Recent advances in empirical regression schemes for BSDEs

emmanuel.gobet@polytechnique.edu

Centre de Mathématiques Appliquées and FiME,
Ecole Polytechnique and CNRS

With the support of La FONDATION du RISQUE

Based on joint works with T. Ben Zineb and P. Turkedjiev.

Main issue: for solving BSDE using empirical regressions, how to *optimally* tune

- ✓ the number of discretization dates,
- ✓ the approximation spaces,
- ✓ the number of simulations?

Agenda

- ✓ Different discrete-time Dynamic Programming (DP) Equations:
  - **ODP**: One step forward DP equation [BT04][GLW05]
  - **MDP**: Multi-step forward DPE ($\approx$ [BD07] without Picard iterations)
  - **Mal.MDP**: Malliavin MDP (alternative representation of Z)

  Pros and cons: error norms and stability, independent clouds of simulations, basis functions, managing constraints...
- ✓ Handling irregular/quadratic BSDE
- ✓ Generic variance reductions
- ✓ Conclusion, perspectives, works in progress, open questions

BSDE Setting

Generalized BSDE with *fixed terminal time* T:

$$\mathbf{Y_t} = \xi + \int_{\mathbf{t}}^{\mathbf{T}} \mathbf{f}(\mathbf{s}, \mathbf{Y_s}, \mathbf{Z_s})\mathbf{ds} - \int_{\mathbf{t}}^{\mathbf{T}} \mathbf{Z_s dW_s} - (\mathbf{L_T} - \mathbf{L_t}),$$

under various assumptions, for instance:

- ✓ driving noise = Brownian Motion W and Poisson measure,
- ✓ L martingale orthogonal to W,
- ✓ quadratic driver, ...

but under **Markovian assumptions**: $\mathbf{f}(\mathbf{s}, \omega, \mathbf{y}, \mathbf{z}) = \mathbf{f}(\mathbf{s}, \mathbf{X_s}, \mathbf{y}, \mathbf{z})$, $\xi = \mathbf{g}(\mathbf{X_T})$, X is a jump-diffusion ➠ $\mathbf{Y_t} = \mathbf{u}(\mathbf{t}, \mathbf{X_t}), \mathbf{Z_t} = \nabla\mathbf{u}(\mathbf{t}, \mathbf{X_t})\sigma(\mathbf{t}, \mathbf{X_t})$.

Multidimensional: $X \in \mathbb{R}^d$, $Y \in \mathbb{R}$, $Z \in \mathbb{R}^q$.

Simulating BSDE = 2 problems:

1. computing u and ∇u (hard)
2. simulate the path of X (easy)

Conditional expectations representations

$$Y_t = \mathbb{E}^{\mathcal{F}_t}\Big(\xi + \int_t^T f(s, X_s, Y_s, Z_s)ds\Big).$$

Solving the BSDE requires the computation of nested conditional expectations.

Advantages of the empirical approach:

✓ black box algorithm (no need to know the model):
input = model simulations ➠ output = BSDE solutions.

✓ uniform controls w.r.t. the model, models may be degenerate, machine learning techniques. But presumably too conservative estimates (worst-case).

Overview of global error decomposition:

$$\text{quadratic error} \le \underbrace{\text{discretization error}}_{\underset{N\to+\infty}{\longrightarrow} 0} + \underbrace{\text{approximation error}}_{\underset{K\to+\infty}{\longrightarrow} 0, \quad \underset{N\to+\infty}{\longrightarrow} +\infty}$$

$$+ \underbrace{\text{statistical error}}_{\underset{M\to+\infty}{\longrightarrow} 0, \quad \underset{N\to+\infty}{\longrightarrow} +\infty, \quad \underset{K\to+\infty}{\longrightarrow} +\infty} + \underbrace{\text{interdependency error}}_{\underset{M\to+\infty}{\longrightarrow} 0, \quad \underset{N\to+\infty}{\longrightarrow} +\infty, \quad \underset{K\to+\infty}{\longrightarrow} +\infty}.$$

TIME DISCRETIZATION OF $Y_t = \xi + \int_t^T f(s, Y_s, Z_s)ds - \int_t^T Z_s dW_s - (L_T - L_t)$

Standard discretization along deterministic time grid $\pi := \{0 = t_0 < \ldots < t_N = T\}$:

- ✓ $(i+1)$-th time-step is $\Delta_i = t_{i+1} - t_i$;
- ✓ mesh size $|\pi| := \max_{0 \le i < N} \Delta_i$;
- ✓ related Brownian motion increments $\Delta W_i := W_{t_{i+1}} - W_{t_i}$.

Discrete time BSDE (Y, Z):

$$\begin{cases} Y_i & = \mathbb{E}_i \left(Y_{i+1} + f_i(Y_{i+1}, Z_i)\Delta_i \right), \quad 0 \le i < N, \\ \Delta_i Z_i & = \mathbb{E}_i \left(Y_{i+1} \Delta W_i^\top \right), \quad 0 \le i < N, \\ Y_N & = \xi, \end{cases}$$

where $\mathbb{E}_i(\cdot) := \mathbb{E}(\cdot | \mathcal{F}_{t_i})$.

- ✓ Because of $f_i(\mathbf{Y_{i+1}}, \ldots)$, **explicit scheme.**
- ✓ Differences with implicit scheme have not been really studied.

1) ODP scheme

One-step forward **D**ynamic **P**rogramming equation

$$\begin{cases} Y_i &= \mathbb{E}_i\left(Y_{i+1} + f_i(Y_{i+1}, Z_i)\Delta_i\right), \quad 0 \le i < N, \qquad Y_N = \xi. \\ \Delta_i Z_i &= \mathbb{E}_i\left(Y_{i+1}\Delta W_i^\top\right), \quad 0 \le i < N. \end{cases} \qquad \textbf{(ODP)}$$

X could be approximated by a path-wise approximation (the Euler scheme for SDE).

We do not discuss here the L_2-convergence of the discrete appproximation $(X_i, Y_i, Z_i)_{0\le i<N}$ towards the limit $(X_t, Y_t, Z_t)_{0\le t\le T}$, see [Zha04][GLW05][BT04].

✓ Usually, the L_2-rate is equal to $N^{\frac{1}{2}}$.

✓ The speed $N^{\frac{1}{2}}$ is achieved by taking appropriate choice of times grids according to fractional smoothness of ξ: see [GM10][GGG12]...

2) MDP scheme

From

$$Y_i = \mathbb{E}_i \left(Y_{i+1} + f_i(Y_{i+1}, Z_i)\Delta_i\right), \quad \Delta_i Z_i = \mathbb{E}_i \left(Y_{i+1}\Delta W_i^\top\right),$$

replugging Y_{i+1} and iterating over i until N gives the **M**ulti-Step forward **D**ynamic **P**rogramming equation:

$$\begin{cases} Y_i & = \mathbb{E}_i \left(\xi + \sum_{k=i}^{N-1} f_k(Y_{k+1}, Z_k)\Delta_k\right), \\ \Delta_i Z_i & = \mathbb{E}_i \left([\xi + \sum_{k=i+1}^{N-1} f_k(Y_{k+1}, Z_k)\Delta_k]\Delta W_i^\top\right). \end{cases} \qquad \textbf{(MDP)}$$

If no extra approximation is incorporated, then **ODP$\Longleftrightarrow$MDP.**

? differences regarding regression schemes?

3) Malliavin scheme: Mal.MDP

Based on Ma-Zhang representation theorem [MZ02] (Bismut type formula for the gradient): **under ellipticity conditions**, we have

$$\mathbf{Z_t} = \mathbb{E}^{\mathcal{F}_t}\left(\mathbf{g}(\mathbf{X_T})\mathbf{I_{t,T}} + \int_{\mathbf{t}}^{\mathbf{T}} \mathbf{f}(\mathbf{s}, \mathbf{X_s}, \mathbf{Y_s}, \mathbf{Z_s})\mathbf{I_{t,s}}\mathbf{ds}\right)$$

for some explicit stochastic integral $I_{t,s}$.

✓ In the case $X = BM$, $I_{t,s} = \frac{(W_s - W_t)^\top}{s-t}$.

✓ In general, $|I_{t,s}|_{L_2} \le c(s-t)^{-\frac{1}{2}}$ (singular weights but integrable).

✓ Leads to a discrete-time version

$$\begin{cases} Y_i &= \mathbb{E}_i\left(\xi + \sum_{k=i}^{N-1} f_k(Y_{k+1}, Z_k)\Delta_k\right), \\ Z_i &= \mathbb{E}_i\left(\xi I_{i,N} + \sum_{k=i}^{N-1} f_k(Y_{k+1}, Z_k) I_{i,k}\Delta_k\right). \end{cases} \qquad (\textbf{Mal.MDP})$$

✓ Discretization error analysis performed by Turkedjiev (2013): usual convergence rates.

To sum up, 3 DP equations

$$\begin{cases} Y_i &= \mathbb{E}_i\left(Y_{i+1} + f_i(Y_{i+1}, Z_i)\Delta_i\right), \quad 0 \le i < N, \qquad Y_N = \xi. \\ \Delta_i Z_i &= \mathbb{E}_i\left(Y_{i+1}\Delta W_i^\top\right), \quad 0 \le i < N. \end{cases} \qquad \textbf{(ODP)}$$

$$\begin{cases} Y_i &= \mathbb{E}_i\left(\xi + \sum_{k=i}^{N-1} f_k(Y_{k+1}, Z_k)\Delta_k\right), \\ \Delta_i Z_i &= \mathbb{E}_i\left([\xi + \sum_{k=i+1}^{N-1} f_k(Y_{k+1}, Z_k)\Delta_k]\Delta W_i^\top\right). \end{cases} \qquad \textbf{(MDP)}$$

$$\begin{cases} Y_i &= \mathbb{E}_i\left(\xi + \sum_{k=i}^{N-1} f_k(Y_{k+1}, Z_k)\Delta_k\right), \\ Z_i &= \mathbb{E}_i\left(\xi I_{i,N} + \sum_{k=i}^{N-1} f_k(Y_{k+1}, Z_k)I_{i,k}\Delta_k\right). \end{cases} \qquad \textbf{(Mal.MDP)}$$

First clear and intuitive differences:

- ✓ ODP computationally simpler (iteration and memory)
- ✓ Mal.MDP requires more to simulate under ellipticity

? Other differences: L_2-stability, empirical regression versions,...

Standing assumptions. f is (locally in time) Lipschitz in (y, z):
$\mathbf{|f_k(y,z) - f_k(y',z')| \le \frac{L_f}{(T-t_k)^{(1-\theta)/2}}(|y-y'| + |z-z'|)}$ for some $\theta \in (0, 1]$.

1ST COMPARISON: STABILITY RESULTS

Computation of theoretical regression functions (so far, no empirical projections, $M = +\infty$)

We are given a **family of projection operator** $\mathcal{P}_i^Y, \mathcal{P}_i^Z$ from $\mathbb{L}_2(\mathcal{F}_T)$ into a linear vector space of $\mathbb{L}_2(\mathcal{F}_{t_i})$:

$$\mathcal{S}_\mathbf{i} = \mathrm{Span}(\mathbf{\Phi_{k,i}} : \mathbf{1} \le \mathbf{k} \le \mathbf{K}_\mathcal{F}),$$

where $\Phi_{k,i} \in \mathbb{L}_2(\mathcal{F}_{t_i})$. **Property (usual iterated projection).** Projecting the r.v. U on $\mathcal{S}_i$ or its conditional expectation $\mathbb{E}_i(U)$ is the same.

PROOF.

$$\begin{aligned}\mathcal{P}_\mathbf{i}(\mathbf{U}) := \arg\inf_{U_i \in \mathcal{S}_i} |U - U_i|^2_{\mathbb{L}_2(\mathbb{P})} &= \arg\inf_{U_i \in \mathcal{S}_i} (|U - \mathbb{E}_i(U)|^2_{\mathbb{L}_2(\mathbb{P})} + |\mathbb{E}_i(U) - U_i|^2_{\mathbb{L}_2(\mathbb{P})}) \\ &= \arg\inf_{U_i \in \mathcal{S}_i} |\mathbb{E}_i(U) - U_i|^2_{\mathbb{L}_2(\mathbb{P})} = \mathcal{P}_\mathbf{i}(\mathbb{E}_\mathbf{i}(\mathbf{U})).\end{aligned}$$

□

Propagation of approximation errors in DP equations

$$\begin{cases} \widehat{Y}_i &= \mathcal{P}_{\mathbf{i}}^{\mathbf{Y}}\big(\widehat{Y}_{i+1} + \Delta_i f_i(\widehat{Y}_{i+1}, \widehat{Z}_i)\big), \\ \Delta_i \widehat{Z}_{i+1} &= \mathcal{P}_{\mathbf{i}}^{\mathbf{Z}}\big(\widehat{Y}_{i+1}\Delta W_i^\top\big), \end{cases} \qquad (\textbf{ODP+regression } M = +\infty)$$

$$\begin{cases} \check{Y}_i &= \mathcal{P}_{\mathbf{i}}^{\mathbf{Y}}\big(\xi + \sum_{k=i}^{N-1} \Delta_k f_k(\check{Y}_{k+1}, \check{Z}_k)\big), \\ \Delta_i \check{Z}_i &= \mathcal{P}_{\mathbf{i}}^{\mathbf{Z}}\big([\xi + \sum_{k=i+1}^{N-1} \Delta_k f_k(\check{Y}_{k+1}, \check{Z}_k)]\Delta W_i^\top\big). \end{cases} \qquad (\textbf{MDP+regression } M = +\infty)$$

Proposition. Consider the $L_\infty(L_2(\Omega), \{0 : N-1\})$ and $L_1(L_2(\Omega), \{0 : N-1\})$ norms: $\mathcal{E}_\infty(U) = \sup_{0\le i\le N-1} \mathbb{E}|U_i|^2$ and $\mathcal{E}_1(U) = \sum_{i=0}^{N-1} \mathbb{E}|U_i|^2\Delta_i$.

$$\textbf{(ODP)}\ \mathcal{E}_\infty(\widehat{Y} - Y) + \mathcal{E}_1(\widehat{Z} - Z) \le_c \sum_{\mathbf{0\le i\le N-1}} \mathbb{E}|Y_i - \mathcal{P}_i^Y(Y_i)|^2 + \sum_{i=0}^{N-1} \mathbb{E}|Z_i - \mathcal{P}_i^Z(Z_i)|^2\Delta_i,$$

$$\textbf{(MDP)} \begin{cases} \mathcal{E}_1(\check{Y} - Y) + \mathcal{E}_1(\check{Z} - Z) \le_c \sum_{\mathbf{0\le i\le N-1}} \mathbb{E}|Y_i - \mathcal{P}_i^Y(Y_i)|^2\boldsymbol{\Delta_i} + \sum_{i=0}^{N-1} \mathbb{E}|Z_i - \mathcal{P}_i^Z(Z_i)|^2\Delta_i, \\ \mathbb{E}|\check{\mathbf{Y}}_\mathbf{i} - \mathbf{Y}_\mathbf{i}|^2 \le_\mathbf{c} \mathbb{E}|\mathbf{Y}_\mathbf{i} - \mathcal{P}_\mathbf{i}^\mathbf{Y}(\mathbf{Y}_\mathbf{i})|^2 + \mathcal{E}_\mathbf{1}(\check{\mathbf{Y}} - \mathbf{Y}) + \mathcal{E}_\mathbf{1}(\check{\mathbf{Z}} - \mathbf{Z}). \end{cases}$$

$$\textbf{(Mal. MDP)}\begin{cases}\widetilde{Y}_i &= \mathcal{P}^{\mathbf{Y}}_{\mathbf{i}}\big(\xi + \sum_{k=i}^{N-1}\Delta_k f_k(\widetilde{Y}_{k+1},\widetilde{Z}_k)\big),\\ \widetilde{Z}_i &= \mathcal{P}^{\mathbf{Z}}_{\mathbf{i}}\big(\xi I_{i,N} + \sum_{k=i+1}^{N-1}\Delta_k f_k(\widetilde{Y}_{k+1},\widetilde{Z}_k)I_{i,k}\big).\end{cases}$$

Proposition. We have

$$\mathcal{E}_1(\widetilde{Y}-Y)+\mathcal{E}_1(\widetilde{Z}-Z) \le_c \sum_{\mathbf{0\le i\le N-1}} \mathbb{E}|Y_i - \mathcal{P}^Y_i(Y_i)|^2 \boldsymbol{\Delta}_{\mathbf{i}} + \sum_{i=0}^{N-1}\mathbb{E}|Z_i - \mathcal{P}^Z_i(Z_i)|^2\Delta_i,$$

$$\mathbb{E}|\widetilde{Y}_i - Y_i|^2 \le_c \mathbb{E}|Y_i - \mathcal{P}^Y_i(Y_i)|^2 + \mathcal{E}_1(\widetilde{Y}-Y) + \mathcal{E}_1(\widetilde{Z}-Z),$$

$$\mathbb{E}|\widetilde{\mathbf{Z}}_{\mathbf{i}} - \mathbf{Z}_{\mathbf{i}}|^2 \le_{\mathbf{c}} \mathbb{E}|\mathbf{Z}_{\mathbf{i}} - \mathcal{P}^{\mathbf{Z}}_{\mathbf{i}}(\mathbf{Z}_{\mathbf{i}})|^2 + \sum_{\mathbf{j=i+1}}^{\mathbf{N-1}} \frac{\mathbb{E}|\mathbf{Y}_{\mathbf{j}} - \mathcal{P}^{\mathbf{Y}}_{\mathbf{j}}(\mathbf{Y}_{\mathbf{j}})|^2 + \mathbb{E}|\mathbf{Z}_{\mathbf{j}} - \mathcal{P}^{\mathbf{Z}}_{\mathbf{j}}(\mathbf{Z}_{\mathbf{j}})|^2}{\sqrt{\mathbf{t}_{\mathbf{j}} - \mathbf{t}_{\mathbf{i}}}}\boldsymbol{\Delta}_{\mathbf{j}}.$$

Proof. Used unusual Gronwall lemma with non-bounded weights:

Lemma. Let $\alpha \ge 0, \beta > 0$. Assume for two sequences $\{u_l\}_{l\ge k}$ and $\{w_l\}_{l\ge k}$

$$u_j \le w_j + C\sum_{l=j+1}^{N-1}\frac{u_l\Delta_l}{(T-t_l)^{\frac12-\beta}(t_l-t_j)^{\frac12-\alpha}}.$$

Then, $u_j \le C'w_j + C'\sum_{l=j+1}^{N-1}\frac{w_l\Delta_l}{(T-t_l)^{\frac12-\beta}(t_l-t_j)^{\frac12-\alpha}}$. □

To sum up

✓ Different DP equations measure the error on (Y, Z) in different norms:

- average over i
- uniformly in i

✓ Different DP equations average differently the projection approximation error:

ODP $<$ **MDP** $<$ **Mal. MDP**

✓ But **Mal.MDP** requires ellipticity and additional weights to simulate.

2nd comparison: ordinary least-squares (OLS) (regression method)

To simplify the problem, computation of $\mathbb{E}(H|X = x)$ with

- ✓ $X \in \mathbb{R}^d$ (random observation=design) and $H \in \mathbb{R}$ (random response),
- ✓ data sample $D_N := (H_i, X_i)_{1\le i\le N}$, **i.i.d. realizations** of (H, X),
- ✓ approximation vector space: $\mathcal{F} = \mathrm{Span}(\boldsymbol{\Phi}_\mathbf{k} : \mathbf{1} \le \mathbf{k} \le \mathbf{K}_\mathcal{F})$,

➥ Nonparametric estimation of $m(x) = \mathbb{E}(H|X = x)$: **[Gyorfi *et al.* 2002, Tsybakov 2009, . . .]**, machine learning (distribution-free estimates).

- ✓ Empirical Regression function: $\mathbf{m_N} = \arg \inf_{\mathbf{f}\in\mathcal{F}} \frac{\mathbf{1}}{\mathbf{N}} \sum_{\mathbf{i=1}}^{\mathbf{N}} |\mathbf{H_i} - \mathbf{f}(\mathbf{X_i})|^{\mathbf{2}}.$

Theorem ([GKKW02]). Assume $\Sigma^2 = \sup_{x\in\mathbb{R}^{d_x}} \mathbb{V}\mathrm{ar}(H|X = x) < +\infty$. Then

$$\mathbb{E}\left[\|m_N - m\|^2_{L_2(\mu_N)}\right] \le \underbrace{\Sigma^2 \frac{K_\mathcal{F}}{N}}_{\text{estimation error}} + \underbrace{\inf_{f\in\mathcal{F}} \mathbb{E}\left[(f(X) - m(X))^2\right]}_{\text{approximation error}}.$$

Consequences: what to expect for the OLS+DP equations?

▷ For instance, in the MDP scheme, the global error estimates
$\mathcal{E}_1(Y-\check{Y})+\mathcal{E}_1(Z-\check{Z}) \leq_c \sum_{0\leq i\leq N-1}\mathbb{E}|Y_i-\mathcal{P}_i^Y(Y_i)|^2\Delta_i+\sum_{i=0}^{N-1}\mathbb{E}|Z_i-\mathcal{P}_i^Z(Z_i)|^2\Delta_i$
become (*in the best case*) for Y

$$\sum_{\mathbf{i=0}}^{\mathbf{N-1}}\mathbb{E}\Big(\frac{\mathbf{1}}{\mathbf{M}}\sum_{\mathbf{m=1}}^{\mathbf{M}}(\mathbf{y_i^M}(\mathbf{X_i^{i,m}})-\mathbf{y_i}(\mathbf{X_i^{i,m}}))^{\mathbf{2}}\Big)\mathbf{\Delta_i} \leq_{\mathbf{c}} \sum_{\mathbf{i=0}}^{\mathbf{N-1}}\left(\frac{\mathbf{K_i^Y}}{\mathbf{M}}+\inf_{\mathbf{f}\in\mathcal{F}_\mathbf{i}^\mathbf{Y}}\mathbb{E}|\mathbf{y_i}(\mathbf{X_i})-\mathbf{f}(\mathbf{X_i})|^{\mathbf{2}}\right)\mathbf{\Delta_i}$$

and similarly for Z.

Parameters tuning:

✓ If $y_i \in C^k$: using local polynomials of deg. $k-1$ on hypercubes of size δ gives

- app. error: $\inf_{f\in\mathcal{F}_i^Y}\mathbb{E}|y_i(X_i)-f(X_i)|^2 \leq c\delta^{2k}$ + tail-truncation error.
- $K_i^Y \sim \delta^{-d}$.

✓ To get N^{-1} for global error, $\delta \sim \mathbf{N^{-1/(2k)}}$ and $\mathbf{M} \sim \mathbf{NK_i^Y} \sim \mathbf{N^{1+d/(2k)}}$

✓ **Complexity** $\sim \mathbf{NM} \sim \mathbf{N^{2+d/(2k)}}$: trade-off dimension/smoothness

▷ As a comparison, using ODP, the global error estimates read (for Y)

$$\sum_{i=0}^{N-1} \mathbb{E}\Big(\frac{1}{M}\sum_{m=1}^{M}(y_i^M(X_i^{i,m}) - y_i(X_i^{i,m}))^2\Big) \le_c \sum_{i=0}^{N-1}\Big(\frac{K_i^Y}{M} + \inf_{f\in\mathcal{F}_i^Y}\mathbb{E}|y_i(X_i) - f(X_i)|^2 + \dots\Big)$$

Parameters tuning

With local polynomials of deg. $k-1$ on hypercubes of size δ and C^k-solution y_i:

✓ app. error: $\inf_{f\in\mathcal{F}_i^Y}\mathbb{E}|y_i(X_i) - f(X_i)|^2 \le c\delta^{2k}$ + tail-truncation error.

✓ $K_i^Y \sim \delta^{-d}$

✓ To get N^{-1} for global error, $\delta \sim N^{-\mathbf{2}/(2k)}$ and $M \sim N^{\mathbf{2}} K_i^Y \sim N^{\mathbf{2}+\mathbf{2}d/(2k)}$

➡ **Similar to double the dimension!! Huge impact.**

✓ Complexity $\sim NM \sim N^{1+\mathbf{2}+\mathbf{2}d/(2k)}$.

✓ The trade-off accuracy/complexity is worse compared to MDP.

What does "best case" means?

- ✓ Simulations used for regression at i have to be independent of other ones (no interdependency error).
  - ▸ N independent cloud of simulations
  - ▸ worse complexity $\sim N^{\mathbf{2}}M \sim N^{3+d/(2k)}$ for MDP or $N^{4+\mathbf{2}d/(2k)}$ for ODP
- ✓ What is the *interdependency cost* of having a unique cloud of independent simulations for all regressions at once? Depends on DP equations.

Theorem (uniform large/small deviation estimate).
Let $T_B\mathcal{F} := \{-B \vee f(\cdot) \wedge B : f \in \text{vector space } \mathcal{F}\}$ the truncated $\mathcal{F}$.
Then, there is a universal constant $c > 0$ s.t. for any $\mathcal{X}_1, \dots, \mathcal{X}_M$ i.i.d.r.v.

$$\mathbb{E}\Big[\sup_{g\in T_B\mathcal{F}}\Big(\int_{\mathbb{R}^d} g^2(x)\mathbb{P}\circ\mathcal{X}_1^{-1}(dx) - \frac{\mathbf{2}}{M}\sum_{m=1}^{M} g^2(\mathcal{X}_m)\Big)_+\Big] \le cB^2\frac{(\dim(\mathcal{F})+1)\log(cM)}{M},$$

$$\mathbb{E}\Big[\sup_{g\in T_B\mathcal{F}}\Big(\int_{\mathbb{R}^d} g^2(x)\mathbb{P}\circ\mathcal{X}_1^{-1}(dx) - \frac{\mathbf{1}}{M}\sum_{m=1}^{M} g^2(\mathcal{X}_m)\Big)_+\Big] \le cB^2\sqrt{\frac{(\dim(\mathcal{F})+1)\log(cM)}{M}}.$$

Since $\dim(\mathcal{F})/M$ should be $\sim N^{-1}$, it may much deteriorates the global error.

The case of quadratic driver

Principle

- ✓ plugging of a priori PDE estimates to force the Lipschitzianity
- ✓ transfer of space irregularity to time singularity

Assumptions

- ✓ For a given constant $c \geq 0$,

$$|f(t,x,y,z)| \leq c\,(1+|y|+|z|^2),$$
$$|f(t,x,y,z)-f(t,x,y',z')| \leq c\,(1+|z|+|z'|)(|y-y'|+|z-z'|)$$

for any $(t,x,y,y',z,z') \in [0,T]\times\mathbb{R}^d\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^d$.

- ✓ The terminal function g is θ-Hölder continuous and bounded.

Theorem ([DG06]). The associated semi-linear PDE is s.t.
$$\mathbf{(T-t)^{(1-\theta)/2}|\nabla u(t,x)\sigma(t,x)| \leq C_u}, \quad \forall \mathbf{(t,x) \in [0,T)\times\mathbb{R}^d}.$$

Corollary. Define the new driver

$$\bar{f}(t,x,y,z) := f\big(t,x,y,\varphi_t(z_1)\ \dots,\varphi_t(z_d)\big)$$

where $\varphi_t : \zeta \in \mathbb{R} \mapsto \varphi_t(\zeta) = \text{sign}(\zeta)\min\big(|\zeta|, \frac{C_u}{(T-t)^{(1-\theta)/2}}\big)$.

Then $\bar{f}(t, X_t, Y_t, Z_t) = f(t, X_t, Y_t, Z_t)$.

➠ equivalent to solve the BSDE with driver f or $\bar{f}$. in practice $C_u =?$

BSDE with globally Lipschitz driver locally in time. The driver $\bar{f}$ is now globally Lipschitz in y, z with a time-dependent constant:

$$|\bar{f}(t,x,y,z) - \bar{f}(t,x,y',z')| \le \frac{L_f}{(T-t)^{(1-\theta)/2}}(|y-y'| + |z-z'|),$$

for any $(y, y', z, z') \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^q \times \mathbb{R}^q$.

All previous discussions and error estimates apply to that setting.

In that analysis, the time grid can be of the form $t_k = T - T(1-k/N)^{1/\theta_\pi}$ with $\theta_\pi \in (0,1]$, like for fractional smoothness conditions.

Variance reductions

- ✓ Usually, the variance of OLS-ODP scheme is lower than OLS-MDP.
- ✓ Variance reduction techniques are complementary tools to speed-up any OLS schemes.

Two ways:

- ✓ taking a proxy (inspired by what is done in PDE)
- ✓ preliminary control variates (for automatic and data-driven improvement)

Numerical tests in progress (or ask Plamen!)

Shifting the driver around a proxy

Assumption (from user *expertise*). The BSDE solution (Y_t, Z_t) is close to $(v(t, X_t), \nabla v(t, X_t)\sigma(t, X_t))$, where v is explicit.

Then, v captures a significant part of the solution and it remains to solve numerically the BSDE residual $(Y_t^0, Z_t^0) := (Y_t - v(t, X_t), Z_t - \nabla v(t, X_t)\sigma(t, X_t))$ with data

- ✓ terminal function: $g(.) - v(T, .)$
- ✓ driver:

$$f^0(t, x, y, z) := f(t, x, y + v(t, x), z + \nabla v(t, x)\sigma(t, x)) - \partial_t v(t, x) - \mathcal{L}v(t, x).$$

Example. If $g(x) = (x - K)_+$ and the driver f comes from the two-interest rates BSDE [Ber95][EPQ97], take for v the Black-Scholes price with a given volatility and a given interest rate.

Example. v can be the solution with zero-driver $\partial_t v(t, x) + \mathcal{L}v(t, x) = 0$. It is known [GM10] that the time-regularity Z^0 behaves better than Z as $t \to T$.

Using Preliminary Control Variates

Generic method for speeding-up the regression computations.

We explain it in the context $\mathbb{E}(H|X)$ with $H = h(U)$.

Assumption. Some regression functions $\mathbb{E}(P_k(U)|X) = m_k(X)$ are known (called PCV): w.l.o.g. $\forall \mathbf{1} \leq \mathbf{k} \leq \mathbf{K}_{\text{pcv}} : \mathbb{E}[\mathbf{P_k}(\mathbf{U})|\mathbf{X}] = \mathbf{0}$.

▷ Heuristics about PCV

✓ No modification of the regression function: for any α

$$\mathbb{E}(H - \sum_{k=1}^{K_{\text{pcv}}} \alpha_k P_k(U)|X) = m(X).$$

✓ Variance reduction:

$$\begin{aligned}(\widehat{\alpha}_k)_{1\leq k\leq K_{\text{pcv}}} &= \arg\inf_{\alpha} \mathbb{E}\left[\left|H - \alpha \cdot P(U)\right|^2\right] \\ &= \arg\inf_{\alpha} \mathbb{V}\text{ar}\big(H - \alpha \cdot P(U)\big) = \arg\inf_{\alpha} \mathbb{E}\left[\mathbb{V}\text{ar}(\mathbf{H} - \alpha \cdot \mathbf{P}(\mathbf{U})|\mathbf{X})\right].\end{aligned}$$

Exemples of PCV inspired by stochastic processes

▷ **Localized functions $P_k(\cdot)$**

✓ **Dimension $d_z = d_x = 1$ and $\bigwedge$-functions.** Consider a Brownian motion W, and let $X := W_t$, $U := (W_t, W_T)$ for $0 < t < T$. Suppose we are interesting to compute $\mathbb{E}[h(W_T)|W_t]$ for some function h and for $t \leq T$. For $K_{\text{pcv}} = 2l + 1$, define

$$\mathbf{p^1_{x_k,\Delta}(x) = \left(1 - \left|\frac{x - x_k}{\Delta}\right|\right)_+,} \qquad \mathbf{x_k = (k - l - 1)\Delta,} \qquad \mathbf{\Delta = \frac{2\sqrt{T}}{l+1}}.$$

Define
$$\mathbf{P_k(W_t, W_T) := p^1_{x_k,\Delta}(W_T)} - \underbrace{\mathbb{E}\left[p^1_{x_k,\Delta}(W_T)|W_t\right]}_{\text{explicit formula}}.$$

✓ **Dimension $d_z = d_x > 1$ and $\bigwedge$-functions.** Immediate extension to BM $W = (W^1, \ldots, W^d)$:

$$\mathbf{P_{k_1,\ldots,k_d}(W_t, W_T) = \prod_{i=1}^{d} p^1_{x_{k_i},\Delta}(W^i_T) - \prod_{i=1}^{d} \mathbb{E}\left[p^1_{x_{k_i},\Delta}(W^i_T)|W^i_t\right]}.$$

▷ Non-localized functions $P_k(\cdot)$: polynomials and martingales

✓ Let W be a scalar **Brownian Motion** and let $(H_k)_k$ be the Hermite polynomials: set $X = W_t, U = (W_t, W_T)$ and

$$\mathbf{P_k(U) = T^{k/2} H_k\left(\frac{W_T}{\sqrt{T}}\right) - t^{k/2} H_k\left(\frac{W_t}{\sqrt{t}}\right).}$$

Straightforward multidimensional extension.

✓ Let N be a **Poisson process** and let $(C_k)_k$ be the Charlier polynomials: set $X = N_t, U = (N_t, N_T)$ and

$$\mathbf{P_k(U) = C_k\left(N_T, T\right) - \left(\frac{t}{T}\right)^k C_k\left(N_t, t\right).}$$

✓ For many other distributions and stochastic processes (**affine processes, processes with quadratic diffusion coefficients, Lévy-driven SDEs with affine vector fields...**), see **[Schoutens '01, Cucheiro *et al.* '12]**

✓ ...

The PCV empirical regression algorithm

Define

✓ the **PCV parameter** set $\mathcal{A}$ (non empty closed convex),

✓ **PCV functions**: $\mathcal{G}^{\mathcal{A}} = \left\{\sum_{k=1}^{K_{\rm pcv}} \alpha_k P_k : \alpha \in \mathcal{A}\right\}$ (non empty closed convex),

✓ the **PCV-modified response**: $H^{\alpha} = H - \sum_{k=1}^{K_{\rm pcv}} \alpha_k P_k$ for $\alpha \in \mathcal{A}$.

▷ **Two-steps algorithm:**

Step 1. Variance Reduction: $(\widetilde{\alpha}_k)_{1\leq k\leq K_{\rm pcv}} = \arg\inf_{\alpha\in\mathcal{A}} \frac{1}{N}\sum_{i=1}^{N}\left|H_i - \sum_{k=1}^{K_{\rm pcv}} \alpha_k P_k(U_i)\right|^2.$

Step 2. Least Squares Regression:

$$(\widetilde{\beta}_k)_{1\leq k\leq K_{\mathcal{F}}} = \arg\inf_{(\beta_k)_k} \frac{1}{N}\sum_{i=1}^{N}\left|H_i - \sum_{k=1}^{K_{\rm pcv}} \widetilde{\alpha}_k P_k(U_i) - \sum_{k=1}^{K_{\mathcal{F}}} \beta_k \Phi_k(X_i)\right|^2.$$

➠ set $\widetilde{m}_N = \sum_{k=1}^{K_{\mathcal{F}}} \widetilde{\beta}_k \Phi_k.$

Technical assumption

- ✓ $\|P_k\|_\infty \le 1$ and $\exists L \ge 1 : \|h\|_\infty \le L$.
- ✓ We choose $\mathcal{A} := \{\alpha \in \mathbb{R}^{K_{\mathrm{pcv}}} : \sum_{i=1}^{K_{\mathrm{pcv}}} |\alpha_i| \le L\}$.

Theorem. Denote by $\widehat{\alpha}$ the optimal PCV parameter and set $\Sigma^2(\widehat{\alpha}) = \sup_{x\in\mathbb{R}^{d_x}} \mathbb{V}\mathrm{ar}(H^{\widehat{\alpha}}|X = x) < +\infty$. Then, for any $\rho > 0$,

$$\mathbb{E}\left[\|\widetilde{\mathbf{m}}_{\mathbf{N}} - \mathbf{m}\|^{\mathbf{2}}_{\mu_{\mathbf{N}}}\right] \le (\mathbf{1}+\rho^{-\mathbf{1}})\mathbf{L}^{\mathbf{4}}\left\{\frac{\mathbf{c_1} + (\mathbf{c_2} + \mathbf{c_3}\log(\mathbf{N}))(\mathbf{K}_{\mathrm{pcv}} + \mathbf{1})}{\mathbf{N}}\right\}$$
$$+ (1+\rho)\frac{K_{\mathcal{F}}}{N}\boldsymbol{\Sigma}^{\mathbf{2}}(\widehat{\alpha}) + (1+\rho)\mathbb{E}\left\{\inf_{f\in\mathcal{F}_N}\|f - m\|^2_{\mu_N}\right\}$$

for some universal constants c_1, c_2 et c_3.

- ✓ We still achieve the best approximation error up to the factor $(1+\rho)$.
- ✓ Reduction of the estimation error : $\boldsymbol{\Sigma}^{\mathbf{2}}(\mathbf{0}) \Rightarrow \boldsymbol{\Sigma}^{\mathbf{2}}(\widehat{\alpha})$.
- ✓ Additional term (error estimation on $\widetilde{\alpha}$): $\frac{\mathbf{c_1}}{\mathbf{N}} + \frac{(\mathbf{c_2}+\mathbf{c_3}\log(\mathbf{N}))(\mathbf{K}_{\mathrm{pcv}}+\mathbf{1})}{\mathbf{N}}$.
- ✓ Better choice: $K_{\mathrm{pcv}} \ll K_{\mathcal{F}}$.

Optimal variance for piecewise constant functions

Theorem. If Φ_k are piecewise constants on statistically equivalent blocks (containing approximately the same number of data) or approximately equi-probabilistic blocks (defined by a constant $c_I \geq 1$), then

$$\begin{aligned}\mathbb{E}[\|\widetilde{m}_N - m\|^2_{\mu_N}] \leq &(1+\rho^{-1})L^4 \left\{\frac{c_1 + (c_2 + c_3 \log(N))(K_{\text{pcv}} + 1)}{N}\right\} \\ &+ (1+\rho)c_I \frac{K_{\mathcal{F}}}{N} \inf_{\alpha \in \mathcal{A}} \mathbb{E}\left[\mathbb{V}\text{ar}(\mathbf{H}^\alpha | \mathbf{X})\right] \\ &+ (1+\rho)\mathbb{E}\left\{\inf_{\Phi \in \mathcal{F}_N} \|\Phi - m\|^2_{\mu_N}\right\}.\end{aligned}$$

Example in dimension $d_x = 2$

- **Goal:** estimate $m(x) = \mathbb{E}[h(W_2, B_2)|W_1 = x, B_1 = x]$ where

$$\mathbf{h(W_2, B_2) = e^{-\frac{W_2^2+B_2^2+\rho W_2 B_2}{2}}} \qquad \text{with } \rho = \mathbf{0.5}.$$

- **Model:** $U = (W_1, B_1, W_2, B_2)$ with (W, B) BM.

- **PCV:** choose $K_{\text{pcv}} = (2l + 1)(2l + 1)$ and set $\Psi(x) = (1 - |x|)_+$and define

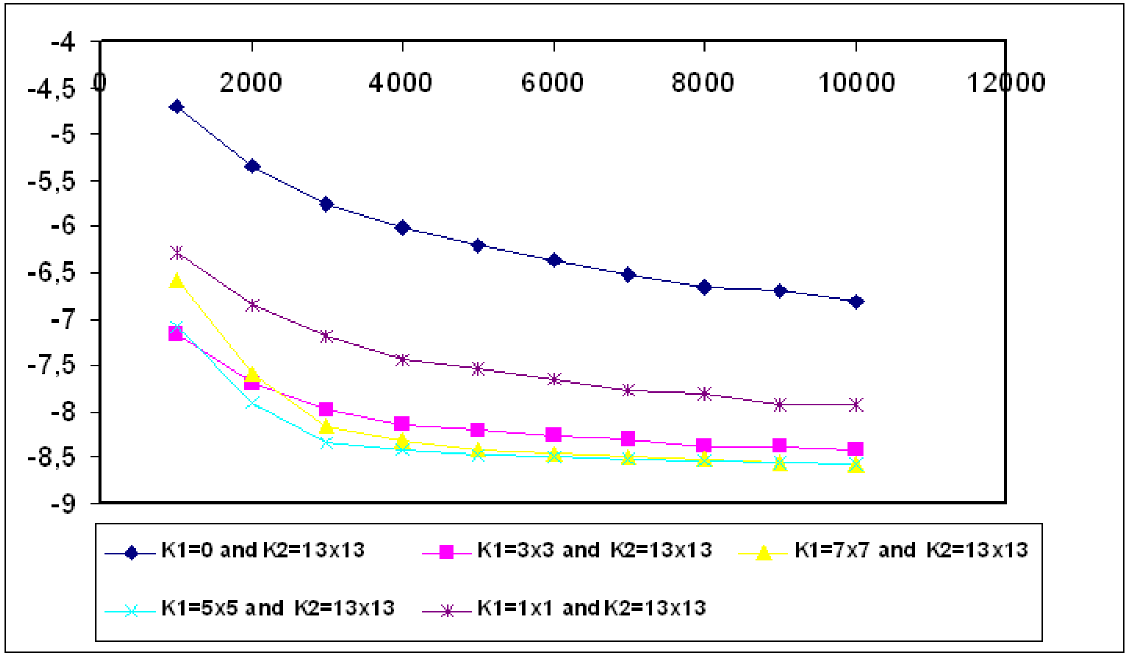
$$\mathbf{P_k(W_1, W_2) = \Psi(\frac{W_2 - (k - l - 1)\Delta}{\Delta}) - \mathbb{E}[\Psi(\frac{W_2 - (k - l)\Delta}{\Delta})|W_1]},$$

where $\Delta = \frac{2\sqrt{2}}{l+1}$. Define $\mathbf{Q_{i,j}(U) = P_i(W_1, W_2)P_j(B_1, B_2)}$, for $1 \leq i, j \leq 2l + 1$.

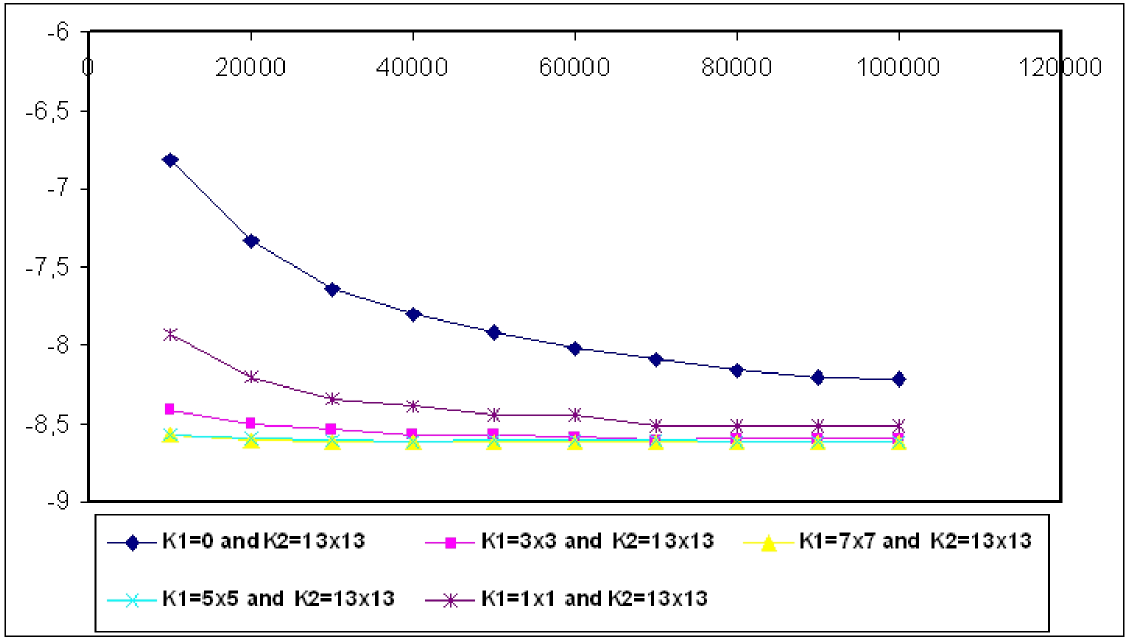
- **Regression basis functions:**

$$\mathbf{\Phi_{i,j}(W_1, B_1) = \Psi(\frac{W_1 - (i - r - 1)\Delta}{\Delta})\Psi(\frac{B_1 - (j - r - 1)\Delta}{\Delta})}$$

for $1 \leq i, j \leq 2r + 1$ with $K_{\mathcal{F}} = (2r + 1)(2r + 1)$.



Empirical error (in log-scale) as a function of $N \leq 12000$.



Empirical error (in log-scale) as a function of $N \geq 10000$.

$PCV\ (N = 1000) \approx Standard\ (N = 20000)$

➠ **Efficiency improvement $\approx$ 20**

Conclusion, perspectives, open problems

▷ **Mathematical aspects.**

✓ Schemes much sensitive to the dimension and the regularity of solution (estimated by a priori PDE estimates).

✓ Most efficient (theoretically) schemes are those based on MDP, but the effect of larger variance is not yet deeply analyzed.

▷ **Programming and algorithmic aspects.**

✓ Local polynomials can be implemented very efficiently by taking advantage of local basis. Crucial trick to make it fast.

✓ Storing in computer memory all coefficients for MDP may become a issue, more critical than for ODP.

✓ Good idea to improve schemes by incorporating theoretical information about the true solution (proxy, upper bound to stabilize the estimates, refined hypercubes near singularity)

✓ Data-driven basis (see experiments in [Lem05], [BW12]). Not yet fully covered by theoretical results.

✓ Parallel computations (Labart-Lelong '13)

▷ **Works in progress.**

✓ Mal. MDP with one single cloud of simulations.

✓ Large dimension and effective dimension of a BSDE regression problems.

✓ Non-linear least-squares regression and sparse representations.

✓ Jump components.

▷ **Open problems.**

✓ How to take advantage of the knowledge of fractional smoothness conditions?

✓ How to design optimal stochastic discretization grids for BSDEs? see [GL12] for optimal discretization stochastic integrals.

✓ BSDE with space constraints (RBSDE and switching, random terminal time).

References

[BD07] C. Bender and R. Denk. A forward scheme for backward SDEs. *Stochastic Processes and their Applications*, 117(12):1793–1823, 2007.

[Ber95] Y.Z. Bergman. Option pricing with differential interest rates. *Rev. of Financial Studies*, 8(2):475–500, 1995.

[BT04] B. Bouchard and N. Touzi. Discrete time approximation and Monte Carlo simulation of backward stochastic differential equations. *Stochastic Processes and their Applications*, 111:175–206, 2004.

[BW12] B. Bouchard and X. Warin. Monte-Carlo valuation of American options: facts and new algorithms to improve existing methods. Carmona, René A. (ed.) et al., Numerical methods in finance. Selected papers based on the presentations at the workshop, Bordeaux, France, June 2010. Berlin: Springer. Springer Proceedings in Mathematics 12, 215-255 (2012)., 2012.

[CKRT12] C. Cuchiero, M. Keller-Ressel, and J. Teichmann. Polynomial processes and their applications to mathematical finance. *Finance and Stochastics*, 2012.

[DG06] F. Delarue and G. Guatteri. Weak existence and uniqueness for forward-backward SDEs. *Stochastic Processes Appl.*, 116(12):1712–1742, 2006.

[EPQ97] N. El Karoui, S.G. Peng, and M.C. Quenez. Backward stochastic differential equations in finance. *Math. Finance*, 7(1):1–71, 1997.

[GGG12] C. Geiss, S. Geiss, and E. Gobet. Generalized fractional smoothness and L_p-variation of BSDEs with non-Lipschitz terminal condition. *Stochastic Processes and their*

Applications, 122(5):2078–2116, 2012.

[GKKW02] L. Gyorfi, M. Kohler, A. Krzyzak, and H. Walk. *A distribution-free theory of nonparametric regression*. Springer Series in Statistics, 2002.

[GL12] E. Gobet and N. Landon. Almost sure optimal hedging strategy. *In revision, Annals of Applied Probability*, 2012.

[GLW05] E. Gobet, J.P. Lemor, and X. Warin. A regression-based Monte Carlo method to solve backward stochastic differential equations. *Annals of Applied Probability*, 15(3):2172–2202, 2005.

[GM10] E. Gobet and A. Makhlouf. L_2-time regularity of BSDEs with irregular terminal functions. *Stochastic Processes and their Applications*, 120:1105–1132, 2010.

[Lem05] J.P. Lemor. *Approximation par projections et simulations de Monte-Carlo des équations différentielles stochastiques rétrogrades.* PhD thesis, Ecole Polytechnique, http://www.imprimerie.polytechnique.fr/Theses/Files/lemor.pdf, 2005.

[MZ02] J. Ma and J. Zhang. Representation theorems for backward stochastic differential equations. *Ann. Appl. Probab.*, 12(4):1390–1418, 2002.

[Sch00] W. Schoutens. *Stochastic Processes and Orthogonal Polynomials*, volume 146 of *Lecture Notes in Statistics*. Springer-Verlag, New-York, 2000.

[Tsy09] A.B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

[Zha04] J. Zhang. A numerical scheme for BSDEs. *Ann. Appl. Probab.*, 14(1):459–488, 2004.