A limit theorem for likelihoods in the LAQ case

Alexander Gushchin[1] Esko Valkeila[2]

[1]Steklov Mathematical Institute

[2]Aalto University

SAPS VIII, Le Mans, March 21, 2011

Outline

1. **Objectives**
2. Previous results
3. Our result
4. Example

Outline

Outline

1. Objectives
2. Previous results
3. Our result
4. Example

Outline

Outline

Density processes

Let

$$\mathbb{E}^n = (\Omega^n, \mathscr{F}^n, \mathbb{F}^n = (\mathscr{F}^n_t)_{0\leqslant t\leqslant 1}, (\mathsf{P}^{n,\vartheta})_{\vartheta\in\Theta\subseteq\mathbb{R}^k})$$

be a sequence of filtered statistical models, $\mathscr{F}^n = \mathscr{F}^n_1$.
If T is an $\mathbb{F}^n$-stopping time then $\mathsf{P}^{n,\vartheta}_T := \mathsf{P}^{n,\vartheta}\big|_{\mathscr{F}^n_T}$.
For simplicity we shall assume that, for every n, $\mathsf{P}^{n,\vartheta}$ coincide on $\mathscr{F}^n_0$ for different ϑ.
Let a point $\vartheta_0 \in \operatorname{int}\Theta$ and a sequence of nonsingular (normalizing) $k \times k$-matrices $\varphi_n \to 0$ be given.
Let $Z^{n,\vartheta}$ be the density process of $\mathsf{P}^{n,\vartheta_0+\varphi_n\vartheta}$ with respect to $\mathsf{P}^n := \mathsf{P}^{n,\vartheta_0}$, i.e. a càdlàg $\mathbb{F}^n$-adapted process $(Z^{n,\vartheta}_t)_{t\leqslant 1}$ with values in $\mathbb{R}_+$ such that, for any $\mathbb{F}^n$-stopping time T, $Z^{n,\vartheta}_T$ is the density of the absolutely continuous part of $\mathsf{P}^{n,\vartheta_0+\varphi_n\vartheta}_T$ with respect to P^n_T. In general, $Z^{n,\vartheta}$ is a P^n-supermartingale.

Objectives

It is important in statistics to find a normalizing sequence of matrices φ_n such that the random function $\big(Z_1^{n,\vartheta},\ \vartheta \in \varphi_n^{-1}(\Theta - \{\vartheta_0\})\big)$ converges in distribution (wrt P^n) to a nondegenerate random function $(Z_1^{\vartheta},\ \vartheta \in \mathbb{R}^k)$ (at least, finite-dimensional convergence).

We are interested in the case where we have weak convergence of $\big(Z_t^{n,\vartheta},\ t \in [0,1],\ \vartheta \in \varphi_n^{-1}(\Theta - \{\vartheta_0\})\big)$ (wrt P^n) to $(Z_t^{\vartheta},\ t \in [0,1],\ \vartheta \in \mathbb{R}^k)$ (uniform finite-dimensional in ϑ and functional in t), and

$$Z_t^{\vartheta} = \exp\Big(\vartheta^\top M_t - \frac{1}{2}\vartheta^\top \langle M\rangle_t \vartheta\Big),$$

where $M = (M_t)$ is a continuous local martingale and is not a Gaussian or a conditionally Gaussian martingale.

Outline

Jacod and Shiryaev (1987)

$(\Omega^n, \mathscr{F}^n, \mathbb{F}^n, (\mathsf{P}^n, \mathsf{P}'^n))$ sequence of binary experiments
Z^n is the density process of P'^n wrt P^n.

Assumptions on the limiting model

- $\Omega = \mathbb{D}(\mathbb{R}) = \mathbb{D}(\mathbb{R}_+; \mathbb{R})$ is the Skorokhod space with the Borel σ-field $\mathscr{F}$ and the filtration $\mathbb{F} = \mathscr{D}(\mathbb{R})$ generated by the canonical process denoted by X.
- C is an adapted continuous increasing process with $C_0 = 0$, defined on $(\Omega, \mathscr{F}, \mathbb{F})$.
- There is a continuous increasing function $t \rightsquigarrow F_t$ with $F_0 = 0$, such that $F - C(\alpha)$ is nondecreasing for all $\alpha \in \Omega$.
- $\alpha \rightsquigarrow C_t(\alpha)$ is Skorokhod-continuous for all $t \in \mathbb{R}_+$.
- There is a unique probability measure P on $(\Omega, \mathscr{F})$ under which $M := X + C/2$ is a continuous local martingale with $M_0 = 0$ and $\langle M \rangle = C$.

Assumptions on convergence

Under these assumptions there exists a unique measure P' on $(\Omega, \mathscr{F})$ under which $M' := X - C/2$ is a continuous local martingale with $M'_0 = 0$ and $\langle M' \rangle = C$. Moreover, $\mathrm{P}' \overset{\text{loc}}{\ll} P$, and the density process of P' with respect to P is e^X. In particular, e^X is a P-martingale.

Assumptions on convergence

- $h^n_t - \frac{1}{8} C_t \circ \log(Z^n \vee \frac{1}{n}) \xrightarrow{\mathrm{P}^n} 0$, $n \to \infty$, $t \in \mathbb{R}_+$, where h^n is the Hellinger process of order $1/2$ for P^n and P'^n;
- Lindeberg-type condition on the jumps of Z^n;
- $\iota^n_t \xrightarrow{\mathrm{P}^n} 0, \quad n \to \infty$, for all $t \in \mathbb{R}_+$, where ι^n is the Hellinger process of order 0 for P^n and P'^n.

Theorem X.1.59 in Jacod and Shiryaev (1987)

Theorem

Let assumptions on the limiting model be satisfied. If assumptions on convergence hold true, then

$$\mathscr{L}(Z^n \mid \mathsf{P}^n) \xrightarrow{\mathbb{D}(\mathbb{R})} \mathscr{L}(e^X \mid \mathsf{P}), \quad n \to \infty.$$

Conversely, if

$$\mathscr{L}(Z^n \mid \mathsf{P}^n) \xrightarrow{d_f} \mathscr{L}(e^X \mid \mathsf{P}), \quad n \to \infty,$$

then assumptions on convergence are satisfied.

Gushchin and Valkeila (2003)

A similar result was proved in Gushchin and Valkeila (2003). The differences in assumptions on the limiting model are:

- Now the canonical process is denoted by M.
- There is a unique probability measure P on $(\Omega, \mathscr{F})$ under which M is a continuous local martingale with $M_0 = 0$ and $\langle M \rangle = C$.

Under these assumptions there exists a unique measure P' on $(\Omega, \mathscr{F})$ under which $M' := M - C$ is a continuous local martingale with $M'_0 = 0$ and $\langle M' \rangle = C$. Moreover, $\mathrm{P}' \overset{\text{loc}}{\ll} P$, and the density process of P' with respect to P is $e^{M - C/2}$.
The statement of the theorem is essentially the same.

Luschgy (1994)

- A semimartingale $X = (X_t)_{t \in \mathbb{R}_+}$ with predictable characteristics $T(\vartheta) = (B(\vartheta), C, \nu(\vartheta))$ is observed on $[0, T]$. It is quasi-left-continuous. Asymptotics: $T \to \infty$.
- The density process of P^η with respect to P^ϑ is expressed (in a standard way) via characteristics.
- Characteristics are “asymptotically differentiable” in ϑ in accordance with a normalizing sequence $\{\varphi_n\}$, and an “asymptotic score process” (at ϑ) is defined and is a locally square-integrable martingale wrt P^ϑ. As a consequence, this asymptotic score process provides a linear-quadratic approximation for the log-likelihood processes.
- It is assumed that, after rescaling in time, the asymptotic score process weakly converges to a continuous local martingale.

Outline

Prerequisites for the limiting model. 1

- $\Omega = \mathbb{D}(\mathbb{R}^q)$ is the Skorokhod space with the Borel σ-field $\mathscr{F}$ and the filtration $\mathbb{F} = \mathscr{D}(\mathbb{R}^q)$ generated by the canonical process denoted by B.
- $C = (C^{ij})_{i,j\leqslant q}$ is an adapted continuous increasing process with values in the set $\mathbb{M}^q_+$ of all symmetric positive semidefinite $q \times q$ matrices, defined on $(\Omega, \mathscr{F}, \mathbb{F})$, $C_0 = 0$.
- There is a continuous and deterministic increasing function $t \rightsquigarrow F_t$ with $F_0 = 0$, such that $F - \sum_{i=1}^q C^{ii}(\alpha)$ is nondecreasing for all $\alpha \in \Omega$.
- $\alpha \rightsquigarrow C_t(\alpha)$ is Skorokhod-continuous for all $t \in \mathbb{R}_+$.
- There is a unique probability measure P on $(\Omega, \mathscr{F})$ under which B is a continuous local martingale with $B_0 = 0$ and $\langle B \rangle = C$.

Prerequisites for the limiting model. 2

The second component for constructing our limiting model is given by

- There are adapted càdlàg processes $G^n = (G^{n,ij})_{i\leqslant k,\, j\leqslant q}$ and $G = (G^{ij})_{i\leqslant k,\, j\leqslant q}$ with values in $\mathbb{R}^{k\times q}$, defined on $(\Omega, \mathscr{F}, \mathbb{F})$.
- If $\alpha_n \to \alpha \in \mathbb{C}(\mathbb{R}^q)$ in the Skorokhod topology on $\mathbb{D}(\mathbb{R}^q)$ (i.e. locally uniformly), then $G^n(\alpha_n) \to G(\alpha)$ in the Skorokhod topology on $\mathbb{D}(\mathbb{R}^{k\times q})$.

Some notation

There is a predictable process $c = (c^{ij})_{i,j\leqslant q}$ with values in $\mathbb{M}_{+}^{q}$, defined on $(\Omega, \mathscr{F}, \mathbb{F})$, such that $C_t^{ij} = c^{ij} \cdot F_t$, $i, j \leqslant q$, and $\operatorname{trace}(c_t) \leqslant 1$ for all $t \in \mathbb{R}_+$ P-a.s.

In what follows we use the notation $G^n \circ B^n$, which is understood as the composition of the mappings $B^n\colon \Omega^n \to \Omega$ and $G^n\colon \Omega \to \mathbb{D}(\mathbb{R}^{k\times q})$. Note that $G \circ B = G$.

For a $(k \times q)$-dimensional adapted càdlàg process $H = (H^{ij})_{i\leqslant k,\, j\leqslant q}$ and a q-dimensional locally square-integrable martingale $N = (N^j)_{j\leqslant q}$ the process $Y := H_- \cdot N$ is understood as a k-dimensional locally square-integrable martingale $Y = (Y^i)_{i\leqslant k}$ such that $Y^i = \sum_{j=1}^{q} H_-^{ij} \cdot N^j$.

Limiting process

Now we are in a position to describe our limiting process. Put

$$K := G \circ B = G$$

and

$$M := K_- \cdot B,$$

then M is a k-dimensional locally square-integrable martingale with the quadratic characteristic

$$\langle M \rangle = (G_- c G_-^\top) \cdot F.$$

Now we define the limiting process Z^ϑ by

$$Z_t^\vartheta = \exp\Big(\vartheta^\top M_t - \frac{1}{2}\vartheta^\top \langle M \rangle_t \vartheta\Big).$$

Some extra notation. 1

Recall that we deal with a sequence

$$\mathbb{E}^n = (\Omega^n, \mathscr{F}^n, \mathbb{F}^n = (\mathscr{F}^n_t)_{t\leqslant 1}, (\mathsf{P}^{n,\vartheta})_{\vartheta\in\Theta\subseteq\mathbb{R}^k})$$

of filtered statistical models, and $Z^{n,\vartheta}$ is the density process of $\mathsf{P}^{n,\vartheta_0+\varphi_n\vartheta}$ with respect to $\mathsf{P}^n := \mathsf{P}^{n,\vartheta_0}$.
Put $Y^{n,\vartheta} := \sqrt{Z^{n,\vartheta}}$, then the process $Y^{n,\vartheta}$ is a P^n-supermartingale.
Define now processes $y^{n,\vartheta}$, $m^{n,\vartheta}$, and $h^{n,\vartheta}$ on the predictable interval $\Gamma^{n,\vartheta} := \{Z^{n,\vartheta}_- > 0\}$ by:

$$y^{n,\vartheta} := (1/Y^{n,\vartheta}_-) \cdot Y^{n,\vartheta},$$

Some extra notation. 2

$m^{n,\vartheta}$ and $h^{n,\vartheta}$ are a local martingale and a predictable increasing process respectively in the Doob–Meyer decomposition

$$y^{n,\vartheta} = m^{n,\vartheta} - h^{n,\vartheta}$$

of a local supermartingale $y^{n,\vartheta}$. The process $m^{n,\vartheta}$ is, in fact, a P^n-locally square-integrable martingale on $\Gamma^{n,\vartheta}$, and $h^{n,\vartheta}$ is the Hellinger processes $h(\frac{1}{2}; P^n, P^{n,\vartheta})$ of order $1/2$ for P^n and $P^{n,\vartheta}$.

Assumptions on convergence. 1

In the theorem below we assume also that

- For every n there is given a locally square-integrable martingale $B^n = (B^{n,j})_{j \leqslant q}$, $B^n_0 = 0$, on $(\Omega^n, \mathscr{F}^n, \mathbb{F}^n, P^n)$ with values in $\mathbb{R}^q$.
- $$\langle B^n, B^n \rangle_t - C_t \circ B^n \xrightarrow{\mathrm{P}^n} 0, \quad n \to \infty, \quad \text{for all} \quad t \in S,$$
  where S is a dense subset of $[0, 1]$ containing 0 and 1.
- $$\|x\|^2 1_{\{\|x\| > \varepsilon\}} \star \nu_1^{B^n} \xrightarrow{\mathrm{P}^n} 0, \quad n \to \infty, \quad \text{for all } \varepsilon > 0,$$

Assumptions on convergence. 2

- Let $K^n := G^n \circ B^n$ and $M^n := K^n_- \cdot B^n$. Then

$$\left\langle m^{n,\vartheta_n} - \frac{1}{2}\vartheta_n^\top M^n \right\rangle_1 \xrightarrow{\mathsf{P}^n} 0, \quad n \to \infty,$$

for each bounded sequence $\{\vartheta_n\}$.

Main result

Theorem

Let all the above assumptions be satisfied. Then, as $n \to \infty$,

$$\mathscr{L}(M^n, \langle M^n \rangle | \mathrm{P}^n) \longrightarrow \mathscr{L}(M, \langle M \rangle | \mathrm{P}), \quad \text{in } \mathbb{D}([0,1], \mathbb{R}^k),$$

$$(P^n) \lhd (P^{n,\vartheta_n}),$$

and

$$\sup_{t \leqslant 1} \left| \log Z_t^{n,\vartheta_n} - \left(\vartheta_n^\top M_t^n - \frac{1}{2} \vartheta_n^\top \left\langle M^n \right\rangle_t \vartheta_n^\top \right) \right| \xrightarrow{\mathrm{P}^n} 0,$$

for each bounded sequence $\{\vartheta_n\}$.

Additional remarks

If G is a Gaussian process on $(\Omega, \mathscr{F}, P)$, then

$$E \exp\Big(\vartheta^\top M_1 - \frac{1}{2}\vartheta^\top \langle M \rangle_1 \vartheta\Big) = 1$$

for any $\vartheta \in \mathbb{R}^k$ and

$$(P^n) \lhd \rhd (P^{n,\vartheta_n})$$

for each bounded sequence $\{\vartheta_n\}$. In particular, this is true if B is a Gaussian martingale (equivalently, C is deterministic) and G is a linear transformation on the corresponding set of continuous functions.
If, additionally, the matrix $\langle M \rangle_t$ is P-a.s. nonsingular, then the family $P_t^{n,\vartheta}$ is locally asymptotically quadratic (LAQ) at ϑ_0.

Outline

AR(p) model

Let us consider an autoregressive AR(p) model

$$y_n = \vartheta_1 y_{n-1} + \cdots + \vartheta_p y_{n-p} + \varepsilon_n,$$

where $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_p)^\top$ is an unknown parameter, ε_n are i.i.d. random variables with $\mathsf{E}\varepsilon_n = 0$ and $\mathsf{E}\varepsilon_n^2 = 1$. We also assume that ε_n have a Lebesgue density $f(x)$ with a finite Fisher information $\boldsymbol{I}$. More precisely, there is a function $v(x)$ (the score function) such that

$$\boldsymbol{I} := \int v^2(x) f(x)\, dx < \infty$$

and

$$\int \Big(\sqrt{f(x-u)} - \sqrt{f(x)} - \frac{1}{2} v(x) \sqrt{f(x)} \Big)^2 dx \to 0, \quad , u \to 0.$$

Multiple unit root

Note that

$$\int xv(x)f(x)\,dx = 1 \quad \text{and} \quad \boldsymbol{I} \geqslant 1.$$

Let

$$\phi(z) = \phi(z;\boldsymbol{\vartheta}) := 1 - \vartheta_1 z - \cdots - \vartheta_p z^p.$$

We assume that the true value $\boldsymbol{\vartheta}_0$ of the parameter is such that

$$\phi(z;\boldsymbol{\vartheta}_0) = (1-z)^p.$$

Prerequisites for asymptotic analysis. 1

Put

$$\boldsymbol{y}_n = (y_n, y_{n-1}, \dots, y_{n-p+1})^\top.$$

Let $\mathbb{B}$ be the backshift operator. We assume that ϑ_0 is a true value of the parameter. Then

$$(1-\mathbb{B})^p y_n = \varepsilon_n.$$

Put

$$y_n(j) = (1-\mathbb{B})^{p-j} y_n, \quad j = 0, \dots, p,$$

$$U_n = \big(y_n(p), \dots, y_n(1)\big)^\top.$$

Prerequisites for asymptotic analysis. 2

The introduced objects are connected by

$$M\boldsymbol{y}_n = U_n,$$

where

$$M := \begin{pmatrix} 1 & 0 & \multicolumn{3}{c}{\dots\dots\dots\dots\dots\dots\dots} & 0 \\ 1 & -1 & 0 & & \dots & 0 \\ \multicolumn{6}{c}{\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots} \\ 1 & -\binom{p-1}{1} & \dots & & (-1)^{p-2}\binom{p-1}{p-2} & (-1)^{p-1} \end{pmatrix},$$

and

$$y_n(j-1) = (1-\mathbb{B})y_n(j) = y_n(j) - y_{n-1}(j),$$

hence

$$y_n(j) = \sum_{k=1}^{n} y_k(j-1), \quad j = 1, \dots, p.$$

Normalizing matrix

Take

$$J_n := \begin{pmatrix} n^{-p} & 0 & \dots & 0 \\ 0 & n^{1-p} & \dots & 0 \\ \hdotsfor{4} \\ 0 & 0 & \dots & n^{-1} \end{pmatrix} M, \quad \varphi_n := J_n^\top.$$

Here

$$Z_t^{n,\boldsymbol{\vartheta}} = \frac{d\mathrm{P}_{[nt]}^{n,\boldsymbol{\vartheta}_0+\varphi_n\boldsymbol{\vartheta}}}{d\mathrm{P}_{[nt]}^{n,\boldsymbol{\vartheta}_0}}.$$

Approximation for the likelihood

It can be shown that

$$\sup_{t\leqslant 1}\left|\log Z_t^{n,\vartheta_n} - \left(\vartheta_n^\top M_t^n - \frac{1}{2}\vartheta_n^\top \langle M^n\rangle_t \vartheta_n^\top\right)\right| \xrightarrow{\mathsf{P}^n} 0,$$

where

$$\begin{aligned} M_t^n &= J_n \sum_{k=1}^{[nt]} \boldsymbol{y}_{k-1} v(\varepsilon_n) \\ &= \sum_{k=1}^{[nt]} n^{1/2-j} U_{k-1}(j) \frac{v(\varepsilon_k)}{n^{1/2}}. \end{aligned}$$

Construction of B^n, B.

This suggests to take $q = 2$,

$$B^n(1) = \sum_{k=1}^{[nt]} \frac{v(\varepsilon_k)}{n^{1/2}}, \quad B^n(2) = \sum_{k=1}^{[nt]} \frac{\varepsilon_k}{n^{1/2}},$$

which converges to a two-dimensional Gaussian martingale $B = (B(1), B(2))^\top$ with the quadratic characteristic

$$\langle B \rangle_t = \begin{pmatrix} \boldsymbol{I} & 1 \\ 1 & 1 \end{pmatrix} t.$$

Construction of G^n, G.

Next, put for

$$G^n(i,2) \equiv 0, \quad i = 1, \dots, p,$$

$$H^n(\alpha) = \int_0^t \alpha\Big(\frac{[ns]}{n}\Big)\, ds,$$

$$G^n(1,1)(\alpha) = \alpha,$$

$$G^n(i,1) = H^n \circ G^n(i,1), \quad i = 2, \dots, p.$$

Then

$$M^n = K^n_- \cdot B^n,$$

The limiting process

thus

$$M = \Big(M(i),\ i = p, p-1, \dots, 1\Big)$$

with

$$M(i)_t = \int_0^t F(i)_s \, dB(1)_s, \quad i = 1, \dots, p,$$

$$F(1) = B(2),$$

$$F(i)_t = \int_0^t F(i-1)_s \, ds, \quad i = 2, \dots, p.$$

Jacod, J. and Shiryaev, A. N.
Limit Theorems for Stochastic Processes,
Springer-Verlag, Berlin, 1987.

Luschgy, H.
Asymptotic inference for semimartingale models with singular parameter points.
J. Statist. Plann. Inference, 39 (1994), no. 2, 155–186.

Gushchin, A. A. and Valkeila, E.
Approximations and limit theorems for likelihood ratio processes in the binary case.
Statist. Decisions, 21 (2003), no. 3, 219–260.

Gushchin, A. A. and Valkeila, E.
Quadratic approximation for log-likelihood ratio processes.
In preparation.