Asymptotic results discrete time martingales and stochastic algorithms

Bernard Bercu

Bordeaux University, France

IFCAM Summer School

Bangalore, India, July 2015

Outline

Outline

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space with a filtration $\mathbb{F} = (\mathcal{F}_n)$ where $\mathcal{F}_n$ is the σ-algebra of events occurring up to time n.

Definition

Let (M_n) be a sequence of integrable random variables defined on $(\Omega, \mathcal{A}, \mathbb{P})$ such that, for all $n \geqslant 0$, M_n is $\mathcal{F}_n$-measurable.

1. (M_n) is a martingale **MG** if for all $n \geqslant 0$,
$$\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] = M_n \qquad \text{a.s.}$$
2. (M_n) is a submartingale **sMG** if for all $n \geqslant 0$,
$$\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] \geqslant M_n \qquad \text{a.s.}$$
3. (M_n) is a supermartingale **SMG** if for all $n \geqslant 0$,
$$\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] \leqslant M_n \qquad \text{a.s.}$$

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space with a filtration $\mathbb{F} = (\mathcal{F}_n)$ where $\mathcal{F}_n$ is the σ-algebra of events occurring up to time n.

Definition

Let (M_n) be a sequence of integrable random variables defined on $(\Omega, \mathcal{A}, \mathbb{P})$ such that, for all $n \geqslant 0$, M_n is $\mathcal{F}_n$-measurable.

1. (M_n) is a martingale **MG** if for all $n \geqslant 0$,

$$\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] = M_n \qquad \text{a.s.}$$

2. (M_n) is a submartingale **sMG** if for all $n \geqslant 0$,

$$\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] \geqslant M_n \qquad \text{a.s.}$$

3. (M_n) is a supermartingale **SMG** if for all $n \geqslant 0$,

$$\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] \leqslant M_n \qquad \text{a.s.}$$

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space with a filtration $\mathbb{F} = (\mathcal{F}_n)$ where $\mathcal{F}_n$ is the σ-algebra of events occurring up to time n.

Definition

Let (M_n) be a sequence of integrable random variables defined on $(\Omega, \mathcal{A}, \mathbb{P})$ such that, for all $n \geqslant 0$, M_n is $\mathcal{F}_n$-measurable.

1. (M_n) is a martingale **MG** if for all $n \geqslant 0$,
$$\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] = M_n \qquad \text{a.s.}$$
2. (M_n) is a submartingale **sMG** if for all $n \geqslant 0$,
$$\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] \geqslant M_n \qquad \text{a.s.}$$
3. (M_n) is a supermartingale **SMG** if for all $n \geqslant 0$,
$$\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] \leqslant M_n \qquad \text{a.s.}$$

Martingales with sums

Example (Sums)

Let (X_n) be a sequence of integrable and independent random variables such that, for all $n \geqslant 1$, $\mathbb{E}[X_n] = m$. Denote

$$S_n = \sum_{k=1}^{n} X_k.$$

We clearly have

$$S_{n+1} = S_n + X_{n+1}.$$

Consequently, (S_n) is a sequence of integrable random variables with

$$\begin{aligned} \mathbb{E}[S_{n+1} \mid \mathcal{F}_n] &= S_n + \mathbb{E}[X_{n+1} \mid \mathcal{F}_n], \\ &= S_n + \mathbb{E}[X_{n+1}], \\ &= S_n + m \end{aligned}$$

Martingales with sums

Example (Sums)

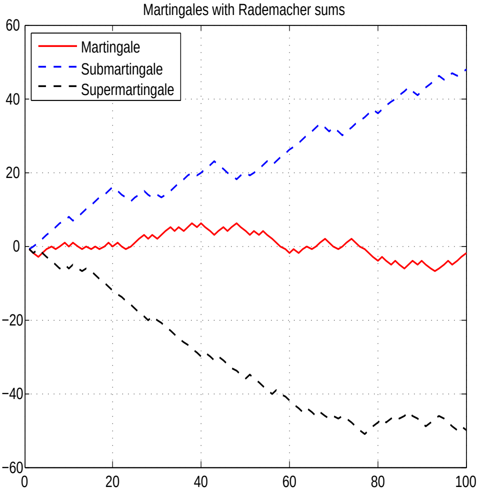
$$\mathbb{E}[S_{n+1} \mid \mathcal{F}_n] = S_n + m.$$

- (S_n) is a **martingale** if $m = 0$,
- (S_n) is a **submartingale** if $m \geqslant 0$,
- (S_n) is a **supermartingale** if $m \leqslant 0$.

$\longrightarrow$ It holds for Rademacher $\mathcal{R}(p)$ distribution with $0 < p < 1$ where

$$m = 2p - 1.$$

Martingales with Rademacher sums



Martingales with products

Example (Products)

Let (X_n) be a sequence of positive, integrable and independent random variables such that, for all $n \geqslant 1$, $\mathbb{E}[X_n] = m$. Denote

$$P_n = \prod_{k=1}^{n} X_k.$$

We clearly have

$$P_{n+1} = P_n X_{n+1}.$$

Consequently, (P_n) is a sequence of integrable random variables with

$$\begin{aligned}
\mathbb{E}[P_{n+1} \mid \mathcal{F}_n] &= P_n \mathbb{E}[X_{n+1} \mid \mathcal{F}_n], \\
&= P_n \mathbb{E}[X_{n+1}], \\
&= m P_n
\end{aligned}$$

Martingales with products

Example (Products)

$$\mathbb{E}[P_{n+1} \mid \mathcal{F}_n] = mP_n.$$

- (P_n) is a **martingale** if $m = 1$,
- (P_n) is a **submartingale** if $m \geqslant 1$,
- (P_n) is a **supermartingale** if $m \leqslant 1$.

$\longrightarrow$ It holds for Exponential $\mathcal{E}(\lambda)$ distribution with $\lambda > 0$ where

$$m = \frac{1}{\lambda}.$$

Stability

Theorem (Stability)

1. *If* (M_n) *is a* **SMG**, *then* $(-M_n)$ *is a* **sMG**.
2. *If* (M_n) *and* (N_n) *are two* **sMG** *and*

$$S_n = \sup(M_n, N_n)$$

$\longrightarrow$ (S_n) *is a* **sMG**.

3. *If* (M_n) *and* (N_n) *are two* **SMG** *and*

$$I_n = \inf(X_n, Y_n)$$

$\longrightarrow$ (I_n) *is a* **SMG**.

Stability, continued

Theorem (Stability)

1. *If* (M_n) *and* (N_n) *are two* **MG**, $a, b \in \mathbb{R}$ *and*

$$S_n = aM_n + bN_n$$

$\longrightarrow$ (S_n) *is a* **MG**.

2. *If* (M_n) *is a* **MG** *and* F *is a convex real function such that, for all* $n \geqslant 1$, $F(M_n) \in L^1(\mathbb{R})$ *and if*

$$F_n = F(M_n)$$

$\longrightarrow$ (F_n) *is a* **sMG**.

Doob's convergence theorem

- Every bounded above increasing sequence converges to its supremum,
- Every bounded bellow decreasing sequence converges to its infimum.

$\longrightarrow$ The stochastic analogous of this result is due to Doob.

Theorem (Doob)

1. *If (M_n) is a* **sMG** *bounded* **above** *by some constant M, then (M_n) converges a.s.*
2. *If (M_n) is a* **SMG** *bounded* **below** *by some constant m, then (M_n) converges a.s.*

Doob's convergence theorem, continued

Theorem (Doob)

Let (M_n) *be a* **MG**, **sMG**, *or* **SMG** *bounded in* $\mathbb{L}^1$ *which means*

$$\sup_{n \geqslant 0} \mathbb{E}[|M_n|] < +\infty.$$

$\longrightarrow$ (M_n) *converges a.s. to an integrable random variable* M_∞.

Joseph Leo Doob

Convergence of martingales

Theorem

Let (M_n) be a **MG** *bounded in $\mathbb{L}^p$ with $p \geqslant 1$, which means that*

$$\sup_{n \geqslant 0} \mathbb{E}[|M_n|^p] < +\infty.$$

1. *If $p > 1$, (M_n) converges a.s. a random variable M_∞. The convergence is also true in $\mathbb{L}^p$.*
2. *If $p = 1$, (M_n) converges a.s. to a random variable M_∞. The convergence holds in $\mathbb{L}^1$ as soon as (M_n) is uniformly integrable that is*

$$\lim_{a \to \infty} \sup_{n \geqslant 0} \mathbb{E}\left[|M_n| \mathrm{I}_{\{|M_n| \geqslant a\}}\right] = 0.$$

Chow's Theorem

Theorem (Chow)

Let (M_n) *be a* **MG** *such that for* $1 \leqslant a \leqslant 2$ *and for all* $n \geqslant 1$,

$$\mathbb{E}[|M_n|^a] < \infty.$$

Denote, for all $n \geqslant 1$, $\Delta M_n = M_n - M_{n-1}$ *and assume that*

$$\sum_{n=1}^{\infty} \mathbb{E}[|\Delta M_n|^a | \mathcal{F}_{n-1}] < \infty \qquad a.s.$$

$\longrightarrow$ (M_n) *converges a.s. to a random variable* M_∞.

Exponential Martingale

Example (Exponential Martingale)

Let (X_n) be a sequence of independent random variable sharing the same $\mathcal{N}(0,1)$ distribution. For all $t \in \mathbb{R}^*$, let $S_n = X_1 + \cdots + X_n$ and denote

$$M_n(t) = \exp\Big(tS_n - \frac{nt^2}{2}\Big).$$

It is clear that $(M_n(t))$ is a **MG** which converges a.s. to zero. However, $\mathbb{E}[M_n(t)] = \mathbb{E}[M_1(t)] = 1$ which means that $(M_n(t))$ does not converge in $\mathbb{L}^1$.

Autoregressive Martingale

Example (Autoregressive Martingale)

Let (X_n) be the autoregressive process given for all $n \geqslant 0$ by

$$\boldsymbol{X_{n+1} = \theta X_n + (1-\theta)\varepsilon_{n+1}}$$

where $X_0 = p$ with $0 < p < 1$ and the parameter $0 < \theta < 1$. Assume that $\mathcal{L}(\varepsilon_{n+1}|\mathcal{F}_n)$ is the Bernoulli $\mathcal{B}(X_n)$ distribution.
We can show that $0 < X_n < 1$ and (X_n) is a **MG** such that

$$\lim_{n\to\infty} X_n = X_\infty \qquad \text{a.s.}$$

The convergence also holds in $\mathbb{L}^p$ for all $p \geqslant 1$. Finally, X_∞ has the Bernoulli $\mathcal{B}(p)$ distribution.

Stopping time theorem

Definition

We shall say that a random variable T is a **stopping time** if T takes its values in $\mathbb{N} \cup \{+\infty\}$ and, for all $n \geqslant 0$, the event

$$\{T = n\} \in \mathcal{F}_n.$$

Theorem

Assume that (M_n) *is a* **MG** *and let* T *be a* **stopping time** *adapted to* $\mathbb{F} = (\mathcal{F}_n)$*. Then,* $(M_{n \wedge T})$ *is also a* **MG**.

Proof of the stopping time theorem

Proof.

First of all, it is clear that for all $n \geqslant 0$, $(M_{n \wedge T})$ is integrable as

$$M_{n \wedge T} = M_T \mathrm{I}_{\{T<n\}} + M_n \mathrm{I}_{\{T \geqslant n\}}.$$

In addition, $\{T \geqslant n\} \in \mathcal{F}_{n-1}$ as **its complementary** $\{T < n\} \in \mathcal{F}_{n-1}$. Then, for all $n \geqslant 0$,

$$\begin{aligned}
\mathbb{E}[M_{(n+1)\wedge T} | \mathcal{F}_n] &= \mathbb{E}[M_T \mathrm{I}_{\{T<n+1\}} + M_{n+1} \mathrm{I}_{\{T \geqslant n+1\}} | \mathcal{F}_n], \\
&= M_T \mathrm{I}_{\{T<n+1\}} + \mathrm{I}_{\{T \geqslant n+1\}} \mathbb{E}[M_{n+1} | \mathcal{F}_n], \\
&= M_T \mathrm{I}_{\{T<n+1\}} + M_n \mathrm{I}_{\{T \geqslant n+1\}}, \\
&= M_T \mathrm{I}_{\{T<n\}} + M_n \mathrm{I}_{\{T=n\}} + M_n \mathrm{I}_{\{T \geqslant n\}} - M_n \mathrm{I}_{\{T=n\}}, \\
&= M_T \mathrm{I}_{\{T<n\}} + M_n \mathrm{I}_{\{T \geqslant n\}}, \\
&= M_{n \wedge T}.
\end{aligned}$$

□

Kolmogorov's inequality

Theorem (Kolmogorov's inequality)

Assume that (M_n) *is a* **MG**. *Then, for all* $a > 0$,

$$\mathbb{P}(M_n^{\#} > a) \leqslant \frac{1}{a}\mathbb{E}\big[|M_n| \mathrm{I}_{\{M_n^{\#} > a\}}\big]$$

where

$$M_n^{\#} = \max_{0 \leqslant k \leqslant n} |M_k|.$$

As (M_n) is a **MG**, we clearly have that $(|M_n|)$ is a **sMG**. The proof relies on the entry time T_a of the **sMG** $(|M_n|)$ into the interval $[a, +\infty[$,

$$T_a = \inf\{n \geqslant 0, |M_n| \geqslant a\}.$$

Proof.

First of all, we clearly have for all $n \geqslant 0$,

$$\{T_a \leqslant n\} = \{\max_{0 \leqslant k \leqslant n} |M_k| \geqslant a\} = \{M_n^\# > a\}.$$

Since $|M_{T_a}| \geqslant a$, it leads to

$$\mathbb{P}(M_n^\# > a) = \mathbb{P}(T_a \leqslant n) = \mathbb{E}\left[\mathrm{I}_{\{T_a \leqslant n\}}\right] \leqslant \frac{1}{a}\mathbb{E}\left[|M_{T_a}|\mathrm{I}_{\{T_a \leqslant n\}}\right].$$

However, we have for all $k \leqslant n$, $|M_k| \leqslant \mathbb{E}[|M_n||\mathcal{F}_k]$ a.s. Therefore,

$$\begin{aligned}
\mathbb{E}\left[|M_{T_a}|\mathrm{I}_{\{T_a \leqslant n\}}\right] &= \sum_{k=0}^{n} \mathbb{E}\left[|M_k|\mathrm{I}_{\{T_a = k\}}\right] \leqslant \sum_{k=0}^{n} \mathbb{E}\left[\mathbb{E}\left[|M_n||\mathcal{F}_k\right]\mathrm{I}_{\{T_a = k\}}\right], \\
&\leqslant \sum_{k=0}^{n} \mathbb{E}\left[|M_n|\mathrm{I}_{\{T_a = k\}}\right] = \mathbb{E}\left[|M_n|\mathrm{I}_{\{T_a \leqslant n\}}\right],
\end{aligned}$$

which completes the proof of Kolmogorov's inequality. □

Doob's inequality

Theorem (Doob's inequality)

Assume that (M_n) is a **MG** *bounded in $\mathbb{L}^p$ with $p > 1$. Then, we have*

$$\mathbb{E}\big[|M_n|^p\big] \leqslant \mathbb{E}\big[(M_n^{\#})^p\big] \leqslant \Big(\frac{p}{p-1}\Big)^p \mathbb{E}\big[|M_n|^p\big].$$

In particular, for $p = 2$,

$$\mathbb{E}\big[|M_n|^2\big] \leqslant \mathbb{E}\big[(M_n^{\#})^2\big] \leqslant 4\mathbb{E}\big[|M_n|^2\big].$$

The proof relies on the elementary fact that for any positive random variable X and for all $p \geqslant 1$,

$$\mathbb{E}\big[X^p\big] = \int_0^{\infty} pa^{p-1}\mathbb{P}(X > a)\,da.$$

Proof of Doob's inequality

Proof.

It follows from Kolmogorov's inequality and Fubini's theorem that

$$\begin{aligned}
\mathbb{E}\big[(M_n^{\#})^p\big] &= \int_0^\infty pa^{p-1}\mathbb{P}\big(M_n^{\#} > a\big)\,da, \\
&\leqslant \int_0^\infty pa^{p-2}\mathbb{E}\big[|M_n|\mathrm{I}_{\{M_n^{\#}>a\}}\big]\,da, \\
&= \mathbb{E}\Big[|M_n|\int_0^\infty pa^{p-2}\mathrm{I}_{\{M_n^{\#}>a\}}\,da\Big], \\
&= \Big(\frac{p}{p-1}\Big)\mathbb{E}\big[|M_n|(M_n^{\#})^{p-1}\big].
\end{aligned}$$

Finally, via Holder's inequality,

$$\mathbb{E}\big[|M_n|(M_n^{\#})^{p-1}\big] \leqslant \Big(\mathbb{E}\big[|M_n|^p\big]\Big)^{1/p}\Big(\mathbb{E}\big[(M_n^{\#})^p\big]\Big)^{(p-1)/p}$$

which completes the proof of Doob's inequality. □

Outline

We start with two useful lemmas in stochastic analysis.

Lemma (Toeplitz)

Let (a_n) *be a sequence of positive real numbers satisfying*

$$\sum_{n=1}^{\infty} a_n = +\infty.$$

In addition, let (x_n) *be a sequence of real numbers such that*

$$\lim_{n \to \infty} x_n = x.$$

Then, we have

$$\lim_{n \to \infty} \left(\sum_{k=1}^{n} a_k \right)^{-1} \sum_{k=1}^{n} a_k x_k = x.$$

Kronecker's Lemma

Lemma (Kronecker)

Let (a_n) be a sequence of positive real numbers strictly increasing to infinity. Moreover, let (x_n) be a sequence of real numbers such that

$$\sum_{n=1}^{\infty} \frac{x_n}{a_n} = \ell$$

exists and is finite. Then, we have

$$\lim_{n \to \infty} a_n^{-1} \sum_{k=1}^{n} x_k = 0.$$

Increasing process

Definition

Let (M_n) be a square integrable **MG** that is for all $n \geqslant 1$,

$$\mathbb{E}[M_n^2] < \infty.$$

The **increasing process** associated with (M_n) is given by $< M >_0 = 0$ and, for all $n \geqslant 1$,

$$< M >_n = \sum_{k=1}^{n} \mathbb{E}[\Delta M_k^2 | \mathcal{F}_{k-1}]$$

where $\Delta M_k = M_k - M_{k-1}$.

$\longrightarrow$ If (M_n) is a square integrable MG and $N_n = M_n^2 - < M >_n$, then (N_n) is a MG.

Example (Increasing Process)

Let (X_n) be a sequence of square integrable and independent random variables such that, for all $n \geqslant 1$, $\mathbb{E}[X_n] = m$ and $\text{Var}(X_n) = \sigma^2 > 0$. Denote

$$M_n = \sum_{k=1}^{n} (X_k - m)$$

Then, (M_n) is a martingale and its increasing process

$$< M >_n = \sigma^2 n.$$

Moreover, if $N_n = M_n^2 - \sigma^2 n$, (N_n) is a MG.

Theorem (Robbins-Siegmund)

Let (V_n), (A_n) and (B_n) be three positive sequences adapted to $\mathbb{F} = (\mathcal{F}_n)$. Assume that V_0 is integrable and, for all $n \geqslant 0$,

$$\mathbb{E}[V_{n+1}|\mathcal{F}_n] \leqslant V_n + A_n - B_n \qquad a.s.$$

Denote

$$\Gamma = \Big\{ \sum_{n=0}^{\infty} A_n < +\infty \Big\}.$$

1. *On Γ, (V_n) converges a.s. to a finite random variable V_∞.*
2. *On Γ, we also have*

$$\sum_{n=0}^{\infty} B_n < +\infty \qquad a.s.$$

$\longrightarrow$ If $A_n = 0$ and $B_n = 0$, then (V_n) is a positive SMG which converges a.s. to V_∞ thanks to Doob's theorem.

Proof.

For all $n \geqslant 1$, denote

$$\boldsymbol{M_n = V_n - \sum_{k=0}^{n-1}(A_k - B_k).}$$

We clearly have, for all $n \geqslant 0$, $\mathbb{E}[M_{n+1}|\mathcal{F}_n] \leqslant M_n$. For any positive a, let T_a be the stopping time

$$T_a = \inf\Big\{n \geqslant 0, \sum_{k=0}^{n}(A_k - B_k) \geqslant a\Big\}.$$

We deduce from the stopping time theorem that $(M_{n\wedge T_a})$ is a SMG bounded below by $-a$. It follows from Doob's theorem that $(M_{n\wedge T_a})$ converges a.s. to M_∞. Consequently, on the set $\{T_a = +\infty\}$, (M_n) converges a.s. to M_∞. In addition, we also have

$$M_{n+1} + \sum_{k=0}^{n} A_k = V_{n+1} + \sum_{k=0}^{n} B_k \geqslant \sum_{k=0}^{n} B_k.$$

Proof of Robbins-Siegmund's theorem, continued

Proof.

Hence, on the set $\Gamma \cap \{T_a = +\infty\}$, we obtain that

$$\sum_{n=0}^{\infty} B_n < +\infty \qquad \text{a.s.}$$

and (V_n) converges a.s. to a finite random variable V_∞. Finally, as (B_n) is a sequence of positive random variables, we have on Γ,

$$\sum_{k=0}^{n} (A_k - B_k) \leqslant \sum_{k=0}^{n} A_k < +\infty \qquad \text{a.s.}$$

It means that

$$\Gamma \subset \bigcup_{p=0}^{\infty} \{T_p = +\infty\}, \qquad \Gamma = \bigcup_{p=0}^{\infty} \Gamma \bigcap \{T_p = +\infty\}$$

which completes the proof of Robbins-Siegmund's theorem. □

Corollary

Let (V_n), (A_n) and (B_n) be three positive sequences adapted to $\mathbb{F} = (\mathcal{F}_n)$. Let (a_n) be a positive increasing sequence adapted to $\mathbb{F} = (\mathcal{F}_n)$. Assume that V_0 is integrable and, for all $n \geqslant 0$,

$$\mathbb{E}[V_{n+1}|\mathcal{F}_n] \leqslant V_n + A_n - B_n \qquad \text{a.s.}$$

Denote

$$\Lambda = \Big\{ \sum_{n=0}^{\infty} \frac{A_n}{a_n} < +\infty \Big\}.$$

1. *On $\Gamma \cap \{a_n \longrightarrow a_\infty\}$, (V_n) converges a.s. to V_∞.*
2. *On $\Gamma \cap \{a_n \longrightarrow +\infty\}$, $V_n = o(a_n)$ a.s., $V_{n+1} = o(a_n)$ a.s. and*

$$\sum_{k=0}^{n} B_k = o(a_n) \qquad \text{a.s.}$$

$\longrightarrow$ This result is the keystone for the SLLN for martingales.

Strong law of large numbers for martingales

Theorem (Strong Law of large numbers)

Let (M_n) *be a square integrable* **MG** *and denote by* $< M >_n$ *its increasing process.*

1. *On* $\{< M >_n \longrightarrow < M >_\infty\}$, (M_n) *converges a.s. to a square integrable random variable* M_∞.
2. *On* $\{< M >_n \longrightarrow +\infty\}$, *we have*

$$\lim_{n \to \infty} \frac{M_n}{< M >_n} = 0 \qquad a.s.$$

More precisely, for any positive γ,

$$\left(\frac{M_n}{< M >_n}\right)^2 = o\left(\frac{(\log < M >_n)^{1+\gamma}}{< M >_n}\right) \qquad a.s.$$

$\longrightarrow$ If it exists a positive sequence (a_n) increasing to infinity such that $< M >_n = O(a_n)$, then we have $M_n = o(a_n)$ a.s.

Easy example

Let (X_n) be a sequence of square integrable and independent random variables such that, for all $n \geqslant 1$, $\mathbb{E}[X_n] = m$ and $\mathrm{Var}(X_n) = \sigma^2 > 0$. We already saw that

$$M_n = \sum_{k=1}^{n} (X_k - m)$$

is square integrable **MG** with $< M >_n = \sigma^2 n$. It follows from the **SLLN** for martingales that $M_n = o(n)$ a.s. which means that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} X_k = m \qquad \text{a.s.}$$

More precisely, for any positive γ,

$$\left(\frac{M_n}{n}\right)^2 = \left(\frac{1}{n} \sum_{k=1}^{n} X_k - m\right)^2 = o\left(\frac{(\log n)^{1+\gamma}}{n}\right) \qquad \text{a.s.}$$

Proof of the strong Law of large numbers

Proof.

For any positive a, let T_a be the stopping time

$$T_a = \inf\Big\{n \geqslant 0, <M>_{n+1} \geqslant a\Big\}.$$

It follows from the stopping time theorem that $(M_{n\wedge T_a})$ is a MG. It is bounded in $\mathbb{L}^2$ as

$$\sup_{n\geqslant 0} \mathbb{E}[(M_{n\wedge T_a})^2] = \sup_{n\geqslant 0} \mathbb{E}[<M>_{n\wedge T_a}] < a.$$

We deduce from Doob's convergence theorem that $(M_{n\wedge T_a})$ converges a.s. to a square integrable random variable M_∞. Hence, on the set $\{T_a = +\infty\}$, (M_n) converges a.s. to M_∞. However,

$$\{<M>_\infty < +\infty\} = \bigcup_{p=1}^{\infty} \{T_p = +\infty\}$$

which completes the proof of the first part of the theorem. □

Proof.

Let $V_n = M_n^2$, $A_n =<M>_{n+1} - <M>_n$ and $B_n = 0$. We clearly have

$$\mathbb{E}[V_{n+1}|\mathcal{F}_n] \leqslant V_n + A_n - B_n \qquad \text{a.s.}$$

For any positive γ, denote

$$a_n =<M>_{n+1} (\log <M>_{n+1})^{1+\gamma}.$$

On $\{<M>_n \longrightarrow +\infty\}$, (a_n) is a positive increasing sequence adapted to $\mathbb{F} = (\mathcal{F}_n)$, which goes to infinity a.s. Hence, for n large enough, $a_n \geqslant \alpha > 1$ and it exists a positive finite random variable β such that

$$\sum_{n=0}^{\infty} \frac{A_n}{a_n} \leqslant \int_{\alpha}^{\infty} \frac{1}{x(\log x)^{1+\gamma}} dx + \beta < +\infty \qquad \text{a.s.}$$

Finally, $V_{n+1} = o(a_n)$ a.s. which achieves the proof of the theorem. □

Central limit theorem for martingales

Theorem (Central Limit Theorem)

Let (M_n) be a square integrable **MG** *and let (a_n) be a sequence of positive real numbers increasing to infinity. Assume that*

1. *It exists a deterministic limit $\ell \geqslant 0$ such that*

$$\frac{<M>_n}{a_n} \xrightarrow{\mathcal{P}} \ell.$$

2. **Lindeberg's condition.** *For all $\varepsilon > 0$,*

$$\frac{1}{a_n} \sum_{k=1}^{n} \mathbb{E}[|\Delta M_k|^2 \mathrm{I}_{\{|\Delta M_k| \geqslant \varepsilon \sqrt{a_n}\}} | \mathcal{F}_{k-1}] \xrightarrow{\mathcal{P}} 0$$

where $\Delta M_k = M_k - M_{k-1}$.

Central limit theorem for martingales

Theorem (Central Limit Theorem)

Let (M_n) be a square integrable **MG** *and let (a_n) be a sequence of positive real numbers increasing to infinity. Assume that*

1. *It exists a deterministic limit $\ell \geqslant 0$ such that*

$$\frac{<M>_n}{a_n} \xrightarrow{\mathcal{P}} \ell.$$

2. **Lindeberg's condition.** *For all $\varepsilon > 0$,*

$$\frac{1}{a_n}\sum_{k=1}^{n} \mathbb{E}[|\Delta M_k|^2 \mathrm{I}_{\{|\Delta M_k| \geqslant \varepsilon\sqrt{a_n}\}} | \mathcal{F}_{k-1}] \xrightarrow{\mathcal{P}} 0$$

where $\Delta M_k = M_k - M_{k-1}$.

Central limit theorem fro martingales, continued

Theorem (Central Limit Theorem)

Then, we have

$$\frac{1}{\sqrt{a_n}} M_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \ell).$$

Moreover, if $\ell > 0$*, we also have*

$$\sqrt{a_n}\Big(\frac{M_n}{< M >_n}\Big) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \ell^{-1}).$$

$\longrightarrow$ Lyapunov's condition implies Lindeberg's condition. $\exists\, \alpha > 2$,

$$\sum_{k=1}^{n} \mathbb{E}[|\Delta M_k|^{\alpha} | \mathcal{F}_{k-1}] = O(a_n) \qquad \text{a.s.}$$

Outline

Stable autoregressive processes

Consider the stable autoregressive process

$$X_{n+1} = \theta X_n + \varepsilon_{n+1}, \qquad |\theta| < 1$$

where (ε_n) is a sequence of **iid** $\mathcal{N}(0, \sigma^2)$ random variables. Assume that X_0 is independent of (ε_n) with $\mathcal{N}(0, \sigma^2/(1-\theta^2))$ distribution.

- (X_n) is a **centered stationary Gaussian process,**
- (X_n) is a **positive recurrent process.**

Goal

$\longrightarrow$ Estimate the unknown parameter θ.

Least squares estimator

Let $\widehat{\theta}_n$ be the **least squares** estimator of the unknown parameter θ

$$\widehat{\theta}_n = \frac{\displaystyle\sum_{k=1}^{n} X_k X_{k-1}}{\displaystyle\sum_{k=1}^{n} X_{k-1}^2}.$$

We have

$$\begin{aligned}
\widehat{\theta}_n - \theta &= \frac{\sum_{k=1}^{n} X_k X_{k-1} - \theta \sum_{k=1}^{n} X_{k-1}^2}{\sum_{k=1}^{n} X_{k-1}^2}, \\
&= \frac{\sum_{k=1}^{n} X_{k-1}(X_k - \theta X_{k-1})}{\sum_{k=1}^{n} X_{k-1}^2}, \\
&= \frac{\sum_{k=1}^{n} X_{k-1}\varepsilon_k}{\sum_{k=1}^{n} X_{k-1}^2}.
\end{aligned}$$

Least squares estimator

Consequently,

$$\widehat{\theta}_n - \theta = \sigma^2 \frac{M_n}{<M>_n}$$

$$M_n = \sum_{k=1}^{n} X_{k-1}\varepsilon_k \qquad \text{and} \qquad <M>_n = \sigma^2 \sum_{k=1}^{n} X_{k-1}^2.$$

The sequence (M_n) is a **square integrable martingale** such that

$$\lim_{n\to\infty} \frac{<M>_n}{n} = \ell \qquad \text{a.s.}$$

where

$$\ell = \frac{\sigma^4}{1-\theta^2}.$$

Stable autoregressive processes

Theorem

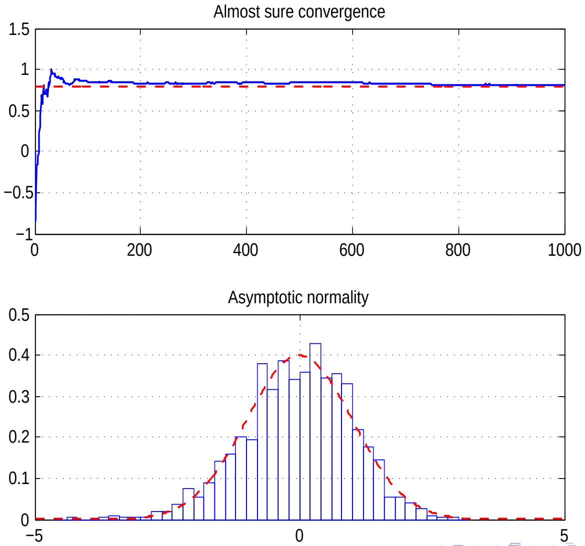
We have the almost sure convergence

$$\lim_{n\to\infty} \widehat{\theta}_n = \theta \qquad \text{a.s.}$$

In addition, we also have the asymptotic normality

$$\sqrt{n}\Big(\widehat{\theta}_n - \theta\Big) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1-\theta^2).$$

Stable autoregressive processes

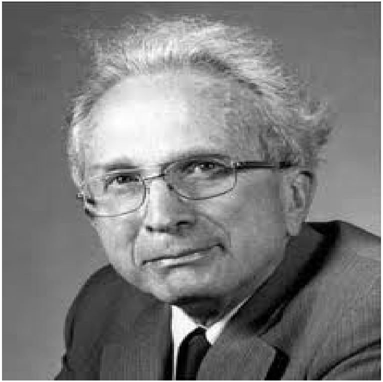


Stochastic approximation

Herbert Robbins

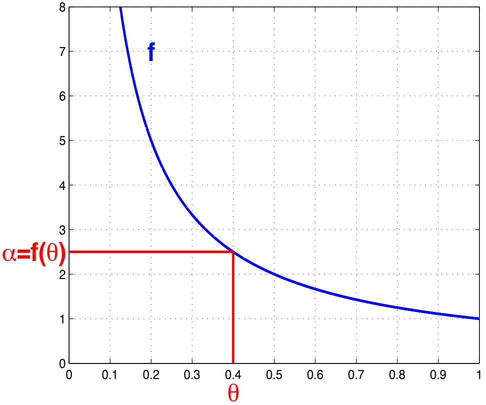
Stochastic approximation

Jack Kiefer



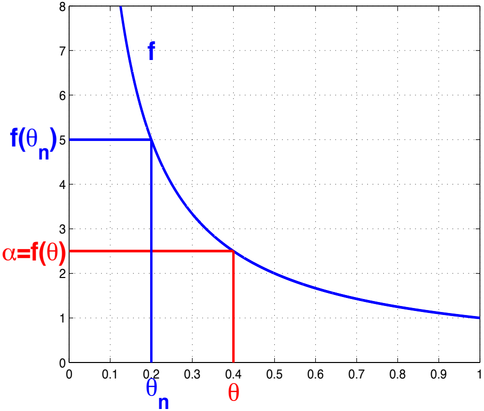
Jacob Wolfowitz

Stochastic approximation



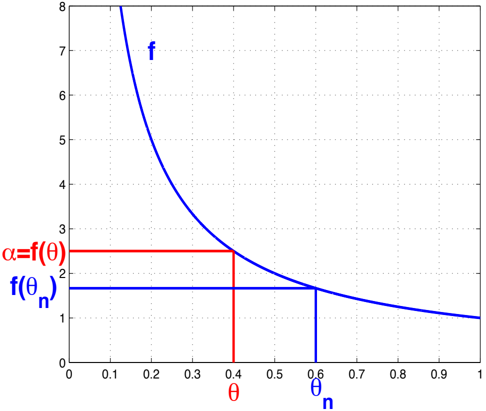
Goal

$\longrightarrow$ Find the value $\boldsymbol{\theta}$ without any knowledge on the function $\boldsymbol{f}$.



Basic Idea

At time n, if you are able to say that $f(\widehat{\theta}_n) > \alpha$, then increase the value of $\widehat{\theta}_n$.



Basic Idea

At time n, if you are able to say that $f(\widehat{\theta}_n) < \alpha$, then decrease the value of $\widehat{\theta}_n$.

Stochastic approximation

Let (γ_n) be a decreasing sequence of positive real numbers

$$\sum_{n=1}^{\infty} \gamma_n = +\infty \qquad \text{and} \qquad \sum_{n=1}^{\infty} \gamma_n^2 < +\infty.$$

For the sake of simplicity, we shall make use of

$$\gamma_n = \frac{1}{n}.$$

Robbins-Monro algorithm

$$\widehat{\theta}_{n+1} = \widehat{\theta}_n + \gamma_{n+1}\Big(T_{n+1} - \alpha\Big)$$

where T_{n+1} is a random variable such that $\mathbb{E}[T_{n+1}|\mathcal{F}_n] = f(\widehat{\theta}_n)$.

Stochastic approximation

Theorem (Robbins-Monro, 1951)

Assume that f is a decreasing function. Then, we have the almost sure convergence

$$\lim_{n \to \infty} \widehat{\theta}_n = \theta \quad \text{a.s.}$$

In addition, as soon as $-2f'(\theta) > 1$, *we also have the asymptotic normality*

$$\sqrt{n}\Big(\widehat{\theta}_n - \theta\Big) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \xi^2(\theta))$$

where the asymptotic variance $\xi^2(\theta)$ *can be explicitely calculated.*

Kernel density estimation

Let (X_n) be a sequence of **iid** random variables with **unknown density function f**. Let K be a positive and bounded function, called **kernel**, such that

$$\int_{\mathbb{R}} K(x)\,dx = 1, \qquad \int_{\mathbb{R}} xK(x)\,dx = 0,$$

$$\int_{\mathbb{R}} K^2(x)\,dx = \sigma^2.$$

Goal

$\longrightarrow$ Estimate the unknown density function f.

Choice of the Kernel

- **Uniform kernel**

$$K_a(x) = \frac{1}{2a}\mathrm{I}_{\{|x|\leqslant a\}},$$

- **Epanechnikov kernel**

$$K_b(x) = \frac{3}{4b}\left(1 - \frac{x^2}{b^2}\right)\mathrm{I}_{\{|x|\leqslant b\}},$$

- **Gaussian kernel**

$$K_c(x) = \frac{1}{c\sqrt{2\pi}}\exp\left(-\frac{x^2}{2c^2}\right).$$

Choice of the Kernel

- **Uniform kernel**

$$K_a(x) = \frac{1}{2a} \mathrm{I}_{\{|x| \leqslant a\}},$$

- **Epanechnikov kernel**

$$K_b(x) = \frac{3}{4b}\left(1 - \frac{x^2}{b^2}\right) \mathrm{I}_{\{|x| \leqslant b\}},$$

- **Gaussian kernel**

$$K_c(x) = \frac{1}{c\sqrt{2\pi}} \exp\left(-\frac{x^2}{2c^2}\right).$$

Choice of the Kernel

- **Uniform kernel**

$$K_a(x) = \frac{1}{2a} \mathrm{I}_{\{|x| \leqslant a\}},$$

- **Epanechnikov kernel**

$$K_b(x) = \frac{3}{4b} \left(1 - \frac{x^2}{b^2}\right) \mathrm{I}_{\{|x| \leqslant b\}},$$

- **Gaussian kernel**

$$K_c(x) = \frac{1}{c\sqrt{2\pi}} \exp\left(-\frac{x^2}{2c^2}\right).$$

The Wolverton-Wagner estimator

We estimate the density function f by

The Wolverton-Wagner estimator

$$\widehat{f}_n(x) = \frac{1}{n}\sum_{k=1}^{n} W_k(x)$$

where

$$W_k(x) = \frac{1}{h_k} K\Big(\frac{X_k - x}{h_k}\Big).$$

The **bandwidth** (h_n) is a sequence of positive real numbers, $h_n \searrow 0$, $nh_n \to \infty$. For $0 < \alpha < 1$, we can make use of

$$h_n = \frac{1}{n^\alpha}.$$

Kernel density estimation

We have

$$\widehat{f}_n(x) - f(x) = \frac{1}{n}\sum_{k=1}^{n} W_k(x) - f(x),$$
$$= \frac{1}{n}\sum_{k=1}^{n}(W_k(x) - \mathbb{E}[W_k(x)]) + \frac{1}{n}\sum_{k=1}^{n}(\mathbb{E}[W_k(x)] - f(x)).$$

Consequently,

$$\widehat{f}_n(x) - f(x) = \frac{M_n(x)}{n} + \frac{R_n(x)}{n}$$

where

$$M_n(x) = \sum_{k=1}^{n}(W_k(x) - \mathbb{E}[W_k(x)]).$$

Kernel density estimation

We have

$$
\begin{aligned}
M_n(x) &= \sum_{k=1}^{n} (W_k(x) - \mathbb{E}[W_k(x)]), \\
< M(x) >_n &= \sum_{k=1}^{n} \mathrm{Var}(W_k(x)).
\end{aligned}
$$

The sequence $(M_n(x))$ is a **square integrable martingale** such that

$$
\lim_{n \to \infty} \frac{< M(x) >_n}{n^{1+\alpha}} = \ell \qquad \text{a.s.}
$$

where

$$
\ell = \frac{\sigma^2 f(x)}{1 + \alpha}.
$$

Kernel density estimation

Theorem

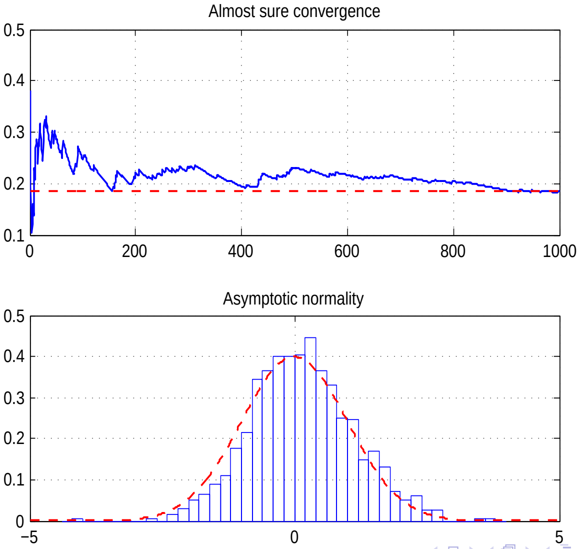
For all $x \in \mathbb{R}$, we have the pointwise almost sure convergence

$$\lim_{n \to \infty} \widehat{f}_n(x) = f(x) \qquad \text{a.s.}$$

In addition, as soon as $1/5 < \alpha < 1$, we have, for all $x \in \mathbb{R}$, the asymptotic normality

$$\sqrt{nh_n}\left(\widehat{f}_n(x) - f(x)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2 f(x)}{1+\alpha}\right).$$

Kernel density estimation




!!!! Many thanks for your attention !!!!