Stable convergence with applications

Emil Aas Stoltenberg

Trial lecture

October 30, 2020

Department of Mathematics
University of Oslo

What is it?

A form of convergence of random variables. In statistics, one typically deals with two forms of convergence of random variables.

Convergence in probability $X_n \xrightarrow{p} X$. For every $\varepsilon > 0$,

$$\Pr(|X_n - X| \geq \varepsilon) \to 0.$$

Convergence in distribution $X_n \xrightarrow{d} X$. For every bounded and continuous function f,

$$\mathrm{E}\, f(X_n) \to \mathrm{E}\, f(X).$$

Stable convergence $X_n \xrightarrow{\text{st.}} X$, is weaker than convergence in probability, and stronger than convergence in distribution.

$$X_n \xrightarrow{p} X \;\Rightarrow\; X_n \xrightarrow{\text{st.}} X \;\Rightarrow\; X_n \xrightarrow{d} X.$$

Why do we need it?

(i) The Cramér–Slutsky rules don't apply when the limit in probability of the denominator is a proper random variable.

(ii) Conditionality principle considerations: Often, we want to bring prelimiting information into the limit distribution.

(iii) Conditioning on a path might mess up the probabilistic structure that is used in the derivation of the asymptotic distribution, for example independence or conditional independence.

(iv) Simplifies measure change. Derive large-sample results under one ('easy') probability measure, then adjust the limiting distribution back to the true probability measure.

(v) Localisation. Large-sample results that apply to a stopped (localised) process, apply almost immediately to the full process. For example, we can assume coefficients are bounded, even when they need not be.

(i) The Cramér–Slutsky rules

If $X_1, \ldots, X_n$ are i.i.d. random variables with expectation θ and variance σ^2. The central limit theorem

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{d} \mathrm{N}(0, \sigma^2).$$

Cramér–Slutsky rules: If $A_n \to_d A$, and $B_n \to_p b$, then

$$A_n B_n \to_d Ab.$$

So if $\widehat{\sigma}_n \to_p \sigma$,

$$\frac{\sqrt{n}(\bar{X}_n - \theta)}{\widehat{\sigma}_n} \xrightarrow{d} (1/\sigma)\mathrm{N}(0, \sigma^2) = \mathrm{N}(0, 1),$$

and we can base inference on θ on the approximation

$$\Pr(\sqrt{n}(\bar{X}_n - \theta) \leq z) \approx \Phi(z/\widehat{\sigma}_n).$$

If σ is a proper random variable, the central limit theorem above is not strong enough for this conclusion.

... might fail when b is a random variable

Let $U_1, U_2, \ldots$ be i.i.d. Unif$(0,1)$, $(a_n)_{n\geq 1} \subset [0, 1/2]$. Set

$$A_n = \begin{cases} 1, & U_n \in [a_n, a_n + 1/2], \\ 0, & \text{otherwise.} \end{cases}$$

Then $A_1, A_2, \ldots$ are i.i.d. Bernoulli$(1/2)$. Let $B \sim \text{Unif}(0,1)$. Then $A_n \to_d A \sim \text{Bernoulli}(1/2)$, and $B \to_p B$.

$$A_n B = \begin{cases} B, & U_n \in [a_n, a_n + 1/2], \\ 0, & \text{otherwise.} \end{cases}$$

Let $f(x) = \max\{\min(x, 1), 0\}$, and note that f is bounded and continuous. Now,

$$\mathrm{E}\, f(A_n B) = \int_{a_n}^{a_n + 1/2} v \,\mathrm{d}v = \frac{1}{2}\big(a_n + \frac{1}{4}\big).$$

This example is from Häusler and Luschgy (2015).

(ii) The conditionality principle[1]

Sir David Cox (1994, p. 442):

> *How does the long run become relevant to a particular set of data? Well, by being suitably conditioned. The arguments for this seem to me absolutely overwhelming* [...]

Famous example from Cox (1958, p. 360): Flip a fair coin X and sample

$$Y \sim \mathrm{N}(\theta, \sigma_X^2),$$

where $\sigma_0^2 < \sigma_1^2$. Suppose $X = 1$. What is the 'correct' variance of $\widehat{\theta} = Y$,

$$\sigma_1^2, \quad \text{or} \quad \frac{\sigma_0^2 + \sigma_1^2}{2}.$$

[1] See Barndorff-Nielsen and Cox (1994, p. 34) for a precise definition.

A toy example[2]

Let $\varepsilon_1, \varepsilon_2, \ldots$ be i.i.d., $\Pr(\varepsilon_1 = -1) = \Pr(\varepsilon_1 = 1) = 1/2$. For some $\rho \in (-1, 1)$ set $\sigma_1 = \rho\varepsilon_1$, and $\sigma_i = \sum_{j=1}^{i-1} \rho^j \varepsilon_j$ for $i \geq 2$. We observe

$$X_i = \theta + \sigma_i \varepsilon_i, \quad \text{for } i = 1, 2, \ldots,$$

and seek to make inference on θ. By Doob's convergence theorem

$$\sigma_n = \sum_{j=1}^{n} \rho^j \varepsilon_j \to \sigma_\infty = \sum_{j=1}^{\infty} \rho^j \varepsilon_j,$$

almost surely, with σ_∞ a random variable. Set

$$\mathrm{E}\exp\{it\sqrt{n}(\bar{X}_n - \theta)\} \to \mathrm{E}\exp\big(-\frac{t^2\sigma_\infty^2}{2}\big)$$

thus X_n tends in distribution to a mixed normal limit,

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{d} \mathrm{N}(0, \sigma_\infty^2).$$

[2] Adapted from Hall and Heyde (1980).

Toy example contd.

But

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{d} \mathrm{N}(0, \sigma_\infty^2),$$

cannot be directly used for inference on θ.

(i) Even though $\widehat{\sigma}_n \to_p \sigma_\infty$, we cannot conclude that $\sqrt{n}(\bar{X}_n - \theta)/\widehat{\sigma}_n$ tends to a standard normal;

(ii) Averaging out σ_∞ breaches the conditionality principle;

(iii) Condition on σ_∞? But then we fiddle with the independence of the $\varepsilon_1, \varepsilon_2, \ldots$.

Definition of stable convergence

A probability space $(\Omega, \mathcal{F}, \Pr)$, and $\mathcal{G} \subset \mathcal{F}$, on which we have a sequence $(X_n)_{n \geq 1}$ with values in a Polish space $(\mathcal{X}, \mathcal{B})$.[3] Say that X_n converges $\mathcal{G}$-stably, and write

$$X_n \Rightarrow X, \quad \mathcal{G}\text{-stably},$$

if

$$\mathrm{E}\, Y f(X_n) \to \int_\Omega \int_\mathcal{X} Y(\omega) f(x) Q(\omega, \mathrm{d}x) \Pr(\mathrm{d}\omega),$$

as $n \to \infty$, for all bounded $\mathcal{G}$-measurable random variables Y, and all bounded and continuous functions f.

[3] A complete (all Cauchy sequences converge) and separable (has a countable and dense subset) metric space.

What does this mean? I

Have a sequence $X_1, X_2, \dots$ on $(\Omega, \mathcal{F}, \Pr)$.

Convergence in distribution:

$$\Pr(X_n \in B) = P_n(B) \to P(B), \quad \text{for all } P\text{-cont. } B,$$

and we 'realise' the limit P with a random variable $X \sim P$. Since $\Pr(X_n \in B) = \Pr(X_n \in B \mid \{\Omega, \emptyset\})$, the distributions of $(X_n)_{n \geq 1}$ given the trivial σ-algebra converge.

Stable convergence: Condition on a larger σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ ('bring more information into the limit'), and

$$\Pr(X_n \in B \mid \mathcal{G}) = Q_n(\cdot, B) \to Q(\cdot, B),$$

in the sense above. Can regard stable convergence as convergence of conditional distributions.

What does this mean? II

As usual, we would like to 'realise' the limiting distribution by a random variable. Construct an extension of the original probability space,

$$\widetilde{\Omega} = \Omega \times \mathbb{R}, \quad \widetilde{\mathcal{F}} = \mathcal{F} \otimes \mathcal{B}, \quad \widetilde{\Pr}(\mathrm{d}\omega, \mathrm{d}x) = Q(\omega, \mathrm{d}x)\Pr(\mathrm{d}\omega).$$

Then define a random variable $Y(\omega, x)$ on the extension, such that

$$\Pr(Y \leq y \mid \mathcal{G})(\omega) = Q(\omega, (-\infty, y\,]).$$

In the toy example, that $\sqrt{n}(\bar{X}_n - \theta)$ converges $\mathcal{F}$-stably to normally distributed Y, means that

$$Y(\omega, \cdot) \sim Q(\omega, (-\infty, y]) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}\sigma_\infty(\omega)} \exp\big(-\frac{z^2}{2\sigma_\infty^2(\omega)}\big)\,\mathrm{d}z.$$

Consequences

If Y_n converges $\mathcal{G}$-stably to Y, then $Y_n \to_d Y$ (set $\xi = 1$).

Proposition VIII.5.33 in Jacod and Shiryaev (2003, p. 513). There is equivalence

(1) $Y_n \Rightarrow Y$ $\mathcal{G}$-stably;
(2) $(Y_n, X) \to_d (Y, X)$ for all $\mathcal{G}$-measurable X;
(3) $(Y_n, X) \Rightarrow (Y, X)$ $\mathcal{G}$-stably for all $\mathcal{G}$-measurable X;
(4) If $Y_n = (Y_{n,1}, \ldots, Y_{n,p})^{\mathrm{t}}$ take values in $\mathbb{R}^p$, then

$$\mathrm{E}\, I_A \exp(iu^{\mathrm{t}} Y_n) \to \mathrm{E}\, I_A \exp(iu^{\mathrm{t}} Y), \quad \text{for all } A \in \mathcal{G}.$$

'Stable' Cramér–Slutsky. If A_n converges $\mathcal{G}$-stably to A, and $B_n \to_p B$, for $\mathcal{G}$-measurable B, then

$$A_n B_n \to_d AB.$$

Pf: By (2), $(A_n, B_n) = (A_n, B) + o_p(1) \to_d (A, B)$, & cont. mapping.

Convergence in distribution, but not stably

(1) Let X_1, X_2 be independent with distribution function F. Set

$$Y_n = \begin{cases} X_1, & \text{for } n \text{ odd,} \\ X_2, & \text{for } n \text{ even.} \end{cases}$$

Then $Y_n \to_d F$, but Y_n does not converge stably. If $A = \{X_1 \leq a\}$, then ,

$$\mathrm{E}\, I_A f(Y_n) = \begin{cases} \mathrm{E}\, I_A\, g(X_1), & \text{for } n \text{ odd,} \\ F(a)\, \mathrm{E}\, g(X_2), & \text{for } n \text{ even.} \end{cases}$$

(2) Let $X_1 \sim \mathrm{N}(0,1)$, independent of $X_2, X_3, \ldots$, that are i.i.d. with $\mathrm{E}\, X_2 = 0$ and $\mathrm{Var}(X_2) = 1$. Set $\mathcal{F} = \sigma(X_1, X_2, X_3, \ldots)$, and

$$Y_n = \frac{1}{\sqrt{2}} X_1 + \frac{1}{\sqrt{2}\sqrt{n}} \sum_{i=2}^{n} X_i.$$

Then

$$Y_n \xrightarrow{d} \mathrm{N}(0,1),$$

but Y_n does not converge $\mathcal{F}$-stably to a $\mathrm{N}(0,1)$.

Applications

(1) Supercritical Galton–Watson processes.

(2) Critical AR(1) processes.

(3) Stochastic volatility models with leverage effect.

$(Y_{n,j})_{n,j\geq 1}$ i.i.d. Poisson(θ). Suppose that $X_0 = 1$, and set

$$X_n = \sum_{j=1}^{X_{n-1}} Y_{n,j}.$$

Assume that $\theta > 1$.[4] Using that $X_1, X_2, \ldots$ is a Markov chain, the log-likelihood is

$$\ell_n(\theta) = \sum_{j=1}^{n} \{x_j(\log\theta + \log x_{j-1}) - \theta x_{j-1} - \log x_j!\},$$

and by solving $\partial \ell_n(\theta)/\partial\theta = 0$ we find

$$\widehat{\theta}_n = \frac{\sum_{j=1}^n X_j}{\sum_{j=1}^n X_{j-1}}, \quad \text{so} \quad \widehat{\theta}_n - \theta = \frac{\sum_{j=1}^n (X_j - \theta X_{j-1})}{\sum_{j=1}^n X_{j-1}}.$$

[4] If $\theta \leq 1$, then $\lim_{n\to\infty} X_n = 0$ almost surely (Williams, 1991, Ch. 0).

Supercritical Galton–Watson processes contd.

The conditional variance process is

$$\sum_{j=1}^{n} \mathrm{E}\left\{(X_j - \theta X_{j-1}) \mid \mathcal{F}_{j-1}\right\} = \sum_{j=1}^{n} \mathrm{Var}(X_j \mid \mathcal{F}_{j-1}) = \theta \sum_{j=1}^{n} X_{j-1}.$$

Since $\theta^{-n} X_n$ is a martingale (and $\sup_n \mathrm{E}\,|\theta^{-n} X_n|^2 < \infty$),

$$\frac{\sum_{j=1}^{n} X_{j-1}}{\sum_{j=1}^{n} \theta^{j-1}} = \frac{\sum_{j=1}^{n} \theta^{j-1} \theta^{-(j-1)} X_{j-1}}{\sum_{j=1}^{n} \theta^{j-1}} \xrightarrow{p} M_\infty,$$

by Doob's convergence theorem, and Toeplitz lemma. But M_∞ is a proper random variable. Since $\sum_{j=1}^{n} \theta^{j-1} \sim \theta^n/(\theta - 1)$,

$$\frac{\theta - 1}{\theta^n} \sum_{j=1}^{n} \mathrm{E}\left\{(X_j - \theta X_{j-1}) \mid \mathcal{F}_{j-1}\right\} \xrightarrow{p} M_\infty.$$

Critical AR(1)-type process

Consider $X_j = \theta(j/n)X_{j-1} + \varepsilon_j$ for $j = 1, \ldots, n$, with $\theta(t)$ some function on $[0, 1]$, $X_0 = 0$, and $\varepsilon_1, \varepsilon_2, \ldots$ i.i.d. with $\mathrm{E}\,\varepsilon_1 = 0$, $\mathrm{Var}(\varepsilon_1) = 1$, and $\mathrm{E}\,\varepsilon_1^4 < \infty$. Test $H_0\colon \theta(t) = 1$.

For $t \in (0, 1]$, the least squares estimator is

$$\widehat{\theta}_n(t) = \frac{\sum_{j/n \le t} X_{j-1}X_j}{\sum_{j/n \le t} X_{j-1}^2}.$$

Under H_0

$$n(\widehat{\theta}_n(t) - 1) = \frac{\sum_{j/n \le t}^{n} X_{j-1}\varepsilon_j}{\sum_{j=1}^{n} X_{j-1}^2},$$

and using the Skorokhod embedding

$$\frac{1}{n} \sum_{j/n \le t} X_{j-1}\varepsilon_j = \int_0^{t_*/n} W_n(s)\,\mathrm{d}W_n(s) + o_p(1),$$

where $W_n(t) = B(tn)/\sqrt{n} \overset{d}{=} B_t$, and $t_* = \max\{t_i\colon t_i \le nt\}$ for stopping times $t_1, t_2, \ldots$.

Critical AR(1)-type process contd.

By an application of Itô's lemma $\mathrm{d}W(s)^2 = 2W(s)\,\mathrm{d}W(s) + s$,

$$\frac{1}{n}\sum_{j/n\le t} X_{j-1}\varepsilon_j = \frac{W_n(t_*/n)^2 - t_*/n}{2} + o_p(1).$$

Since $n^{-1}\mathrm{E}\,t_* = n^{-1}\mathrm{Var}(\sum_{i/n\le t}\varepsilon_i) = [nt]/n \to t$, and $n^{-2}\mathrm{E}\,t_*^2 \le n^{-2}4\,\mathrm{E}\,\varepsilon_1^4$, we get $t_*/n \to_p t$. By continuity of $t \mapsto W_n(t)$

$$\frac{1}{n}\sum_{j/n\le t} X_{j-1}\varepsilon_j \xrightarrow{d} \frac{B_t^2 - t}{2},$$

If we show that this convergence is stable (which it is), then

$$\Big(\frac{1}{n}\sum_{j/n\le t} X_{j-1}\varepsilon_j, \frac{1}{n^2}\sum_{i/n\le t} X_{i-1}^2\Big) \xrightarrow{d} \Big(\frac{B_t^2 - t}{2}, \int_0^t B_s^2\,\mathrm{d}s\Big),$$

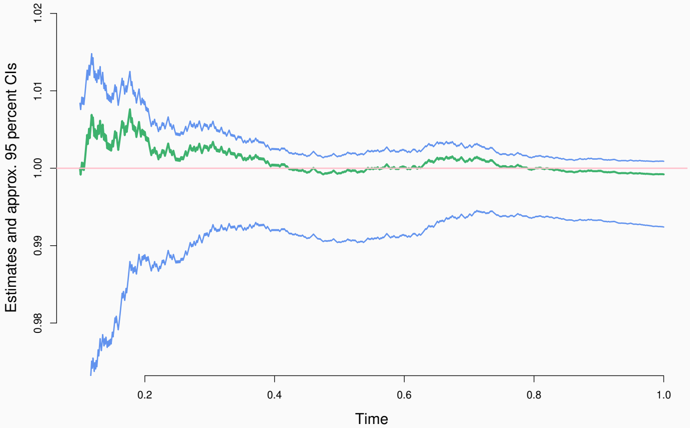
and we can conclude

$$n(\widehat{\theta}_n(t) - 1) \xrightarrow{d} \frac{B_t^2 - t}{2\int_0^t B_s^2\,\mathrm{d}s}.$$

Simulations

With $n = 1\,000$, $\sigma = 1$, and H_0 true $\theta(t) = 1$, and $\varepsilon_1, \varepsilon_2, \ldots$ i.i.d.

$$\Pr(\varepsilon_i = 1) = \Pr(\varepsilon_i = -1) = 1/2.$$



Volatility estimation

Consider the process

$$\mathrm{d}X_t = \mu_t\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t, \quad t \in [0,T],$$

where $X_0 = x_0$, W_t is a one-dimensional Wiener process, and the volatility σ_t^2 is itself a non-negative continuous Itô-process that might in part be driven by W_t. X_t is observed at times

$$0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T,$$

and $t_{j+1} - t_j = T/n$ for all j. Want inference on the *integrated volatility*

$$\theta_t = \int_0^t \sigma_s^2\,\mathrm{d}s, \quad \text{at time } T.$$

That the *realised volatility*

$$\widehat{\theta}_t^n = \sum_{t_{j+1}\leq t} (X_{t_{j+1}} - X_{t_j})^2 \xrightarrow{p} \int_0^t \sigma_s^2\,\mathrm{d}s = \theta_t, \tag{1}$$

is a fundamental fact about (semi-)martingales (Jacod and Shiryaev, 2003, Theorem I.4.47, p. 52).

Volatility estimation contd.

One finds that

$$\widehat{\theta}_t^n - \theta_t = M_t^n + o_p(n^{-1/2}),$$

where M_t^n is the continuous time martingale

$$M_t^n = 2 \sum_{t_{j+1} \le t} \int_{t_j}^{t_{j+1}} (X_s - X_{t_j})\, \mathrm{d}X_s + 2 \int_{t_*}^{t} (X_s - X_{t_*})\, \mathrm{d}X_s,$$

with $t_* = \max\{t_{j+1} \colon t_{j+1} \le t\}$.

Heuristic argument:

$$(X_s - X_{t_j})^2 = (\int_{t_j}^{s} \sigma_u \, \mathrm{d}W_u)^2 \approx (s - t_j)\sigma_{t_j}^2.$$

The predictable quadratic variation,

$$\begin{aligned}\langle M^n, M^n \rangle_{(t_j, t_{j+1}]} &= \int_{t_j}^{t_{j+1}} (X_s - X_{t_j})^2 \sigma_s^2 \, \mathrm{d}s \\ &\approx \int_{t_j}^{t_{j+1}} (s - t_j)\sigma_{t_j}^4 \, \mathrm{d}s = \frac{(t_{j+1} - t_j)^2}{2} \sigma_{t_j}^4.\end{aligned}$$

Volatility estimation contd.

$$n\langle M^n, M^n\rangle_t \xrightarrow{p} 2T\int_0^t \sigma_s^4\,\mathrm{d}s, \quad \text{for all } t$$

and by a martingale CLT (Mykland and Zhang, 2012, p. 152),

$$n^{1/2}(\widehat{\theta}_T^n - \theta_T) \to \big(2T\int_0^T \sigma_t^4\,\mathrm{d}t\big)^{1/2} Z$$

stably in distribution, where $Z \sim \mathrm{N}(0,1)$ is independent of $\int_0^T \sigma_t^4\,\mathrm{d}t$. Due to the stability of this convergence

$$\frac{n^{1/2}(\widehat{\theta}_T^n - \theta_T)}{c_n} \xrightarrow{d} \mathrm{N}(0,1),$$

where c_n^2 is a consistent estimator of $2T\int_0^T \sigma_t^4\,\mathrm{d}t$ (see Mykland and Zhang (2012, Theorem 2.28, pp. 137–138) for such an estimator).

Martingale central limit theorems

Two theorems for continuous martingales, here stated for continuous martingales, both extend to the càdlàg case.

Theorem 3.6. (Helland, 1982, p. 88) *Let M^n be a sequence of continuous local martingales on $[0,T]$. Suppose that there is a measurable function f such that*

$$\langle M^n, M^n \rangle_t \xrightarrow{p} \int_0^t f^2(s)\,\mathrm{d}s, \quad \textit{for all } t.$$

Then $M^n \Rightarrow \int f\,\mathrm{d}W$, where W_t is a one-dimensional Wiener process.

(G): $\mathcal{F}_t$ is generated by independent Wiener processes $W_t^{(1)}, \ldots, W_t^{(p)}$, for some $p \geq 1$.

Theorem 3.7. (Zhang (2001), Mykland and Zhang (2012, p. 152)) *Assume* (G). *Let M^n be a sequence of continuous local martingales on $[0,T]$. Suppose that there is a $\mathcal{F}_t$-adapted process f_t such that*

$$\langle M^n, M^n \rangle_t \xrightarrow{p} \int_0^t f^2(s)\,\mathrm{d}s, \quad \textit{for all } t,$$

and that for $j = 1, \ldots, p$

$$\langle M^n, W^{(j)} \rangle_t \xrightarrow{p} 0, \quad \textit{for all } t. \tag{8}$$

Then $M^n \Rightarrow \int f\,\mathrm{d}W$ $\mathcal{F}$-stably, where W_t is a one-dimensional Wiener process defined on an extension.

The independent-of-data condition

Stable convergence is weak convergence conditionally on (parts) of the data. Need something more concrete than a σ-algebra to represent the data:

$$\mathcal{F}_t = \sigma(W_t^{(1)}, \ldots, W_t^{(p)}) = \sigma(\text{indep. Wiener processes}),$$

is sufficient when dealing with continuous processes. When trying to show that $M^n \Rightarrow M = \int f \, \mathrm{d}W'$, we must ensure that M^n tends to something that is uncorrelated with the $W^{(1)}, \ldots, W^{(p)}$, that is, $\langle M^n, W^{(j)} \rangle_t \to_p 0$ for all j.

The Lévy-characterisation (Karatzas and Shreve, 1991, p. 157): The following are equivalent

1) $(X^{(1)}, \ldots, X^{(k)})$ is a standard Wiener process;
2) $X_t^{(i)} X_t^{(j)} - \delta_{i,j} t$ is a local martingale for $1 \le i, j \le k$;
3) $[X^{(i)} X^{(j)}]_t = \delta_{i,j} t$ for $1 \le i, j \le k$,

where $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ otherwise, the Kronecker delta.

Check the independence condition

Let $\mathrm{d}X_t = \sigma_t \, \mathrm{d}W_t$, for $t \in [0,1]$, with W a 1-dim. Wiener process, and set

$$M_t^n = 2n^{1/2} \sum_{t_{j+1} \le t} \int_{t_j}^{t_{j+1}} (X_s - X_{t_j}) \, \mathrm{d}X_s + 2n^{1/2} \int_{t_*}^{t} (X_s - X_{t_*}) \, \mathrm{d}X_s,$$

The tricicity

$$n^{1/2}[X,X,X]_t^n = n^{1/2} \sum_{t_{n,i+1} \le t} (X_{t_{n,i+1}} - X_{t_{n,i}})^3,$$

is consistent for $[M^n, X]_t$. If $t_{n,i+1}$ are, for example, fixed and equidistant times $t_{n,i+1} - t_{n,i} = 1/n$, then

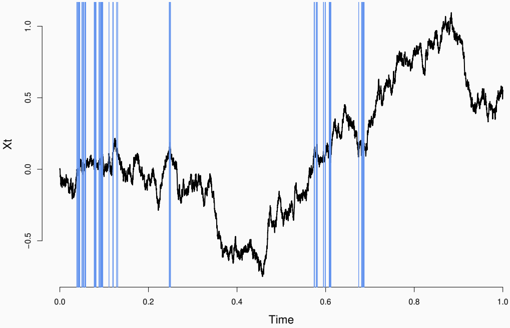
$$n^{1/2}[X,X,X]_t^n = \sum_{t_{n,i+1} \le t} (X_{t_{n,i+1}} - X_{t_{n,i}})^3 \approx \frac{1}{n} \sum_{t_{n,i+1} \le t} \mathrm{N}(0, \sigma_{t_{n,i}}^2)^3,$$

which tends to zero in probability, $\mathrm{E}\,\mathrm{N}(0,\sigma^2)^3 = 0$. Thus, $[X,X,X]_t^n \to_p v_t \neq 0$, is closely related to the skewness of the increments $X_{t_{n,i+1}} - X_{t_{n,i}}$.

Endogenous observation times

Let X_t be a 1-dim. Wiener process. We observe X_t at $t_{n,0} = 0$, and

$$t_{n,i+1} = \text{smallest } t > t_{n,i} \text{ s.t.} \begin{cases} X_t - X_{t_{n,i-1}} = n^{-1/2}a, \text{ or} \\ X_t - X_{t_{n,i-1}} = -n^{-1/2}b, \end{cases}, \ a, b > 0$$



A version of Example 4 is in Li et al. (2014, p. 590).

Endogenous observation times contd.

Then

$$X_{t_{n,i}} - X_{t_{n,i-1}} \overset{d}{=} n^{-1/2} Y, \quad \text{where} \quad Y = \begin{cases} a, & \text{with prob. } \frac{b}{a+b}, \\ -b, & \text{with prob. } \frac{a}{a+b}. \end{cases},$$

independent. Importantly,

$$\mathrm{E}\left(X_{t_{n,i}} - X_{t_{n,i-1}}\right)^3 = \frac{a-b}{n^{3/2}},$$

so non-zero skewness when $a - b \neq 0$, and $n^{1/2}[X, X, X]_t^n \xrightarrow{p} a - b$. Can 'fix' this by constructing a martingale (finding a process g_s)

$$\widetilde{M}_t^n = M_t^n - \int_0^t g_s \, \mathrm{d}X_s, \quad \text{so that} \quad \langle \widetilde{M}^n, X \rangle_t \xrightarrow{p} 0,$$

and adjusting back to get

$$M_t^n \Rightarrow \int_0^t a_s \, \mathrm{d}X_s + \int_0^t b_s \, \mathrm{d}W_s', \quad \mathcal{F}\text{-stably}.$$

which is a $\mathcal{F}$-conditional Gaussian martingale (Jacod and Shiryaev, 2003, p. 130).

Relationship to standard asymptotics

Suppose that $Z_1, Z_2, \ldots$ are i.i.d. random variables with

$$\mathrm{E}\, Z_1 = 0, \quad \mathrm{Var}(Y) = \sigma^2, \quad \zeta = \frac{\mathrm{E}\, Z^3}{\sigma^3}, \quad \kappa = \frac{\mathrm{E}\, Z^4}{\sigma^4}.$$

Estimator $\widehat{\sigma}_n^2 = n^{-1} \sum_{i=1}^n Z_i^2$, and

$$\sqrt{n}(\widehat{\sigma}_n^2 - \sigma^2) \xrightarrow{d} \mathrm{N}\{0, \sigma^4(\kappa - 1)\}.$$

Consider

$$\widetilde{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n Z_i^2 - \widehat{c}_n \frac{1}{n}\sum_{i=1}^n Z_i, \quad \text{with} \quad \widehat{c}_n = \frac{\sum_{i=1}^n Z_i^3}{\sum_{i=1}^n Z_i^2} \xrightarrow{p} \frac{\mathrm{E}\, Z_1^3}{\mathrm{E}\, Z_1^2} =: c,$$

where c minimises

$$\mathrm{Var}(Z_1^2 - cZ_1) = \sigma^4(\kappa - 1) + c^2 \mathrm{E}\,[Z_1^2] - 2c\mathrm{E}\,[Z_1^3]$$

$$\begin{aligned} \sqrt{n}(\widetilde{\sigma}_n^2 - \sigma^2) &= \frac{1}{n}\sum_{i=1}^n (Z_i^2 - cZ_i) - (\widehat{c}_n - c)\frac{1}{n}\sum_{i=1}^n Z_i \\ &= \frac{1}{n}\sum_{i=1}^n (Z_i^2 - cZ_i) + o_p(1) \xrightarrow{d} \mathrm{N}\{0, \sigma^4(\kappa - 1 - \zeta^2)\}. \end{aligned}$$

Measure change and stable convergence

Suppose

$$\mathrm{d}X_t = \mu_t \,\mathrm{d}t + \sigma_t \,\mathrm{d}W_t, \quad X_0 = x_0,$$

with W a Wiener process under P, and that we observe X_t at discrete time $0 \le t_{n,0}, \dots, t_{n,n} \le T$. It can be easier to derive large-sample results when X_t is a martingale, rather than a semimartingale, that is

$$\mathrm{d}X_t = \sigma_t \,\mathrm{d}W'_t, \quad X_0 = x_0,$$

with W' a Wiener process under P'.

The probabilities P and P' are mutually absolutely continuous (see Girsanov's theorem, Karatzas and Shreve (1991, Corollary 3.5.2, p. 192)) and the Radon–Nikodym derivative $\mathrm{d}P/\mathrm{d}P'$ is $\mathcal{F}$-measurable. Let $Y_n = \sqrt{n}(\widehat{\theta}_T^n - \theta_T)$, as above, and assume that $Y_n \Rightarrow Y$ $\mathcal{F}$-stably under P'. For bounded & cont. g, and bounded $\mathcal{F}$-meas. ξ,

$$\mathrm{E}_P \,\xi g(Y_n) = \mathrm{E}_{P'} \frac{\mathrm{d}P}{\mathrm{d}P'} \xi g(Y_n) \to \mathrm{E}_{P'} \frac{\mathrm{d}P}{\mathrm{d}P'} \xi g(Y) = \mathrm{E}_P \,\xi g(Y),$$

see Mykland and Zhang (2009, Prop. 1, p. 1408).

Localisation and stable convergence

An example from Mykland and Zhang (2012, pp. 156–161): When proving that

$$Y_n = n^{1/2}(\widehat{\theta}_T^n - \theta_T) \to \big(2T \int_0^T \sigma_t^4 \,\mathrm{d}t\big)^{1/2} Z = Y, \quad \text{stably,}$$

it is convenient to assume that $\sigma_t^2 \leq \sigma_+^2$, for all t, where σ_+^2 is some constant. Stable convergence makes it possible to relax this assumption, and instead assume that σ_t^2 is locally bounded. That is, there is a sequence $\tau_1 < \tau_2 < \tau_2 < \cdots$ of stopping times such that

$$\Pr(\lim_{m\to\infty} \tau_m = T) = 1, \quad \text{and} \quad \sigma_t^2 \leq \sigma_{m,+}^2, \text{ for } 0 \leq t \leq \tau_m.$$

For if $Y_n \Rightarrow T$ $\mathcal{F}$-stably, then $\xi I_{\{\tau_m \leq T\}}$ is $\mathcal{F}$-measurable

$$\mathrm{E}\,\xi I_{\{\tau_m \leq T\}} f(Y_n) \to \mathrm{E}\,\xi I_{\{\tau_m \leq T\}} f(Y),$$

and

$$\begin{aligned} &|\mathrm{E}\,\xi f(Y_n) - \mathrm{E}\,\xi f(Y)| \\ &\quad \leq |\mathrm{E}\,\xi I_{\{\tau_m \leq T\}} f(Y_n) - \mathrm{E}\,\xi I_{\{\tau_m \leq T\}} f(Y)| + \max_y |f(y)| \Pr(\tau_m > T). \end{aligned}$$

References

Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics*. Chapman & Hall, Boca Raton, FL.

Billingsley, P. (1995). *Probability and Measure*. John Wiley & Sons, New York.

Cox, D. (1958). Some problems connected with statistical inference. *The Annals of Mathematical Statistics*, 29:357–372.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.

Häusler, E. and Luschgy, H. (2015). *Stable Convergence and Stable Limit Theorems*. Springer, Heidelberg.

Helland, I. S. (1982). Central limit theorems for martingales with discrete or continuous time. *Scandinavian Journal of Statistics*, 9:79–94.

Jacod, J. and Shiryaev, A. (2003). *Limit Theorems for Stochastic Processes. Second Edition*. Springer, Berlin.

Karatzas, I. and Shreve, S. (1991). *Brownian Motion and Stochastic Calculus. Second Edition.* Springer, New York.

Li, Y., Mykland, P. A., Renault, E., Zhang, L., and Zheng, X. (2014). Realized volatility when sampling times are possibly endogenous. *Econometric Theory*, 30:580–605.

Mykland, P. A. and Zhang, L. (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica*, 77:1403–1445.

Mykland, P. A. and Zhang, L. (2012). The econometrics of high frequency data. In Kessler, M., Lindner, A., and Sørensen, M., editors, *Statistical Methods for Stochastic Differential Equations*, pages 109–190. CRC Press.

Reid, N. (1994). A conversation with Sir David Cox. *Statistical Science*, 9:439–455.

Williams, D. (1991). *Probability with Martingales.* Cambridge University Press, Cambridge.

Zhang, L. (2001). From martingales to ANOVA: Implied and realized volatility. The University of Chicago, Department of Statistics.

Some additional details on the AR(1) example i

The denominator in the expression for $\widehat{\theta}_n(t)$ is

$$Z_n(t) = \frac{1}{n} \sum_{j/n \le t} X_{j-1}\varepsilon_j.$$

Since $X_0 = 0$, for $j = 1, \dots, n$, we have $X_j = \sum_{i=1}^{j} \varepsilon_i$, and by the Skorokhod embedding, there are stopping times $t_1 \le t_2 \le \cdots$, so that

$$X_j = \sum_{i=1}^{j} \varepsilon_i = B(t_j), \quad \text{and} \quad \varepsilon_j = B(t_j) - B(t_{j-1}), \text{ for } j = 1, 2, \dots,$$

for a Brownian motion B, where we take $t_0 = 0$. Set $W_n(t) = B(nt)/\sqrt{n} \overset{d}{=} B_t$ by Brownian scaling. We can now write

$$\begin{aligned} Z_n(t) &= \sum_{j/n \le t} \frac{1}{\sqrt{n}} B(t_{j-1}) \frac{1}{\sqrt{n}} \{B(t_j) - B(t_{j-1})\} \\ &= \sum_{j/n \le t} W_n(t_{j-1}/n)\{W_n(t_j/n) - W_n(t_{j-1}/n)\} \\ &= \int_0^{t_*/n} W_n(s)\, \mathrm{d}W_n(s) + o_p(1), \end{aligned}$$

Some additional details on the AR(1) example ii

where $t_* = \max\{t_j : t_j \leq nt\}$, and the $o_p(1)$ term is

$$\sum_{j/n \leq t} \int_{t_{j-1}/n}^{t_j/n} \{W_n(s) - W_n(t_{j-1}/n)\} \, \mathrm{d}W_n(s) = o_p(1),$$

By Itô's formula

$$Z_n(t) = \int_0^{t_*/n} W_n(s) \, \mathrm{d}W_n(s) + o_p(1) = \frac{W_n(t_*/n)^2 - t_*/n}{2} + o_p(1).$$

Since $t_*/n \to_p t$, and $t \mapsto W_n(t)$ is continuous,

$$Z_n(t) \xrightarrow{d} \frac{B_t^2 - t}{2}.$$

For the claims in the slides above, it is also important to argue that

$$(n^{-1} \sum_{j/n \leq t} X_{j-1}\varepsilon_j, n^{-2} \sum_{j/n \leq t} X_{j-1}^2),$$

converges jointly. Then finish up the proof of finite-dim. convergence, and check tightness, to get full process convergence.