Non-intrusive stratified resampling and regression method for dynamic programming problem

Gang Liu
gang.liu@polytechnique.edu

joint work with
Emmanuel Gobet and Jorge P. Zubelli

CMAP, Ecole Polytechnique

March 25, 2016

Overview

Introduction

Dynamic programming problem:

$$Y_i = \mathbb{E}\left[g_i(Y_{i+1}, \ldots, Y_N, X_i, \ldots, X_N) \mid X_i\right], \quad i = N-1, \ldots, 0,$$
$$Y_N = g_N(X_N),$$

We want to estimate the function y_i such that $Y_i = y_i(X_i)$.

Examples

Optimal Stoppting: $V_i = \operatorname{ess\,sup}_{\tau \in \mathcal{T}_{i,N}} \mathbb{E}\left[f_\tau(X_\tau)\right]$

$$
\begin{aligned}
Y_i &= \mathbb{E}\left[V_{i+1} | X_i\right] \\
Y_i &= \mathbb{E}\left[\max(Y_{i+1}, f_{i+1}(X_{i+1})) \mid X_i\right] \\
Y_{N-1} &= \mathbb{E}\left[f_N(X_N) | X_{N-1}\right]
\end{aligned}
$$

BSDE:

$$
\begin{aligned}
\mathcal{Y}_t &= g(W_1) + \int_t^1 f(s, \mathcal{Y}_s, W_s)\mathrm{d}s - \int_t^1 \mathcal{Z}_s \mathrm{d}W_s \\
\mathcal{Y}_t &= \mathbb{E}\left[g(W_1) + \int_t^1 f(s, \mathcal{Y}_s, W_s)\mathrm{d}s \mid W_t\right] \\
Y_i &= \mathbb{E}\left[g(X_N) + \frac{1}{N}\sum_{j=i+1}^{N} f_j(Y_j, X_j) | X_i\right]
\end{aligned}
$$

Numerical resolution

Usual approach:

- given $(\hat{y}_j)_{j=i+1,\cdots,N}$, we want to find $\hat{y}_i$ estimating y_i
- use a large number of simulation(or resimulation) of $X_{i:n}$
- $\hat{y}_i$ is an estimator for the conditional expectation

$$\mathbb{E}\left[g_i(\hat{y}_{i+1}(X_{i+1}), \ldots, \hat{y}_N(X_N), X_i, \ldots, X_N) \mid X_i\right]$$

  using a global statistical regression with a large function dictionnary

Q: What if model paramters for X are (partially) unknown and only a small set of historical data is available?
A: path reconstruction(resampling) + stratification

Stratified resampling method: stratification

We divide the whole space into K strata $(\mathcal{H}_k)_{1\le k\le K}$ such that

$$\mathcal{H}_k \cap \mathcal{H}_l = \emptyset \quad \text{for } k \neq l, \qquad \bigcup_{k=1}^{K} \mathcal{H}_k = \mathbb{R}^d.$$

Given a probability measure ν on $\mathbb{R}^d$, denote its restriction on $\mathcal{H}_k$ by

$$\nu_k(\mathrm{d}x) := \frac{1}{\nu(\mathcal{H}_k)} 1_{\mathcal{H}_k}(x)\nu(\mathrm{d}x).$$

We shall use ν_k to sample initial points on $\mathcal{H}_k$ and perform local regression on $\mathcal{H}_k$ to get estimation of $y_i(x)1_{x\in\mathcal{H}_k}$

Stratified resampling method: resampling

We are given M independent observation of X, with M small

$$\text{Data: } (X_i^m : 0 \leq i \leq N)_{1 \leq m \leq M}$$

Noise extraction and resampling

- We assume $X_j^{i,x} = \theta_{i,j}(x, U)$, $\theta_{i,j}$ are known such that we can extract out the value of U
- to simulate a trajectory of $X_{i:N}$ starting at x', we compute $\theta_{i,j}(x', U)$

Remark: to apply this procedure demands much less information than the full detail of the underlying model

Stratified resampling method: resampling

- $X_t = x_0 + \int_0^t \mu_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}W_s$

$$U = (X_{i+1} - X_i)_{0 \le i \le N-1} \qquad \theta_{ij}(x, U) = x + \sum_{i \le k < j} U_k$$

- $X_t = x_0 \exp\left(\int_0^t \mu_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}W_s\right)$

$$U = \left(\log(\frac{X_{i+1}}{X_i})\right)_{0 \le i \le N-1} \qquad \theta_{ij}(x, U) := x \prod_{i \le k < j} \exp(U_k)$$

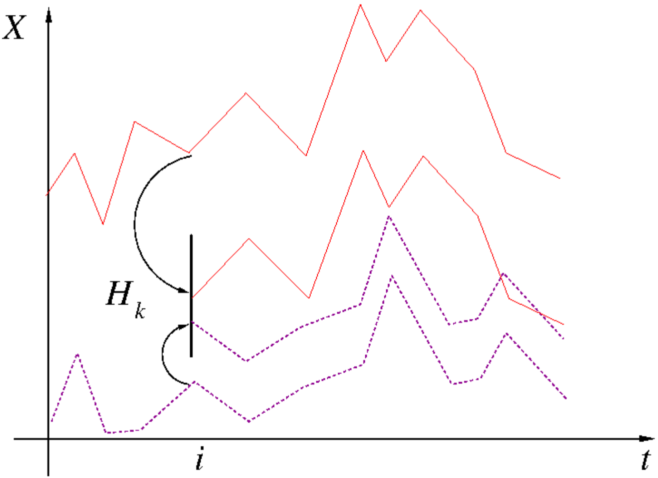
- $X_t = x_0 - \int_0^t a(X_s - \bar{X}_s)\mathrm{d}s + \sigma_s W_t$

$$U_{i,j} = X_j - e^{-a(t_j - t_i)} X_i \qquad \theta_{ij}(x, U) := e^{-a(t_j - t_i)} x + U_{i,j}$$

$\mu_s, \sigma_s, \bar{X}_s$ are unknown deterministic function of s, X_i stands for X_{t_i}

Stratified resampling method: resampling

Using given data to generate paths starting from a given point, in an additive model(e.g. Lévy process)



Ordinary Least Sqaure regression

Definition of OLS(Ordinary Least Square operator): for a given function f, a function dictionary $\mathcal{L}_k$ and M points $(Y_m)_{1\leq m\leq M}$

$$\mathrm{OLS}(f, \mathcal{L}_k, (Y_m)_{1\leq m\leq M}) := \arg\inf_{\varphi\in\mathcal{L}_k} \frac{1}{M}\sum_{m=1}^{M} |f(Y_m) - \varphi(Y_m)|^2.$$

OLS is linear and contracting w.r.t L^2 norm, and it interchanges with conditional expectation(in a sense to be precised).

Stratified resampling method

$$Y_i = \mathbb{E}\left[g_i(Y_{i+1}, \ldots, Y_N, X_i, \ldots, X_N) \mid X_i\right], \quad i = N-1, \ldots, 0,$$
$$Y_N = g_N(X_N),$$

Suppose we already have an estimation $y_{i+1}^{(M)}$ for y_{i+1}. For each $\mathcal{H}_k$, we will do the following:

- sample M i.i.d. copy of $(X_i^m)_{1\leq m\leq M}$ according to the law ν_k
- constructing M paths starting from these points, denoted $X_{i:N}^{1:M}$, using the resampling formula
- compute $\psi_i^{(M),k} = \mathrm{OLS}(S^{(M)}, \mathcal{L}_k, X_{i:N}^{1:M})$ with $S^{(M)}(x_{i:N}) := g_i(y_{i+1}^{(M)}(x_{i+1}), \ldots, y_N^{(M)}(x_N), x_{i:N})$
- set $y_i^{(M),k} = \mathcal{T}_{|y_i|_\infty}(\psi_i^{(M),k})$

Finally we get an estimation $y_i^{(M)}$ for y_i: $y_i^{(M)} = \sum_{k=1}^{K} y_i^{(M),k} 1_{\mathcal{H}_k}$

Error analysis: $g_i = g_i(Y_{i+1}, X_{i:N})$

All the paths used in our method are reconstructed from the initially given root sample, which complicates the error analysis.

We shall need several assumptions:

- g_i is bounded and is Lipschitz w.r.t. y_{i+1}
- $\exists C_\nu$ s.t. $\int_{\mathbb{R}^d} \mathbb{E}\left[\varphi^2(X_{i+1}^{i,x})\right] \nu(\mathrm{d}x) \leq C_\nu \int_{\mathbb{R}^d} \varphi^2(x)\nu(\mathrm{d}x)$.
- assumptions on covering number of the function dictionary $\mathcal{L}_k$

Remarks on assumptions

- If g_i is not bounded, we may use truncation number $\to +\infty$
- Under mild conditions on the model of X, we can find appropriate ν
- Assumption on covering number is mainly used to link error under empirical measure to error under exact sampling measure.

Error analysis: $g_i = g_i(Y_{i+1}, X_{i:N})$

Define $T_{i,k} := \inf_{\varphi\in\mathcal{L}_k} |y_i - \varphi|^2_{\nu_k}$ and take $\dim(\mathcal{L}_k) = \dim(\mathcal{L})$

With previous assumptions we can prove that

Theorem

$$\mathbb{E}\left[|y_i^{(M)} - y_i|^2_\nu\right]$$

$$\leq 4(1+\varepsilon)L^2_{g_i} C_\nu \mathbb{E}\left[|y_{i+1}^{(M)} - y_{i+1}|^2_\nu\right] + 2\sum_{k=1}^{K} \nu(H_k) T_{i,k} + C\frac{dim(\mathcal{L})\log(M)}{M}$$

Error analysis: multi-step scheme for BSDE

Assumptions:

- f_{t_i} is bounded and is Lipschitz w.r.t. y, g is bounded
- $\exists C_\nu$ s.t. $\int_{\mathbb{R}^d} \mathbb{E}\left[\varphi^2(X_j^{i,x})\right] \nu(\mathrm{d}x) \leq C_\nu \int_{\mathbb{R}^d} \varphi^2(x)\nu(\mathrm{d}x)$.
- assumptions on covering number of the function dictionary $\mathcal{L}_k$

$$\mathbb{E}\left[|y_i^{(M)} - y_i|_\nu^2\right] \leq C\frac{\log(M)\dim(\mathcal{L})}{M} +$$

$$2\sum_{k=1}^{K} \nu(H_k)\, T_{i,k} + 8\frac{1}{N}\sum_{j=i+1}^{N-1} L_{f_j}^2 \left(C_\nu \mathbb{E}\left[|y_j^{(M)} - y_j|_\nu^2\right]\right)$$

Travel agency problem

A travel agency wants to lanch a promotion, its profit is affected by the temperature and the exchange rate. We want to compute $v(X_0^1, X_0^2)$ defined by

$$
\begin{aligned}
&\operatorname*{ess\,sup}_{\tau\in\mathcal{T}} \mathbb{E}\left[q((\tau-0.25)^2\times 240+X_\tau^1)e^{-|\tau-1/6|}\left(\underline{c}-c(e^{X^2_{\tau+1/12}})\right)\right]\\
=&\operatorname*{ess\,sup}_{\tau\in\mathcal{T}} \mathbb{E}\left[q((\tau-0.25)^2\times 240+X_\tau^1)e^{-|\tau-1/6|}\left(\underline{c}-\mathbb{E}\left[c(e^{X^2_{\tau+1/12}})\mid X_\tau^2\right]\right)\right]
\end{aligned}
$$

where $\mathcal{T}=\{\frac{k}{48}, k=0,1,\cdots,24\}$

$$
\begin{aligned}
\mathrm{d}X_t^1 &= -aX_t^1\mathrm{d}t+\sigma_1\mathrm{d}W_t, X_0^1=0,\\
X_t^2 &= -\frac{\sigma_2^2}{2}t+\sigma_2 B_t.
\end{aligned}
$$

Travel agency problem

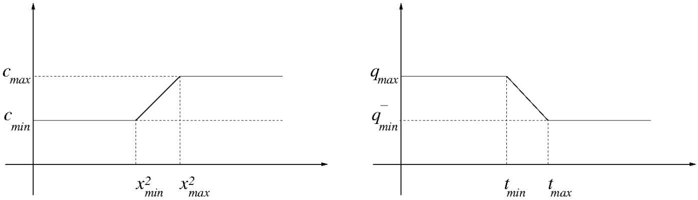


Figure: Pictorial description of the cost function c (left) and of the campaign effectiveness q (right).

Parameters: $a = 2, \sigma_1 = 10, \sigma_2 = 0.2, \underline{c} = 3, x^2_{min} = e^{-0.5}, x^2_{max} = e^{0.5}, c_{min} = 1, c_{max} = c_{min} + x^2_{max} - x^2_{min}, t_{min} = 0, t_{max} = 15, q_{min} = 1, q_{max} = 4.$

$\mu(\mathrm{d}x) = \frac{k}{2}(1 + |x|)^{-k-1}, k = 6$ to resample points for X^1
$\nu(dx) = \frac{1}{2}e^{-|x|}dx$ to resample points for X^2.

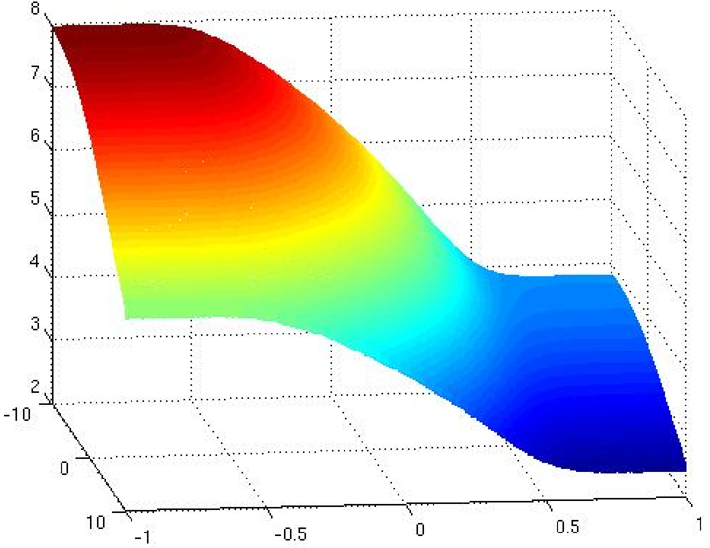


Figure: Piecewise constant estimation with $M = 320$, $K = 200$

Travel agency problem

The squared $L^2(\mu \otimes \nu)$ norm of our reference estimation is 32.0844

simple regression	$K = 10$	$K = 20$	$K = 50$	$K = 100$
$M = 20$	0.1827	0.0512	0.0349	0.0269
$M = 40$	0.1982	0.0361	0.0249	0.0114
$M = 80$	0.2063	0.0325	0.0051	0.0047
$M = 160$	0.1928	0.0264	0.0058	0.0067

Average squared L^2 errors with 20 macro runs.

nested regression	$K = 10$	$K = 20$	$K = 50$	$K = 100$
$M = 20$	0.1711	0.0458	0.0436	0.0252
$M = 40$	0.1648	0.0361	0.0130	0.0169
$M = 80$	0.1534	0.0273	0.0109	0.0085
$M = 160$	0.1510	0.0296	0.0048	0.0058

Average squared L^2 errors with 20 macro runs.

Population distribution: FKPP equation

$$\partial_t u + \partial_{xx}^2 u + au(1-u) = 0 \ , \quad u(T,x) = h(x),\, x \in \mathbb{R}, t \le T \ ,$$

A travelling wave solution:

$$h(x) := \frac{1}{\left(1 + C\exp\left(\pm\frac{\sqrt{6a}}{6}x\right)\right)^2}$$

then

$$u(t,x) = \frac{1}{\left(1 + C\exp\left(\frac{5a}{6}(t-T) \pm \frac{\sqrt{6a}}{6}x\right)\right)^2}$$

The probabilistic formulation:

$$\begin{aligned} dP_s &= \sqrt{2}dW_s \ , \\ dY_s &= -f(Y_s)ds + Z_s dW_s, \text{ where} f(x) = ax(1-x) \\ Y_T &= u(T, P_T) = h(P_T) \ . \end{aligned}$$

Then, the process $Y_t = \mathbb{E}\left[Y_T + \int_t^T f(Y_s)ds | P_t\right]$ satisfies $Y_t = u(t, P_t)$.

Results in dimension 1

$T = 1$, time discretization $t_i = \frac{i}{N} T, 0 \leq i \leq N$ with $N = 10$. We divide the real line $\mathbb{R}$ into K subintervals $(I_i)_{1 \leq i \leq K}$, piecewise constant estimation on each interval. We approximate the squared $L^2(\nu)$ error of our estimation by

$$\sum_{1 \leq k \leq K} |u(0, y_k) - \hat{u}(0, y_k)|^2 \nu(I_k)$$

We take $\nu(dx) = \frac{1}{2} e^{-|x|} dx$. The squared $L^2(\nu)$ norm of $u(0, y)$ is around 0.25.

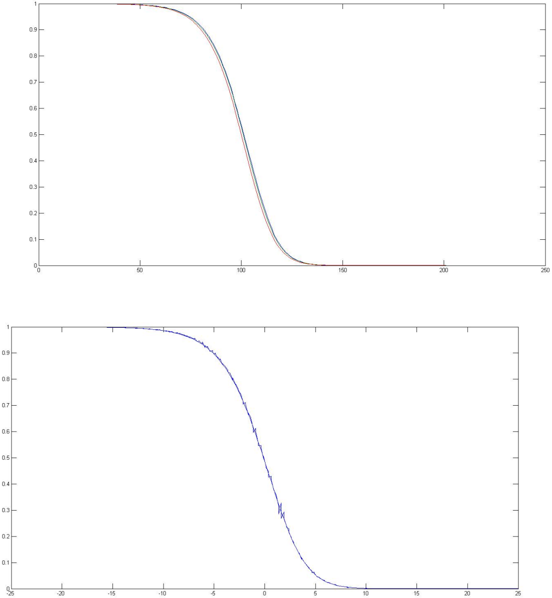


Figure: Piecewise constant and linear estimation, $M = 50$ and $K = 200$

Results in dimension 1

one-step	$K = 10$	$K = 20$	$K = 50$	$K = 100$	K = 200	K = 400
$M = 20$	0.0993	0.0253	0.0038	0.0014	0.0014	0.0019
$M = 40$	0.0997	0.0252	0.0034	9.01e-04	5.16e-04	6.17e-04
$M = 80$	0.0993	0.0249	0.0029	6.15e-04	3.92e-04	3.91e-04
$M = 160$	0.0990	0.0248	0.0029	3.15e-04	1.57e-04	1.71e-04
$M = 320$	0.0990	0.0248	0.0028	2.47e-04	1.02e-04	1.19e-04
$M = 640$	0.0990	0.0246	0.0028	2.26e-04	5.46e-05	4.94e-05

Average squared L^2 errors with 50 macro runs.

multi-step	$K = 10$	$K = 20$	$K = 50$	$K = 100$	K = 200	K = 400
$M = 20$	0.0484	0.0066	0.0017	0.0015	0.0011	0.0013
$M = 40$	0.0488	0.0058	8.45e-04	5.81e-04	6.35e-04	5.68e-04
$M = 80$	0.0478	0.0053	4.33e-04	2.96e-04	3.45e-04	4.06e-04
$M = 160$	0.0481	0.0051	2.98e-04	2.23e-04	1.71e-04	1.08e-04
$M = 320$	0.0479	0.0051	1.79e-04	6.48e-05	8.38e-05	1.04e-04
$M = 640$	0.0478	0.0050	1.50e-04	6.49e-05	6.66e-05	5.70e-05

Average squared L^2 errors with 50 macro runs.

Two dimensional case

$$\partial_t W + \sum_{1 \le i,j \le d} A_{ij} \partial_{y_i} \partial_{y_j} W + aW(1-W) = 0 \ , \ t \le T \ \text{,and} \ y \in \mathbb{R}^d \ .$$

A is *positive-definite* constant $d \times d$ matrix. With the final condition

$$W(T,y) = h(y' \mathbf{\Sigma}^{-1} \theta) \ ,$$

where $\mathbf{\Sigma} = \mathbf{\Sigma}' = \sqrt{A}$ and θ is arbitrary unit vector., the solution is

$$W(t,y) := u(t, y' \mathbf{\Sigma}^{-1} \theta)$$

Two dimensional case

$T = 1$ and $t_i = \frac{i}{N}T, 0 \le i \le N$ with $N = 10$. The real line $\mathbb{R}$ is divided into K subintervals. We take $\Sigma = [1, \beta; \beta, 1]$ with $\beta = 0.25$ and $\theta = \frac{[1;1]}{\sqrt{2}}$. We implement piecewise constant estimation on each rectangle $I_i \times I_j$. Then finally we get an estimator $\hat{W}(0, y)$. Then we approximate the squared $L^2(\nu \otimes \nu)$ error by

$$\sum_{1 \le k_1 \le K, 1 \le k_2 \le K} |W(0, y_{k_1}, y_{k_2}) - \hat{W}(0, y_{k_1}, y_{k_2})|^2 \nu \otimes \nu(I_{k_1} \times I_{k_2})$$

one-step	$K = 10$	$K = 20$	$K = 50$	$K = 100$	K $= 200$
$M = 20$	0.0592	0.0167	0.0027	0.0018	0.0010
$M = 40$	0.0588	0.0163	0.0022	5.34e-04	5.00e-04
$M = 80$	0.0588	0.0160	0.0019	3.74e-04	2.98e-04
$M = 160$	0.0586	0.0160	0.0018	3.08e-04	9.16e-05
$M = 320$	0.0586	0.0159	0.0017	1.1e-04	9.24e-05

Average squared L^2 errors with 50 macro runs.

Thank you