Chapter 2
Asymptotic Expansions for Stochastic Processes

Nakahiro Yoshida

2.1 Introduction

The central limit theorems are the basis of the large sample statistics. In estimation theory, the asymptotic efficiency is evaluated by the asymptotic variance of estimators, and in testing statistical hypotheses, the critical region of a test is determined by the normal approximation.

Though asymptotic properties of statistics are based on central limit theorems, the accuracy of their approximation is not necessarily sufficient in practice, especially in the case not many observations are available. Even then, we experienced possibility of getting more precise approximation by the asymptotic expansion methods.

The asymptotic expansion has theoretical importance. This method is today recognized as basis of various branches of theoretical statistics like higher order inferential theory, prediction, model selection, resampling methods, information geometry, and so on. For example, the Akaike Information Criterion (AIC) for statistical model selection is a statistic that incorporates higher-order behavior of the maximum log likelihood.

In the recent four decades, intensive studies have been done for statistics of semimartingales. See, e.g., Kutoyants [54, 55, 56], Basawa and Prakasa Rao [8], Küchler and Soerensen [51], and Prakasa Rao [80, 79]. Since large sample theoretical approaches are inevitable to semimartingales, the development was in exact timing interactively with that of limit theorems.

The counterpart of traditional independent observations is the class of stochastic processes with ergodic property. Laws of large numbers were often deduced from mixing properties or from ergodic theorems through Markovian structures of processes, and various central limit theorems have been produced in the mixing framework and in the martingale framework. Thus, after developments of the first order statistics, it was natural that
a part of studies of limit theorems for stochastic processes was directed to higher-order asymptotics. This trend entailed generalization of techniques applicable to dependency.

The emphasize of this survey is put on central limit theorems and asymptotic expansion applied to statistics for semimartingales. The results can essentially apply to Markov chains, therefore so-called nonlinear time series models. On the other hand, it should be remarked that quite a few techniques invented in classical higher-order limit theorems, such as smoothing inequalities, work as fundamentals of the theory of asymptotic expansion for semimartingales.

Since non-normality of the limit distribution of statistical estimators, even in regular experiments, emerged rather early [95, 4], the non-ergodic statistics was commonly recognized and established in the 70s. There appear limit theorems that have a mixture of normal distributions as the limit distribution. Intuitively, the Fisher information or the energy of the martingale of the score function does not converge to a constant like classical statistics, but does to a random variable. Then the error becomes asymptotically conditionally normal given the random Fisher information. The non-ergodic statistics required developments in limit theorems and raises a problem about asymptotic expansion. These topics will be discussed in Section 2.5.

2.2 Refinements of Central Limit Theorems

Let \((\xi_j)_{j \in \mathbb{N}}\) be a sequence of \(d\)-dimensional independent and identically distributed (i.i.d.) random vectors with \(E[\xi_1] = 0\) and \(\text{Cov}[(\xi_1)] = I_d\), the identity matrix.

2.2.1 Rate of Convergence of the Central Limit Theorem

The central limit theorem states \(S_n = n^{-1/2} \sum_{j=1}^{n} \xi_j \rightarrow^d N_d(0, I_d)\), namely, for any bounded continuous function \(g\) on \(\mathbb{R}^d\), \(\int_{\mathbb{R}^d} gd(Q_n - \Phi) \rightarrow 0\) as \(n \rightarrow \infty\), where \(Q_n\) is the distribution of \(S_n\) and \(\Phi = N_d(0, I_d)\).

Let \(\beta_{s,i} = E[|\xi_1^{(i)}|^s]\) and \(\beta_s = \sum_{i=1}^{d} \beta_{s,i}, \xi_1^{(i)}\) being the \(i\)-th element of \(\xi_1\). For a function \(g\) on \(\mathbb{R}^d\), let \(\omega_g(A) = \sup\{|g(x) - g(y)|; x, y \in A\}\) and let \(\omega_g(x; \epsilon) = \omega_g(B(x, \epsilon))\) for \(B(x, \epsilon) = \{y; |x - y| < \epsilon\}\). The existence of third order moment gives a refinement of the central limit theorem. For example, under the assumption \(\beta_3 < \infty\), it holds that for every real valued bounded measurable function \(g\) on \(\mathbb{R}^d\),

\[
\left| \int_{\mathbb{R}^d} gd(Q_n - \Phi) \right| \leq c_0 \omega_g(\mathbb{R}^d) \beta_3 n^{-1/2} + \int_{\mathbb{R}^d} \omega_g(c; c_2 \beta_3 n^{-1/2} \log n) d\Phi
\]

(2.1)

if \(\beta_3 < c_1 n^{1/2}(\log n)^{-d}\), where \(c_0, c_1,\) and \(c_2\) are constants depending on \(d\) (Theorem 4.2 of Bhattacharya [15]). See also Bhattacharya [13, 14] for the origin of this result. Bhattacharya and Ranga Rao [20] give a comprehensive exposition and generalizations.
2.2.2 Cramér-Edgeworth Expansion

The \( \nu \)-th cumulant of \( \xi_1 \) is denoted by \( \chi_\nu \) for a multi-index \( \nu \in \mathbb{Z}_+^d \), \( \mathbb{Z}_+ = \{0, 1, \ldots\} \). That is, for the characteristic function \( \varphi_{\xi_1} \) of \( \xi_1 \),

\[
\log \varphi_{\xi_1}(u) = \sum_{\nu: 2 \leq |\nu| \leq s} \frac{X_\nu}{\nu!} (iu)^\nu + o(|u|^s) \quad (u \to 0)
\]

where \( |\nu| = \nu_1 + \cdots + \nu_d \) and \( u^\nu = (u_1)^{\nu_1} \cdots (u_d)^{\nu_d} \) for \( \nu = (\nu_1, \ldots, \nu_d) \in \mathbb{Z}_+^d \) and \( u = (u^1, \ldots, u^d) \in \mathbb{R}^d \).

Let \( S_n = n^{-1/2} \sum_{j=1}^n \xi_j \). Then independency yields

\[
\varphi_{S_n}(u) = e^{-|u|^2/2} \exp \left[ \sum_{\nu: 3 \leq |\nu| \leq s} \frac{X_\nu}{\nu!} (iu)^\nu n^{-|\nu|-2}/2 \right] \times [1 + o(n^{-(s-2)/2})]
\]

as \( n \to \infty \) for every \( u \in \mathbb{R}^d \). The last expression is rewritten as

\[
\varphi_{S_n}(u) = e^{-|u|^2/2} \left[ 1 + \sum_{r=1}^{s-2} n^{-r/2} \tilde{P}_r (iu) \right] + o(n^{-(s-2)/2}). \tag{2.2}
\]

Here each \( \tilde{P}_r \) is a certain polynomial whose coefficients are written in \( \chi_\nu \)'s. The first term on the right-hand side of (2.2) will be denoted by \( \tilde{P}_n \).

The \((s-2)\)-th order Edgeworth expansion of the distribution of \( S_n \) is given by the Fourier inversion \( p_n = \mathcal{F}^{-1} [\tilde{P}_n] \) of \( \tilde{P}_n \). Asymptotic expansion gives higher-order approximation of the distribution of \( S_n \). This method goes back to Tchebycheff, Edgeworth, and Cramér.

Regularity of the distribution is often supposed to obtain an asymptotic expansion of the distribution. Otherwise, this approximation is not necessarily valid. In fact, for the Bernoulli trials \( \xi_j \) \((j \in \mathbb{N})\), i.e., these random variables are independent and \( P[\xi_j = -1] = P[\xi_j = 1] = 1/2 \). We denote by \( F_n \) the distribution function of \( n^{-1/2} \sum_{j=1}^n \xi_j \). Then for even \( n \in \mathbb{N} \),

\[
F_n(0) - F_n(0-) = \mathbb{P} \left[ \sum_{j=1}^n \xi_j = 0 \right] = \binom{n}{n/2} \left( \frac{1}{2} \right)^n \sim \sqrt{2/\pi} n^{-1/2}
\]

and hence for any sequence of continuous functions \( \Phi_n \),

\[
\lim inf_{n \to \infty} (2n)^{1/2} \sup_{x \in \mathbb{R}} \left| F_{2n}(x) - \Phi_n(x) \right| > 0.
\]

Therefore the ordinary Edgeworth expansion always fails to give a first-order asymptotic expansion to \( F_n \).

The Cramér condition

\[
\lim sup_{|u| \to \infty} |\varphi_{\xi_1}(u)| < 1 \tag{2.3}
\]
is effective to deduce the decay of the characteristic function of $S_n$. If the distribution $\mathcal{L}[\xi_1]$ has a nonzero absolutely continuous part of the Lebesgue decomposition, then Condition (2.3) holds.

Under (2.3), combining the estimate (2.6) with (2.5) below, it is possible to evaluate the error of the asymptotic expansion. Let $s$ be an integer greater than 2. Let $M_r(f) = \sup_{x \in \mathbb{R}^d} (1 + |x|^r)^{-1} |f(x)|$ for measurable function $f$ on $\mathbb{R}^d$. Let $s' \leq s$. Then, under (2.3),

$$\left| \int_{\mathbb{R}^d} f dQ_n - \int_{\mathbb{R}^d} f_p dx \right| \leq M_{s'}(f) \epsilon_n + c(s,d) \int_{\mathbb{R}^d} |\omega_f(x;2e^{-cn})\Phi(dx)|$$

(2.4)

where $c$ is a positive constant, $c(s,d)$ is a constant depending on $(s,d)$, and $\epsilon_n = o(n^{-(s-2)/2})$ as $n \to \infty$. This result is Theorem 20.1 of Bhattacharya and Ranga Rao [20]. We refer the reader to Cramér [26], Bhattacharya [14], Petrov [75], and other papers mentioned therein for results in the early days.

### 2.2.3 Smoothing Inequality

The so-called smoothing inequality plays an essential role in validation of the above refinements (2.1) and (2.4) of the central limit theorem. Let $p$ be an integer with $p \geq 3$. Consider a probability measure $\mathcal{K}$ on $\mathbb{R}^d$ and a constant $\alpha$ such that $\alpha := K_e(B(0,\alpha)) > 1/2$. The scaled measure $\mathcal{K}_e$ is defined by $\mathcal{K}_e(A) = \mathcal{K}(e^{-1}A)$ for Borel sets $A$. Given a finite measure $P$ and a finite signed measure $Q$ on $\mathbb{R}^d$, let $\gamma_f(e) = \|f^*\|_{\infty} \int_{\mathbb{R}^d} h(|x|)/\mathcal{K}_e \ast (P - Q)(dx)$, $\xi_f(e) = \|f^*\|_{\infty} \int_{\{x:|x|\geq 2ae\}} h(|x|)\mathcal{K}(dx)$, and $\tau(t) = \sup_{x:|x| \leq 5ae} \int_{\mathbb{R}^d} |\omega_f(x + y, 2ae)Q^+(dy)|$, where $f^*(x) = f(x)/h(|x|)$, $h(r) = 1 + r^{p_0}$ ($p_0 = 2(p/2)$), and $Q^+$ is the positive part of $Q$. Among many versions, Sweeting’s smoothing inequality [88] is given by

$$\left| (P - Q)[f] \right| \leq \frac{1}{2 \alpha - 1} \left[ A_0 \gamma_f(e) + A_1 \xi_f(e'/e) + \tau(t) \right] + \left( \frac{1 - \alpha}{\alpha} \right) A_2 \|f^*\|_{\infty}$$

(2.5)

for $\epsilon, \epsilon'$ satisfying $0 < \epsilon < \epsilon' < a^{-1}$ and $t \in \mathbb{N}$ ($ae't \leq 1$), where $A_0, A_1,$ and $A_2$ are some constants depending on $p, d, (P + |Q|)[h(|\cdot|)]$. See Bhattacharya [13, 14, 15] and Bhattacharya and Rao [20] for more information of smoothing inequalities.

There exists a constant $C_d$ such that

$$\int_{\mathbb{R}^d} |f(x)| dx \leq C_d \max_{m \in \mathbb{N}^d_{|t|=0,d+1}} \int_{\mathbb{R}^d} |\partial^m F[f](u)| du$$

(2.6)

for all measurable functions $f : \mathbb{R}^d \to \mathbb{R}$ satisfying $\int_{\mathbb{R}^d} (1 + |x|^{d+1})|f(x)| dx < \infty$; see [19, 20]. Thus, the comparison between two measures comes down to the integrability of their Fourier transforms and estimation of the gap between them.
2.2.4 Applications to Statistics

Asymptotic expansion has been a basis of modern theoretical statistics. Bhattacharya and Ghosh [17] established validity of the Edgeworth expansion of functionals of independent random variables, and it was applied to various statistical problems by many authors; see Google Scholar for citing papers. The reader finds related works in Bhattacharya and Denker [11]. Bootstrap methods obtain their basis on the asymptotic expansion (Hall [36]). Information geometry introduced $\alpha$-connection and gave an interpretation of the higher-order efficiency of the maximum likelihood estimators by the curvature of the fiber associated with the estimator (Amari [1]). Asymptotic expansion was also applied to construction of information criteria for model selection as well as prediction problems; e.g., Konishi and Kitagawa [50], Uchida and Yoshida [91], Komaki [49].

2.3 Asymptotic Expansion for Mixing Processes

As a generalization from independency, central limit theorems and asymptotic expansion were developed under mixing properties; Ibragimov [39] among many others for a central limit theorem. Error bounds were given in Tikhomirov [89], Stein [86], and others. Nagaev [71, 72] presented rates of convergence and asymptotic expansions for Markov chains. Doukhan [29] gives exposition of mixing properties and related central limit theorems.

The class of diffusion processes is of importance as the intersection of the Markovian processes and the processes for which the ergodicity can be successfully treated. Bhattacharya [16], Bhattacharya and Ramasubramanian [18], and Bhattacharya and Wasiełak [12] provided ergodicity of multidimensional diffusion processes and related limit theorems. Also see the textbook by Meyn and Tweedie [67] for a general exposition of ergodicity, and a series of papers of Meyn and Tweedie [64, 65, 66]. Kusuoka and Yoshida discussed mixing property of possibly degenerate diffusion processes in [53]. Masuda [61] gave mixing bounds for jump diffusion processes.

Under assumption of mixing property, Götze and Hipp [34] gave asymptotic expansions for sums of weakly dependent processes that are approximated by a Markov chain. The smoothing inequality discussed in Section 2.2 was applied together with inventive estimates of the characteristic function. A Cramér type estimate was assumed for a conditional characteristic function of local increments of the process. Götze and Hipp [35] carried out their scheme for more concrete time series.

The Markovian property in practice plays an essential role in estimation of the characteristic function of an additive functional of the underlying process. Mixing property is deeply related to the ergodicity especially in Markovian contexts. Therefore it is practically natural to approach Edgeworth expansion through mixing.

Given a probability space $(\Omega, \mathcal{F}, P)$, let $Y = (Y_t)_{t \in \mathbb{R}^\times}$ be a $d_2$-dimensional càdlàg process and let $X = (X_t)_{t \in \mathbb{R}^+}$ be a $d_1$-dimensional càdlàg process with independent increments in the sense that $\mathcal{B}^{XY}_{[0,r]}$ is independent of $\mathcal{B}^{IX}_{[r,\infty)}$ for $r \in \mathbb{R}^+$, where $\mathcal{B}^{XY}_{[0,r]} = \sigma[X_t, Y_t; t \in [0, r]] \vee \mathcal{N}$ and $\mathcal{B}^{IX}_I = \sigma[X_t - X_s; s, t \in I] \vee \mathcal{N}$, $I \subset \mathbb{R}^+$, with $\mathcal{N}$ being the null-$\sigma$-field. Suppose that $Y$ is an $\epsilon$-Markov process driven by $X$. That is, there exists a nonnegative constant $\epsilon$ such
that \( Y_t \) is \( \mathcal{B}^Y_{[s-t,a]} \) \( \vee \mathcal{B}^{dX}_{[s-t]} \) measurable for all \( t \geq s \geq \epsilon \), where \( \mathcal{B}^Y_I = \sigma[Y_t; t \in I] \vee \mathcal{N} \). Let \( \mathcal{B} I = \sigma[X_t - X_s, Y_s, t \in I] \vee \mathcal{N} \) for \( I \in \mathbb{R}_+ \).

An \( \alpha \)-mixing condition for \( Y \) is expressed by the inequality

\[
E[|Eg^Y_{[s-t]}(f) - E[f]|] \leq \tilde{a}_Y(s,t)||f||_\infty
\]

for \( s \leq t \) and bounded \( \mathcal{B}^Y_{[s,t]} \)-measurable functions \( f \). Let \( \alpha(s,t) = \tilde{a}_Y(s,t-\epsilon) \) if \( s \leq t - \epsilon \) and \( 1 \) if \( s > t - \epsilon \). Let \( \alpha(h) = \sup_{h' \geq h, a \in \mathbb{R}_+} \alpha(s, s + h') \), we shall assume exponential rate, namely, there exists a constant \( a > 0 \) such that \( \alpha(h) \leq a^{-1} e^{-ah} \) for all \( h > 0 \). This condition can be relaxed but the exponential rate is assumed for simplicity.

We consider a \( d \)-dimensional process \( Z = (Z_t)_{t \in \mathbb{R}_+} \) satisfying that \( Z_0 = 0 \) is \( \mathcal{B}^{[0]} \)-measurable and that \( Z_t - Z_s \) is \( \mathcal{B}^{[t,s]} \)-measurable for every \( t \geq s \geq \epsilon \). Given an integer \( p \geq 3 \), we assume that there exists \( h_0 > 0 \) such that

\[
E[|Z_0|^{p+1}] + \sup_{t,h \in \mathbb{R}_+, 0 \leq h \leq h_0} E[|Z+h-Z|^p] < \infty,
\]

and that \( E[Z_t] = 0 \) for all \( t \in \mathbb{R}_+ \).

Suppose that there exists a sequence of intervals \( I(j) = [u(j), v(j)] \) \( (j = 1, \ldots, n(T)) \) such that \( \lim_{T \to \infty} n(T)/T > 0 \) and \( 0 < \delta \leq v(j) - u(j) \leq \tilde{\delta} < \infty \) for some \( \delta \) and \( \tilde{\delta} \), and that for each \( j \), some \( \sigma \)-field \( \mathcal{B}^{[v(j)-\delta, v(j)]}_I \) of \( \mathcal{B}^{[v(j)-\delta, v(j)]} \) satisfies \( E_{\mathcal{B}^{[v(j)-\epsilon, v(j)]}_I}[h] = E_{\mathcal{B}^{[v(j)-\epsilon, v(j)]}}[h] \) for all bounded \( \mathcal{B}^{[v(j), \infty]} \)-measurable functions \( h \). Let \( \hat{C}(j) = \mathcal{B}_{[u(j)-\epsilon, u(j)]} \vee \mathcal{B}^{[v(j)-\epsilon, v(j)]} \). Denote by \( Z_j \) the increment of \( Z \) over the interval \( J \). Moreover, suppose that

\[
\lim_{B \to \infty} \lim_{T \to \infty} n(T)^{-1} \sum_j E[\sup_{u \in B} |E_{\hat{C}(j)}[e^{iuZ_j}]|] = 0 \quad (2.7)
\]

and \( \lim_{T \to \infty} n(T)^{-1} \sum_j E[|\psi_j|] > 0 \) for some \( [0, 1] \)-valued measurable functionals \( \psi_j \).

These conditions work as a kind of Cramér’s condition. Thus, in this situation, we obtain an Edgeworth expansion of \( T^{-1/2}Z_T \) as follows. The cumulant functions \( \chi_{T,r}(u) \) of \( T^{-1/2}Z_T \) are defined by \( \chi_{T,r}(u) = (\partial^r \chi)'_{\epsilon=0} \log E[\exp(iu \cdot T^{-1/2}Z_T)] \) for \( u \in \mathbb{R}^d \). Next define \( \hat{P}_{T,r}(u) \) by the formal expansion

\[
\exp\left( \sum_{r=2}^{\infty} \frac{(r!)^{-1} e^{-2}\chi_{T,r}(u)}{r!} \right) = \exp(2^{-1}\chi_{T,2}(u)) + \sum_{r=1}^{\infty} e^{T^{-r/2}} \hat{P}_{T,r}(u).
\]

Let \( \Psi_{T,b} = \mathcal{F}^{-1}[\hat{\Psi}_{T,b}] \) for \( \hat{\Psi}_{T,b}(u) = \exp(2\chi_{T,2}(u)) + \sum_{r=1}^{\infty} e^{T^{-r/2}} \hat{P}_{T,r}(u) \). Then if the covariance matrix \( \text{Cov}[T^{-1/2}Z_T] \) converges to a regular matrix as \( T \to \infty \), then it is possible to show that a similar estimate to (2.4), and the error \( |E[f(T^{-1/2}Z_T)] - \Psi_{T,b}[f]| \) becomes \( o(T^{-(p-2)/2}) \) ordinarily in applications. See Kusuoka and Yoshida [53] and Yoshida [99].

In order to validate the asymptotic expansion, it suffices to find good truncation functionals \( \psi_j \) and \( \sigma \)-fields \( \mathcal{B}^{[v(j)-\epsilon, v(j)]}_I \) as well as intervals \( I(j) \) for which (2.7) is satisfied. For example, we shall consider a system of stochastic integral equations
\[ Y_t = Y_0 + \int_0^t A(Y_{s-})ds + \int_0^t B(Y_{s-})dw_s + \int_0^t \int C(Y_{s-}, x)\tilde{\mu}(ds, dx) \]

\[ Z_t = Z_0 + \int_0^t A'(Y_{s-})ds + \int_0^t B'(Y_{s-})dw_s + \int_0^t \int C'(Y_{s-}, x)\tilde{\mu}(ds, dx) \]

where \( Z_0 \) is \( \sigma[Y_0] \)-measurable, \( A \in C^\infty(\mathbb{R}^d; \mathbb{R}^d) \), \( B \in C^\infty(\mathbb{R}^d; \mathbb{R}^d \otimes \mathbb{R}^m) \), \( C \in C^\infty(\mathbb{R}^d \times E; \mathbb{R}^d) \), and similarly \( A' \in C^\infty(\mathbb{R}^d; \mathbb{R}^d) \), \( B' \in C^\infty(\mathbb{R}^d; \mathbb{R}^d \otimes \mathbb{R}^m) \), \( C' \in C^\infty(\mathbb{R}^d \times E; \mathbb{R}^d) \), where \( w \) is an \( m \)-dimensional Wiener process, \( E \) is an open set in \( \mathbb{R}^b \), and \( \tilde{\mu} \) is a compensated Poisson random measure on \( \mathbb{R}_+ \times E \) with intensity \( dt \times dx \). Under standard regularity conditions, \((Y_t, Z_t)\) can be regarded as smooth functionals over the canonical space. In this case, the process \( X_t \) can be chosen as \( X_t = (w_t, \mu_t(g_t); t \in \mathbb{N}) \) for a countable measure determining family over \( E \), and \( Y \) is a 0-Markov process (i.e., a Markovian process). Though there are several versions of the Malliavin calculus for jump processes, we consider a classical version based on diffusive intensive measure for example by Bichteler et al. [22]. Then it is possible to make truncation functionals \( \psi_j \) by using local non-degeneracy of the Malliavin covariance matrix of the system. See Kusuoka and Yoshida [53] and Yoshida [99] for details of this case. The local non-degeneracy of the Malliavin covariance of the functional to be expanded plays a similar role as the Cramér condition in independent cases, assisted by the support theorem for stochastic differential equations.

Since typical statistics are expressed as a Bhattacharya-Ghosh [17] transform of a multidimensional additive functional that admits the Edgeworth expansion, it is possible to obtain Edgeworth expansions for them. This enables us to construct higher-order statistics for stochastic processes. For example, consider a sequence of discrete-time \( L^2 \)-martingales \( M^n = (M^n_t)_{t=0,1,\ldots,T_n} \) for \( t = 0, 1, \ldots, T_n \). Then a classical martingale central limit theorem is stated as follows. Suppose that (i) \( \sum_{t=1}^{T_n} E^n[(\xi^n_t)^2|\mathcal{F}_{t-1}] \to^p \sigma^2 \) as \( n \to \infty \) for some constant \( \sigma^2 \), and that for \( \epsilon > 0 \), \( \sum_{t=1}^{T_n} E^n[(\xi^n_t - \epsilon)^2 1_{|\xi^n_t| > \epsilon}|\mathcal{F}_{t-1}] \to^p 0 \) as \( n \to \infty \). Then \( M^n_{T_n} \to^d N(0, \sigma^2) \) as \( n \to \infty \). Here \( E^n \) denotes the expectation with respect to \( P^n \), and the convergence \( \to^p \) is naturally defined along the sequence \((P^n)_{n \in \mathbb{N}}\). For this result, see B. M. Brown [25], Dvoretzky [30], McLeish [63], Rebolledo [82], Hall and Heyde [37]. Functional type convergence results also hold.
Various extensions were made to limit theorems for semimartingales. Among them, a version of the central limit theorem for semimartingales is as follows. Consider a sequence of stochastic processes $X^n, n \in \mathbb{N}$, each of which is a semimartingale defined on a stochastic basis $\mathcal{F}^n$ with a filtration $\mathbf{F}^n = (\mathcal{F}^n_t)_{t \in \mathbb{R}_+}$, and has the local characteristics $(B^n, C^n, \nu^n)$, where $B^n$ is the finite variation part with respect to the truncation by the function $x_1 \mathbb{1}_{\{|x| \leq 1\}}$, $C^n$ is the predictable covariation process for the continuous local martingale part $X^{n,c}$ of $X^n$, and $\nu^n$ is the compensator of the integral-valued random measure $\mu^n$ of jumps of $X^n$. Denote by $M = (M_t)_{t \in \mathbb{R}_+}$ a continuous Gaussian martingale with a (deterministic) quadratic variation $\langle M \rangle$. Suppose that $X^n_0 = 0$ and the following conditions are fulfilled for every $t > 0$ and $\epsilon > 0$ as $n \to \infty$: (i) $\int_0^t \int_{|x| > \epsilon} \nu^n(ds, dx) \to^p 0$, (ii) $B^{n,c} + \sum_{s \leq t} \int_{|x| \leq \epsilon} \nu^n((s), dx) \to^p 0$, $B^{n,c}$ being the continuous part of $B^n$, and (iii) $C^n + \int_0^t \int_{|x| > \epsilon} x^2 \nu^n(ds, dx) - \sum_{s \leq t} \left(\int_{|x| \leq \epsilon} \nu^n((s), dx)\right)^2 \to^p \langle M \rangle_t$. Then the finite-dimensional convergence $X^n \to^d M$ holds. Moreover, under (i), (ii), and (iii) $\sup_{t \in [0, t]} \left|B^{n,c} + \sum_{s \leq t} \int_{|x| \leq \epsilon} \nu^n((s), dx)\right| \to^p 0$ as $n \to \infty$ for every $t > 0$ and $\epsilon > 0$, in place of (ii), one has the functional convergence $X^n \to^d M$ in $\mathcal{D}([0, t])$ as $n \to \infty$. See Liptser and Shiryaev [59], Jacod et al. [42], Jacod and Shiryaev [43], and Liptser and Shiryaev [60]. Developments of the central limit theorems for martingales and convergences to processes with independent increments are owed to many authors. We refer the reader to the bibliographical comments to Chapter VIII of Jacod and Shiryaev [43].

The simplest case is the central limit theorem for continuous local martingales. Let $M^n = (M^n_t)_{t \in [0, 1]}$ be a continuous local martingale defined on $\mathcal{F}^n$. If $\langle M^n \rangle_1 \to^p C_\infty$ as $n \to \infty$ for some constant $C_\infty$, then

$$M^n_1 \to^d N(0, C_\infty) \quad \text{as} \quad n \to \infty. \quad (2.8)$$

For later discussions, it is worth recalling the derivation of the central limit theorem (2.8). Let $C^n_t = \langle M^n \rangle_t$. We have a trivial decomposition of the characteristic function of $M^n_t$:

$$E[e^{iuM^n_t}] = T_0 + T_1 + T_2 \quad (2.9)$$

for $u \in \mathbb{R}$, where $T_0 = E[e^{-2^{-1}C_\infty u^2}]$, $T_1 = E[e^{iuM^n_1}(1 - e^{2^{-1}(C^n_1 - C_\infty)u^2})]$ and $T_2 = E[(e^{iuM^n_1} + 2C^n_1u - 1)e^{-2^{-1}C_\infty u^2}]$. If necessary, we replace $M^n$ by a suitably stopped process to validate integrability of variables. By the convergence of $C^n_t$, the tangent $T_1$ tends to 0. Moreover, the torsion $T_2$ vanishes thanks to the martingale property of the exponential martingale since $C_\infty$ is deterministic. Thus, $E[e^{iuM^n_1}] \to E[e^{-2^{-1}C_\infty u^2}] = e^{-2^{-1}C_\infty u^2}$, which proves (2.8).

For martingales with jumps, a uniformity condition such as the conditional type Linderberg condition is necessary to obtain central limit theorems. Otherwise, processes with independent increments can appear as the limit.

### 2.4.2 Berry-Esseen Bounds

Berry-Esseen type bounds are in Bolthausen [24] and Häsler [38]. Rate of convergence in the central limit theorem for semi-martingales is in Liptser and Shiryaev [58, 60]. In other frames of dependent structures, error bounds are found in Bolthausen [23] for functionals.

### 2.4.3 Asymptotic Expansion of Martingales

Consider a sequence of random variables \((Z_n)_{n \in \mathbb{N}}\) having a stochastic expansion

\[
Z_n = M_n + r_n N_n,
\]

where for each \( n \in \mathbb{N} \), \( M_n \) denotes the terminal random variable \( M^n_1 \) of a continuous martingale \((M^n_t)_{t \in [0,1]}\) with \( M^n_0 = 0 \), on a stochastic basis \( \mathcal{B}^n = (\Omega^n, \mathcal{F}^n, \mathbb{P}^n, P^n) \). \( \mathbb{P}^n = (\mathcal{F}^n_t)_{t \in [0,1]} \). The variable \( N_n \) is a random variable on \( \mathcal{B}^n \) but no specific structure like adaptiveness is assumed, and \( (r_n) \) is a sequence of positive numbers tending to zero as \( n \to \infty \). Suppose that \( \langle M^n \rangle_1 \to \rho \) as \( n \to \infty \) for the quadratic variation \( \langle M^n \rangle \) of \( M^n \). Then the martingale central limit theorem (2.8) ensures the convergence \( M_n \to^d N(0,1) \) as \( n \to \infty \).

The effect of the tangent \( T_1 \) appears in the asymptotic expansion of the law \( L(Z_n) \). We suppose that \( (M_n, \xi_n, N_n) \to^d (Z, \xi, \eta) \) as \( n \to \infty \) for \( \xi_n = r_n^{-1}(\langle M^n \rangle_1 - 1) \). Define the density \( p_n \) by

\[
p_n(z) = \phi(z) + \frac{1}{2} r_n \partial_z^2 (E[\xi|Z = z] \phi(z)) - r_n \partial_z (E[\eta|Z = z] \phi(z)),
\]

where \( \phi \) is the standard normal density. Furthermore, we assume that each \((\Omega^n, \mathcal{F}^n, P^n)\) is equipped with a Malliavin calculus and random variables are differentiable in Malliavin’s sense. Then the derivatives in (2.11) exist, and for any \( \alpha \in \mathbb{Z}_+, \ p > 1 \) and \( q > 2/3 \), we obtain the estimate

\[
\left| E[f(Z_n)] - \int f(z)p_n(z)dz \right| \leq C(\|f(Z_n)\|_{L^{p'}} + \|f\|_{L^{1+(1+|\xi|)^{q/2}}}d\xi) \\
\times (r_n^{1-q}P[\sigma_{M_n} < s_n]^{1/p} + \epsilon_n)
\]

for any measurable function \( f \) satisfying \( E[|f(Z_n)|] < \infty \) and \( \int |f(x)|p_n(z)dz < \infty \), where \( \sigma_{M_n} \) is the Malliavin covariance of \( M_n \), \( s_n \) are positive smooth functionals with complete non-degeneracy \( \sup_{n \in \mathbb{N}} E[s_n^{-n}] < \infty \) for any \( m > 1 \), \( p' = p/(p-1) \), \( \epsilon_n = o(r_n) \), and \( C \) is a constant independent of \( f \). Assumption of full non-degeneracy for \( \sigma_{M_n} \) is not realistic in statistical applications, nor necessary in asymptotic expansion.

The central limit theorem for the functional of the form \( \int_0^T T^{-1/2}a_t dw_t \) for a random process \( a_t \) is indispensable to deduce asymptotic normality of the estimators in the likelihood analysis of the drift parameter of ergodic diffusion processes. Then it is natural to seek for asymptotic expansion for martingales to formulate higher-order statistical inference for diffusion processes. As a matter of fact, the martingale expansion went ahead of the mixing method, as for semimartingales. The second-order mean-unbiased maximum
likelihood estimator $\hat{\theta}^*_T$ of the drift parameter $\theta$ of an ergodic diffusion process has the Edgeworth expansion
\[
P\left[ \sqrt{IT}(\hat{\theta}^*_T - \theta) \leq x \right] = \Phi(x) + \frac{I^{-1/3}}{2I^{3/2}} (x^2 - 1)\phi(x) + o(T^{-1/2})
\]
where $I$ is the Fisher information at $\theta$ and $\Phi$ is the standard normal distribution function. $I^{-1/3}$ is the coefficient of the Aamari-Chentsov affine $\alpha$-connection for $\alpha = -1/3$ [97].

See [97] for details of this subsection. A similar asymptotic expansion formula exits for general martingales $M^n$ with jumps. In that case, we take $\xi_n = \varrho^{-1}_n(\Xi_n - 1)$ with $\Xi_n = \frac{1}{3}[M^n] + \frac{2}{3}\langle M^n \rangle_1$. A Malliavin calculus on Wiener-Poisson space is used to quantify the non-degeneracy of $M_n$ [98].

Mykland [68, 69, 70] provided asymptotic expansion of moments. The author was inspired by his pioneering work.

The mixing approach gives in general more efficient way to asymptotic expansion if one treats functionals of $\epsilon$-Markov processes with mixing property like the above example. However, the martingale approach still has advantages of wide applicability. For example, an estimator of volatility in finite time horizon, non-Gaussianity appears in the higher-order term of the limit distribution even if the statistic is asymptotically normal. Such phenomena cannot be handled by mixing approach; however, the martingale expansion still gives asymptotic expansion.

### 2.5 Non-ergodic Statistics and Asymptotic Expansion

#### 2.5.1 Non-central Limit of Estimators in Non-ergodic Statistics

The non-ergodic statistics features asymptotic mixed normality of estimators. Non-normality of the maximum likelihood estimators was observed quite many years ago: White [95], Anderson [4], Rao [81], Keiding [46, 47].

Extension of the classical asymptotic decision theory was required to formulate non-ergodic statistics: Basawa and Koul [7], Basawa and Prakasa Rao [8], Jeganathan [45], and Basawa and Scott [9]. From aspects of limit theorems, the notion of stable convergence is fundamental since the Fisher information is random even in the limit. The nesting condition with Rényi mixing is a standard argument there. In this trend, Feigin [31] proved stable convergence for semimartingales.

Statistical inference for high frequency data has been attracting attention since around 1990. Huge volume of literature is available today: Prakasa Rao [77, 78], Dacunha-Castelle and Florens-Zmirou [27], Florens-Zmirou [32], Yoshida [96, 100], Genon-Catalat and Jacob [33], Bibby and Soerensen [21], Kessler [48], Andersen and Bollerslev [2], Andersen et al. [3], Barndorff-Nielsen and Shephard [5, 6], Shimizu and Yoshida [85], Uchida [90], Ogihara and Yoshida [73, 74], Uchida and Yoshida [92, 93], and Masuda [62] among many others. Recently a great interest is in estimation of volatility. The scaled error of a volatility estimator admits a stable convergence to a mixed normal distribution, that is, typically for a volatility estimator $\hat{\theta}_n$, $\sqrt{n}(\hat{\theta}_n - \theta) \to^d I^{-1/2} \zeta$ where $I$ is the random Fisher information and $\zeta \sim N(0, 1)$ independent of $I$. It is possible to apply the martingale problem method.
2 Asymptotic Expansions for Stochastic Processes

as in Genon-Catalot and Jacod [33], Jacod [41], or convergence of stochastic integrals in Jakubowski et al. [44] and Kurtz and Protter [52] to obtain stable convergence.

2.5.2 Non-ergodic Statistics and Martingale Expansion

To go beyond the first order\(^1\) asymptotic statistical theory, we need to develop asymptotic expansion of functionals. However, the potential (Doléans-Dade exponential\(^{-1}\)) that makes a local martingale from \(\exp(uM^n_t)\) no longer has a deterministic limit, and this breaks a usual way to asymptotic expansion. In other words, the exponential martingale in \(\mathcal{T}_2\) is not a martingale under the measure \(E[e^{-C_nu^2/2}]/E[e^{-C_nu^2/2}]\), and the torsion of this shift on the martingale appears in the expansion.

We will consider a \(d\)-dimensional random variable \(Z_n\) that admits the stochastic expansion (2.10) on \((\Omega, \mathcal{F}, \mathbb{P}, P)\), \(\mathbb{P} = (\mathcal{F}_t)_{t\in[0,1]}\). \(M^n\) is a \(d\)-dimensional continuous local martingale with \(M^n_0 = 0\), and \(N_n\) is a \(d\)-dimensional random variable. Let \(C^n_1 = (M^n_1)_1\), \(\mathbb{R}^d \otimes \mathbb{R}^d\)-valued random matrix. A \(d_1\)-dimensional reference variable is denoted by \(F_n\). For example, \(F_n\) is the Fisher information matrix. We shall present an expansion of the joint law \(\mathcal{L}((Z_n, F_n))\).

The tangent vectors are given by \(\dot{C}^n = r_n(C^n - C^n_1)\) and \(\dot{F}_n = r_n(F_n - F^n_1)\). Suppose that \((M^n, N_n, \dot{C}^n, \dot{F}_n) \to_{d,(\mathcal{F})} (M^n, N^n, C^n_1, F^n_1)\) and \(M^n_1 \sim N_1(0, C^n_1)\). These limit variables are defined on the extension \((\Omega, \mathcal{F}, \mathbb{P}) = (\Omega \times \omega, \mathcal{F} \times \mathcal{F}, \mathcal{P} \times \mathcal{P})\) of \((\Omega, \mathcal{F}, P)\).

Let \(\mathcal{F} = \mathcal{F} \vee \sigma[M^n_1]\). Random function \(\dot{C}^n(\omega, z) = \tilde{C}(\omega, z)\) is a matrix-valued random function satisfying \(\dot{C}(\omega, M^n_1) = E[\dot{C}^n | \mathcal{F}]\). Similarly, let \(\dot{F}_\infty(\omega, M^n_1) = E[\dot{F}^n | \mathcal{F}]\) and \(\dot{N}_\infty(\omega, M^n_1) = E[\dot{N}^n | \mathcal{F}]\).

To make an expansion formula, we need two kinds of random symbols: the adaptive random symbol and the anticipative random symbol. The adaptive random symbol is defined by

\[
\sigma(z, iu, iv) = \frac{1}{2} \tilde{C}(z)[(iu)^{\otimes 2}] + \dot{N}(z)[iu] + \dot{F}(z)[iv]
\]

for \(u \in \mathbb{R}^d\) and \(v \in \mathbb{R}^{d_1}\). Here the brackets mean a linear functional. This random symbol is corresponding to the correction term of the classical asymptotic expansion. Let \(\Psi_{\infty}(u, v) = \exp(-\frac{1}{2} \dot{C}(z)[u^{\otimes 2}] + iF_{\infty}(v))\), \(C_{\infty} := C^n_1\) and let \(L^n_t(u) = \exp(iM^n_t(u) + \frac{1}{2}C^n_t(u^{\otimes 2})) - 1\). Then the anticipative random symbol \(\tilde{\sigma}(iu, iv) = \sum_{j} c_j(iu)^{m_j}(iv)^{n_j}\) (multi-index) is specified by

\[
\lim_{n \to \infty} r_n^{-1} E[L^n_t(u) \Psi_{\infty}(u, v) \psi_n] = E[\Psi_{\infty}(u, v) \tilde{\sigma}(iu, iv)],
\]

where \(\psi_n \sim 1\) is a truncation functional a suitable choice of which reflects the local non-degeneracy of \((Z_n, F_n)\).

\(^1\) The order of the central limit theorem is referred to as the first order in asymptotic decision theory, differently from the numbering of terms in asymptotic expansion.
For the full random symbol $\sigma = \sigma + \overline{\sigma}$, the asymptotic expansion formula is defined by

$$p_n(z, x) = E[\phi(z; 0, C_{\infty})\delta_z(F_{\infty})] + r_nE[\sigma(z, \partial_x, \partial_\infty)[\phi(z; 0, C_{\infty})\delta_z(F_{\infty})]],$$

where $\phi(z; 0, C)$ is the normal density with mean 0 and covariance matrix $C$, and $\delta_z(F_{\infty})$ is Watanabe’s delta function; cf. Watanabe [94], Ikeda and Watanabe [40]. The adjoint $\sigma(z, \partial_z, \partial_\infty)^*$ is naturally defined as $\overline{\sigma}(z, \partial_z, \partial_\infty)^*[\phi(z; 0, C_{\infty})\delta_z(F_{\infty})] = \sum_j (-\partial_x)^{n_j} (-\partial_\infty)^{m_j} s_j(z; 0, C_{\infty})\delta_z(F_{\infty})$ and similarly for $\sigma$. The density formula gives a concrete expression since $E[\psi\delta_z(F)] = E[\psi|F = x]p_F(x)$ for functionals $\psi$ and $F$.

Under certain non-degeneracy conditions, for any positive numbers $B$ and $\gamma$,

$$\sup_{f \in \mathbb{E}(B, \gamma)} \left| E[f(Z_n, F_n)] - \int_{\mathbb{R}^{d_1}} f(z, x)p_n(z, x)dzdx \right| = o(r_n) \quad (2.13)$$

as $n \to \infty$, where $\mathbb{E}(B, \gamma)$ is the set of measurable functions $f : \mathbb{R}^{d+d_1} \to \mathbb{R}$ satisfying $|f(z, x)| \leq B(1 + |z| + |x|)^\gamma$ for all $(z, x) \in \mathbb{R}^d \times \mathbb{R}^{d_1}$. Details are given in [102].

The martingale expansion (2.13) was applied to the realized volatility in [101]. The martingale part $M_n$ is a sum of double Skorokhod integrals. The anticipative random symbol $\overline{\sigma}$ specified by the integration-by-parts formula at (2.12) has expression involving the Malliavin derivatives. Recently Podolskij and Yoshida [76] obtained expansions for $p$-variations. Construction of higher order statistical inference is a theme of the non-ergodic statistics today.

Acknowledgements This work was in part supported by Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research No. 24340015 (Scientific Research), Nos. 24650148 and 26540011 (Challenging Exploratory Research); CREST Japan Science and Technology Agency; and by a Cooperative Research Program of the Institute of Statistical Mathematics.

References

variance: high frequency based covariance, regression, and correlation in financial 

[7] Ishwar V. Basawa and Hira L. Koul. Asymptotic tests of composite hypotheses 
1979.


Esseen bounds for statistics of weakly dependent samples. *Bernoulli*, 3:329–349, 
1997.


[12] Rabi Bhattacharya and Aramian Wasielak. On the speed of convergence of multi-
dimensional diffusions to equilibrium. *Stochastics and Dynamics*, 12(1), 2012.


[14] Rabi N. Bhattacharya. Rates of weak convergence and asymptotic expansions for 


[16] Rabi N. Bhattacharya. On classical limit theorems for diffusions. *Sankhyā (Statis-


[18] Rabi N. Bhattacharya and Sundareswaran Ramasubramanian. Recurrence and erg-

the 1976 original.


[21] Bo Martin Bibby and Michael Sørensen. Martingale estimation functions for dis-

[22] Klaus Bichteler, Jean-Bernard Gravereaux, and Jean Jacod. *Malliavin Calculus for 
Processes with Jumps*, volume 2 of *Stochastics Monographs*. Gordon and Breach 

[23] Erwin Bolthausen. The Berry-Esseen theorem for functionals of discrete Markov 


Rabi N. Bhattacharya
Selected Papers
Denker, M.; Waymire, E.C. (Eds.)
2016, XXI, 711 p. 1 illus., Hardcover
ISBN: 978-3-319-30188-4
A product of Birkhäuser Basel