}_T(\theta^*)}\mathbb{Z}_T(u;\theta^*)\,\mathrm{d}u\right)^{-1}\right]\Bigg\}+C(\pi).$$

The last expression is $O(1)$ as $T\to\infty$. Here we used the fact that the ratio in the brace is bounded by 1 to obtain (13) from (12). [If necessary, we can make p_1 slightly bigger at the beginning of this proof since $D_j-d_j-p>0$.] □

Remark 8 Condition (11) is a condition in the central area because we have put other conditions in neighborhoods of the ∞. A sufficient condition for (10) is that

$$\sup_{\substack{\theta^*\in\mathcal{K}\\ T\ge T_0}}P_{\theta^*}\left[\left(\int_{|u|\le\delta}\mathbb{Z}_T(u;\theta^*)\mathrm{d}u\right)^{-1}\right]<\infty \tag{14}$$

for some positive δ and T_0. This condition can be reduced to an easily verified continuity condition of the random field near $u=0$. See Lemma 2.

Even if r^ϵ in (11) is replaced by $\exp(r^\epsilon)$, it is possible to give a similar result to Theorem 8. We omit the details.

Here, we briefly mention another approach. Define new random fields $\overline{\mathbb{Z}}_T^k(u_k;\theta_k^*)$ $(k=1,\ldots,K)$ by

$$\overline{\mathbb{Z}}_T^k(u_k;\theta_k^*)=\exp\left(\overline{\mathbb{H}}_T^k(\theta_k^\dagger)-\overline{\mathbb{H}}_T^k(\theta_k^*)\right),\quad\theta_k^\dagger=\theta_k^*+a_T^k(\theta_k^*)u_k$$

with

$$\overline{\mathbb{H}}_T^k(\theta_k)=\log\left\{\int\exp\left(\mathbb{H}_T(\underline{\theta}_{k-1},\theta_k,\overline{\theta}_{k+1})\right)\pi(\underline{\theta}_{k-1},\overline{\theta}_{k+1}|\theta_k)\,d\underline{\theta}_{k-1}d\overline{\theta}_{k+1}\right\},$$

where $\pi(\underline{\theta}_{k-1},\overline{\theta}_{k+1}|\theta_k)=\pi(\underline{\theta}_{k-1},\theta_k,\overline{\theta}_{k+1})/\pi_k(\theta_k)$, $\pi_k(\theta_k)=\int\pi(\underline{\theta}_{k-1},\theta_k,\overline{\theta}_{k+1})\,\mathrm{d}\underline{\theta}_{k-1}\mathrm{d}\overline{\theta}_{k+1}$. Then for the simultaneous Bayes type estimator $\tilde{\theta}_T^+$ for the prior density π, $\tilde{u}_T^+ = (\tilde{u}_{k,T}^+) = a_T(\theta^*)^{-1}(\tilde{\theta}_T^+ - \theta^*)$ is expressed as

$$\tilde{u}_{k,T}^+ = \left\{ \int_{\mathbb{U}_T^k(\theta_k^*)} \overline{\mathbb{Z}}_T^k(u_k; \theta_k^*)\, \pi_k(\theta_k^\dagger)\, \mathrm{d}u_k \right\}^{-1} \int_{\mathbb{U}_T^k(\theta_k^*)} u_k\, \overline{\mathbb{Z}}_T^k(u_k; \theta_k^*)\, \pi_k(\theta_k^\dagger)\, \mathrm{d}u_k.$$

Let

$$\tilde{u}_k^+ = \left\{ \int_{\mathbb{R}^{d_k}} \overline{\mathbb{Z}}^k(u_k; \theta_k^*)\, \mathrm{d}u_k \right\}^{-1} \int_{\mathbb{R}^{d_k}} u_k\, \overline{\mathbb{Z}}^k(u_k; \theta_k^*)\, \mathrm{d}u_k,$$

where $\overline{\mathbb{Z}}^k(u_k; \theta_k^*)$ will be the limit random field of $\overline{\mathbb{Z}}_T^k(u_k; \theta_k^*)$. Let $\overline{\mathbb{Z}}_T(u; \theta^*) = \left(\overline{\mathbb{Z}}_T^k(u_k; \theta_k^*)\right)$.

Theorem 9 *Let $\mathcal{K} \subset \Theta$. Suppose that the following conditions are satisfied:*

(i) *For every $k = 1, \ldots, K$ and $R > 0$, $\overline{\mathbb{Z}}_T^k(\cdot; \theta_k^*) \to^{d(P_{\theta^*})} \overline{\mathbb{Z}}^k(\cdot; \theta_k^*)$ in $C(\{u_k; |u_k| \leq R\})$ uniformly in $\theta^* \in \mathcal{K}$ as $T \to \infty$.*
(ii) *For some $L > 1$ and some constants $D_k > d_k + p$,*

$$\sup_{\theta^* \in \mathcal{K}} P_{\theta^*}\left[\sup_{u_k \in V_T^k(r, \theta_k^*)} \overline{\mathbb{Z}}_T^k(u_k; \theta_k^*) \geq \frac{C_1}{r^{D_k}} \right] \leq \frac{C_2}{r^L}$$

for all $T > 0$, $r > 0$ and $k = 1, \ldots, K$.

Then the following assertions hold:

(a) *For any $f \in C(\mathbb{R}^{d_k}; \mathbb{R}^s)$ satisfying $\sup_{u_k \in \mathbb{R}^d}(1 + |u_k|^p)^{-1}|f(u_k)| < \infty$,*

$$\int f(u_k)\overline{\mathbb{Z}}_T^k(u_k; \theta_k^*)\pi_k(\theta_k^\dagger)\mathrm{d}u_k \to^{d(P_{\theta^*})} \int f(u_k)\overline{\mathbb{Z}}^k(u_k; \theta_k^*)\pi_k(\theta_k^*)\mathrm{d}u_k$$

uniformly in $\theta^* \in \mathcal{K}$ as $T \to \infty$.
(b) *When $p \geq 1$, if each family $\mathcal{L}\{(\int_{\mathbb{R}^{d_k}} \overline{\mathbb{Z}}^k(u_k; \theta_k^*)\mathrm{d}u_k)^{-1} | P_{\theta^*}\}$ $(\theta^* \in \mathcal{K})$ is uniformly tight, then for the simultaneous Bayes type estimators, $\tilde{u}_{k,T}^+ \to^{d(P_{\theta^*})} \tilde{u}_k^+$ uniformly in $\theta^* \in \mathcal{K}$ as $T \to \infty$ for each $k = 1, \ldots, K$. Moreover, if for every $R > 0$, $\bar{\mathbb{Z}}_T(u; \theta^*) \to^{d(P_{\theta^*})} \bar{\mathbb{Z}}(u; \theta^*)$ in $C(\{u; |u| \leq R\})$ uniformly in $\theta^* \in \mathcal{K}$ as $T \to \infty$, then for the simultaneous Bayes type estimator, $\tilde{u}_T^+ \to^{d(P_{\theta^*})} \tilde{u}^+$ uniformly in $\theta^* \in \mathcal{K}$ as $T \to \infty$.*
(c) *Let $p \geq 1$ and $L > p + 1$. Suppose that For some $T_0 > 0$,*

$$\sup_{\substack{\theta^* \in \mathcal{K} \\ T \geq T_0}} P_{\theta^*}\left[\left(\int_{\mathbb{U}_T^k(\theta_k^*)} \overline{\mathbb{Z}}_T^k(u_k; \theta_k^*)\pi_k(\theta_k^\dagger)\mathrm{d}u \right)^{-1} \right] < \infty.$$

Then

$$P_{\theta^*}\left[f(\tilde{u}^+_{k,T})\right] \to \mathbb{P}\left[f(\tilde{u}^+_k)\right]$$

uniformly in $\theta^* \in \mathcal{K}$ *as* $T \to \infty$, *for any continuous function* f *satisfying* $\limsup_{|u_k|\to\infty} |u_k|^{-p}|f(u_k)| < \infty$.

Theorem 9 can be proved as those of Theorems 6, 7 and 8. It is rather simple and so omitted. Extension to the convergence of the joint moment of $\tilde{u}^+_T$ in the above theorem is obvious.

The simultaneous Bayes type estimator puts the same order of weights on all grades of parameters. Since the convergence rates of the estimators are different in different classes, the significance of the prior information for an estimator differs with estimators in other classes. So a more natural estimator would be the **adaptive Bayes type estimator** we shall discuss below. The adaptive Bayes approach seems to be better than the simultaneous Bayes approach from a computational point of view because it diminishes the dimension of the integration areas.

Given an estimator $\underline{\widetilde{\theta}}_{k-1,T}$ for $\underline{\theta}_{k-1}$, we define a Bayes type estimator $\tilde{\theta}_{k,T}$ by

$$\begin{aligned}\widetilde{\theta}_{k,T} = &\left\{\int_{\Theta_k} \exp\left(\mathbb{H}_T(\underline{\widetilde{\theta}}_{k-1,T}, \theta_k, \overline{\theta}^\star_{k+1})\right) \pi_{k,T}(\theta_k)\, \mathrm{d}\theta_k\right\}^{-1}\\ &\cdot \int_{\Theta_k} \theta_k \exp\left(\mathbb{H}_T(\underline{\widetilde{\theta}}_{k-1,T}, \theta_k, \overline{\theta}^\star_{k+1})\right) \pi_{k,T}(\theta_k)\, \mathrm{d}\theta_k,\end{aligned}$$

where $\overline{\theta}^\star_{k+1}$ is a known dummy value of $\overline{\theta}_{k+1}$, and $\pi_{k,T}(\theta_k)\,\mathrm{d}\theta_k$ is a prior distribution of θ_k possibly depending on T; by convention, we neglect $\underline{\widetilde{\theta}}_{0,T}$ and $\overline{\theta}^\star_{K+1,T}$. The prior densities $\pi_k \equiv \pi_{k,T}$ are assumed to be uniformly (in T) continuous and to satisfy $0 < \inf_{\theta_k,T,\omega} \pi_{k,T} \le \sup_{\theta_k,T,\omega} \pi_{k,T} < \infty$, for simplicity. Usually the value of $\overline{\theta}_{k+1}$ is not informative in estimation of θ_k, so that a choice of $\overline{\theta}^\star_{k+1}$ does not have much meaning asymptotically. We recall the random field $\mathbb{Z}^k_T$ defined in (8):

$$\mathbb{Z}^k_T(u_k; \underline{\theta}_{k-1}, \theta^*_k, \overline{\theta}_{k+1}) = \exp\left(\mathbb{H}_T(\underline{\theta}_{k-1}, \theta^\dagger_k, \overline{\theta}_{k+1}) - \mathbb{H}_T(\underline{\theta}_{k-1}, \theta^*_k, \overline{\theta}_{k+1})\right).$$

Then the normalized Bayes estimators $\tilde{u}_{k,T} = a^k_T(\theta^*_k)^{-1}(\tilde{\theta}_{k,T} - \theta^*_k)$ are given by

$$\begin{aligned}\tilde{u}_{k,T} = &\left(\int_{\mathbb{U}^k_T(\theta^*_k)} \mathbb{Z}^k_T(u_k; \underline{\widetilde{\theta}}_{k-1,T}, \theta^*_k, \overline{\theta}^\star_{k+1})\pi_k(\theta^\dagger_k)\, \mathrm{d}u_k\right)^{-1}\\ &\times \int_{\mathbb{U}^k_T(\theta^*_k)} u_k \mathbb{Z}^k_T(u_k; \underline{\widetilde{\theta}}_{k-1,T}, \theta^*_k, \overline{\theta}^\star_{k+1})\pi_k(\theta^\dagger_k)\, \mathrm{d}u_k\end{aligned}$$

and their targets by

$$\tilde{u}_k = \left(\int_{\mathbb{R}^{d_k}} \mathbb{Z}^k(u_k; \underline{\theta}^*_{k-1}, \theta^*_k, \overline{\theta}^*_{k+1})\, \mathrm{d}u_k\right)^{-1} \int_{\mathbb{R}^{d_k}} u_k \mathbb{Z}^k(u_k; \underline{\theta}^*_{k-1}, \theta^*_k, \overline{\theta}^*_{k+1})\, \mathrm{d}u_k.$$

We denote $(\tilde{u}_{k,T})$ by $\tilde{u}_T$ and $(\tilde{u}_k)$ by $\tilde{u}$. Also, $\widetilde{\theta}_{k,T}$ will be denoted by $\widetilde{\theta}_k$ for notational simplicity. Then we have

Theorem 10 *Let $\mathcal{K} \subset \Theta$. Suppose that the following conditions are satisfied:*

(i) *For every $k = 1, \ldots, K$ and $R > 0$, $\mathbb{Z}_T^k(\cdot; \underline{\widetilde{\theta}}_{k-1}, \theta_k^*, \overline{\theta}^\star_{k+1}) \to^{d(P_{\theta^*})} \mathbb{Z}^k(\cdot; \underline{\theta}^*_{k-1}, \theta_k^*, \overline{\theta}^*_{k+1})$ in $C(\{u_k; |u_k| \le R\})$ uniformly in $\theta^* \in \mathcal{K}$ as $T \to \infty$.*

(ii) *For some $L > 1$ and some constants $D_k > d_k + p$,*

$$\sup_{\theta^* \in \mathcal{K}} P_{\theta^*}\left[\sup_{u_k \in V_T^k(r,\theta_k^*)} \mathbb{Z}_T^k(u_k; \underline{\widetilde{\theta}}_{k-1}, \theta_k^*, \overline{\theta}^\star_{k+1}) \ge \frac{C_1}{r^{D_k}}\right] \le \frac{C_2}{r^L}$$

for all $T > 0$, $r > 0$ and $k = 1, \ldots, K$.

Then the following assertions hold:

(a) *For any $f \in C(\mathbb{R}^{d_k}; \mathbb{R}^s)$ satisfying $\sup_{u_k \in \mathbb{R}^d}(1 + |u_k|^p)^{-1}|f(u_k)| < \infty$,*

$$\int f(u_k)\mathbb{Z}_T^k(u_k; \underline{\widetilde{\theta}}_{k-1}, \theta_k^*, \overline{\theta}^\star_{k+1})\pi_k(\theta_k^\dagger)\mathrm{d}u_k \to^{d(P_{\theta^*})} \int f(u_k)\mathbb{Z}^k(u_k; \theta^*)\pi_k(\theta_k^*)\mathrm{d}u_k$$

uniformly in $\theta^* \in \mathcal{K}$ as $T \to \infty$.

(b) *When $p \ge 1$, if each family $\mathcal{L}\{(\int_{\mathbb{R}^{d_k}} \mathbb{Z}^k(u_k; \theta^*)\mathrm{d}u_k)^{-1}|P_{\theta^*}\}$ $(\theta^* \in \mathcal{K})$ is uniformly tight, then for the adaptive Bayes type estimators, $\tilde{u}_{k,T} \to^{d(P_{\theta^*})} \tilde{u}_k$ uniformly in $\theta^* \in \mathcal{K}$ as $T \to \infty$ for each $k = 1, \ldots, K$. Moreover, if for every $R > 0$, the joint random vector $(\mathbb{Z}_T^k(u_k; \underline{\widetilde{\theta}}_{k-1}, \theta_k^*, \overline{\theta}^\star_{k+1}))_{k=1,\ldots,K} \to^{d(P_{\theta^*})} (\mathbb{Z}^k(u_k; \theta^*))_{k=1,\ldots,K}$ in $C(\{u; |u| \le R\})$ uniformly in $\theta^* \in \mathcal{K}$ as $T \to \infty$, then the adaptive Bayes type estimator $\tilde{u}_T \to^{d(P_{\theta^*})} \tilde{u}$ uniformly in $\theta^* \in \mathcal{K}$ as $T \to \infty$.*

(c) *Let $p \ge 1$ and $L > p + 1$. Suppose that For some $T_0 > 0$,*

$$\sup_{\substack{\theta^* \in \mathcal{K} \\ T \ge T_0}} P_{\theta^*}\left[\left(\int_{\mathbb{U}_T^k(\theta_k^*)} \mathbb{Z}_T^k(u_k; \underline{\widetilde{\theta}}_{k-1}, \theta_k^*, \overline{\theta}^\star_{k+1})\pi_k(\theta_k^\dagger)\mathrm{d}u\right)^{-1}\right] < \infty.$$

Then

$$P_{\theta^*}\left[f(\tilde{u}_{k,T})\right] \to \mathbb{P}\left[f(\tilde{u}_k)\right]$$

uniformly in $\theta^* \in \mathcal{K}$ as $T \to \infty$, for any continuous function f satisfying $\limsup_{|u_k| \to \infty} |u_k|^{-p}|f(u_k)| < \infty$. *Moreover,*

$$P_{\theta^*}\left[g(\tilde{u}_T)\right] \to \mathbb{P}\left[g(\tilde{u})\right]$$

for any continuous function g satisfying $\limsup_{|u| \to \infty} |u|^{-p}|g(u)| < \infty$ *under the joint convergence assumption in* (b).

In order to verify the integrability condition in (c) of Theorem 10, the following lemma is convenient. Let $I(\delta) = \{u \in \mathbb{R}^d;\ |u_i| \leq \delta\ (i = 1, \ldots, d)\}$ for $\delta > 0$. Consider a measurable random field $K : \Omega \times I(\delta) \to \mathbb{R}$.

Lemma 2 *Suppose that for some $p > d$, $\delta > 0$ and $C_0 > 0$,*

$$\mathbb{P}\left[|K(u)|^p\right] \leq C_0|u|^p$$

for all $u \in I(\delta)$. Then $K(0) = 0$ a.s. and

$$\mathbb{P}\left[\left(\int_{I(\delta)} e^{K(u)}\,\mathrm{d}u\right)^{-1}\right] \leq C_1$$

for some constant C_1 depending only on p, d, δ, and C_0.

Proof For $h \in (0, \delta]$,

$$\begin{aligned}
\mathbb{P}\left[\int_{I(h)} e^{K(u)}\,\mathrm{d}u < h^d\right] &\leq \mathbb{P}\left[\int_{I(h)} (K(u)+1)\ \mathrm{d}u < h^d\right] \\
&\leq \mathbb{P}\left[\int_{I(h)} K(u)\,\mathrm{d}u < -(2^d-1)h^d\right] \\
&\leq \mathbb{P}\left[\int_{I(h)} |K(u)|\,\mathrm{d}u > (2^d-1)h^d\right] \\
&\leq (1-2^{-d})^{-p}\,\mathbb{P}\left[\left(\frac{1}{(2h)^d}\int_{I(h)} |K(u)|\,\mathrm{d}u\right)^p\right] \\
&\leq C_0(1-2^{-d})^{-p}h^p.
\end{aligned}$$

For $t \geq \delta^{-d}$,

$$\begin{aligned}
\mathbb{P}\left[\left(\int_{I(\delta)} e^{K(u)}\,\mathrm{d}u\right)^{-1} > t\right] &\leq \mathbb{P}\left[\left(\int_{I(t^{-1/d})} e^{K(u)}\,\mathrm{d}u\right)^{-1} > t\right] \\
&\leq C_0(1-2^{-d})^{-p}t^{-p/d}.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\int_0^\infty \mathbb{P}\left[\left(\int_{I(\delta)} e^{K(u)}\,\mathrm{d}u\right)^{-1} > t\right]\mathrm{d}t \\
&\quad\leq \delta^{-d} + C_0(1-2^{-d})^{-p}\left(\frac{p}{d}-1\right)^{-1}\delta^{p-d} =: C_1.
\end{aligned}$$

□

5.3 Convergence of random fields on compacts

Here, we keep the notation thus far and assume that $\mathbb{H}_T$ is of class C^2 in θ. Let $\underline{\vartheta}_{k-1}$ be a sequence of measurable mappings depending on T and taking values in $\prod_{j=1}^{k-1}\Theta_j$, and $\overline{\vartheta}_{k+1}$ a sequence of measurable mappings depending on T and taking values in $\prod_{j=k+1}^{K}\Theta_j$. It should be noted in this notation that, for example, the first elements of $\underline{\vartheta}_2$ and $\underline{\vartheta}_3$ are possibly different.

Theorem 11 *Suppose that the following conditions are satisfied:*

(i) *On a probability space, there exist random variables $\Delta_k(\theta^*)$ taking values in $\mathbb{R}^{d_k}$, $k=1,\ldots,K$, such that for every $u=(u_k)\in\mathbb{R}^d$,*

$$\left(\partial_{\theta_k}\mathbb{H}_T(\underline{\theta}^*_{k-1},\theta^*_k,\overline{\vartheta}_{k+1})[a^k_T(\theta^*_k)u_k]\right)_{k=1}^{K}\to^d\left(\Delta_k(\theta^*)[u_k]\right)_{k=1}^{K}.$$

(ii) *There exist deterministic positive definite symmetric matrices $\Gamma_k(\theta^*)\in\mathbb{R}^{d_k}\otimes\mathbb{R}^{d_k}$ such that for every $u=(u_k)\in\mathbb{R}^d$,*

$$\partial^2_{\theta_k}\mathbb{H}_T(\underline{\theta}^*_{k-1},\theta^*_k,\overline{\vartheta}_{k+1})[(a^k_T(\theta^*_k)u_k)^{\otimes 2}]\to^p-\Gamma_k(\theta^*)[(u_k)^{\otimes 2}].$$

(iii) *For every k, there exist a positive number ζ_k and an random variable $W^k_T(\theta^*)$ such that*
(iii-a) $|\partial_{\underline{\theta}_{k-1}}\partial_{\theta_k}\mathbb{H}_T(\underline{\theta}^*_{k-1},\theta^*_k,\overline{\vartheta}_{k+1})|\le W^k_T(\theta^*)$.
(iii-b) *for $\theta,\theta'\in\Theta$ satisfying $\overline{\theta}_{k+1}=\overline{\theta'}_{k+1}$,*

$$\left|\partial_{\underline{\theta}_k}\partial_{\theta_k}\mathbb{H}_T(\theta)-\partial_{\underline{\theta}_k}\partial_{\theta_k}\mathbb{H}_T(\theta')\right|\le W^k_T(\theta^*)|\underline{\theta}_k-\underline{\theta'}_k|^{\zeta_k},$$

(iii-c) $W^k_T(\theta^*)|a^k_T(\theta^*_k)|^2=O_p(1)$, *i.e., for every $\epsilon>0$, there exist $T_0>0$ and $M>0$ such that $P_{\theta^*}[W^k_T(\theta^*)|a^k_T(\theta^*_k)|^2>M]<\epsilon$ for all $T\ge T_0$.*

(iv) *If $K\ge 2$, $\left|a^k_T(\theta^*_k)^{-1}(\underline{\vartheta}_{k-1}-\underline{\theta}^*_{k-1})\right|\to^p 0$ for each $k\in\{2,\ldots,K\}$.*

Let

$$\mathbb{Z}^k(u_k;\theta^*)=\exp\left(\Delta_k(\theta^*)[u_k]-\frac{1}{2}\Gamma_k(\theta^*)[(u_k)^{\otimes 2}]\right). \tag{15}$$

Let $R>0$. Then

$$\left(\mathbb{Z}^k_T(u_k;\underline{\vartheta}_{k-1},\theta^*_k,\overline{\vartheta}_{k+1})\right)_{k=1}^{K}\to^d\left(\mathbb{Z}^k(u_k;\theta^*)\right)_{k=1}^{K} \tag{16}$$

in $C(B(R);\mathbb{R}^K)$ as $T\to\infty$. If the convergences and estimates in Conditions (i)-(iv) *hold uniformly in a set $\mathcal{K}\subset\Theta$, then so does the convergence* (16).

Proof By definition of $\mathbb{Z}_T^k$,

$$\begin{aligned}
&\log \mathbb{Z}_T^k(u_k; \underline{\vartheta}_{k-1}, \theta_k^*, \overline{\vartheta}_{k+1}) \\
&\quad = \mathbb{H}_T(\underline{\vartheta}_{k-1}, \theta_k^\dagger, \overline{\vartheta}_{k+1}) - \mathbb{H}_T(\underline{\vartheta}_{k-1}, \theta_k^*, \overline{\vartheta}_{k+1}) \\
&\quad = \partial_{\theta_k}\mathbb{H}_T(\underline{\vartheta}_{k-1}, \theta_k^*, \overline{\vartheta}_{k+1})[a_T^k(\theta_k^*)u_k] \\
&\qquad + \int_0^1 (1-s)\partial_{\theta_k}^2\mathbb{H}_T(\underline{\vartheta}_{k-1}, \theta_k^* + sa_T^k(\theta_k^*)u_k, \overline{\vartheta}_{k+1})\,\mathrm{d}s[(a_T^k(\theta_k^*)u_k)^{\otimes 2}].
\end{aligned}$$

Set

$$\mathbb{G}_T^k(u_k) = \int_0^1 (1-s)\partial_{\theta_k}^2\mathbb{H}_T(\underline{\vartheta}_{k-1}, \theta_k^* + sa_T^k(\theta_k^*)u_k, \overline{\vartheta}_{k+1})\,ds[(a_T^k(\theta_k^*)u_k)^{\otimes 2}].$$

Then

$$\begin{aligned}
&\left|\mathbb{G}_T^k(u_k) - \frac{1}{2}\partial_{\theta_k}^2\mathbb{H}_T(\underline{\theta}_{k-1}^*, \theta_k^*, \overline{\vartheta}_{k+1})[(a_T^k(\theta_k^*)u_k)^{\otimes 2}]\right| \\
&\quad \le \int_0^1 (1-s)\left|\partial_{\theta_k}^2\mathbb{H}_T(\underline{\vartheta}_{k-1}, \theta_k^* + sa_T^k(\theta_k^*)u_k, \overline{\vartheta}_{k+1})\right. \\
&\qquad \left. -\partial_{\theta_k}^2\mathbb{H}_T(\underline{\theta}_{k-1}^*, \theta_k^*, \overline{\vartheta}_{k+1})\right|\,ds\;|a_T^k(\theta_k^*)|^2|u_k|^2 \\
&\quad \le \left[\left|\underline{\vartheta}_{k-1} - \underline{\theta}_{k-1}^*\right| + |a_T^k(\theta_k^*)||u_k|\right]^{\zeta_k} W_T^k(\theta^*)\;|a_T^k(\theta_k^*)|^2|u_k|^2 \\
&\quad \to^p 0
\end{aligned}$$

uniformly in $B(R)$, by (iii-c) and (iv). This together with (ii) implies $\{(\mathbb{G}_T^k)\}_{T\ge T_0}$ is C-tight for large T_0 in $\mathcal{P}(C(B(R);\mathbb{R}^K))$.

Similarly,

$$\begin{aligned}
&\left|\partial_{\theta_k}\mathbb{H}_T(\underline{\vartheta}_{k-1}, \theta_k^*, \overline{\vartheta}_{k+1})[a_T^k(\theta_k^*)u_k] - \partial_{\theta_k}\mathbb{H}_T(\underline{\theta}_{k-1}^*, \theta_k^*, \overline{\vartheta}_{k+1})[a_T^k(\theta_k^*)u_k]\right. \\
&\qquad \left. -\partial_{\underline{\theta}_{k-1}}\partial_{\theta_k}\mathbb{H}_T(\underline{\theta}_{k-1}^*, \theta_k^*, \overline{\vartheta}_{k+1})\,[\underline{\vartheta}_{k-1} - \underline{\theta}_{k-1}^*, a_T^k(\theta_k^*)u_k]\right| \\
&\quad = \left|\int_0^1 \partial_{\underline{\theta}_{k-1}}\partial_{\theta_k}\mathbb{H}_T(\underline{\theta}_{k-1}^* + s(\underline{\vartheta}_{k-1} - \underline{\theta}_{k-1}^*), \theta_k^*, \overline{\vartheta}_{k+1})\,ds\,[\underline{\vartheta}_{k-1} - \underline{\theta}_{k-1}^*, a_T^k(\theta_k^*)u_k]\right. \\
&\qquad \left. -\partial_{\underline{\theta}_{k-1}}\partial_{\theta_k}\mathbb{H}_T(\underline{\theta}_{k-1}^*, \theta_k^*, \overline{\vartheta}_{k+1})\,[\underline{\vartheta}_{k-1} - \underline{\theta}_{k-1}^*, a_T^k(\theta_k^*)u_k]\right| \\
&\quad \le \left|\underline{\vartheta}_{k-1} - \underline{\theta}_{k-1}^*\right|^{\zeta_k} W_T^k(\theta^*)\;|a_T^k(\theta_k^*)|^2|u_k|\;|a_T^k(\theta_k^*)^{-1}(\underline{\vartheta}_{k-1} - \underline{\theta}_{k-1}^*)| \\
&\quad \to^p 0
\end{aligned}$$

due to (iii-c) and (iv). By the same reasoning with (iii-a), (iii-c) and (iv),

$$\partial_{\underline{\theta}_{k-1}}\partial_{\theta_k}\mathbb{H}_T(\underline{\theta}_{k-1}^*, \theta_k^*, \overline{\vartheta}_{k+1})\,[\underline{\vartheta}_{k-1} - \underline{\theta}_{k-1}^*, a_T^k(\theta_k^*)u_k] \to^p 0.$$

Therefore, $\{(\partial_{\theta_k}\mathbb{H}_T(\underline{\vartheta}_{k-1}, \theta_k^*, \overline{\vartheta}_{k+1})[a_T^k(\theta_k^*)u_k])\}_{T\geq T_0}$ is C-tight for large T_0 in $\mathcal{P}(C(B(R);\mathbb{R}^K))$.

Finite dimensional convergence was obtained at the same time. Consequently, we have obtained the desired convergence. □

Remark 9 Extension to the stable convergence is straightforward. If the convergence in (i) of Theorem 11 holds $\mathcal{G}$-stably for a σ-field $\mathcal{G}$ on Ω and possibly random $\Gamma_k(\theta^*)$ in (ii) are $\mathcal{G}$-measurable, then the resulting convergence (16) holds $\mathcal{G}$-stably. It is also the case in Theorem 12.

Remark 10 A sufficient condition of (iii-b) is that

$$|a_T^k(\theta_k^*)|^2 \sup_\theta |\partial^2_{\underline{\theta}_k}\partial_{\theta_k}\mathbb{H}_T(\theta)| = O_p(1)$$

as $T\to\infty$. In order to verify this condition or Condition (iii), the inequalities in Section 7 are usually used, especially to deal with *martingale* terms.

Theorem 12 *Suppose that the following conditions are satisfied:*

(i) *On a probability space, there exist random variables $\Delta_k(\theta^*)$ taking values in $\mathbb{R}^{d_k}$, $k=1,\ldots,K$, such that for every $u=(u_k)\in\mathbb{R}^d$,*

 (i-a) $\left(\partial_{\theta_k}\mathbb{H}_T(\underline{\theta}^*_{k-1}, \theta_k^*, \overline{\theta}^*_{k+1})[a_T^k(\theta_k^*)u_k]\right)_{k=1}^K \to^d (\Delta_k(\theta^*)[u_k])_{k=1}^K$.

 (i-b) $\sup_{\overline{\theta}_{k+1}}\Big|\partial_{\theta_k}\mathbb{H}_T(\underline{\theta}^*_{k-1}, \theta_k^*, \overline{\theta}_{k+1})[a_T^k(\theta_k^*)u_k] - \partial_{\theta_k}\mathbb{H}_T(\underline{\theta}^*_{k-1}, \theta_k^*, \overline{\theta}^*_{k+1})[a_T^k(\theta_k^*)u_k]\Big| \to^p 0.$

(ii) *There exist deterministic positive definite symmetric matrices $\Gamma_k(\theta^*)\in\mathbb{R}^{d_k}\otimes\mathbb{R}^{d_k}$ such that for every $u=(u_k)\in\mathbb{R}^d$,*

$$\sup_{\theta_k:|\theta_k-\theta_k^*|\leq R|a_T^k(\theta_k^*)|}\ \sup_{\overline{\theta}_{k+1}}\Big|\partial^2_{\theta_k}\mathbb{H}_T(\underline{\theta}^*_{k-1}, \theta_k, \overline{\theta}_{k+1})[(a_T^k(\theta_k^*)u_k)^{\otimes 2}] + \Gamma_k(\theta^*)[(u_k)^{\otimes 2}]\Big| \to^p 0$$

 for $R>0$.

Then for $\mathbb{Z}^k(u_k;\theta^*)$ given by (15),

$$\left(\mathbb{Z}_T^k(u_k; \underline{\theta}^*_{k-1}, \theta_k^*, \overline{\theta}^\dagger_{k+1})\right)_{k=1}^K \to^d \left(\mathbb{Z}^k(u_k;\theta^*)\right)_{k=1}^K \tag{17}$$

in $C(B(R);\mathbb{R}^K)$ as $T\to\infty$. If the convergences and estimates in Conditions (i) *and* (ii) *hold uniformly in a set $\mathcal{K}\subset\Theta$, then so does the convergence* (17).

In particular, one has the convergence $\mathbb{Z}_T(u;\theta^*)\to^d\mathbb{Z}(u;\theta^*)$ in $C(B(R);\mathbb{R})$, where

$$\mathbb{Z}(u;\theta^*) = \exp\left(\sum_{k=1}^K \Delta_k(\theta^*)[u_k] - \frac{1}{2}\sum_{k=1}^K \Gamma_k(\theta^*)[(u_k)^{\otimes 2}]\right).$$

Proof The argument is similar to the proof of Theorem 11. We replace $\underline{\vartheta}_{k-1}$ by $\underline{\theta}^*_{k-1}$, and $\overline{\vartheta}_{k+1}$ by $\overline{\theta}^\dagger_{k+1}$, but it should be noted that now the variables with † depend on u. Set

$$\mathbb{D}^k_T(u) = \partial_{\theta_k}\mathbb{H}_T(\underline{\theta}^*_{k-1}, \theta^*_k, \overline{\theta}^\dagger_{k+1})[a^k_T(\theta^*_k)u_k]$$

and

$$\mathbb{C}^k_T(u) = \int_0^1 (1-s)\partial^2_{\theta_k}\mathbb{H}_T(\underline{\theta}^*_{k-1}, \theta^*_k + sa^k_T(\theta^*_k)u_k, \overline{\theta}^\dagger_{k+1})\, ds[(a^k_T(\theta^*_k)u_k)^{\otimes 2}].$$

Then $\log \mathbb{Z}^k_T(u_k; \underline{\theta}^*_{k-1}, \theta^*_k, \overline{\theta}^\dagger_{k+1}) = \mathbb{D}^k_T(u) + \mathbb{C}^k_T(u)$. The finite-dimensional convergence of $\mathbb{D}^k_T$ and $\mathbb{C}^k_T$ and their C-tightness follow from Conditions (i) and (ii), respectively. □

6 Estimators for a stochastic differential equation based on sampled data

Let us consider a stationary diffusion process[13] $X = (X_t)_{t\in\mathbb{R}_+}$ adapted to a filtration $\mathbf{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ and satisfying the stochastic differential equation

$$\mathrm{d}X_t = a(X_t, \theta_2)\mathrm{d}t + b(X_t, \theta_1)\mathrm{d}w_t, \quad X_0 = x_0.$$

Here w_t is an r-dimensional standard $\mathbf{F}$-Wiener process independent of the initial value x_0. θ_1 and θ_2 are unknown parameters with $\theta_i \in \Theta_i$, a bounded open set in $\mathbb{R}^{m_i}$ admitting Sobolev's inequalities for embedding $W^{1,p}(\Theta_i) \hookrightarrow C(\overline{\Theta}_i)$.[14] The distribution of x_0 depends on those parameters. The true value of the unknown parameter is denoted by $\theta^* = (\theta^*_1, \theta^*_2)$.

We assume the following condition:

[D1] (i) The mappings $a : \mathbb{R}^d \times \Theta_2 \to \mathbb{R}^d$ and $b : \mathbb{R}^d \times \Theta_1 \to \mathbb{R}^d \otimes \mathbb{R}^r$ have continuous derivatives satisfying

$$\sup_{\theta_2\in\Theta_2} \left|\partial^i_{\theta_2} a(x, \theta_2)\right| \le C(1 + |x|)^C \quad (0 \le i \le 4)$$

and

$$\sup_{\theta_1\in\Theta_1} \left|\partial^j_x \partial^i_{\theta_1} b(x, \theta_1)\right| \le C(1 + |x|)^C \quad (0 \le i \le 4;\ 0 \le j \le 2)$$

for some constant C.

[13] Obviously, stationarity is not essential in our discussion. Extension to the nonstationary case is straightforward.

[14] For a sufficient condition for it, see Adams (1975). On the other hand, it is possible to replace estimates by Sobolev's inequality appearing below by Kolmogorov's type criterion for tightness of random fields. See Sect. 7.

(ii) $B(x, \theta_1) = bb'(x, \theta_1)$ is elliptic uniformly in (x, θ_1).
(iii) For some constant C,

$$\sup_{\theta_2 \in \Theta_2} |a(x_1, \theta_2) - a(x_2, \theta_2)| + \sup_{\theta_1 \in \Theta_1} |b(x_1, \theta_1) - b(x_2, \theta_1)| \\ \leq C|x_1 - x_2| \quad (x_1, x_2 \in \mathbb{R}^d).$$

(iv) $X_0 \in \bigcap_{p>0} L^p(P_{\theta^*})$.

Now we want to estimate the unknown parameters with the discrete-time observations $\mathbf{x}_n = (X_{t_i})_{i=0}^n$, where $t_i = ih$ with $h = h_n$ depending on $n \in \mathbb{N}$. For this purpose, we consider a quasi-likelihood function

$$p_n(\mathbf{x}_n, \theta) = \prod_{i=1}^{n} \frac{1}{(2\pi h)^{d/2} |B(X_{t_{i-1}}, \theta_1)|^{1/2}} \\ \times \exp\left(-\frac{1}{2h} B(X_{t_{i-1}}, \theta_1)^{-1} \left[(\Delta_i X - ha(X_{t_{i-1}}, \theta_2))^{\otimes 2} \right] \right)$$

with $\Delta_i X = X_{t_i} - X_{t_{i-1}}$, and the maximum likelihood type estimator $\hat{\theta}_n = (\hat{\theta}_{1,n}, \hat{\theta}_{2,n})$ that maximizes $p_n(\mathbf{x}_n, \theta)$ in $\theta = (\theta_1, \theta_2) \in \overline{\Theta} = \overline{\Theta_1 \times \Theta_2}$.

Since when $h \downarrow 0$, the ratio $p_n(\mathbf{x}_n, \theta_1, \theta_2')/p_n(\mathbf{x}_n, \theta_1, \theta_2)$ tends to the representation of the likelihood ratio for the second parameter, which is well-known Girsanov's formula, this quasi-likelihood seems natural in estimating the drift parameter. It is the case even for positive but small h because it is the likelihood function for the process generated by the Euler-Maruyama scheme. We mention some literature in Remark 12.

For a function $F(x, \theta)$, we will write $F_i(\theta)$ or more simply F_i for $F(X_{t_i}, \theta)$, and also F_i^* for $F(X_{t_i}, \theta^*)$. Symbol C denotes a generic constant, which varies from line to line. We write C_p for C when we want to emphasize the dependency of C on another parameter p.

We assume a mixing property for X:

[D2] There exists a positive constant a such that

$$\alpha_X(h) \leq a^{-1} e^{-ah} \quad (h > 0),$$

where

$$\alpha_X(h) = \sup_{t \in \mathbb{R}_+} \sup_{\substack{A \in \sigma[X_r;\, r \leq t], \\ B \in \sigma[X_r;\, r \geq t+h]}} |P_{\theta^*}[A \cap B] - P_{\theta^*}[A] P_{\theta^*}[B]|.$$

Ergodic property is a consequence of [D2]: for an invariant probability measure $\nu = \nu_{\theta^*}$ of X_t for θ^*,

$$\frac{1}{T} \int_0^T g(X_t) \mathrm{d}t \to \int_{\mathbb{R}^d} g(x) \nu(\mathrm{d}x) \quad (T \to \infty) \tag{18}$$

in probability for any bounded measurable function g. It is a simple matter to deduce that the convergence (18) takes place in L^p for every $p > 0$ and every measurable function g of at most polynomial growth. For the mixing properties of diffusion processes, we refer the reader to Meyn and Tweedie (1992, 1993a,b), Veretennikov (1987, 1997) and Kusuoka and Yoshida (2000). Assumption of the mixing property is of practical convenience. The exponential rate can be relaxed to sufficiently high order of polynomial rate; however, it would make the presentation more complicated. If a one-dimensional diffusion process is treated, then (18) is sufficient and we can go without a mixing assumption because we have an explicit expression of the Green function for the Poisson equation. However, our interest was not necessarily in a one-dimensional diffusion, which admits the local time estimator. Moreover, we should note that, though we will not do it here explicitly, it is possible to show the uniformity of our forthcoming results under uniform version of conditions.

We will assume that $h \to 0$ and $nh^2 \to 0$ as $n \to \infty$.[15] Moreover, we assume that for some positive constant ϵ_0, $nh \geq n^{\epsilon_0}$ for large n.

Take $\mathbb{H}_n$ as

$$\mathbb{H}_n(\theta_1, \theta_2) = \log \left\{ (2\pi h)^{nd/2} p_n(\mathbf{x}_n, \theta) \right\}. \tag{19}$$

We wrote n in place of "T", and here θ_1 is as "θ" and θ_2 as "τ" in our general setting. Then "$b_T(\xi_0)$" should be n at the first stage, and

$$\begin{aligned} \mathbb{Y}_n(\theta_1, \theta_2; \theta^*) &= \frac{1}{n} \left(\mathbb{H}_n(\theta_1, \theta_2) - \mathbb{H}_n(\theta_1^*, \theta_2) \right) \\ &= -\frac{1}{2n} \sum_{i=1}^{n} \left\{ \left(B_{i-1}^{-1}(\theta_1) - B_{i-1}^{-1}(\theta_1^*) \right) [(\Delta_i X - h a_{i-1}(\theta_2))^{\otimes 2}] h^{-1} \right. \\ &\quad \left. + \log \frac{|B_{i-1}(\theta_1)|}{|B_{i-1}(\theta_1^*)|} \right\}. \end{aligned} \tag{20}$$

Define $\mathbb{Y}(\theta_1; \theta^*) = \mathbb{Y}(\theta_1, \theta_2; \theta^*)$ by

$$\mathbb{Y}(\theta_1; \theta^*) = -\frac{1}{2} \int_{\mathbb{R}^d} \left\{ \mathrm{Tr} \left(B(x, \theta_1)^{-1} B(x, \theta_1^*) - I_d \right) + \log \frac{|B(x, \theta_1)|}{|B(x, \theta_1^*)|} \right\} \nu(\mathrm{d}x).$$

The sum in $\{\ldots\}$ on the right-hand side of the above equation is obviously non-negative. We assume the following identifiability condition:

[D3] There exists a positive constant $\chi(\theta^*)$ such that $\mathbb{Y}(\theta_1; \theta^*) \leq -\chi(\theta^*)|\theta_1 - \theta_1^*|^2$ for all $\theta_1 \in \Theta_1$.

That the parametric model can be extended continuously to a compact set including Θ_1 and $\mathbb{Y} \neq 0$ for $\theta_1 \neq \theta_1^*$ on it is a sufficient condition for [D3].

[15] The last one is the so-called condition for rapidly increasing experimental design. Relaxing this condition is possible, but we do not do it here for simplicity. We will comment on this point later.

$\Delta_n(\theta_2;\theta^*)$ for "$\Delta_T(\tau;\xi_0)$" and $\Gamma_n(\theta_1,\theta_2;\theta^*)$ for "$\Gamma_T(\theta,\tau;\xi_0)$" are given by

$$\Delta_n(\theta_2;\theta^*)[u_1] = -\frac{1}{2n^{1/2}}\sum_{i=1}^{n}\left\{\partial_{\theta_1}B_{i-1}^{-1}(\theta_1^*)[u_1,(\Delta_i X - ha_{i-1}(\theta_2))^{\otimes 2}]h^{-1} + \frac{\partial_{\theta_1}|B_{i-1}|(\theta_1^*)[u_1]}{|B_{i-1}(\theta_1^*)|}\right\}$$

and

$$\Gamma_n(\theta_1,\theta_2;\theta^*)[u_1,u_1] = \frac{1}{2n}\sum_{i=1}^{n}\left\{\partial_{\theta_1}^2 B_{i-1}^{-1}(\theta_1)[u_1^{\otimes 2},(\Delta_i X - ha_{i-1}(\theta_2))^{\otimes 2}]h^{-1} + \partial_{\theta_1}^2 \log\frac{|B_{i-1}(\theta_1)|}{|B_{i-1}(\theta_1^*)|}[u_1^{\otimes 2}]\right\}$$

for $u_1 \in \mathbb{R}^{m_1}$.

First, we want to obtain a polynomial-type large deviation inequality. For this, we need some preliminary estimates.

Lemma 3 *For every $p > 1$,*

$$\sup_{n\in\mathbb{N}} P_{\theta^*}\left[\sup_{\theta_2\in\Theta_2}|\Delta_n(\theta_2;\theta^*)|^p\right] < \infty.$$

Proof Since

$$\Delta_i X - ha_{i-1}(\theta_2) = b_{i-1}(\theta_1^*)\Delta_i w + \int_{t_{i-1}}^{t_i}(b(X_t,\theta_1^*) - b(X_{t_{i-1}},\theta_1^*))\mathrm{d}w_t + \int_{t_{i-1}}^{t_i}(a(X_t,\theta_2^*) - a(X_{t_{i-1}},\theta_2))\mathrm{d}t,$$

we obtain a decomposition

$$\Delta_n(\theta_2;\theta^*)[u_1] = M_n + R_n,$$

where

$$M_n = -\frac{1}{2n^{1/2}}\sum_{i=1}^{n}\left\{\partial_{\theta_1}B_{i-1}^{-1}(\theta_1^*)[u_1,(b_{i-1}(\theta_1^*)\Delta_i w)^{\otimes 2}]h^{-1} + \frac{\partial_{\theta_1}|B_{i-1}|(\theta_1^*)[u_1]}{|B_{i-1}(\theta_1^*)|}\right\} \tag{21}$$

and R_n is the residual. It is quite routine to show that for every $p \geq 1$,

$$\sup_{\theta_2\in\Theta_2}\|R_n\|_p \leq C|u_1|(\sqrt{nh^2}+\sqrt{h}) \to 0$$

if one uses the orthogonality and the estimate $\|X_t - X_s\|_p \leq C\sqrt{|t-s|}$.

Due to

$$P_{\theta^*}\left[\left\{\partial_{\theta_1} B_{i-1}^{-1}(\theta_1^*)[u_1, (b_{i-1}(\theta_1^*)\Delta_i w)^{\otimes 2}]h^{-1} + \frac{\partial_{\theta_1}|B_{i-1}|(\theta_1^*)[u_1]}{|B_{i-1}(\theta_1^*)|}\right\} \middle| \mathcal{F}_{t_{i-1}}\right]$$
$$= \partial_{\theta_1} B_{i-1}^{-1}(\theta_1^*)[u_1, B_{i-1}(\theta_1^*)] + \frac{\partial_{\theta_1}|B_{i-1}|(\theta_1^*)[u_1]}{|B_{i-1}(\theta_1^*)|}$$
$$= 0,$$

we observe that $n^{1/2}M_n$ is the terminal value of a discrete-time martingale with respect to the filtration $(\mathcal{F}_{t_i})_{i=0,1,\dots,n}$. Thus Burkholder's inequality implies $\sup_{n\in\mathbb{N}} \|M_n\|_p < \infty$. Therefore,

$$\sup_{n\in\mathbb{N}} \sup_{\theta_2\in\Theta_2} \|\Delta_n(\theta_2;\theta^*)\|_p < \infty$$

for every p.

In a similar fashion, from

$$\partial_{\theta_2}\Delta_n(\theta_2;\theta^*)[u_2,u_1]$$
$$= \frac{1}{n^{1/2}} \sum_{i=1}^n \partial_{\theta_1} B_{i-1}^{-1}(\theta_1^*)[u_1, \Delta_i X - ha_{i-1}(\theta_2), \partial_{\theta_2} a_{i-1}(\theta_2)[u_2]],$$

we have

$$\sup_{\theta_2\in\Theta_2} \left\|\partial_{\theta_2}\Delta_n(\theta_2;\theta^*)\right\|_p \le C_p(\sqrt{h} + \sqrt{nh^2})$$

for every p.

By Sobolev's inequality,

$$P_{\theta^*}\left[\sup_{\theta_2\in\Theta_2} |\Delta_n(\theta_2;\theta^*)|^p\right] \le P_{\theta^*}\left[C\int_{\Theta_2} \left\{|\Delta_n(\theta_2;\theta^*)|^p + |\partial_{\theta_2}\Delta_n(\theta_2;\theta^*)|^p\right\} \mathrm{d}\theta_2\right]$$
$$\le C_{\Theta_2}\left\{\sup_{\theta_2\in\Theta_2} P_{\theta^*}\left[|\Delta_n(\theta_2;\theta^*)|^p\right] + \sup_{\theta_2\in\Theta_2} P_{\theta^*}\left[|\partial_{\theta_2}\Delta_n(\theta_2;\theta^*)|^p\right]\right\},$$

where $p > m_2$. The right-hand side of the above inequality is bounded uniformly in $n \in \mathbb{N}$. □

Remark 11 In the above proof, we did not use the mixing property because the principal part of $\Delta_n(\theta_2;\theta^*)$ was a martingale. If we use the mixing assumption, another proof is possible.

Lemma 4 *Let F_j ($j \in \mathbb{N}$) be a (stationary) centered process and suppose that for some $h > 0$ and $C > 0$,*

$$\sup_{j\in\mathbb{N}} \sup_{A\in\sigma[F_l;l\le j],\ B\in\sigma[F_l;l\ge j+k]} |P[A\cap B] - P[A]P[B]| \le C\exp(-hk)$$

for all $k \in \mathbb{N}$ and that for every $p \ge 2$, $\sup_j \|F_j\|_p \le C_p$ for some constant C_p depending on p but independent of $h > 0$. Then for some constant $C' = C'(C, p, C_{p+1}) < \infty$ independent of h and the sequence F_j,

$$P\left[\sup_{j=1,\dots,n}\left|\sum_{i=1}^{j} F_i\right|^p\right] \le C'\left[(nh^{-1})^{p/2} + nh^{1-p}\right]$$

for all $n \in \mathbb{N}$.

Proof We use a Rosenthal type inequality for mixing sequences (Rio 1994 or Doukhan et al. 2003). Stationarity is not necessary, as a matter of fact. Thanks to the inequality on p.68 of Doukhan et al. (2003), in our case we know that the constant $M_{p,\alpha}$ in their notation admits the estimate (with a small positive constant C)

$$\begin{aligned} M_{p,\alpha} &\le \left(\int_0^1 (1 - h^{-1}\log(Ct))^{(p+1)(p-1)}\,\mathrm{d}t\right)^{1/(p+1)} \sup_j \|F_j\|_{p+1}^p \\ &\le C'_p h^{1-p}. \end{aligned}$$

Thus the Rosenthal-type inequality yields the desired inequality. □

Lemma 5 *Let $\epsilon_1 = \epsilon_0/2$. Then for every $p > 0$,*

$$\sup_{n\in\mathbb{N}} P_{\theta^*}\left[\left(\sup_{\theta\in\Theta} n^{\epsilon_1}\left|\mathbb{Y}_n(\theta_1,\theta_2;\theta^*) - \mathbb{Y}(\theta_1;\theta^*)\right|\right)^p\right] < \infty.$$

Proof Let

$$\mathbb{Y}_n^\dagger(\theta_1;\theta^*) = -\frac{1}{2n}\sum_{i=1}^{n}\left\{\left(B_{i-1}^{-1}(\theta_1) - B_{i-1}^{-1}(\theta_1^*)\right)[B_{i-1}(\theta_1^*)] + \log\frac{|B_{i-1}(\theta_1)|}{|B_{i-1}(\theta_1^*)|}\right\}.$$

Define $R_n^\dagger$ by

$$R_n^\dagger = \mathbb{Y}_n(\theta_1,\theta_2;\theta^*) - \mathbb{Y}_n^\dagger(\theta_1;\theta^*) - M_n^\dagger$$

for

$$M_n^\dagger = -\frac{1}{2n}\sum_{i=1}^{n}\left(B_{i-1}^{-1}(\theta_1) - B_{i-1}^{-1}(\theta_1^*)\right)[(b_{t_{i-1}}(\theta_1^*)\Delta_i w)^{\otimes 2}h^{-1} - B_{i-1}(\theta_1^*)].$$

As in the proof of Lemma 3, we obtain

$$\sup_{n\in\mathbb{N}}\sup_{\theta\in\Theta}\|n^{\frac{1}{2}}R_n^{\dagger}\|_p<\infty \quad\text{and}\quad \sup_{n\in\mathbb{N}}\sup_{\theta_1\in\Theta_1}\|n^{\frac{1}{2}}M_n^{\dagger}\|_p<\infty$$

for every p. Moreover, following the same procedure, we obtain

$$\sup_{n\in\mathbb{N}}\sup_{\theta\in\Theta}\|n^{\frac{1}{2}}\partial_\theta R_n^{\dagger}\|_p<\infty \quad\text{and}\quad \sup_{n\in\mathbb{N}}\sup_{\theta_1\in\Theta_1}\|n^{\frac{1}{2}}\partial_{\theta_1} M_n^{\dagger}\|_p<\infty,$$

after all,

$$\sup_{n\in\mathbb{N}}\left\|\sup_{\theta\in\Theta}|n^{\frac{1}{2}}R_n^{\dagger}|\right\|_p<\infty \quad\text{and}\quad \sup_{n\in\mathbb{N}}\left\|\sup_{\theta_1\in\Theta_1}|n^{\frac{1}{2}}M_n^{\dagger}|\right\|_p<\infty$$

for every p.

Let $\mathcal{B}_{[s,t]}$ be the σ-field generated by X_u and w_v-w_u $(s\leq u\leq v\leq t)$. Lemma 1 of Kusuoka and Yoshida (2000) or Lemma 1 of Yoshida (2004) ensures that the exponential mixing property [D2] implies the exponential mixing property for $\mathcal{B}_I$ $(I\subset\mathbb{R}_+)$. Obviously, under the assumption that $nh\geq n^{\epsilon_0}$ for large n,

$$\begin{aligned}&(n^{-1+\frac{1}{2}\epsilon_0})^p\left[(nh^{-1})^{p/2}+nh^{1-p}\right]\\&\quad\leq[n^{\epsilon_0}(nh)^{-1}]^{p/2}[1+(nh)^{1-p/2}]\leq 1+(nh)^{1-p/2}\end{aligned}$$

for large n. We may assume that $p\geq 2$. Since $\mathbb{Y}_n^{\dagger}(\theta_1;\theta^*)-\mathbb{Y}(\theta_1;\theta^*)$ is a centered functional, it follows from the mixing property and Lemma 4 that

$$\sup_{n\in\mathbb{N}}\sup_{\theta_1\in\Theta_1}\left\|n^{\epsilon_1}\left(\mathbb{Y}_n^{\dagger}(\theta_1;\theta^*)-\mathbb{Y}(\theta_1;\theta^*)\right)\right\|_p<\infty.$$

Quite similarly, we can obtain

$$\sup_{n\in\mathbb{N}}\sup_{\theta_1\in\Theta_1}\left\|n^{\epsilon_1}\partial_{\theta_1}\left(\mathbb{Y}_n^{\dagger}(\theta_1;\theta^*)-\mathbb{Y}(\theta_1;\theta^*)\right)\right\|_p<\infty.$$

and consequently

$$\sup_{n\in\mathbb{N}}\left\|\sup_{\theta_1\in\Theta_1}n^{\epsilon_1}|\mathbb{Y}_n^{\dagger}(\theta_1;\theta^*)-\mathbb{Y}(\theta_1;\theta^*)|\right\|_p<\infty$$

for every p. Thus we obtain the desired inequality. □

Set

$$\begin{aligned}\Gamma_1(\theta^*)[u_1,u_1] &:= \Gamma(\theta_2;\theta^*)[u_1,u_1]\\ &:= \frac{1}{2}\int\left\{\partial_{\theta_1}^2 B^{-1}(x,\theta_1)[u_1^{\otimes 2}, B(x,\theta_1^*)]\right.\\ &\quad\left.+\partial_{\theta_1}^2 \log\frac{|B(x,\theta_1)|}{|B(x,\theta_1^*)|}[u_1^{\otimes 2}]\right\}\Big|_{\theta_1=\theta_1^*}\nu(\mathrm{d}x)\\ &= \frac{1}{2}\int \mathrm{Tr}\left\{B^{-1}(\partial_{\theta_1}B)B^{-1}(\partial_{\theta_1}B)(x,\theta_1^*)[u_1^{\otimes 2}]\right\}\ \nu(\mathrm{d}x)\end{aligned}$$

for $u_1\in\mathbb{R}^{m_1}$; we dared keep the denominator in the logarithm.

Condition [D3] implies [B1] and [B2]. We make L sufficiently large as all necessary estimates in the sequel can be obtained. Inequalities in [A6] with a sufficiently large $\beta_2\in(0,1/2)$ follow from Lemmas 3 and 5. Next, we shall verify [A1$''$].

Lemma 6 *For any $M_3>0$,*

$$\sup_{n\in\mathbb{N}} P_{\theta^*}\left[\left(n^{-1}\sup_{\theta\in\Theta}\left|\partial_{\theta_1}^3\mathbb{H}_n(\theta_1,\theta_2)\right|\right)^{M_3}\right]<\infty.$$

Proof From (20), we have the following 3 and 4-linear forms:

$$\begin{aligned}&n^{-1}\partial_{\theta_1}^3\mathbb{H}_n(\theta_1,\theta_2)\\ &\quad=-\frac{1}{2n}\sum_{i=1}^n\left\{\partial_{\theta_1}^3 B_{i-1}^{-1}(\theta_1)[(\Delta_i X-ha_{i-1}(\theta_2))^{\otimes 2}]h^{-1}+\partial_{\theta_1}^3\log|B_{i-1}(\theta_1)|\right\}\end{aligned}$$

and

$$n^{-1}\partial_{\theta_2}\partial_{\theta_1}^3\mathbb{H}_n(\theta_1,\theta_2)=-\frac{1}{n}\sum_{i=1}^n\partial_{\theta_1}^3 B_{i-1}^{-1}(\theta_1)[\Delta_i X-ha_{i-1}(\theta_2),\partial_{\theta_2}a_{i-1}(\theta_2)].$$

Then the L^p-boundedness of those random multi-linear forms is easy to obtain. □

Lemma 7 *For any $M_4>0$,*

$$\sup_{n\in\mathbb{N}} P_{\theta^*}\left[\sup_{\theta_2\in\Theta_2}\left(n^{\epsilon_1}\left|\Gamma_n(\theta_1^*,\theta_2;\theta^*)-\Gamma_1(\theta^*)\right|\right)^{M_4}\right]<\infty.$$

Proof By definition,

$$
\begin{aligned}
&\Gamma_n(\theta_1^*, \theta_2; \theta^*)[u_1, u_1] - \Gamma_1(\theta^*)[u_1, u_1] \\
&= \frac{1}{2n}\sum_{i=1}^{n}\Bigg\{\partial_{\theta_1}^2 B_{i-1}^{-1}(\theta_1)[u_1^{\otimes 2}, (\Delta_i X - ha_{i-1}(\theta_2))^{\otimes 2}]h^{-1} \\
&\quad + \partial_{\theta_1}^2 \log\frac{|B_{i-1}(\theta_1)|}{|B_{i-1}(\theta_1^*)|}[u_1^{\otimes 2}]\Bigg\}\Bigg|_{\theta_1=\theta_1^*} \\
&\quad -\frac{1}{2}\int\Bigg\{\partial_{\theta_1}^2 B^{-1}(x, \theta_1)[u_1^{\otimes 2}, B(x, \theta_1^*)] \\
&\quad + \partial_{\theta_1}^2 \log\frac{|B(x, \theta_1)|}{|B(x, \theta_1^*)|}[u_1^{\otimes 2}]\Bigg\}\Bigg|_{\theta_1=\theta_1^*} \nu(\mathrm{d}x).
\end{aligned}
$$

Then one obtains the result in a similar way as the proof of Lemmas 6 and 5. □

The function $\theta \mapsto p_n(\mathbf{x}_n, \theta)$ can be continuously extended to the boundary of Θ. Conditions [A1″] and [A4′] follow from Lemmas 6 and 7. Let

$$
V_n^1(r, \theta_1^*) = \{u_1 \in \mathbb{U}_n^1(\theta_1^*);\ r \le |u_1|\},
$$

where

$$
\mathbb{U}_n^1(\theta_1^*) = \{u_1 \in \mathbb{R}^{m_1};\ \theta_1^* + n^{-1/2}u_1 \in \Theta_1\}.
$$

From Theorem 3, we obtain the polynomial type large deviation inequality

$$
P_{\theta^*}\left[\sup_{(u_1,\theta_2)\in V_n^1(r,\theta_1^*)\times\Theta_2} \mathbb{Z}_n^1(u_1; \theta_1^*, \theta_2) \ge e^{-r}\right] \le \frac{C_L}{r^L} \tag{22}
$$

for all $r > 0$ and $n \in \mathbb{N}$, where

$$
\mathbb{Z}_n^1(u_1; \theta_1^*, \theta_2) = \exp\left\{\mathbb{H}_n(\theta_1^* + n^{-1/2}u_1, \theta_2) - \mathbb{H}_n(\theta_1^*, \theta_2)\right\}.
$$

In particular, applying (22) for arbitrary $L > 0$, we have

$$
\sup_{n\in\mathbb{N}} P_{\theta^*}\left[|\sqrt{n}(\hat{\theta}_1 - \theta_1^*)|^p\right] < \infty
$$

for every $p > 0$. See Proposition 1.

We proceed to the second step of estimates. In this step, we focus our attention to the statistical random field

$$
\mathbb{Z}_n^2(u_2; \hat{\theta}_1, \theta_2^*) = \exp\left(\mathbb{H}_n(\hat{\theta}_1, \theta_2^* + (nh)^{-1/2}u_2) - \mathbb{H}_n(\hat{\theta}_1, \theta_2^*)\right).
$$

Here it should be noted that "$\mathbb{H}_T$" in Sect. 2 is in this case $\theta_2 \mapsto \mathbb{H}_n(\hat{\theta}_1, \theta_2)$ without the parameter "τ". Moreover, $\mathbb{Y}_n(\theta_2; \theta^*)$ for $\mathbb{Y}_T(\theta, \tau; \xi_0)$ in Sect. 2 is given by

$$\begin{aligned}\mathbb{Y}_n(\theta_2; \theta^*) &= (nh)^{-1} \Big(\mathbb{H}_n(\hat{\theta}_1, \theta_2) - \mathbb{H}_n(\hat{\theta}_1, \theta_2^*)\Big) \\ &= (nh)^{-1} \Bigg\{ \sum_{i=1}^n B_{i-1}^{-1}(\hat{\theta}_1)[a_{i-1}(\theta_2) - a_{i-1}(\theta_2^*), \Delta_i X] \\ &\quad - \frac{1}{2} \sum_{i=1}^n h\, B_{i-1}^{-1}(\hat{\theta}_1)[a_{i-1}(\theta_2)^{\otimes 2} - a_{i-1}(\theta_2^*)^{\otimes 2}] \Bigg\}.\end{aligned}$$

Define $\mathbb{Y}(\theta_2; \theta^*)$[16] by

$$\mathbb{Y}(\theta_2; \theta^*) = -\frac{1}{2} \int_{\mathbb{R}^d} B(x, \theta_1^*)^{-1}[(a(x, \theta_2) - a(x, \theta_2^*))^{\otimes 2}]\, \nu(\mathrm{d}x). \tag{23}$$

In place of $\Delta_T(\tau; \xi_0)[u]$ of Sect. 3, we consider

$$\begin{aligned}\Delta_n(\theta^*)[u_2] &= \partial_{\theta_2} \mathbb{H}_n(\hat{\theta}_1, \theta_2^*)[(nh)^{-1/2} u_2] \\ &= (nh)^{-1/2} \sum_{i=1}^n B_{i-1}^{-1}(\hat{\theta}_1)[\partial_{\theta_2} a_{i-1}(\theta_2^*) u_2, \Delta_i X - h a_{i-1}(\theta_2^*)]\end{aligned}$$

for $u_2 \in \mathbb{R}^{m_2}$.

Lemma 8 (a) *For every $p > 0$,*

$$\sup_{n\in\mathbb{N}} \left\| \sup_{\theta_2 \in \Theta_2} (nh)^{\epsilon_1} |\mathbb{Y}_n(\theta_2; \theta^*) - \mathbb{Y}(\theta_2; \theta^*)| \right\|_p < \infty.$$

(b) *For every $p > 0$,*

$$\sup_{n\in\mathbb{N}} P_{\theta^*} \left[|\Delta_n(\theta^*)|^p \right] < \infty.$$

Proof $\mathbb{Y}_n(\theta_2; \theta^*)$ has a decomposition

$$\mathbb{Y}_n(\theta_2; \theta^*) = \mathcal{M}_n(\hat{\theta}_1, \theta_2) + \mathcal{R}_n(\hat{\theta}_1, \theta_2) + \widetilde{\mathbb{Y}}_n(\hat{\theta}_1, \theta_2; \theta^*)$$

[16] This is different from $\mathbb{Y}(\theta_1; \theta^*) = \mathbb{Y}(\theta_1, \theta_2; \theta^*)$ before.

where

$$\mathcal{M}_n(\theta_1,\theta_2) = (nh)^{-1}\sum_{i=1}^{n} B_{i-1}^{-1}(\theta_1)[a_{i-1}(\theta_2)-a_{i-1}(\theta_2^*), \int_{t_{i-1}}^{t_i} b(X_t,\theta_1^*)\mathrm{d}w_t],$$

$$\mathcal{R}_n(\theta_1,\theta_2) = (nh)^{-1}\sum_{i=1}^{n} B_{i-1}^{-1}(\theta_1)[a_{i-1}(\theta_2)-a_{i-1}(\theta_2^*), \int_{t_{i-1}}^{t_i}\big(a(X_t,\theta_2^*)-a(X_{t_{i-1}},\theta_2^*)\big)\,\mathrm{d}t]$$

and

$$\widetilde{\mathbb{Y}}_n(\theta_1,\theta_2;\theta^*) = -\frac{1}{2n}\sum_{i=1}^{n} B_{i-1}^{-1}(\theta_1)[(a_{i-1}(\theta_2)-a_{i-1}(\theta_2^*))^{\otimes 2}].$$

Estimating the norms of $\partial_\theta^j \mathcal{M}_n(\theta_1,\theta_2)$ $(j=0,1)$ with orthogonalities, we see that for every $p>0$,

$$\sup_{n\in\mathbb{N}} P_{\theta^*}\left[\sup_{\theta\in\Theta}|(nh)^{\frac{1}{2}}\mathcal{M}_n(\theta_1,\theta_2)|^p\right]<\infty;$$

in particular,

$$\sup_{n\in\mathbb{N}} P_{\theta^*}\left[\sup_{\theta_2\in\Theta_2}|(nh)^{\frac{1}{2}}\mathcal{M}_n(\hat\theta_1,\theta_2)|^p\right]<\infty.$$

We use L^p-triangular inequality to estimate $\partial_{\theta_2}^j\mathcal{R}_n(\hat\theta_1,\theta_2)$ $(j=0,1)$ and obtain

$$P_{\theta^*}\left[\sup_{\theta_2\in\Theta_2}|(nh)^{\frac{1}{2}}\mathcal{R}_n(\hat\theta_1,\theta_2)|^p\right]^{1/p} = O(\sqrt{n}\,h)=o(1).$$

Since the L^p-boundedness of $\sqrt{n}(\hat\theta_1-\theta_1^*)$ has been established, we can show that

$$\begin{aligned}
&\left\| n^{1/2}\sup_{\theta_2\in\Theta_2}\left|\widetilde{\mathbb{Y}}_n(\hat\theta_1,\theta_2;\theta^*)-\widetilde{\mathbb{Y}}_n(\theta_1^*,\theta_2;\theta^*)\right|\right\|_p\\
&\le \frac{n^{1/2}}{2n}\sum_{i=1}^{n}\left\| B_{i-1}^{-1}(\hat\theta_1)-B_{i-1}^{-1}(\theta_1^*)\right\|_{2p}\left\|\sup_{\theta_2\in\Theta_2}|a_{i-1}(\theta_2)-a_{i-1}(\theta_2^*)|^2\right\|_{2p}\\
&= O(1)
\end{aligned}$$

for every $p > 1$. Finally, since $\tilde{\mathbb{Y}}_n(\theta_1^*, \theta_2; \theta^*) - \mathbb{Y}(\theta_2; \theta^*)$ is a centered random variable, in a similar way to the proof of Lemma 5, we can show that

$$\sup_{n\in\mathbb{N}} \left\| \sup_{\theta_2\in\Theta_2} n^{\epsilon_1} |\widetilde{\mathbb{Y}}_n(\theta_1^*, \theta_2; \theta^*) - \mathbb{Y}(\theta_2; \theta^*)| \right\|_p < \infty$$

for every p. It is possible to replace n^{ϵ_1} in the above inequality by $(nh)^{\epsilon_1}$ to conclude the proof of (a).

If we decompose $\partial_{\theta_2}\mathbb{H}_n(\theta_1, \theta_2^*)[(nh)^{-1/2}u_2]$ into a "martingale" part and a residual part, it is not difficult to show

$$\sup_{n\in\mathbb{N}} P_{\theta^*}\left[\sup_{\theta_1\in\Theta_1} |\partial_{\theta_2}\mathbb{H}_n(\theta_1, \theta_2^*)[(nh)^{-1/2}u_2]|^p \right] < \infty$$

so that

$$\sup_{n\in\mathbb{N}} P_{\theta^*}\left[|\Delta_n(\theta^*)|^p\right] < \infty.$$

□

From Lemma 8, we see that Condition [A6] is satisfied for the random field in question.

We have

$$\begin{aligned}
&(nh)^{-1}\partial^3_{\theta_2}\mathbb{H}_n(\hat{\theta}_1, \theta_2^*)\\
&\quad = (nh)^{-1}\sum_{j=0}^{2} C(j) \sum_{i=1}^{n} B_{i-1}^{-1}(\hat{\theta}_1)[\partial_{\theta_2}^{j+1} a_{i-1}(\theta_2)[\,\cdot\,], \partial_{\theta_2}^{2-j}\{\Delta_i X - h a_{i-1}(\theta_2)\}[\,\cdot\,]]\Big|_{\theta_2=\theta_2^*}
\end{aligned}$$

and

$$\begin{aligned}
\Gamma_n(\theta_2; \theta^*)[u_2^{\otimes 2}] &:= -(nh)^{-1}\partial^2_{\theta_2}\mathbb{H}_n(\hat{\theta}_1, \theta_2)[u_2^{\otimes 2}]\\
&= (nh)^{-1}\sum_{i=1}^{n} B_{i-1}^{-1}(\hat{\theta}_1)[\partial_{\theta_2} a_{i-1}(\theta_2)[u_2], \partial_{\theta_2} h a_{i-1}(\theta_2)[u_2]]\\
&\quad -(nh)^{-1}\sum_{i=1}^{n} B_{i-1}^{-1}(\hat{\theta}_1)[\partial^2_{\theta_2} a_{i-1}(\theta_2)[u_2^{\otimes 2}], \Delta_i X - h a_{i-1}(\theta_2)].
\end{aligned}$$

Let

$$\Gamma_2(\theta^*)[u_2^{\otimes 2}] = \int_{\mathbb{R}^d} B(x, \theta_1^*)^{-1}[\partial_{\theta_2} a(x, \theta_2^*)[u_2], \partial_{\theta_2} a(x, \theta_2^*)[u_2]]\, \nu(\mathrm{d}x).$$

We need the following identifiability condition:

[D4] There exists a positive constant $\chi'(\theta^*)$ such that $\mathbb{Y}(\theta_2;\theta^*) \le -\chi'(\theta^*)|\theta_2-\theta_2^*|^2$ for all $\theta_2 \in \Theta_2$. Here $\mathbb{Y}(\theta_2;\theta^*)$ is given in (23).

We can prove the following lemma in a similar manner as Lemma 8:

Lemma 9 (a) *For every $M_3 > 0$,*

$$\sup_{n\in\mathbb{N}} P_{\theta^*}\left[\left((nh)^{-1}\sup_{\theta_2\in\Theta_2}\left|\partial_{\theta_2}^3\mathbb{H}_n(\hat{\theta}_1,\theta_2)\right|\right)^{M_3}\right] < \infty.$$

(b) *For every $M_4 > 0$,*

$$\sup_{n\in\mathbb{N}} P_{\theta^*}\left[\left|(nh)^{\epsilon_1}\left(\Gamma_n(\theta_2^*;\theta^*) - \Gamma_2(\theta^*)\right)\right|^{M_4}\right] < \infty.$$

By using Lemma 9, we see that Conditions [A1″] and [A4′] are satisfied if we choose the parameters adequately. Conditions [B1] and [B2] hold under [D4]. Applying Theorem 3 once again, we obtain the polynomial type large deviation inequality

$$P_{\theta^*}\left[\sup_{u_2\in V_n^2(r,\theta_2^*)} \mathbb{Z}_n^2(u_2;\hat{\theta}_1,\theta_2^*) \ge e^{-r}\right] \le \frac{C_L}{r^L} \tag{24}$$

for all $r > 0$ and $n \in \mathbb{N}$, where $V_n^2(r,\theta_2^*) = \{u_2 \in \mathbb{U}_n^2(\theta_2^*);\ r \le |u_2|\}$ with $\mathbb{U}_n^2(\theta_2^*) = \{u_2 \in \mathbb{R}^{m_2};\ \theta_2^* + (nh)^{-1/2}u_2 \in \Theta_2\}$. Then it follows from Proposition 2 together with Inequalities (22) and (24) that the normalized M-estimator $\hat{u}_n = (\sqrt{n}(\hat{\theta}_1 - \theta_1^*), \sqrt{nh}(\hat{\theta}_2 - \theta_2^*))$ satisfies the polynomial type large deviation inequality

$$P_{\theta^*}\left[|\hat{u}_n| > r\right] \le \frac{C_L'}{r^L} \tag{25}$$

for all $n > 0$ and $r > 0$, for some constant C_L'.

For $u_i \in \mathbb{R}^{m_i}$ $(i = 1, 2)$, let $\theta_1^\dagger = \theta_1^* + n^{-1/2}u_1$ and $\theta_2^\dagger = \theta_2^* + (nh)^{-1/2}u_2$. In the same way as we took in Lemma 9 or directly, we can show that

$$\begin{aligned}
\mathbb{H}_n(\theta_1^\dagger,\theta_2^\dagger) - \mathbb{H}_n(\theta_1^\dagger,\theta_2^*) &= \partial_{\theta_2}\mathbb{H}_n(\theta_1^\dagger,\theta_2^*)[(nh)^{-1/2}u_2] \\
&\quad + \frac{1}{2}\partial_{\theta_2}^2\mathbb{H}_n(\theta_1^\dagger,\theta_2^*)[((nh)^{-1/2}u_2)^{\otimes 2}] + o_p(1) \\
&= \partial_{\theta_2}\mathbb{H}_n(\theta_1^*,\theta_2^*)[(nh)^{-1/2}u_2] \\
&\quad + \frac{1}{2}\partial_{\theta_2}^2\mathbb{H}_n(\theta_1^\dagger,\theta_2^*)[((nh)^{-1/2}u_2)^{\otimes 2}] + o_p(1) \\
&\to^d \Delta_2[u_2] - \frac{1}{2}\Gamma_2(\theta^*)[u_2^{\otimes 2}],
\end{aligned} \tag{26}$$

where $\Delta_2 \sim N_{m_2}(0, \Gamma_2(\theta^*))$, which follows from the central limit theorem for

$$\begin{aligned}&(nh)^{-1/2}\partial_{\theta_2}\mathbb{H}_n(\theta_1^*, \theta_2^*)[u_2]\\&\quad= (nh)^{-1/2}\sum_{i=1}^{n} B_{i-1}^{-1}(\theta_1^*)[\partial_{\theta_2}a_{i-1}(\theta_2^*)u_2,\ \Delta_i X - ha_{i-1}(\theta_2^*)]\\&\quad= (nh)^{-1/2}\sum_{i=1}^{n} B_{i-1}^{-1}(\theta_1^*)[\partial_{\theta_2}a_{i-1}(\theta_2^*)u_2,\ \int_{t_{i-1}}^{t_i} b(X_t, \theta_1^*)dw_t] + O_p(\sqrt{n}\,h).\end{aligned}$$

Furthermore, we already know that the random field $\theta_1 \mapsto \mathbb{H}_n(\theta_1, \theta_2)$ satisfies Condition [A1] so that it admits the locally asymptotic normal structure:

$$\begin{aligned}\mathbb{H}_n(\theta_1^\dagger, \theta_2^*) - \mathbb{H}_n(\theta_1^*, \theta_2^*) &= \Delta_n(\theta_2^*; \theta^*)[u_1] - \frac{1}{2}\Gamma_1(\theta^*)[u_1^{\otimes 2}] + o_p(1)\\&\to^d \Delta_1[u_1] - \frac{1}{2}\Gamma_1(\theta^*)[u_1^{\otimes 2}]\end{aligned} \tag{27}$$

as $n \to \infty$, where $\Delta_1 \sim N_{m_1}(0, \Gamma_1(\theta^*))$: the asymptotic normality is a consequence of the martingale central limit theorem for M_n given in (21). From (26) and (27), we obtain the local asymptotic normality for the full parameter:

$$\begin{aligned}\log \mathbb{Z}_n(u_1, u_2; \theta^*) &= \mathbb{H}_n(\theta_1^\dagger, \theta_2^\dagger) - \mathbb{H}_n(\theta_1^*, \theta_2^*)\\&\to^d \Delta_1[u_1] + \Delta_2[u_2] - \frac{1}{2}\left\{\Gamma_1(\theta^*)[u_1^{\otimes 2}] + \Gamma_2(\theta^*)[u_2^{\otimes 2}]\right\}.\end{aligned} \tag{28}$$

The orthogonality between Δ_1 and Δ_2 is obvious: $\Delta_1 \perp\!\!\!\perp \Delta_2$. Easily we see finite-dimensional convergence, i.e., the joint convergence for finitely many (u_1, u_2)'s, is valid for (28).

Let $C(R) = \{u = (u_1, u_2); |u| \leq R\}$. The tightness of the family $\{\log Z_n(u_1, u_2; \theta^*)|_{C(R)};\ n \in \mathbb{N}\}$ for every $R > 0$ can be proved by a tightness criterion of a moment inequality if one uses L^p-estimates for $\sup_{u\in C(R)} |\partial_u \log Z_n(u_1, u_2; \theta^*)|$, which are similar to the estimates of $\Delta_n(\theta_2; \theta^*)[u_1]$, and $\Delta_n(\theta^*)[u_2]$ together with the estimate of $\mathbb{Y}_n(\theta_2; \theta^*)$ for uniformity in the parameter (see also Theorem 12). Thus we conclude that for every $R > 0$,

$$\mathbb{Z}_n(u_1, u_2; \theta^*) \to^d \mathbb{Z}(u_1, u_2; \theta^*) \quad \text{in } C(R) \tag{29}$$

as $n \to \infty$, where

$$\mathbb{Z}(u_1, u_2; \theta^*) = \exp\left(\Delta_1[u_1] + \Delta_2[u_2] - \frac{1}{2}\left\{\Gamma_1(\theta^*)[u_1^{\otimes 2}] + \Gamma_2(\theta^*)[u_2^{\otimes 2}]\right\}\right).$$

We obtain the following theorem from Theorem 5:

Theorem 13 *Suppose that Conditions* [D1]–[D4] *are fulfilled. Then for any sequence of M-estimators for $\theta = (\theta_1, \theta_2)$, it holds that*

$$\left(\sqrt{n}(\hat{\theta}_1 - \theta_1^*), \sqrt{nh}(\hat{\theta}_2 - \theta_2^*)\right) \to^d (\zeta_1, \zeta_2)$$
$$\sim N_{m_1+m_2}\left(0, \text{diag}\,[\Gamma_1(\theta^*)^{-1}, \Gamma_2(\theta^*)^{-1}]\right)$$

as $n \to \infty$. Moreover, $P_{\theta^*}\left[f(\sqrt{n}(\hat{\theta}_1 - \theta_1^*), \sqrt{nh}(\hat{\theta}_2 - \theta_2^*))\right] \to \mathbb{P}[f(\zeta_1, \zeta_2)]$ as $n \to \infty$ for all continuous functions f of at most polynomial growth.

Remark 12 Prakasa Rao (1983, 1988) presented asymptotic results for an ergodic diffusion process under a sampling scheme. The joint weak convergence in Theorem 13 was given in Yoshida (1992b). These are based on the local Gaussian approximations in their respective settings. Kessler (1997) treated an approximation with higher order correction terms to relax the rate of convergence of h to zero. Shimizu and Yoshida (2003) obtained asymptotic normality for a maximum likelihood type estimator for a diffusion process with jumps. Yoshida (2005a,b) gave an asymptotic expansion formula, as well as the first-order results, for an M-estimator for a diffusion process with (possibly infinitely many) jumps on compacts.

Finally, we consider the Bayesian type estimator. Applying Theorem 10 with the aid of Lemma 2, by using the weak convergence of $(\mathbb{Z}_n^1(u_1; \theta_1^*, \theta_2^\star), \mathbb{Z}_n^2(u_2; \tilde{\theta}_1, \theta_2^*))$ like (27) and (26) (cf. Theorem 11), and the polynomial type large deviation inequalities (22) and the inequality

$$P_{\theta^*}\left[\sup_{u_2 \in V_n^2(r, \theta_2^*)} \mathbb{Z}_n^2(u_2; \tilde{\theta}_1, \theta_2^*) \geq e^{-r}\right] \leq \frac{C_L}{r^L} \quad (r > 0,\ n \in \mathbb{N}), \tag{30}$$

which can be proved in the same way as (24), we obtain the asymptotic results as follows:

Theorem 14 *Suppose that Conditions* [D1]–[D4] *are fulfilled. Then for adaptive Bayes type estimator $(\tilde{\theta}_{1,n}, \tilde{\theta}_{2,n})$ for $\theta = (\theta_1, \theta_2)$, it holds that*

$$\left(\sqrt{n}(\tilde{\theta}_{1,n} - \theta_1^*), \sqrt{nh}(\tilde{\theta}_{2,n} - \theta_2^*)\right) \to^d (\zeta_1, \zeta_2)$$
$$\sim N_{m_1+m_2}\left(0, \text{diag}\,[\Gamma_1(\theta^*)^{-1}, \Gamma_2(\theta^*)^{-1}]\right)$$

as $n \to \infty$. Moreover, $P_{\theta^*}\left[f(\sqrt{n}(\tilde{\theta}_1 - \theta_1^*), \sqrt{nh}(\tilde{\theta}_2 - \theta_2^*))\right] \to \mathbb{P}[f(\zeta_1, \zeta_2)]$ as $n \to \infty$ for all continuous functions f of at most polynomial growth.

Remark 13 From our discussions, we have already had the same asymptotic normality of the simultaneous Bayes estimator.

Remark 14 The obtained asymptotic variance implies that our Bayesian type estimators are first-order efficient.

Remark 15 We confined our attention to estimation by sampled data from a diffusion process. However, the weak convergence of Bayes type estimators, to say nothing of the moment convergence, and the weak plus moment convergence of M-estimators are new at least for multi-dimensional diffusion processes even under continuous observations. Furthermore, it is clear that our method can apply to semimartingales.

The log Likelihood function for a semimartingale has a representation

$$\mathbb{H}_T(\theta) - \mathbb{H}_T(\theta_0) = \delta(\theta) \cdot M^c_T - \frac{1}{2}\delta(\theta)^{\otimes 2} \cdot \langle M^{\otimes 2}\rangle_T + (Y(\theta) - 1) * \tilde{\mu}_T + (\log Y(\theta) - Y(\theta) + 1) * \mu_T$$

under the true model. Obviously, a set of conditions (ergodicity, differentiability, integrability, and global identifiability) implies a polynomial-type large deviation inequality for the statistical random field $\mathbb{Z}_T(u;\theta_0)$, the weak convergence of $\mathbb{Z}_T(u;\theta_0)$ in $\hat{C}(\mathbb{R}^m)$, the asymptotic normality of any sequence of the maximum likelihood estimators $\hat{\theta}_T$ for θ, and the convergence of moments of $\hat{u}_T = \sqrt{T}(\hat{\theta}_T - \theta_0)$. Presenting a set of sufficient conditions for each result in this article applied to this semimartingale model is just an exercise of the semimartingale theory. Details of this and sampling problems are given elsewhere (Yoshida 2005b). A polynomial-type large deviation inequality for so called Z-estimation was discussed by Yoshida (2007).

7 Fundamental inequalities

This section recalls a few results for completeness of our discussion about tightness. Section 7.2 gives a version of GRR inequality. Details will be presented in Yoshida (2006).

7.1 *C*-space

Let $(\mathbb{T}, d)$ be a separable locally compact metric space and $(B, \|\cdot\|)$ a separable Banach space. Denote by $\hat{C}(\mathbb{T}; B)$ the space of continuous mappings $f : \mathbb{T} \to B$ such that $\lim_{t\to\infty} f(t) = 0$. That is, for any $\epsilon > 0$, there exists a compact set $K \subset \mathbb{T}$ such that $\sup_{t\in\mathbb{T}\setminus K} \|f(t)\| < \epsilon$. Then $\hat{C}(\mathbb{T}; B)$ is a separable Banach space with respect to the supremum norm $\|f\|_\infty = \sup_{t\in\mathbb{T}} \|f(t)\|$.

For a set $K \subset \mathbb{T}$ and $\delta > 0$, the modulus of continuity of $f : \mathbb{T} \to B$ on K is defined by $w_K(f, \delta) = \sup_{s,t\in K;\ d(s,t)<\delta} \|f(s) - f(t)\|$. We define a modulus of continuity near "∞" by $\hat{w}_{\mathbb{T}\setminus K}(f) = \sup_{t\in\mathbb{T}\setminus K} \|f(t)\|$. $\mathbb{T}_0$ is a dense subset of $\mathbb{T}$.

Theorem 15 *A sequence $\{X^n\}_{n\in\mathbb{N}}$ of B-valued continuous random fields on $\mathbb{T}$ is tight if and only if the following three conditions are satisfied:*

(i) *For each $t \in \mathbb{T}_0$, the family $\{X^n(t)\}_{n\in\mathbb{N}}$ is relatively compact.*
(ii) *For every compact set $K \subset \mathbb{T}$ and every $\epsilon > 0$,*

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} P^n\left[w_K(X^n;\delta) > \epsilon\right] = 0,$$

(iii) *For any $\epsilon > 0$ and $\eta > 0$, there exists a compact set $K \subset \mathbb{T}$ such that*

$$\limsup_{n\to\infty} P\left[x;\ \hat{w}_{\mathbb{T}\setminus K}(X^n) > \epsilon\right] \le \eta.$$

7.2 GRR inequality and Kolmogorov's tightness criterion

Let $(\mathbb{T}, d)$ be a metric space. Equip $\mathbb{T}$ with the Borel σ-field $\mathcal{B}$. Suppose that ν is a locally finite measure on $(\mathbb{T}, \mathcal{B})$, that is, for every $t \in \mathbb{T}$, there exists an open neighborhood U_t of t such that $\nu(U_t) < \infty$. Also suppose that $\nu(B(t, \epsilon)) > 0$ for any $t \in \mathbb{T}$ and $\epsilon > 0$, where $B(t, \epsilon) = \{s;\ d(t, s) \le \epsilon\}$. Let $(\mathcal{X}, \rho)$ be another metric space. Let $\Phi, \eta : \mathbb{R}_+ \to \mathbb{R}_+$ be continuous, strictly increasing with $\Phi(0) = 0$ and $\eta(0) = 0$, and $\lim_{t\to\infty} \Phi(t) = \infty$.

For a mapping $f : \mathbb{T} \to \mathcal{X}$, we denote

$$f^{\%}(t, s) = \begin{cases} \dfrac{\rho(f(t), f(s))}{\eta(d(t, s))} & \text{if } t \ne s \\ 0 & \text{if } t = s. \end{cases}$$

The following theorem was initiated by Garsia, Rodemich and Rumsey:[17]

Theorem 16 *Suppose that $f : \mathbb{T} \to \mathcal{X}$ is continuous and satisfies*

$$\int_{\mathbb{T}}\int_{\mathbb{T}} \Phi(f^{\%}(t, s))\ \nu(\mathrm{d}t)\nu(\mathrm{d}s) < \infty.$$

Then

$$\rho(f(t), f(s)) \le 8\,\lambda(f^{\%})\ \sup_{z\in\mathbb{T}} \int_0^{4d(t,s)} \Phi^{-1}\left(\frac{4}{\nu(B(z, \frac{\epsilon}{2}))^2}\right)\ \eta(\mathrm{d}\epsilon)$$

for all $t, s \in \mathbb{T}$,[18] *where*

$$\lambda(f^{\%}) = \inf\left\{\lambda > 0;\ \int_{\mathbb{T}}\int_{\mathbb{T}} \Phi\left(\frac{f^{\%}(t, s)}{\lambda}\right)\ \nu(\mathrm{d}t)\nu(\mathrm{d}s) \le 1\right\}.$$

Let $r > 0$ and $\beta > 0$. We apply Theorem 16 to the case where $\mathbb{T}$ is a bounded Borel measurable set in $\mathbb{R}^d$, ν is the Lebesgue measure restricted to $\mathbb{T}$, $\Phi(x) = x^r$, $\eta(x) = x^{\beta+\frac{2d}{r}}$. We assume that there exists a constant $a \in (0, \infty)$ such that

$$\inf_{t\in\mathbb{T}} \nu(B_\epsilon(t)) \ge a(\epsilon^d \wedge 1) \quad (\epsilon > 0). \tag{31}$$

[17] This is a simple generalization of the GRR inequality by Arnold and Imkeller (1996).

[18] The factor in front of $d(t, s)$ is 4. We do not assume the convexity of $\mathbb{T}$. For example, when $\mathbb{T}$ is a subset of $\mathbb{R}^2$, it is possible that the open disc centered at $(t + s)/2$ and of diameter $d(t, s)$ is not included in $\mathbb{T}$.

Since $\mathbb{T}$ is bounded, Condition (31) is equivalent to that there exist positive numbers a and ϵ_0 such that

$$\inf_{t\in\mathbb{T}} \nu(B_\epsilon(t)) \geq a\,\epsilon^d \quad \text{for } \epsilon \in (0, \epsilon_0). \tag{32}$$

It may be said that Condition (31) or (32) restricts the shape of the boundary of $\mathbb{T}$. Then there exists a constant $C = C(d, \beta, r, a, \text{diam}(\mathbb{T}))$ under (31) or $C = C(d, \beta, r, a, \epsilon_0, \text{diam}(\mathbb{T}))$ under (32) such that

$$\rho(f(t), f(s)) \leq C\, A(f)\, |t-s|^\beta \quad (t, s \in \mathbb{T}) \tag{33}$$

for all continuous functions $f : \mathbb{T} \to \mathcal{X}$, where

$$A(f) = \left\{ \int_{\mathbb{T}} \int_{\mathbb{T}} \left(\frac{\rho(f(t), f(s))}{|t-s|^{\beta+\frac{2d}{r}}} \right)^r \mathrm{d}t\mathrm{d}s \right\}^{\frac{1}{r}}.$$

We consider a random field $X : \Omega \times \mathbb{T} \to \mathcal{X}$ for a probability space $(\Omega, \mathcal{F}, P)$. Additionally, suppose that

$$P\left[\rho(X(t), X(s))^r\right] \leq B\, |t-s|^{d+\alpha} \quad (t, s \in \mathbb{T}) \tag{34}$$

for some $\alpha > r\beta$. Then from (33), we obtain

$$P\left[\left(\sup_{t,s\in\mathbb{T}:\, t\neq s} \frac{\rho(X(t), X(s))}{|t-s|^\beta} \right)^r\right] \leq C^r BM, \tag{35}$$

where

$$M = \int_{\mathbb{T}} \int_{\mathbb{T}} |t-s|^{\alpha-r\beta-d}\, \mathrm{d}t\mathrm{d}s < \infty$$

since $\alpha - r\beta > 0$.

In particular, from (35),

$$P\left[\sup_{t,s\in\mathbb{T}:\, t\neq s} \frac{\rho(X(t), X(s))}{|t-s|^\beta} \geq h \right] \leq \frac{C^r BM}{h^r} \tag{36}$$

for all $h > 0$.

References

Adams, R. A. (1975). Sobolev spaces. In *Pure and applied mathematics* (Vol. 65). London, New York: Academic Press.

Arnold, L., Imkeller, P. (1996). Stratonovich calculus with spatial parameters and anticipative problems in multiplicative ergodic theory. *Stochastic Processes and their Applications, 62*, 19–54.

Basawa, I. V., Prakasa Rao, B. L. S. (1980). *Statistical inference for stochastic processes*. London, New York: Academic Press.

Doukhan, P. (1995). Mixing: Properties and examples. In *Lecture notes in statistics* (Vol. 85). Berlin: Springer.

Doukhan, P., Oppenheim, G., Taqqu, M. S. (Eds.). (2003). *Theory and applications of long-range dependence*. Boston: Birkhauser.

Ibragimov, I. A., Has'minskii, R. Z. (1972). The asymptotic behavior of certain statistical estimates in the smooth case. I. Investigation of the likelihood ratio. (Russian) *Teorija Verojatnostei i ee Primenenija, 17*, 469–486.

Ibragimov, I. A., Has'minskii, R. Z. (1973). Asymptotic behavior of certain statistical estimates. II. Limit theorems for a posteriori density and for Bayesian estimates. (Russian) *Teorija Verojatnostei i ee Primenenija, 18*, 78–93.

Ibragimov, I. A., Has'minskii, R. Z. (1981). *Statistical estimation: Asymptotic theory*. New York: Springer.

Kessler, M. (1997). Estimation of diffusion processes from discrete observations. *Scandinavian Journal of Statistics, Theory and Applications, 24*, 211–229.

Kusuoka, S., Yoshida, N. (2000). Malliavin calculus, geometric mixing, and expansion of diffusion functionals. *Probability Theory and Related Fields, 116*, 457–484.

Kutoyants, Y. A. (1984). *Parameter estimation for stochastic processes* (B. L. S. Prakasa Rao, Ed., Trans.). Berlin: Herdermann.

Kutoyants, Y. (1994). *Identification of dynamical systems with small noise*. Dordrecht: Kluwer.

Kutoyants, Y. A. (1998). Statistical inference for spatial Poisson processes. In *Lecture notes in statistics* (Vol. 134). Berlin: Springer.

Kutoyants, Y. A. (2004). Statistical inference for ergodic diffusion processes. In *Springer series in statistics*. London: Springer.

Meyn, S. P., Tweedie, R. L. (1992). Stability of Markovian processes. I. Criteria for discrete-time chains. *Advances in Applied Probability, 24*, 542–574.

Meyn, S. P., Tweedie, R. L. (1993a). Stability of Markovian processes. II. Continuous-time processes and sampled chains. *Advances in Applied Probability, 25*, 487–517.

Meyn, S. P., Tweedie, R. L. (1993b). Stability of Markovian processes. III. Foster–Lyapunov criteria for continuous-time processes. *Advances in Applied Probability, 25*, 518–548.

Prakasa Rao, B. L. S. (1983). Asymptotic theory for non-linear least square estimator for diffusion processes, *Mathematische Operationsforschung und Statistik Series Statistics, 14*, 195–209.

Prakasa-Rao, B. L. S. (1988). Statistical inference from sampled data for stochastic processes, *Contemporary Mathematics, 80*, 249–284.

Rio, E. (1994). Inegalites de moments pour les suites stationnaires et fortement melangeantes. (French) [Moment inequalities for stationary strongly mixing sequences] *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique, 318*, 355–360.

Sakamoto, Y., Yoshida, N. (1999). Higher order asymptotic expansions for a functional of a mixing process and applications to diffusion functionals (unpublished).

Sakamoto, Y., Yoshida, N. (2004). Asymptotic expansion formulas for functionals of epsilon—Markov processes with a mixing property. *Annals of the Institute of Statistical Mathematics, 56*, 545–597.

Shimizu, Y., Yoshida, N. (2002). Estimation of diffusion processes with jumps from discrete observations. *Statistical Inference for Stochastic Processes* (to appear).

Uchida, M., Yoshida, N. (2001). Information criteria in model selection for mixing processes. *Statistical Inference for Stochastic Processes, 4*, 73–98.

Varakin, A. B., Veretennikov, A. Y. (2002). On parameter estimation for "polynomial ergodic" Markov chains with polynomial growth loss functions. *Markov Processes and Related Fields, 8*, 127–144.

Veretennikov, A. Y. (1987). Bounds for the mixing rate in the theory of stochastic equations. *Theory of Probability and its Applications, 32*, 273–281.

Veretennikov, A. Y. (1997). On polynomial mixing bounds for stochastic differential equations. *Stochastic Processes and their Applications, 70*, 115–127.

Yoshida, N. (1988). On asymptotic mixed normality of the maximum likelihood estimator in a multidimensional diffusion process. *Statistical theory and data analysis, II (Tokyo, 1986)* (pp. 559–566). Amsterdam: North-Holland.

Yoshida, N. (1990). Asymptotic behavior of M-estimator and related random field for diffusion process. *Annals of the Institute of Statistical Mathematics, 42*(2), 221–251.

Yoshida, N. (1992a). Asymptotic expansion for small diffusions via the theory of Malliavin–Watanabe. *Probability Theory and Related Fields, 92*, 275–311 (1992).

Yoshida, N. (1992b). Estimation for diffusion processes from discrete observation. *Journal of Multivariate Analysis, 41*, 220–242.
Yoshida, N. (1993). Asymptotic expansion of Bayes estimators for small diffusions. *Probability Theory and Related Fields, 95*, 429–450.
Yoshida, N. (1995). Malliavin calculus and asymptotic expansion for martingales. *Probability Theory and Related Fields, 109*, 301–342.
Yoshida, N. (2004). Partial mixing and conditional Edgeworth expansion. *Probability Theory and Related Fields, 129*, 559–624.
Yoshida, N. (2005a). General M-estimation for stochastic differential equation with jumps by sampled data (in preparation).
Yoshida, N. (2005b). Asymptotic expansion for general M-estimation for stochastic differential equation with jumps. *Zenkin Tenkai 2003*, Hiroshima International University (in preparation).
Yoshida, N. (2006). Polynomial type large deviation inequalities and convergence of statistical random fields. Research Memorandum, The Institute of Statistical Mathematics.
Yoshida, N. (2007). Statistical random field with graded parameters and mighty convergence. In Workshop on Stochastic Analysis and Statistics held at ISM, February 15, 2007.