Summary. We give an overview of recent developments in the theory of statistical inference for stochastic processes.

1 Frame of the first-order asymptotic decision theory

Consider a sequence of statistical experiments $\mathcal{E}^T = (\mathcal{X}^T, A^T, \{P^T_{\theta}\}_{\theta \in \Theta})$ ($T \in \mathbb{R}_+$). Let $\hat{\theta}_T : \mathcal{X}^T \to \Theta$ be a sequence of estimators of the unknown parameter $\theta$. A basic property $\hat{\theta}_T$ should have is the consistency:

$$\hat{\theta}_T \xrightarrow{P^T_{\theta}} \theta \quad (T \to \infty)$$

for every $\theta \in \Theta$. The analyst should not use any estimator without checking this property. For example, if one uses an estimator which is not consistent...
to test a statistical hypothesis, the more data he obtains, the more incorrect the decision becomes: the probability of the type I error tends to one.

As the second step of analysis, it is important to investigate the limit distribution of the estimator to compare efficiencies of consistent estimators. For regular experiments admitting a differentiable structure, the sequence of experiments \( \{E^T\} \) often satisfies the locally asymptotically normal (LAN) condition:

\[
dP_{\theta + r^T u}^T / dP_{\theta}^T = \exp \left( \Delta_T(\theta)[u] - \frac{1}{2} G_T(\theta)[u, u] + \epsilon_T(\theta) \right),
\]

\[
G_T(\theta) \to_{P_{\theta}}^T G(\theta),
\]

\[
\Delta_T(\theta) \to^d G(\theta)^{1/2} Z, \quad Z \sim N(0, 1),
\]

\[
\epsilon_T(\theta) \to_{P_{\theta}}^T 0
\]

where \( r_T \) is a p.d. matrix satisfying \( |r_T| \to 0 \) and \( G(\theta) \) is a deterministic p.d. matrix.

An asymptotically minimax risk bound (Hájek’s inequality) follows from the LAN property. Since the maximum likelihood estimator (MLE) and the Bayes estimator (BE) usually attain this lower bound, they are asymptotically efficient (in the first order). The moment estimator has an asymptotic covariance matrix which is larger than that of the MLE, therefore it is less efficient. From aspects of robust statistics, the moment estimator is unstable, and there is no reason to use it unless it is the only estimator one can apply.

It is possible to show the weak convergence of the random field \( u \mapsto dP_{\theta + r^T u}^T / dP_{\theta}^T \). Then the mapping theorem yields the asymptotic normality of the MLE

\[
r_T^{-1}(\hat{\theta}_T - \theta) \to^d N(0, G(\theta)^{-1}).
\]

A similar result holds for the BE. Ibragimov and Hasminskii’s scheme enables us to obtain \( L^p \)-boundedness of \( r_T^{-1}(\hat{\theta}_T - \theta) \) ([18]).

2 Inference for stochastic differential equations with full observations

Let \( X = (X_t)_{t \in [0, T]} \) be a solution of the following stochastic differential equation:

\[
dX_t = V_0(X_t, \theta)dt + V(X_t, \theta)dw_t,
\]
$X_0 = x_0.$

Here $w$ is a multi-dimensional Wiener process, and $\theta \in \Theta$ denotes an unknown parameter. For continuous observations $(X_t)_{t \in [0,T]}$, mainly the following two cases are studied:

**Ergodic diffusion** $V(x, \theta) = V(x)$, $X$ is ergodic $^1$, $T \to \infty$.

**Small diffusion** $V(x, \theta) = \epsilon V(x)$, $T$ is fixed, $\epsilon \downarrow 0.^2$

The first-order asymptotic theory (consistency, asymptotic normality, LAN and LAMN, asymptotic minimax bounds, convolution theorem, asymptotic efficiency) has been developed. We refer the reader to the following textbooks and references therein: Kutoyants [28, 26, 25], Prakasa Rao [46, 47], Basawa and Prakasa Rao [9], Küchler and Sørensen [21], Basawa and Scott [10].

### 3 Asymptotic expansion

Practically speaking, the normal approximation based on the central limit theorems does not work so well when the amount of Fisher information is not sufficiently large. Asymptotic expansion method provides better approximations even in such a situation, and it is one of the fundamentals of the higher-order asymptotic theory. See: Akahira-Takeuchi [1], Pfanzagl [43], Ghosh [16], Barndorff-Nielsen and Cox [5], Pace and Salvan [42], Akahira and Takeuchi [2], Bhattacharya and Ghosh [11], Siotani et al. [57], Amari [3], Amari et al. [4], Murray and Rice [38], Hall [17], Taniguchi [64], Taniguchi and Kakizawa [65].

#### 3.1 System with small noises

Watanabe theory was applied to derive expansion formulas for statistical functionals of a process with small diffusion. In this direction are the studies: [73, 74] (MLE), [77] (BE), Dermoune and Kutoyants [14] (misspecified

---

$^1$Or “non-ergodic”. In non-ergodic case, the LAN property should be replaced by the LAMN property.

$^2$Random limit models like the partially linear Gaussian model are put in this category, cf. [82].
model), [50] (mixture model), [78] (system with small perturbations), [71] (model selection), [82] (SDE driven by Lévy processes, conditional expansion), [34] (numerical scheme for double Edgeworth expansions), [60] (random limit expansion), Sakamoto [49] (discriminant analysis).

As a byproduct of statistical researches, the asymptotic expansion scheme was proposed to compute the values of options: [76], [61], [22, 23], [60], [62], [19], [63], [72].

### 3.2 Martingale expansion: global approach

In [39, 40, 41], Mykland provided an expansion of the moment $E[g(M_{n,T})]$ for a sequence of martingales $(M_{n,t} : t \in [0,T_n])_{n \in \mathbb{N}}$ and for a $C^2$ function $g$. He proved the validity by Heath’s Skorokhod embedding and Itô’s formula thanks to the differentiability of $g$.

If $g$ is not differentiable, the error bound is not generally correct. However, the expansion formula is valid under nondegeneracy of the Malliavin covariance ([79, 80]). The asymptotic expansion of the maximum likelihood estimator for an ergodic diffusion process followed and it improved the Berry-Esseen bounds by Mishra and Prakasa Rao [36] and Bose [13]. Asymptotic expansion formulas for M-estimators of an ergodic diffusion process were given in [52].

It is more efficient to apply the local approach below if the process admits the mixing property. However the martingale expansion still works even when the local approach does not. It is the case where the first-order term converges in distribution to a normal distribution but the second-order term asymptotically has a non-Gaussian distribution. Volatility estimation is an example ([79]), and estimation of a regression model with a long-memory explanatory variable is another example ([80]). Because one cannot apply a central limit theorem to compute the second-order distribution, a martingale problem approach is needed in the first case, and a non-central limit theorem in the second case.

### 3.3 Expansion for mixing processes: local approach

The $\epsilon$-Markov process is an abstract model which includes Markovian stochastic differential equations (with jumps), certain delay systems, point processes such as cluster processes, and quasi-Markovian non-linear time series models with discrete-time parameter. For a functional of an $\epsilon$-Markov process,
Kusuoka and Yoshida [24] proved the validity of the formal Edgeworth expansion. The local Cramér condition was replaced by the local non-degeneracy of the Malliavin covariance. For example, under the non-degeneracy of the Lie algebra at a point in the support of the stationary distribution, the formal expansion is valid. Paper [81] showed the role of the support theorems for the SDE with jumps in verifying the local non-degeneracy.

Asymptotic expansion for a functional admitting a stochastic expansion are in [51, 53, 54]. Expansion of the maximum likelihood estimator for a general diffusion process follows, as an example.

Recent related results are: model selection problem through asymptotic expansion methods ([69, 70]), expansion of test statistics by Sakamoto [48], [29] for nonparametric estimation, [55] for cluster processes.

### 3.3.1 Stochastic volatility model

An expansion for a functional of the Ornstein-Uhlenbeck process having a Lévy driving term was presented in [35]. Let us consider a system of SDE’s:

\[
\begin{align*}
    dX_t &= -\lambda X_t dt + dZ_t, \\
    dY_t &= (\gamma + \beta X_t) dt + \theta \sqrt{X_t} dw_t + \rho dZ_t, \\
    Y_0 &= 0,
\end{align*}
\]

where \( Z \) denotes a Lévy process, and it is a subordinator when \( \theta \neq 0 \). \( w \) is a Wiener process independent of \( Z \). \( X \) is called an Ornstein-Uhlenbeck process, and it has an invariant measure (under mild conditions), which is necessarily selfdecomposable. It is possible to derive Edgeworth expansion of the expectation \( P \left[ f \left( T^{-1/2} (Y_T - P[Y_T]) \right) \right] \) as \( T \to \infty \). This expansion explains the aggregation Gaussianity of \( Y_T \) for large \( T \), and the divergence from the normality for small \( T \).

Masuda [33] proved the mixing property for the Ornstein-Uhlenbeck process driven by a Lévy process. Related topics are found in Masuda [32]. For this stochastic volatility model, see Barndorff-Nielsen et al. [6, 7, 8].

### 3.3.2 Regression model with a long memory explanatory process

Consider a stochastic regression model

\[
Y_t = X_t \theta + e_t,
\]

where \( X = (X_t) \) is a long-memory explanatory process and \( e = (e_t) \) is an \( m \)-Markov process independent of \( X \). \( \theta \) is an unknown (vector) parameter
and we are interested in estimation of $\theta$ from observations $(X_t, Y_t)_{t=1}^T$. For the simple least square estimator $\hat{\theta}_T$,

$$
\sqrt{T}(\hat{\theta}_T - \theta) = \left( \frac{1}{T} \sum_{t=1}^{T} X'_t X_t \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} X'_t \epsilon_t.
$$

Because of the strong dependency of $X$, the full process $(X, e)$ does not satisfy a usual sufficient condition for asymptotic expansion even though a fast convergence rate is assumed for the $\alpha$-mixing coefficient of $e$. However, it is possible to derive a valid Edgeworth expansion of $P[f(\sqrt{T}(\hat{\theta}_T - \theta))]$ if we apply partial mixing by [81]. The expansion is not standard because the second term has a rate related with the Hurst index (not a usual one of $T^{-1/2}$). For details, see [37] together with a continuous-time analogue.

4 Inference under sampling schemes: first-order and higher-order

We shall view inference under sampling schemes. Let us consider the SDE in Section 2. The diffusion coefficients $V$ may include unknown parameters, and our data is $(X_{t_i})_{i=0}^n$, where $t_i = ih$, so $T = nh$.

4.1 $(h \to 0) + (T \to \infty)$

The first-order asymptotic properties were discussed in Prakasa-Rao [44, 45], Yoshida [75], Kessler [20]. Recently, Shimizu and Yoshida [56] presented an estimating function for a stochastic differential equation with jumps (Section 4.6). The first-order properties have been proved in it. The asymptotic expansions follow from a version of [81] and [53, 54].

4.2 (fixed $h) + (T \to \infty)$

This case was treated in Bibby and Sørensen [12]. Masuda [30, 31] proved the asymptotic normality of the moment estimator for a state space model involving jump noise terms. In each case, the asymptotic expansion will follow as in Section 4.1.
4.3 \((h \to 0) + (\text{fixed } T)\)

The first-order asymptotics for the diffusion coefficient was obtained by Genon-Catalot and Jacod [15]. For the linearly parameterized case, the expansion of the estimator is in [79].

4.4 \((\text{fixed } h) + (\text{fixed } T) + (\text{small diffusion})\)

The first-order inference was discussed in Sørensen [58]. Asymptotic expansion can be deduced, e.g., from [52].

4.5 \((h \to 0) + (\text{fixed } T) + (\text{small diffusion})\)

Recently, the first-order asymptotic results were obtained by Sørensen and Uchida [59], Uchida [66, 67]. Local asymptotic normality was proved by Uchida [68]. The martingale expansion method can apply to derive asymptotic expansions.

Small dispersion parameter (Condition \(\epsilon \downarrow 0\)) can be removed to keep identifiability in the limit by making a Berry-Esseen type bound which one obtains by a similar technique to the martingale method.

4.6 Diffusion with jumps

Let \(E_\gamma\) be an open set in \(\mathbb{R}^b\) for \(\gamma = 1, \ldots, m\), and let \((E_0, E_0)\) be another measurable space. For each \(\gamma = 0, 1, \ldots, m\), \(\mu^\gamma\) denotes a Poisson random measure on \(\mathbb{R}_+ \times E_\gamma\) with compensator \(dt \times X_\theta^\gamma (dv^\gamma)\). \(w^\beta\) denote independent Wiener processes. We consider a \(d\)-dimensional stochastic process \(X = (X_t)\) which satisfies the stochastic integral equation

\[
X_t = X_0 + \int_0^t A(X_t, \theta) dt + \sum_{\beta=1}^{r} \int_0^t B_\beta(X_s, \theta) dw_s^\beta \\
\quad + \sum_{\gamma=0}^{m} \int_0^t \int_{E_\gamma} C_{\gamma}(X_{s-}, v^\gamma, \theta) \tilde{\mu}^\gamma(ds, dz). \tag{1}
\]

\(\theta\) is a vector valued unknown parameter (thus this includes the case where parameters in coefficients are distinct). Processes with infinitely many jumps on compacts are within our scope.

We assume that the data consist of the discrete-time observations \((X_{t_i})_{i=0}^n\) from the complete realization \((X_t)_{t \in \mathbb{R}_+}\). Here \(t_i = ih\) \((i = 0, 1, \ldots, n)\) with
$h = h_n$. We assume the ergodicity of $X$, and then it is possible to construct an M-estimator which is consistent, asymptotic normal under the sampling scheme $h \to 0$ and $nh \to \infty$. Moreover, the asymptotic expansion has been obtained for this estimator ([83]).

The fixed-$h$ case is nothing but a nonlinear time series analysis.

References


Sydney Auckland/New York: Arnold/Oxford University Press 1999

D.C.: Chapman & Hall 1999


107, 53-60, The Institute of Statistical Mathematics, Tokyo (1998)

to diffusion functionals. unpublished (1999)

to diffusion functionals. unpublished (1999)


property: to appear in Annals of the Institute of Statistical Mathematics


Columbus, Ohio: American Sciences Press 1985


Bernoulli, 9, 1051-1069 (2003)


6, 115-151 (1999)


New York: Springer 1991

Springer 2000

[66] Uchida, M.: Estimation for dynamical systems with small noise from discrete observations. Preprint No. 2002-4, Depart-
ment of Statistics and Operations Research, University of Copenhagen. submitted (2002)


Stochastic Processes, 4, 73-98 (2001)


kusatsu.tex