Advanced Monte Carlo and Optimization Methods for Optimal Stopping Problems: Part I

Denis Belomestny

Premolab, Duisburg-Essen University

18.09.2012

Optimal Stopping Problems

- $(X_j)_{j\geq 0}$ is a Markov chain
  - on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_j)_{j\geq 0}, \mathrm{P}_x)$
  - with values in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$
  - starting at x under P_x for some $x \in \mathbb{R}^d$
- $G_j : \mathbb{R}^d \to \mathbb{R}$, $j = 0, \ldots, \mathcal{J}$, is a set of measurable functions that fulfill

$$\mathrm{E}_x \left[\sup_{0\leq j\leq \mathcal{J}} |G_j(X_j)| \right] < \infty$$

Optimal Stopping Problems

Consider the following discrete time optimal stopping problem:

$$Y_0^* = \sup_{\tau \in \{0,,\ldots,\mathcal{J}\}} \mathrm{E}_x \left[G_\tau(X_\tau) \right],$$

where

- τ is a $(\mathcal{F}_j)$-stopping time with values in $\{1, \ldots, \mathcal{J}\}$, i.e. $\{\tau = j\} \in \mathcal{F}_j$

Question

How to approximate Y_0^* in the case when the expectation $\mathrm{E}[G_j(X_\tau)]$ cannot be computed in a closed form ?

Dynamic Programming Principle

- Snell-Envelope Process

$$Y_j^*(X_j) = \sup_{\tau \in \{j,,\ldots,\mathcal{J}\}} \mathrm{E}\left[G_\tau(X_\tau)|X_j\right]$$

- Continuation values

$$C_j^*(x) := \mathrm{E}[Y_{j+1}^*(X_{j+1})|X_j = x], \quad j = 0, \ldots, \mathcal{J} - 1$$

Observation

$$Y_{\mathcal{J}}^*(X_{\mathcal{J}}) = G_{\mathcal{J}}(X_{\mathcal{J}}), \qquad a.s.$$

Dynamic Programming Principle

It holds

$$
\begin{aligned}
C^*_{\mathcal{J}}(x) &= 0, \\
C^*_j(x) &= \mathrm{E}[\max(G_{j+1}(X_{j+1}), C^*_{j+1}(X_{j+1}))|X_j = x]
\end{aligned}
$$

for $j = 0, \ldots, \mathcal{J} - 1$.

Observation

The use of the DPP is relatively straightforward in low dimensions. However, many problems arising in practice are high-dimensional !

Dynamic Programming Principle

The family of stopping rules τ_j^*, $j = 0, \ldots, \mathcal{J}$, defined via

$$\begin{aligned} \tau_{\mathcal{J}}^* &= \mathcal{J}, \\ \tau_j^* &= j \times \mathbf{1}_{\{C_j^*(X_j) \leq G_j(X_j)\}} + \tau_{j+1}^* \times \mathbf{1}_{\{C_j^*(X_j) > G_j(X_j)\}} \end{aligned}$$

for $j = 0, \ldots, \mathcal{J} - 1$ is optimal, i.e.,

$$Y_j^* = \mathrm{E}[G_{\tau_j^*}(X_{\tau_j^*})|X_j], \quad j = 0, \ldots, \mathcal{J}.$$

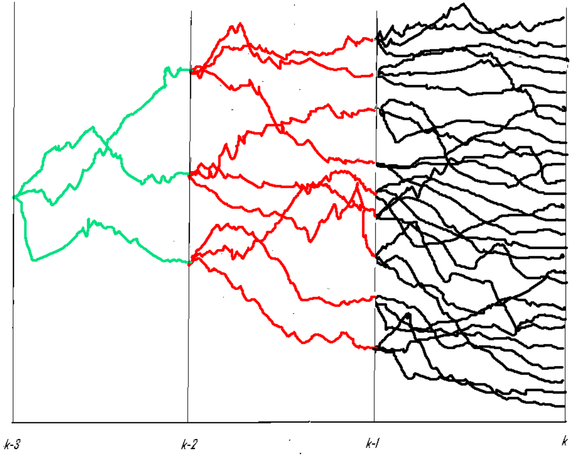
Nested conditional expectations

- Problem: How to approximate the nested conditional expectations in the backward dynamic programming algorithm?

- Naive approach: Average over simulated paths (plain Monte Carlo) as suggested by the Law of Large Numbers.

Infeasible: Computational cost explodes rapidly with the number of exercise dates.

Nested conditional expectations



Regression Methods

1. Simulate M trajectories of the process X
2. Construct estimates $\widehat{C}_{1,M}, \ldots, \widehat{C}_{\mathcal{J},M}$ recursively via backward induction:
   - Put $\widehat{C}_{\mathcal{J},M}(x) \equiv 0$
   - If an estimate $\widehat{C}_{j+1,M}(x)$ is already constructed define $\widehat{C}_{j,M}(x)$ as an estimate of

     $$\mathrm{E}[\max(G_{j+1}(X_{j+1}), \widehat{C}_{j+1,M}(X_{j+1}))|X_j = x],$$

     based on the sample

     $$(X_j^{(m)}, \widehat{C}_{j+1,M}(X_{j+1}^{(m)})), \quad m = 1, \ldots, M.$$

Global Regression

- Fix a vector of basis functions $\psi = (\psi_1, \dots, \psi_K)$
- Let $(\widehat{\alpha}_1, \dots, \widehat{\alpha}_K)$ be a solution of the least squares optimization problem

$$\operatorname*{arginf}_{\alpha \in \mathbb{R}^K} \sum_{m=1}^{M} \left[\widehat{V}_{j+1,M}(X_{j+1}^{(m)}) - \alpha_1 \psi_1(X_j^{(m)}) - \ldots - \alpha_K \psi_K(X_j^{(m)}) \right]^2$$

with $\widehat{V}_{j+1,M}(x) = \max\left\{ G_{j+1}(x), \widehat{C}_{j+1,M}(x) \right\}$

- Define the approximation

$$\widehat{C}_{j,M}(x) = \widehat{\alpha}_1 \psi_1(x) + \ldots + \widehat{\alpha}_K \psi_K(x), \quad x \in \mathbb{R}^d$$

Global Regression

- Define a design $K \times K$ matrix B with entries

$$B_{p,q} = \frac{1}{M} \sum_{m=1}^{M} \psi_p(X_j^{(m)}) \psi_q(X_j^{(m)})$$

- Define a K-dimensional vector b with entries

$$b_p = \frac{1}{M} \sum_{m=1}^{M} \psi_p(X_j^{(m)}) \widehat{V}_{j+1,M}(X_{j+1}^{(m)})$$

Theorem

$$\widehat{C}_{j,M}(x) = (B^{-1} b)^{\top} \psi(x)$$

Local Regression

- Fix a number $l \in \mathbb{N}$ and a point $x \in \mathbb{R}^d$
- Fix a function (kernel) $K \geq 0$ on $\mathbb{R}^d$ with supp $K \subset [-1, 1]^d$
- Let $\mathcal{Q}_{x,M}$ be a polynomial in $\mathbb{R}^d$ of degree l which solves the optimization problem

$$\inf_{\text{quadr. pol. } q} \sum_{m=1}^{M} \left[\widehat{V}_{j+1,M}(X_{j+1}^{(m)}) - q(X_j^{(m)} - x) \right]^2 K\left(\frac{X_j^{(m)} - x}{h} \right)$$

- Define $\widehat{C}_{j,M}(x) = \mathcal{Q}_{x,M}(0)$

Local Regression

Introduce a vector $S = (S_u)_{|u| \le l}$ with

$$S_u = \frac{1}{Mh^d} \sum_{m=1}^{M} \widehat{V}_{j+1,M}(X_{j+1}^{(m)}) \left(\frac{X_j^{(m)} - x}{h} \right)^u K \left(\frac{X_j^{(m)} - x}{h} \right)$$

Let $Z(z) = (z^u)_{|u| \le l}$ be the vector of all monomials of order less than or equal to l and the matrix $\Gamma = (\Gamma_{u_1,u_2})_{|u_1|,|u_2| \le l}$ be defined as

$$\Gamma_{u_1,u_2} = \frac{1}{Mh^d} \sum_{m=1}^{M} \left(\frac{X_j^{(m)} - x}{h} \right)^{u_1+u_2} K \left(\frac{X_j^{(m)} - x}{h} \right).$$

It holds

$$\widehat{C}_{j,M}(x) = Z^\top(0) \Gamma^{-1} S$$

Value Function Estimates

There are two possibilities to estimate $Y_0^* = Y_0^*(x)$

1. Put

$$\widetilde{Y}_{0,M} := \max\left\{G_0(x), \widehat{C}_{0,M}(x)\right\}, \quad x \in \mathbb{R}^d.$$

2. Consider a suboptimal stopping rule

$$\widehat{\tau}_M = \min\left\{0 \le j \le \mathcal{J} : \widehat{C}_{j,M}(X_j) \le G_j(X_j)\right\}$$

and define $\widehat{Y}_{0,M}$ as a Monte Carlo estimate of

$$\mathrm{E}[G_{\widehat{\tau}_M}(X_{\widehat{\tau}_M})|X_0 = x], \quad x \in \mathbb{R}^d$$

based on a new independent set of trajectories.

Question

Which estimate is better ?

Lower Estimates

- $\widehat{Y}_{0,M}$ is low biased estimate, i.e.,

$$\mathrm{E}[\widehat{Y}_{0,M}] = \mathrm{E}[G_{\widehat{\tau}_M}(X_{\widehat{\tau}_M})|X_0 = x] \leq Y_0^*$$

- Both estimates converge to Y_0^*, provided $\widehat{C}_{j,M} \to C_j^*$ as $M \to \infty$

Observation

As was observed by practitioners $\widehat{Y}_{0,M}$ has rather stable behavior with respect to $\widehat{C}_0(x), \ldots, \widehat{C}_{\mathcal{J}-1}(x)$, i.e., even rather poor estimates of continuation values may lead to a good estimate $\widehat{Y}_{0,M}$.

Convergence of value function estimates

Question

Are the convergence rates of $\widehat{Y}_{0,M}$ faster than those of $\widetilde{Y}_{0,M}$?

Answer

They are always faster and may even not depend on the convergence rates of $\widehat{C}_{k,M}(x)$

In fact

- The convergence rates of $\widetilde{Y}_{0,M}$ coincide with ones of $\widehat{C}_{0,M}$

Stoppong Boundary Assumptions

Assume that there exist constants $A_{0,k} > 0$, $\delta_0 > 0$ and $\alpha > 0$ such that

$$\mathrm{P}\left(0 < |C_j^*(X_j) - G_j(X_j)| \leq \delta\right) \leq A_{0,k}\delta^{\alpha}$$

for all $\delta < \delta_0$ and $j = 0, \ldots, \mathcal{J} - 1$.

Remark

This assumption provides a characterization of the behavior of the process X near the exercise boundary $\partial\mathcal{E}$, where

$$\mathcal{E} := \left\{(j, x) : G_j(x) \geq C_j^*(x)\right\}.$$

Convergence Rates (Upper Bounds)

Theorem

Suppose that there exist constants A_1, A_2 and a positive sequence γ_M such that for any $\delta > \delta_0 > 0$

$$\mathrm{P}_{x_0}^{\otimes M}\left(\sup_{x\in\mathcal{X}}|\widehat{C}_{j,M}(x) - C_j(x)| \geq \delta\gamma_M^{-1/2}\right) \leq A_1 \exp(-A_2\delta),$$

where the set $\mathcal{X} \subset \mathbb{R}^d$ fulfills $\mathrm{P}(X_j \in \mathcal{X}) = 1$, $j = 0, \ldots, \mathcal{J} - 1$. If the stopping boundary assumption (SBA) is fulfilled then

$$0 \leq Y_0^* - \mathrm{E}_{x_0}^{\otimes M}[\widehat{Y}_{0,M}] \leq A\left[\sum_{l=0}^{\mathcal{J}-1} A_{0,l}\right]\gamma_M^{-(1+\alpha)/2}$$

with some constant A depending only on α, A_1 and A_2.

Convergence Rates (Upper Bounds)

Theorem

Suppose that there exist constants $A_1,\ A_2$ *and a positive sequence* γ_M *such that for any* $\delta > \delta_0 > 0$

$$\mathrm{P}_{x_0}^{\otimes M}\left(\sup_{x\in\mathcal{X}}|\widehat{C}_{j,M}(x) - C_j(x)| \geq \delta\gamma_M^{-1/2}\right) \leq A_1\exp(-A_2\delta),$$

where the set $\mathcal{X}\subset\mathbb{R}^d$ *fulfills* $\mathrm{P}(X_j\in\mathcal{X}) = 1$, $j = 0,\ldots,\mathcal{J}-1$. *It holds*

$$\mathrm{P}_{x_0}^{\otimes M}\left(|\widetilde{Y}_{0,M} - Y_0^*| \geq \delta\gamma_M^{-1/2}\right) \leq A_1\exp(-A_2\delta),$$

i.e.,

$$|\widetilde{Y}_{0,M} - Y_0^*| = O_{\mathrm{P}}(\gamma_M^{-1/2})$$

Convergence Rates (Low Bounds)

Theorem

Let $\mathcal{J} = 2$. *Fix a function* $G : \mathbb{R}^d \to \{0, 1\}$ *and let* $\mathcal{P}_\alpha$ *be a class of measures such that the SBA is fulfilled with some* $\alpha > 0$. *For any positive sequence* γ_M *satisfying* $\gamma_M^{-1} = o(1)$, $M^{-1}\gamma_M = O(1)$ *there exist* $\mathcal{P}_{\alpha,\gamma} \subset \mathcal{P}_\alpha$ *such that for any stopping rule* $\widehat{\tau}_M$ *and any estimators* $\{\widehat{C}_{j,M}\}$ *measurable w.r.t.* $\mathcal{F}^{\otimes M}$

$$\sup_{\mathrm{P}\in\mathcal{P}_{\alpha,\gamma}} \mathrm{P}^{\otimes M}\left(\sup_{x\in\mathbb{R}^d} |\widehat{C}_{j,M}(x) - C_j(x)| \geq \delta\gamma_M^{-1/2}\right) > 0$$

and

$$\sup_{\mathrm{P}\in\mathcal{P}_{\alpha,\gamma}} \left\{ \sup_{\tau\in\{0,\ldots,\mathcal{J}\}} \mathrm{E}_{\mathrm{P}}[G(X_\tau)] - \mathrm{E}_{\mathrm{P}^{\otimes M}}[\mathrm{E}_{\mathrm{P}}\, G(X_{\widehat{\tau}_M})] \right\} \geq A\gamma_M^{-(1+\alpha)/2}.$$

Convergence Rates ($\alpha = \infty$)

If $\alpha = \infty$, i.e.,

$$\mathrm{P}\left(0 < |C_j(X_j) - G_j(X_j)| \le \delta_0\right) = 0$$

for some $\delta_0 > 0$ and $j = 0, \ldots, \mathcal{J} - 1$, then

$$0 \le Y_0^* - \mathrm{E}_{\mathrm{P}_x^{\otimes M}}[\widehat{Y}_{0,M}] \le A_4 \exp(-A_5 \delta_0 \gamma_M)$$

with some constant A_4 and A_5.

Remark

The convergence rates are exponential in γ_M. So, even the use of inaccurate estimates $\{\widehat{C}_k\}$ would not have dramatic impact on the quality of $\widehat{Y}_{0,M}$ in this case.

Convergence Rates ($\alpha = \infty$)

- Consider a two-period stopping problem with

$$C_0^*(x) = \mathrm{E}[G_1(X_1)|X_0 = x],$$

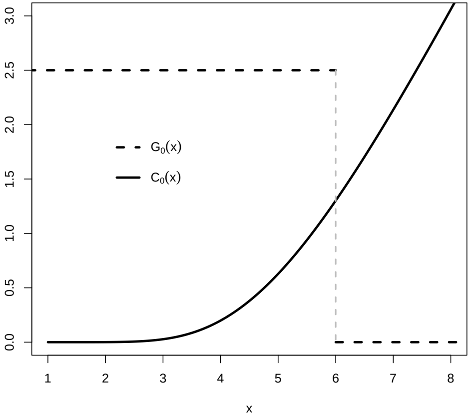
where G_1 is positive and monotone increasing function.

- Define

$$G_0(x) = \begin{cases} C_0^*(x_0) + \delta_0, & x < x_0, \\ C_0^*(x_0) - \delta_0, & x \geq x_0. \end{cases}$$

with some x_0 and $\delta_0 < C_0(x_0)$.

Convergence Rates ($\alpha = \infty$)



Convergence Rates ($\alpha = \infty$)

It is easy to see that

$$\mathrm{P}(0 < |C_0^*(X_0) - G_0^*(X_0)| \leq \delta_0) = 0$$

and

$$\begin{aligned}
\mathcal{C} &= \{x \in \mathbb{R} : C_0^*(x) \geq G_0^*(x)\} = \{x \in \mathbb{R} : x \geq x_0\} \\
\mathcal{E} &= \{x \in \mathbb{R} : C_0^*(x) < G_0^*(x)\} = \{x \in \mathbb{R} : x < x_0\}
\end{aligned}$$

Convergence of cont. values

For any $\beta \in \mathbb{R}_+$ and any function g on $\mathbb{R}^d$ denote

$$g_x(x') = \sum_{|s| \leq \lfloor \beta \rfloor} \frac{(x' - x)^s}{s!} D^s g(x),$$

where $s = (s_1, \ldots, s_d)$ is a multi-index, $|s| = s_1 + \ldots + s_d$ and

$$D^s = \frac{\partial^{s_1 + \ldots + s_d}}{\partial x_1^{s_1} \cdot \ldots \cdot \partial x_d^{s_d}}.$$

Convergence of cont. values

$g \in \Sigma(\beta, H, \mathbb{R}^d)$ (($\beta, H, \mathbb{R}^d$)-Hölder smooth function) if

- g is $\lfloor \beta \rfloor$ times continuously differentiable
- for any $x, x' \in \mathbb{R}^d$

$$|g(x') - g_x(x')| \le H\|x - x'\|^{\beta}, \quad x' \in \mathbb{R}^d$$

Convergence of cont. values

- Assumption

$$C_j^* \in \Sigma(\beta, H, \mathbb{R}^d), \quad j = 0, \ldots, \mathcal{J}.$$

Proposition

There exist positive constants B_1, B_2 and B_3 such that for any h satisfying $B_1 h^\beta < \sqrt{|\log h|/Mh^d}$ and any $\zeta \geq \zeta_0$ with some $\zeta_0 > 0$

$$\mathrm{P}^{\otimes M}\left(\sup_{x\in\mathcal{A}} |\widehat{C}_{j,M}(x) - C_j^*(x)| \geq \zeta\sqrt{\frac{|\log h|}{Mh^d}}\right) \leq B_2 \exp(-B_3\zeta)$$

for $j = 0, \ldots, \mathcal{J} - 1$.

Convergence of cont. values

Corollary

We get with $h = M^{-1/(2\beta+d)}$ *and any* $\zeta \geq \zeta_0 > 0$

$$\mathrm{P}^{\otimes M}\left(\sup_{x\in\mathcal{A}} |\widehat{C}_{k,M}(x) - C_k^*(x)| \geq \frac{\zeta \log^{1/2} M}{M^{\beta/(2\beta+d)}}\right) \leq B_2 \exp(-B_3\zeta)$$

for $j = 0, \ldots, \mathcal{J} - 1$.

Numerical example: Bermudan max call

- Consider d identically distributed assets with dividend yield δ
- The risk-neutral dynamic of assets is given by

$$\frac{dX_k(t)}{X_k(t)} = (r-\delta)dt + \sigma dW_k(t), \quad k = 1, ..., d,$$

where $W_k(t)$, $k = 1, ..., d$, are independent one-dimensional Brownian motions and r, δ, σ are constants.

- At any time $t \in \{t_0, \ldots, t_{\mathcal{J}}\}$ the holder of the option may exercise it and receive the payoff

$$G_j(X_j) = e^{-rt_j}(\max(X_1(t_j), ..., X_d(t_j)) - \kappa)^+$$

with $X_j = (X_1(t_j), ..., X_d(t_j))$

Numerical example: Bermudan max call

- Estimate all c.v. via DPP and Nadaraya-Watson estimator:

$$\widehat{C}_{j,M}(x) = \frac{\sum_{m=1}^{M} V_{j+1}^{(m)} K((x - X_j^{(m)})/h)}{\sum_{m=1}^{M} K((x - X_j^{(m)})/h)}$$

with

$$V_{j+1}^{(m)} = \max\left\{ G(X_{j+1}^{(m)}), \widehat{C}_{j+1,M}(X_{j+1}^{(m)}) \right\}, \quad j = 0, \ldots, \mathcal{J} - 1,$$

where K is a kernel and h is a bandwidth.

Numerical example: Bermudan max call

- The first estimate for the price of the option at time $t_0 = 0$:

$$\widetilde{Y}_{0,M} := \frac{1}{M}\sum_{m=1}^{M} V_1^{(m)}$$

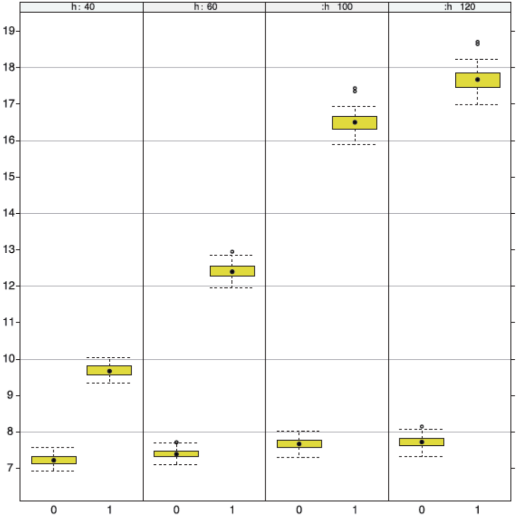
- Construct stopping a policy $\widehat{\tau}$ via

$$\widehat{\tau}^{(n)} = \min\left\{1 \leq j \leq \mathcal{J} : \widehat{C}_{j,M}(X_j^{(M+n)}) \leq G_j(X_j^{(M+n)})\right\}$$

- The second estimate

$$\widehat{Y}_{0,M} = \frac{1}{N}\sum_{n=1}^{N} G_{\widehat{\tau}^{(n)}}(X_{\widehat{\tau}^{(n)}}^{(M+n)})$$

Numerical example: Bermudan max call



Bibliography

Belomestny, D. (2011).
Pricing Bermudan options using regression: optimal rates of convergence for lower estimates. *Finance and Stochastics*, **15**(4), 655–683.