Bootstrap and dependence

P. Bertail (Crest-Laboratoire de Statistique)

Outlines

. The Main Principles in the i.i.d. case : the plug-in rule

. Main asymptotic results : improvement and failures

. Generalization to time-series

  . The moving block approach: stationary times-series

  . The regenerative approach : Markov chains

Main principles : the i.i.d. case

The plug-in rule

- T: P (probability space)→B (separable)
- $(X_1, X_2, \ldots, X_n)$ i.i.d P in P

$$\boxed{P_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}}$$ empirical probability

Empirical counterpart : $\boxed{T(P_n)}$

Ex:

$$\boxed{\begin{array}{l} T(P) = E_P X = \int xP(dx)\ ,T(P_n) = \int xP_n(dx) = \overline{X}_n \\ M - parameter : T(P)\ solution\ of\ E_P\Psi(X, T(P)) = 0 \\ M - estimator : T(P_n)\ solution\ of\ n^{-1}\sum_{i=1}^{n}\Psi(X_i, T(P_n)) = 0 \end{array}}$$

Properties of $T(P_n)$, linked to

. Properties of the empirical process

. Properties of continuity and differentiability of T

→Main tools of robustness and semiparametric approach

(cf Huber(1981), Van der Vaart and Wellner (1996),
Bickel, Ritov, Klassen, Wellner(1993))

Goal of the Bootstrap

Let $T_n=T(X_1,\ldots,X_n)$ be a statistics estimating T(P)

Goal : -giving a good approximation of the distribution of T_n
-constructing good confidence interval (bounds with a small error, intervals with an exact exact level close to the original)

$\boxed{\theta = T(P) \in R^q,\ T_n \in R^q,\ \Sigma_n \in M_{p,p}(R)}$ (standardisation of T_n)

Parameters of interest : distributions

$$\boxed{\mathrm{L}_n(\mathrm{x}, \mathrm{P}) = \Pr_{P^{\otimes n}}(T_n - T(P) < x)}$$

$$\boxed{\mathrm{K}_n(\mathrm{x}, \mathrm{P}) = \Pr_{P^{\otimes n}}(n^{1/2}\Sigma_n^{-1}(T_n - T(P)) < x)}$$

seen as sequence of functionals
$L_n(.,P)$ or $K_n(.,P)$ from $\mathcal{P}$ to some space of distributions B

In general, the exact distributions are unknown :

. First classical approach : $n \rightarrow \infty$: CLT for T_n

. Second order approach : improve over the CLT, asymptotic expansions (Edgeworth expansion).

. Bootstrap : Plug-in rule $P \rightarrow P_n$: $\mathrm{L}_n(x, P_n)$ *and* $K_n(., P_n)$ $\Updownarrow$

Bootstrap distribution= empirical estimator of the distribution of T_n
= random probability measure

$$\hat{\mathrm{K}}_{\mathrm{n}}^{\mathrm{B}}(\mathrm{x}, \mathrm{P}_{\mathrm{n}}) = \hat{\mathrm{P}}\mathrm{r}^{B}_{P_n^{\otimes n}}\left(n^{1/2}(\Sigma_n^{*})^{-1}(T_n^{*} - T(P_n)) < x\right)$$

$$= B^{-1} \sum_{(i_1, i_2, \dots, i_n) \in I_B} 1_{n^{1/2}(\Sigma_n^{-1}(X_{i_1}, X_{i_2}, \dots X_{i_n})\left\{T_n\ (X_{i_1}, X_{i_2}, \dots X_{i_n}) - T_n < x\right\}}$$

$$I_B = subset\, of\ B\ \ n - uplets$$

Empirical distribution of all the values that may be constructed by
Selecting samples with replacement from the original sample

→Very simple algorithm : Montecarlo in the observations.

If n big, only B simulation by drawing with replacement in the observations

Bootstrap Algorithm : Efron(1979),(1981)

$$m = n$$

$(X_1^*, X_2^*, \ldots, X_m^*)$ *i.i.d.* P_n *(Bootstrap sample)*

$$\boxed{\text{Compute } \mathrm{T_m}(X_1^*, X_2^*, \ldots, X_m^*),\ \Sigma_m(X_1^*, X_2^*, \ldots, X_m^*)\ (\textit{Bootstrapped statistics})}$$

Bootstrap dist. $\mathrm{K_m}(\mathrm{x}, \mathrm{P_n}) = \mathrm{Pr}_{P_n^{\otimes n}}(m^{1/2}(\Sigma_m^*)^{-1}(T_m^{\ *} - T_n) < x \mid P_n)$

is approximated by B Monte-Carlo simulations

Step 1 do until $\mathrm{i} > \mathrm{B}$;

Draw a sample $(x_1^{(\mathrm{i})*}, x_2^{(i)*}, \ldots, x_m^{(i)*})$ *with replacement*

from the observations $(x_1, x_2, \ldots, x_n)$;

Compute the values $T_m^{(i)*} = T(x_1^{(\mathrm{i})*}, x_2^{(i)*}, \ldots, x_m^{(i)*}),\ \Sigma_m^{(i)*}$;

endo;

Step 2 use the histograms of the $T_m^{(i)*} - T_n$ *or* $t_m^{(i)*} = m^{1/2}(\Sigma_m^{(i)*})^{-1}(T_m^{\ (i)*} - T_n)$

$i = 1, \ldots, B$ as approximations of $\mathrm{L_n}(\mathrm{x}, \mathrm{P_n})$ and $\mathrm{K_n}(\mathrm{x}, \mathrm{P_n})$

Asymptotic validity (m=n) : linked to the equicontinuity of the sequence $K_n(.,Q)$ at the true P (Bickel &Freedman(1981)).

$$\boxed{\sup_{x \in R^q} \left| K_n(x, P_n) - K_n(x, P) \right| \xrightarrow[n\to\infty]{} 0 \ \ (in\ \Pr.\ or\ a.s.)}$$

Fails on bounderies (if there are some jumps in the limiting distribution)

Ex: . $T_n = \max(X_i)$ (3 domains of attractions, three limiting types)

. Degenerate U stat.

$$T_n = (\overline{X}_n)^2, \quad when\ E_P X^2 < \infty$$

$$\boxed{\begin{array}{l} T(P) = E_P X \neq 0,\ n^{1/2}(T_n - T(P)) \xrightarrow{L} N(0, S^2) \\ T(P) = E_P X = 0,\ n(T_n - T(P))/V_P(X) \xrightarrow{L} \chi^2(1) \end{array}}$$

$$\boxed{In\ general\ \ T(P_n) \neq 0 \Rightarrow Bootstrap\ \ failure\ \ at\ 0}$$

. Statistics with hyperefficiency points: bootstrap fails at hyperefficiency points

Asymptotic Validity : choosing $m<<n$ ($m/n \to 0$) and bootstrapping without replacement always works under minimal assumptions (Subsampling)

Bretagnolle (1983), Ann. Inst. H. P., Politis and Romano(1995),Ann. Stat
Bertail, Politis and Romano(1997), JASA, Bertail, Politis, Rhomari(1997),
Statistics, Bertail(1997), Bernouilli, Bertail, Haeffke,Politis,White(2003?)

LOSS IN EFFICICIENCY but ROBUSTNESS

Second order validity : The bootstrap distribution is better than the asymptotic distribution for smooth functionals.

If T(P) is a function of moments (cf Hall(1986), a third order Frechet differentiable functionals (for some adequate metric) then

$$\sup_{x\in R^q}\left| \mathrm{K}_n(\mathrm{x},\mathrm{P}) - \Phi(\mathrm{x}) \right| \le C/n^{1/2} \quad (Berry-Esséen)$$

$$\sup_{x\in R^q}\left| \mathrm{K}_n(\mathrm{x},\mathrm{P}_n) - \mathrm{K}_n(\mathrm{x},\mathrm{P}) \right| = O_P(n^{-1}) \quad (O(n^{-1}\log\log(n)^{1/2}) \quad a.s.)$$

Conditions : Cramer condition on the first gradient(distribution of the first gradient not lattice), moment conditions on the first and second Gradient.

VERY ACCURATE CONFIDENCE INTERVALS BASED ON $K_n(.,P_n)$

Bootstrap for dependent data

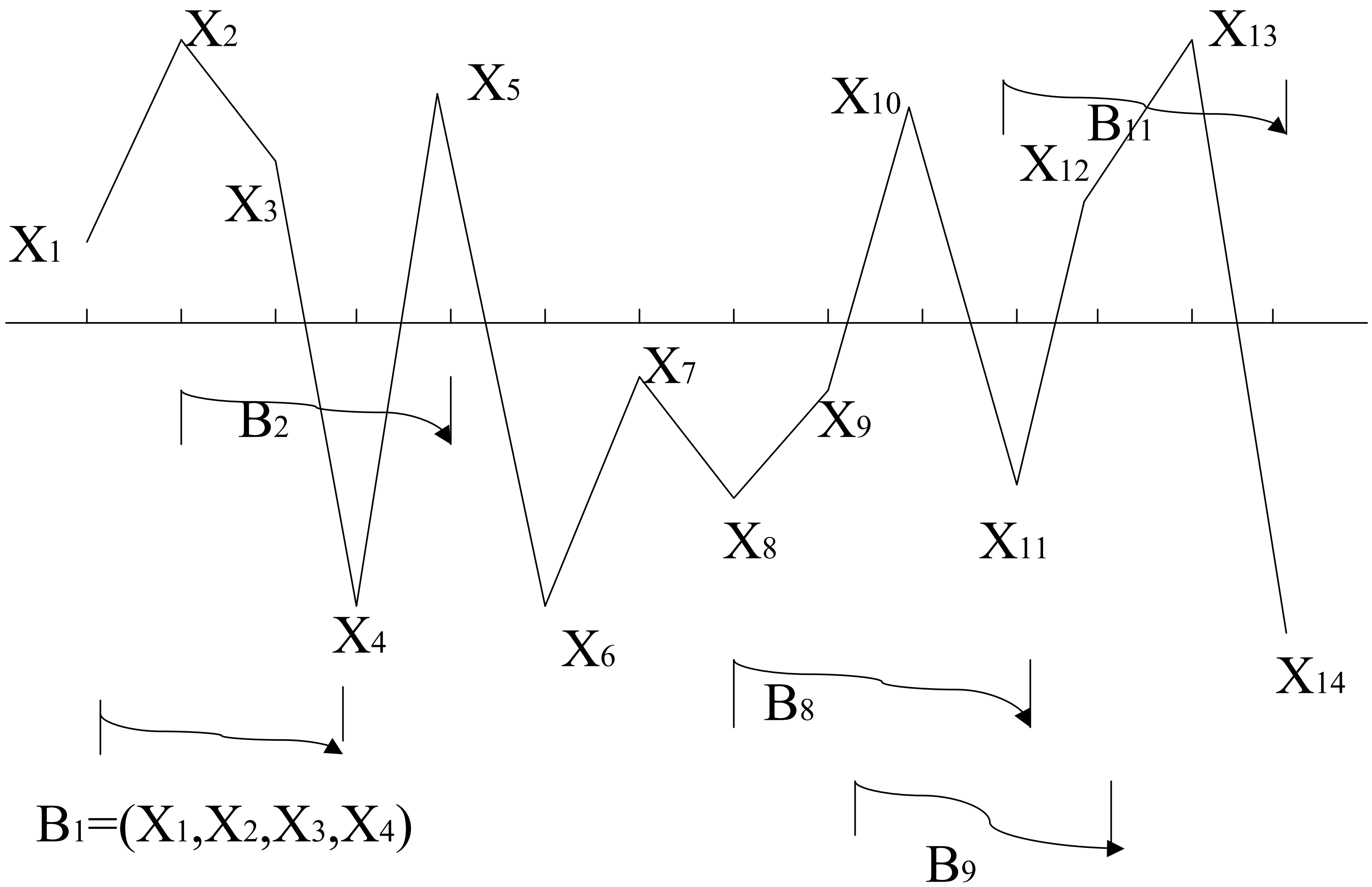
- Moving Block Bootstrap (Künsch (1989)), Ann. Stat, Liu and Singh (1992).
- Model based approach (ARMA model, non-linear AR model) : i.i.d residual reduction (Bose, Ann. Stat. (1988), Mammen and al. (2002), Bernouilli).

    →Sieve approach (approximation by AR model with « infinite order ») : linear times series (Buhlmann(1997), Bernouilli).

•Markov chains : recent works based on generating pseudo-data via a preliminary estimation of the transition probability (Rajarshi (1990), Paparadotis & Politis(2002), JSPI, Horowitz (2004), Econometrica) : discrete MC.

→Regenerative Approach (Bertail & Clémençon (2002)(2003))

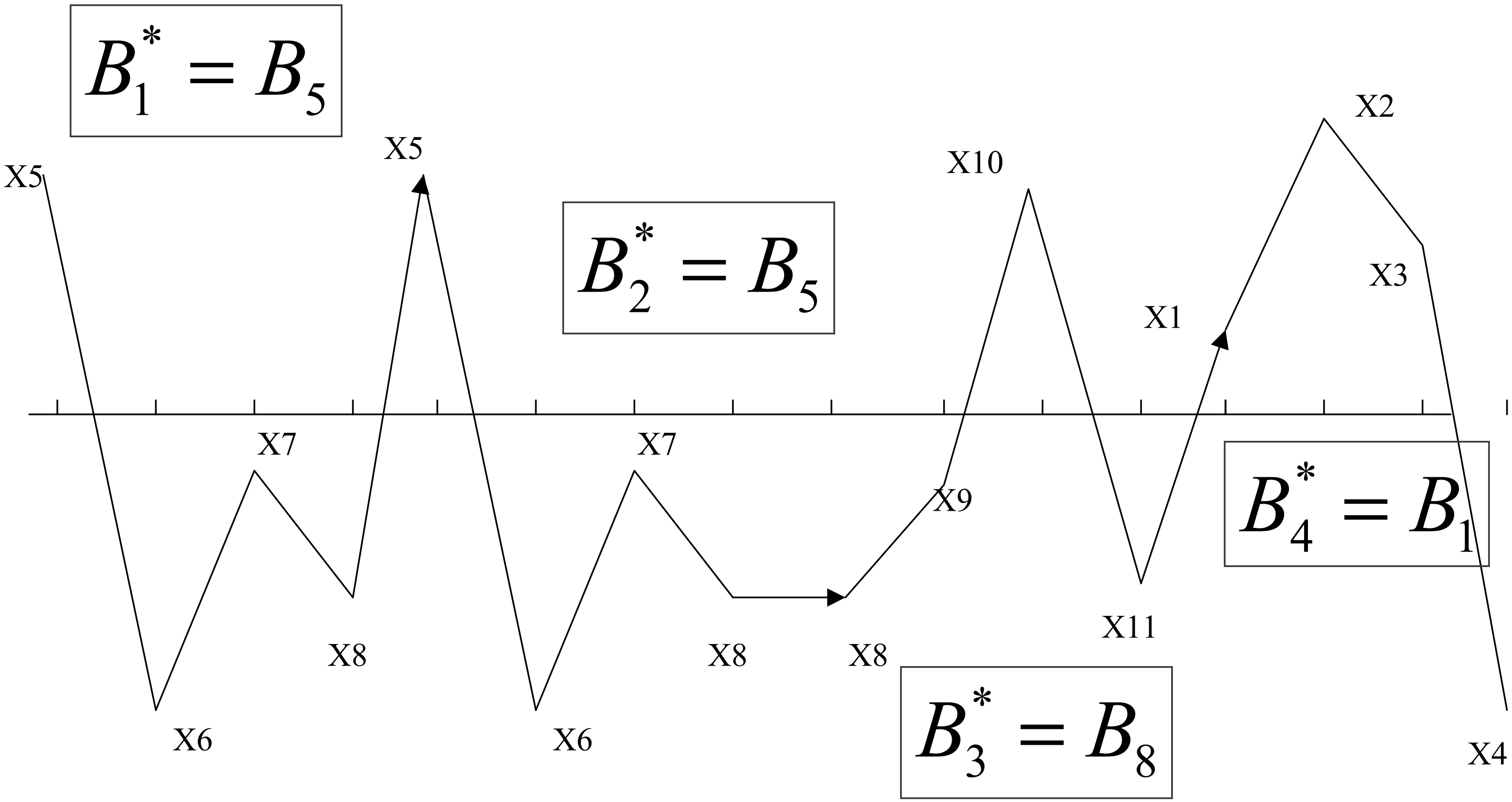
The Moving-Block Bootstrap



$$length\ of\ blocks: b_n = o(n^{1/2})$$

$$(B_1^*, B_2^*, B_3^*, \ldots, B_{[n/b_n]}^*) \quad i.i.d \;\; F_B = \frac{1}{(n-b_n+1)} \sum_{i=1}^{n-b_n+1} \delta_{B_i}$$

Resample blocks with replacement : reconstruct a pseudo times-series



Asymptotic Validity : Kunsch (1989)

Hypotheses : Stationarity + m-dependence or exponentially decreasing strong-mixing coefficients

$$b_n \to \infty, \ \frac{b_n}{\sqrt{n}} \to 0$$

Second order validity of the naive MBB fails for 2 reasons :

-Bias problem (→ recenter by the expected value of the bootstrap mean Lahiri(1992), J. Mult. Anal.)

-The second-order properties are very sensitive to the choice of the standardisation. The variance of the bootstrap distribution has not the correct biais → additional adjustment.

Best achievable rate $\boxed{O_P(n^{-3/4})}$ (the standardisation may be <0)

$\boxed{O_P(n^{-2/3})}$ (for positive standardisation)

Regenerative Bootstrap

- Markov chain with state space $(E, \mathcal{E})$

$$X_0 \sim \nu,$$

$$P(X_{n+1} \in B \mid X_0, \ldots, X_n) = \Pi(X_n, B) \text{ a.s. } .$$

- Hypothesis : The chain is positive Harris recurrent for some measure ψ
  - ψ Irreductible
  - Harris recurrent ($\psi(B) > 0$ $P_x(\sum_{n=1}^{\infty} I\{X_n \in B\} = \infty) = 1$
  - Positive (Invariant measure $\boxed{\mu}$)

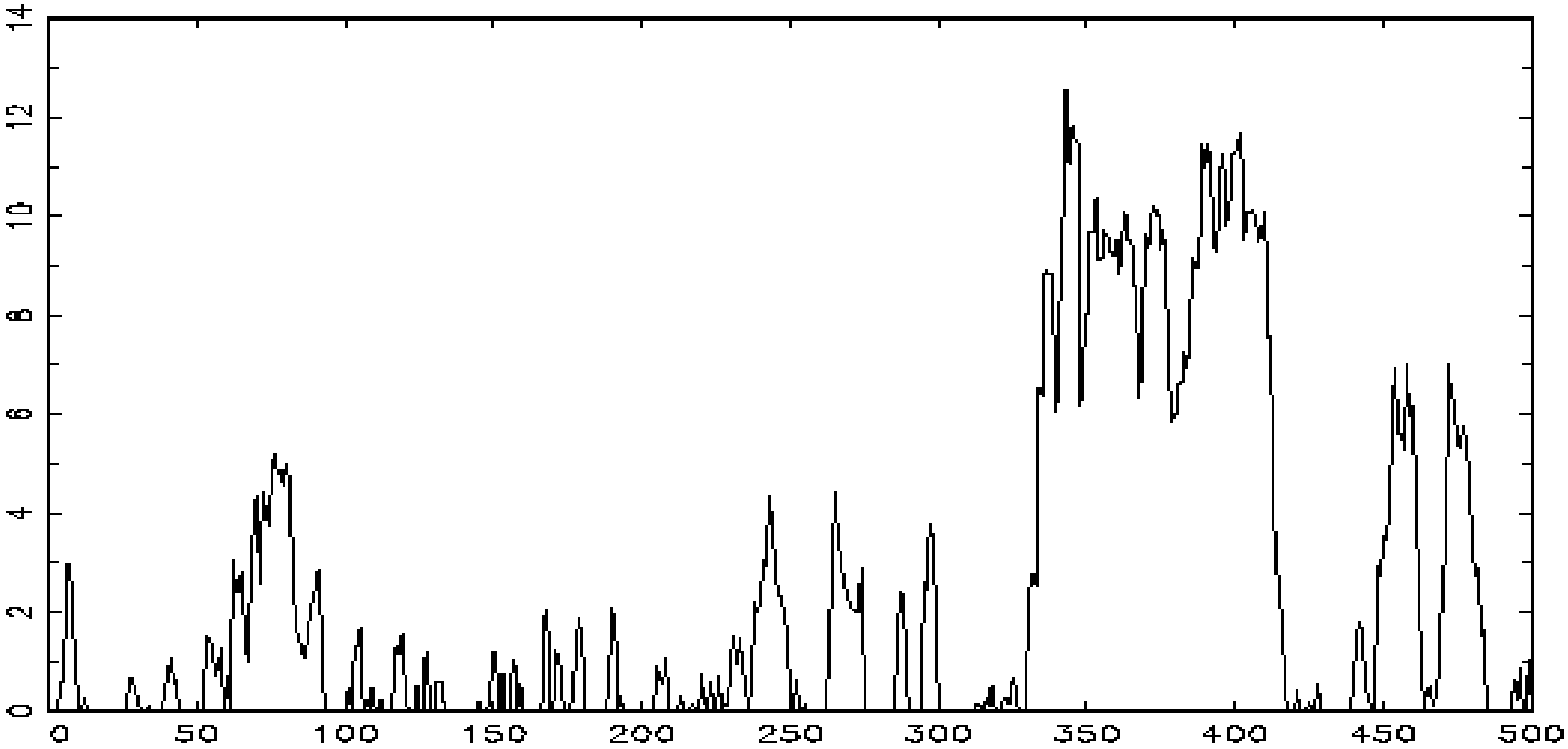
$$\mu\Pi(dy) = \int_{x \in E} \mu(dx)\Pi(x, dy) = \mu(dy)$$

  - The chain has an accessible atom A $\Pi(x, .) = \Pi(y, .)$ and $\psi(A) > 0.$

Example : Queuing system with E(inter-arrivals) < E (inter-service)

$$\boxed{W_{t+1} = (W_t + \Delta\tau_t - \Delta T_t)_+}$$

$$\boxed{\theta = E_\mu W_t}$$



A={0}

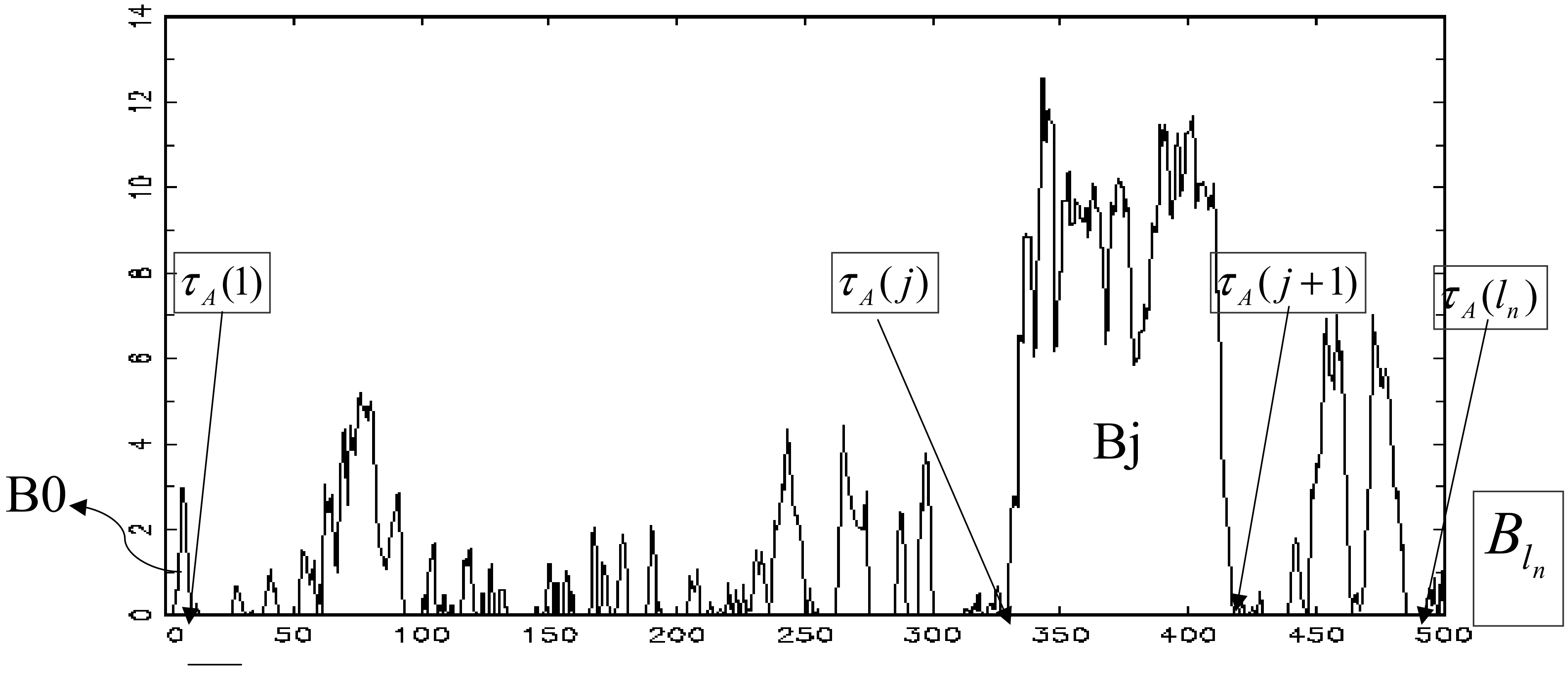
Work-modulated server : $\boxed{\Delta\tau_t}$ Independant conditionally to $\boxed{W_t}$

$$\tau_A = \tau_A(1) = \inf\{n \geq 1,\ X_n \in A\}$$

$$\tau_A(j) = \inf\{n > \tau_A(j-1),\ X_n \in A\} \text{ for } j \geq 2$$

$$\mathcal{B}_1 = \left(X_{\tau_A(1)+1}, \ldots,\ X_{\tau_A(2)}\right), \ldots,\ \mathcal{B}_j = \left(X_{\tau_A(j)+1}, \ldots,\ X_{\tau_A(j+1)}\right), \ldots$$

$l_n = \sum_{i=0}^{n} 1_A(X_i)$ Number of visits into A



By the strong-markov property, the blocks $B_1, \ldots, B_{l_n}$ are independant.

Regenerative Bootstrap

$$X^* = (B_1^*, B_2^*, B_3^*, \ldots, B_{l_n}^*) \quad i.i.d \;\; F_B = \frac{1}{l_n} \sum_{i=0}^{l_n - 1} \delta_{B_i}$$

Asymptotic validity : Datta and McCormick (1993)
(based on Rajarshi (1990) : discrete case)

Hypothesis : Stationarity
Strong-mixing rate (Götze and Hipp (1983))
Moment conditions

This regenerative bootstrap of the mean is NOT second order correct (up to [redacted])

Very Bad finite sample performances

A second order correct regenerative Bootstrap

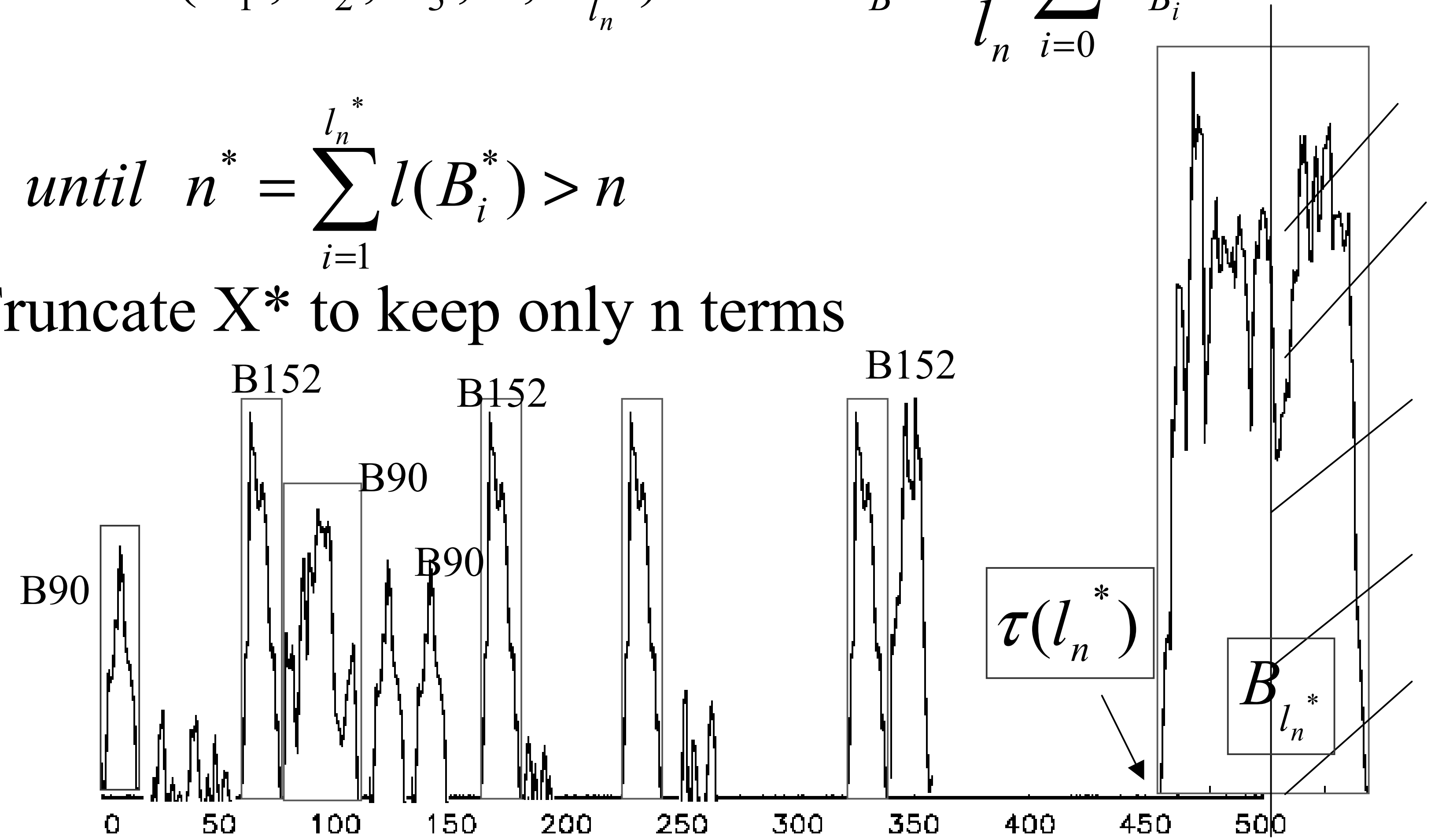
Draw

$\boxed{\text{Rate of approximation : } O_P(n^{-1}\log(n))}$

$$X^* = (B_1^*, B_2^*, B_3^*, \ldots, B_{l_n^*}^*) \quad i.i.d \;\; F_B = \frac{1}{l_n}\sum_{i=0}^{l_n-1}\delta_{B_i}$$

$$until \quad n^* = \sum_{i=1}^{l_n^*} l(B_i^*) > n$$

Truncate X* to keep only n terms



Harris recurrent Markov chain : the general case

Idea : a constructive approximation of the Nummelin splitting technique. Extension of the initial chain to a chain with an atom.

Hypothesis : existence of a minorization condition on a small set

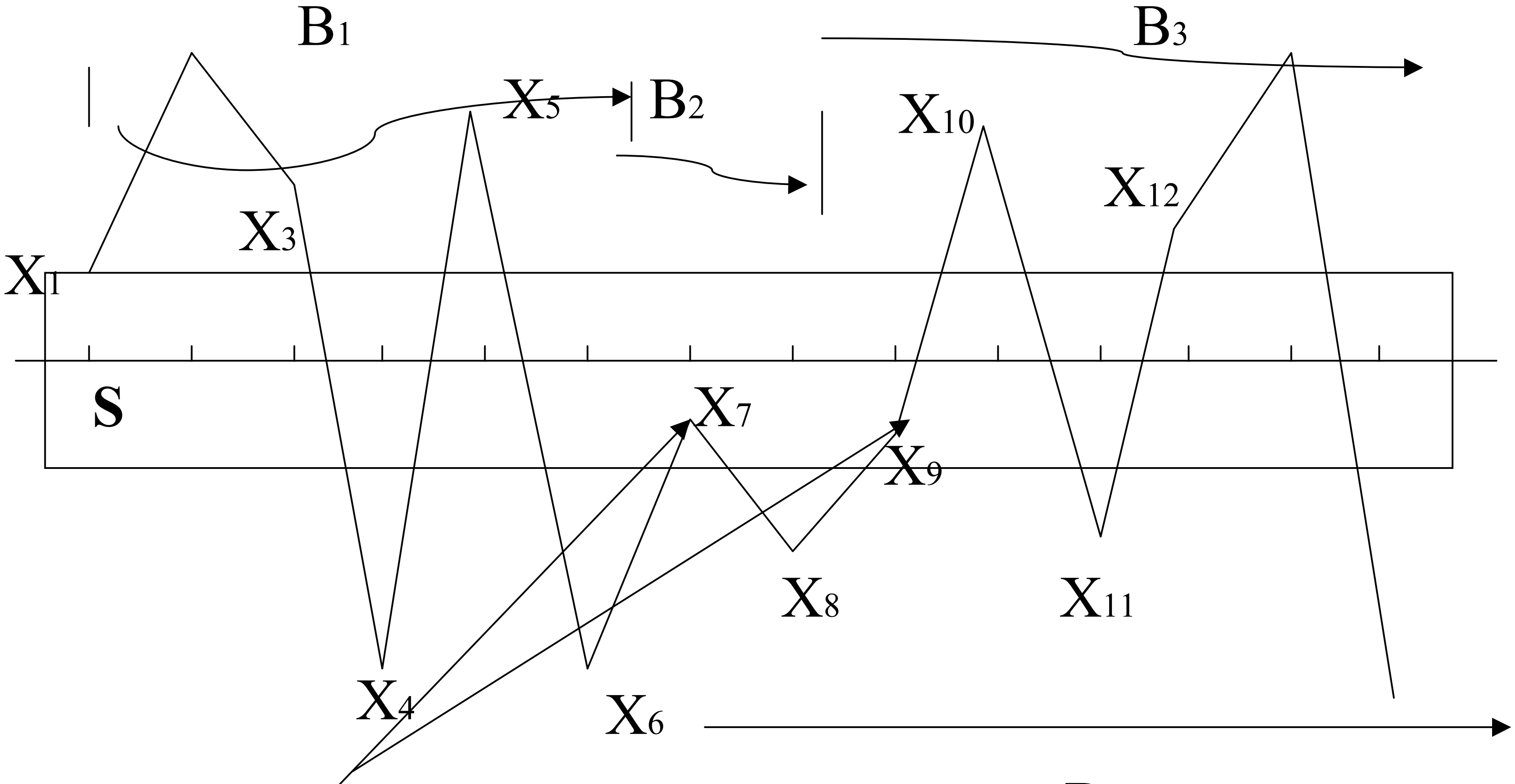
Definition *For a Markov chain valued in a state space $(E, \mathcal{E})$ with transition probability Π, a set $S \in \mathcal{E}$ is said to be small if there exist an integer $m > 0$, a probability measure Φ supported by S, and $\delta > 0$ such that*

$$\forall x \in S, \forall A \in \mathcal{E}, \quad \Pi^m(x, A) \geq \delta \Phi(A),$$

denoting by Π^m the m-th iterate of Π. When this holds, we say that the chain satisfies the minorization condition $\mathcal{M}(m, S, \delta, \phi)$.

$\boxed{m=1}$

Regenerative Bootstrap for the approximated splitted chain



or B2

Randomization $B(1, \delta\phi(x_i) / \hat{p}_n(x_{i+1}, x_i))$

Practical Pb : choice of the small set

Theoretical Pb : error induced by splitting at the wrong place

Theorem : If h is a bounded functionals and

$$E\tau_A^{8+\varepsilon} < \infty \quad for\ \varepsilon > 0$$

if the uniform rate of convergence of $\hat{p}_n$ to p is of order α_n then the second order validity still holds with a reminder

of order $O_P(n^{-1/2}\alpha_n^{-1})$

Proof : Calculation of the moments of the difference between the empirical moments of the f(Bi)'s obtained by splitting at the right theoretical place and the f(Ai)'s obtained by the approximated splitted chain

Use of the linearity of the functional

Remark :

Typically the rate of convergence is of of order $\alpha_n = n^{-s/(2s+1)}$

For very regular chain, we get the i.i.d rate $O_P(n^{-1})$

Bibliography

- **Introductory books**

. Efron (1981), SIAM

.Efron, Tibshirani (1997), Chapman, Hall

. Davison, Hinkley (2000),

- **Theoretical book**

.Hall (1992), Springer

.Shao, Tu (1995),Springer (biblio)

. Barbe, Bertail (1995), Springer

. Giné (1996), in Ecole Proba St Flour, Springer

Subsampling

Politis, Romano, Wolf (2000), Springer