Nonparametric Adaptive Bayesian Stochastic Control

Tao Chen
Department of Mathematics
University of Michigan, Ann Arbor
chenta@umich.edu

Joint work with
J. Myung

Mathematical Finance Colloquia
University of Southern California
October 26 2020

Motivation: model uncertainty in stochastic control problems

- To control the risk due to model uncertainty (error in model estimatation or model misspecification)
- To solve control problems when the true law of the underlying stochastic process is unknown
- Popular applicable methods usually rely on parametric models
- Traditional model-free approach is not easy to implement and is rarely used by practitioners
- Reinforcement learning faces the dilemma of exploration versus exploitation

Main Goals

- To propose and study a **nonparametric adaptive Bayesian approach** for discrete time Markovian control problems subject to Knightian uncertainty, which integrates learning and optimizing.
- To develop a numerical method for efficient and easy implementation of the proposed nonparametric framework.

T. Chen and J. Myung *Nonparametric Adaptive Bayesian Optimal Control with Financial Applications.* In preparation. 2020.

Notations

- $(\Omega, \mathcal{F})$ - measurable space
- $T \in \mathbb{N}$ - fixed time horizon
- $\mathcal{T} = \{0, 1, \ldots, T\}$, $\mathcal{T}' = \{0, 1, \ldots, T-1\}$, and $\mathcal{T}'' = \{1, \ldots, T\}$
- $X = \{X_t, t \in \mathcal{T}\}$ - controlled process taking values in $\mathbb{R}^d$
- $\varphi = \{\varphi_t,\ t \in \mathcal{T}'\}$ - adapted process taking values in a compact set A
- $Z = \{Z_t, t \in \mathcal{T}''\}$ - random process that is driving X
- L - measurable function understood as a loss function
- $\mathbf{\Theta}$ - set of parameters (if a parametric model is assumed for Z)
- $\mathbb{P}_\theta$ - the probablity law of Z corresponding to θ

Example: Dynamic Optimal Portfolio Selection

An investor is deciding on investing in a risky asset and a risk-free banking account by maximizing the expected utility of the terminal wealth.

- r - the constant risk free rate
- e^{Z_t} - the return on the risky asset
- The true law of Z_t is unknown
- The dynamics of the wealth process produced by a s.f. strategy

$$X_{t+1} = X_t(1 + r + \varphi_t(e^{Z_{t+1}} - 1 - r)), \quad t \in \mathcal{T}', \ X_0 = x_0.$$

- To maximize the expected $U(X_T^{\varphi})$ but with respect to what model?

Review of Existing Methods

When a parametric model is assumed for the underlying process, there are several methods can be used.

- (Myopic) adaptive solves the problem

$$\inf_{\varphi \in \mathcal{A}} \mathbb{E}_\theta[L(X, \varphi)]$$

for every $\theta \in \mathbf{\Theta}$ to get $\varphi_t^{\mathsf{MA}}(\theta)$, $t \in \mathcal{T}'$.

- (Static) robust solves the problem

$$\inf_{\varphi \in \mathcal{A}} \sup_{\theta \in \mathbf{\Theta}} \mathbb{E}_\theta[L(X, \varphi)].$$

- Bayesian control solves the Bellman equation

$$V(t, x) = \inf_{a \in A} \int_{\mathbf{\Theta}} \mathbb{E}_\theta[V(t+1, X_{t+1}^{a,\theta}(x, Z_{t+1}))]\pi_t(d\theta).$$

Review of Existing Methods (cont.)

- Strong robust

$$V(t,x) = \inf_{a\in A} \sup_{\theta\in\mathbf{\Theta}} \mathbb{E}_\theta[V(t+1, X_{t+1}^{a,\theta}(x, Z_{t+1}))].$$

- Adaptive robust

$$V(t,y) = \inf_{a\in A} \sup_{\theta\in\mathbf{\Theta}_t(\hat{\theta})} \mathbb{E}_\theta[V(t+1, Y_{t+1}^{a,\theta}(y, Z_{t+1}))],$$

where $y = (x, \hat{\theta})$.

- (Time consistent) adpative

$$V(t,y) = \inf_{a\in A} \mathbb{E}_{\hat{\theta}}[V(t+1, Y_{t+1}^{a,\hat{\theta}}(y, Z_{t+1}))],$$

where $y = (x, \hat{\theta})$.

Comments

The postulated parametric model can be wrong. For example, one usually assumes that the log-return of a risky asset has a normal distribution, but in fact it could be bimodal.

If the assumed parametric model is correct, then

- Myopic adaptive is time inconsistent and suffers from error in estimation.
- Time consistent adaptive still suffers from error in estimation.
- Static and strong robust methods can be overly conservative.

Our goal is to propose a nonparametric methodology that avoids model misspecification, is robust to error in estimation, and easily balances between being aggressive and conservative.

Definition (Dirichlet Process)

Let α be a finite non-null measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, and let $\mathcal{D}$ be a stochastic process indexed by elements in $\mathscr{B}(\mathbb{R})$. We say $\mathcal{D}$ is a Dirichlet process with parameter α and write $\mathcal{D} \in \mathscr{D}(\alpha)$, if for every finite measurable partition $\{B_1, \ldots, B_n\}$ of $\mathbb{R}$, the random vector $(\mathcal{D}(B_1), \ldots, \mathcal{D}(B_n))$ has a Dirichlet distribution with parameter $(\alpha(B_1), \ldots, \alpha(B_n))$.

- $\mathcal{D}$ is a process in space and it is a random probability measure, i.e. every realization of $\mathcal{D}$ is a probability measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.
- The mean of $\mathcal{D}$ is $\alpha/\alpha(\mathbb{R})$.
- The support of $\mathcal{D}$ with respect to the topology of weak convergence is the set of all distributions whose support is contained in the support of α.
- A Dirichlet process is characterized by the parameter α. Through the rest of the talk, we will let $\alpha = cP$ where c is a postive constant and P is a probability measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

Dynamic Learning

Learning of the unknown distribution can be expressed through a sequence of Dirichlet processes. Denote by $\mathbb{P}$ the unknown distribution. Pick a constant c_0 and a probability measure P_0. In a Bayesian manner, we assume that $\mathbb{P} \in \mathscr{D}(c_0 P_0)$.

- Let $Z_1, \ldots, Z_n$ be a sample from $\mathbb{P}$. Then the posterior of $\mathbb{P}$ is given by $\mathscr{D}(c_n \mathcal{P}_n)$ where $c_n = c_0 + n$, and

$$\mathcal{P}_n = \frac{c_0 P_0 + \sum_{i=1}^{n} \delta_{Z_i}}{c_n}.$$

- Online learning of the unknown distribution: $\mathbb{P} \in \mathscr{D}(c_n \mathcal{P}_n)$, where

$$\mathcal{P}_n = \frac{c_{n-1} \mathcal{P}_{n-1} + \delta_{Z_n}}{c_n} =: f_P(n, \mathcal{P}_{n-1}, Z_n)$$

- $\mathcal{P}_n$ is a random probability measure and it is a weighted average of P_0 and the empirical distribution.

Formulation of Adaptive Bayesian Control Problem

Denote by $\mathscr{P}(\mathbb{R})$ the set of all probability measures equipped with Borel σ-algebra corresponding to the Prokhorov metric (i.e. weak convergence).

- Augmented state process $Y_t = (X_t, \mathcal{P}_t) \in \mathbb{R}^d \times \mathscr{P}(\mathbb{R}) =: E_Y$ where $Y_0 = (x_0, P_0)$.
- Dynamics of Y_t: $Y_t = \mathrm{G}(t, Y_{t-1}, \varphi_{t-1}, Z_t)$ such that

$$\begin{aligned} X_t =& f_X(X_{t-1}, \varphi_{t-1}, Z_t) \\ \mathcal{P}_t =& f_P(t, \mathcal{P}_{t-1}, Z_t). \end{aligned}$$

- The mapping G is continuous and Borel-measurable.
- Borel-measurable stochastic kernel on E_Y given $E_Y \times A$

$$\begin{aligned} Q(B \mid t, y, a) =& \int \mathbb{P}(\mathrm{G}(t, y, a, Z_t) \in B)\mathcal{D}_t(d\mathbb{P}) \\ =& P(\mathrm{G}(t, y, a, Z_t) \in B), \quad y = (x, P),\ \mathcal{D}_t \in \mathscr{D}(c_t P). \end{aligned}$$

Formulation of Adaptive Bayesian Control Problem (cont.)

For any $\varphi \in \mathcal{A}$, define a probability measure $\mathbb{Q}$ on the canonical space E_Y^{T+1}:

$$\mathbb{Q}^{\varphi}_{(x_0,P_0)}(B_0 \times \cdots \times B_T) = \int_{B_0} \cdots \int_{B_T} \prod_{t=1}^{T} Q(dy_t | t, y_{t-1}, \varphi_{t-1}) \delta_{(x_0,P_0)}(dy_0).$$

Then, the nonparametric adaptive Bayesian optimal control problem is formulated as

$$\inf_{\varphi \in \mathcal{A}} \mathbb{E}_{\mathbb{Q}^{\varphi}_{(x_0,P_0)}} \left[L(X,\varphi) \right].$$

Solution of the Adaptive Bayesian Control Problem

Consider only the terminal loss ℓ and Bellman equation

$$\begin{aligned} V(T,y) &= \ell(x) \\ V(t,y) &= \inf_{a \in A} \int \mathbb{E}_{\mathbb{P}} \left[V(t+1, \mathrm{G}(t+1, y, a, Z_{t+1})) \right] \mathcal{D}(d\mathbb{P}) \\ &= \inf_{a \in A} \mathbb{E}_{P} \left[V(t+1, \mathrm{G}(t+1, y, a, Z_{t+1})) \right], \end{aligned}$$

for $y = (x, P) \in E_Y$, $\mathcal{D} \in \mathscr{D}(c_t P)$, $t \in \mathcal{T}'$.

Proposition

The functions $V(t, \cdot)$, $t \in \mathcal{T}$, are lower semi-analytic and the universally measurable selector φ^*_t exist:

$$V(t,y) = \mathbb{E}_{P} \left[V(t+1, \mathrm{G}(t+1, y, \varphi^*_t(y), Z_{t+1})) \right], \quad t \in \mathcal{T}'.$$

Solution of the Adaptive Bayesian Control Problem (cont.)

Theorem

The nonparametric adaptive Bayesian optimal control problem is solved by the above Bellman equations:

$$V(0, y_0) = \inf_{\varphi \in \mathcal{A}} \mathbb{E}_{\mathbb{Q}^{\varphi}_{(x_0, P_0)}} \left[L(X, \varphi) \right], \quad y_0 = (x_0, P_0).$$

For any $t \in \mathcal{T}'$, one optimal control is given as φ^*_t and universally measurable.

Outline for Numerical Implementation

We want to find a numerical solver for

$$V(t,y) = \inf_{a \in A} \mathbb{E}_P \left[V(t+1, \mathrm{G}(t+1, y, a, Z_{t+1})) \right], \quad y = (x, P) \in E_Y,\ t \in \mathcal{T}'.$$

What needs to be done:

- Discretization of the state space for Y, accompanied by interpolation in order to evaluate $V(t,y)$.
- Approximation of the integral since integrand is not analytically available.
- Approximation of the optimizers $\varphi^*(y)$ in order to apply it to the out-of-sample paths.
- The key idea is to recursively construct a functional approximation $\hat{V}(t,\cdot)$ that is used for interpolation and prediction.

Infinitely Dimensional State Space

The difficulty arises in this problem is that the augmented state space E_Y is infinitely dimenional. How should we discretize it?

- This is related to the choice of functional approximation $\hat{V}(t,\cdot)$.
- We choose Gaussian process surrogates for construction of $\hat{V}(t,\cdot)$.
- We use the mapping $P \mapsto (\mathbb{E}_P[Z], \dots, \mathbb{E}_P[Z^m])$, where $Z \sim P$, and approximate E_Y by $\mathbb{R}^d \times \mathbb{R}^m$.
- Another way to consider is modifying the kernel function of GP such that it takes into account the Prokhorov distance between probability measures.

Basic Loop

$$\begin{aligned}\tilde{V}(t,\hat{y}) &= \inf_{a\in A}\mathbb{E}_P[\tilde{V}(t+1,\mathrm{G}(t+1,\hat{y},a,Z_{t+1}))] \\ &=: F(\tilde{V}(t+1,\cdot),\hat{y}), \quad \hat{y}\in\mathbb{R}^d\times\mathbb{R}^m\end{aligned}$$

In the spirit of Regression Monte Carlo, we have a $\mathrm{fit}-\mathrm{predict}-\mathrm{optimize}-\mathrm{fit}$ loop:

1. (Assume that the surrogate $\hat{V}(t+1,\cdot)$ has been fitted)
2. Select an experimental design of N sites y^n, $n=1,\ldots,N$;
3. Solve the optimization problem at each y^n, using $\mathrm{predict}(\hat{V}(t,y'))$ for the expectation. This yields the outputs $e^n=F(\hat{V}(t,\cdot),y^n)$ and optimal control a^n at y^n;
4. Fit $\hat{V}(t,\cdot)$ based on data $(y^{1:N},e^{1:N})$ and $\hat{\varphi}_t^*(\cdot)$ based on $(y^{1:N},a^{1:N})$;
5. Goto 1: start the next recursion for $t-1$

Adaptive Bayesian Utility Maximization

- Consider the loss function $\ell(x) = \frac{1-x^{1-\eta}}{1-\eta}$ where $\eta > 1$.
- $X_t = X_{t-1}(1 + r + \varphi_{t-1}(e^{Z_t} - 1 - r)) =: f_W(X_{t-1}, \varphi_{t-1}, r, Z_t)$.
- We assume that the random noise process $Z_1, \dots, Z_T$ is an i.i.d. sequence of random variables.
- Each Z_t has 50% probability to come from $\mathcal{N}(\mu_1, \sigma_1^2)$ and 50% probability to come from $\mathcal{N}(\mu_2, \sigma_2^2)$.
- $$\inf_{\varphi \in \mathcal{A}} \mathbb{E}_{\mathbb{Q}^{\varphi}_{x_0, P_0}} \left[\frac{1 - X_T^{1-\eta}}{1-\eta} \right] = -\sup_{\varphi \in \mathcal{A}} \mathbb{E}_{\mathbb{Q}^{\varphi}_{x_0, P_0}} \left[\frac{X_T^{1-\eta} - 1}{1-\eta} \right].$$

Comparison to Classical Parametric Approaches

Assume that there are 3 types of investors: nonparametric adaptive Bayesian (BA), strong robust (SR), and time consistent adaptive (AD).

- All of them have the same data of the historical log-return: $Z_{-t_0}, \dots, Z_{-1}$.
- SR assumes the distribution of Z to be $\mathcal{N}(\mu, \sigma^2)$, and she solves the robust Bellman equation

$$V^{\mathsf{sr}}(t,x) = \sup_{a \in A} \inf_{(\mu,\sigma^2) \in \boldsymbol{\Theta}(\hat{\mu}_0, \hat{\sigma}_0^2)} \mathbb{E}_{\mu,\sigma^2}\left[V^{\mathsf{sr}}(t+1, f_W(x,a,r,Z_{t+1}))\right],$$

where $\hat{\mu}_0$ and $\hat{\sigma}_0^2$ are estimators based on $Z_{-t_0}, \dots, Z_{-1}$ and $\boldsymbol{\Theta}(\hat{\mu}_0, \hat{\sigma}_0^2)$ is the corresponding confidence region.

Comparison to Classical Parametric Approaches (cont.)

- AD assumes the dsitribution of Z to be $\mathcal{N}(\mu, \sigma^2)$, and she solves the adaptive Bellman equation

$$V^{\mathsf{ad}}(t, y) = \sup_{a \in A} \mathbb{E}_{\hat{\mu}, \hat{\sigma}^2} \left[V^{\mathsf{ad}}(t+1, f_W(x, a, r, Z_{t+1}), f_\Theta(t, \hat{\mu}, \hat{\sigma}^2, Z_{t+1})) \right],$$

  where $y = (x, \hat{\mu}, \hat{\sigma}^2)$.

- BA takes $P_0 = \mathcal{N}(\hat{\mu}_0, \hat{\sigma}_0^2)$, chooses $c_0 > 0$, and solves the Bellman equation

$$V^{\mathsf{ba}}(t, y) = \sup_{a \in A} \mathbb{E}_P \left[V^{\mathsf{ba}}(t+1, \mathrm{G}(t+1, y, a, Z_{t+1})) \right], \quad y = (x, P).$$

We choose $\mu_1 = -0.02/30$, $\sigma_1 = 0.4/\sqrt{30}$, $\mu_2 = 0.13/30$, $\sigma_2 = 0.3/\sqrt{30}$, and randomly generate two cases of $\hat{\mu}_0$ and $\hat{\sigma}_0^2$. For comparison, we apply the computed strategies in both cases on the same set of out-of-sample Z-paths.

Structure of Optimal Control

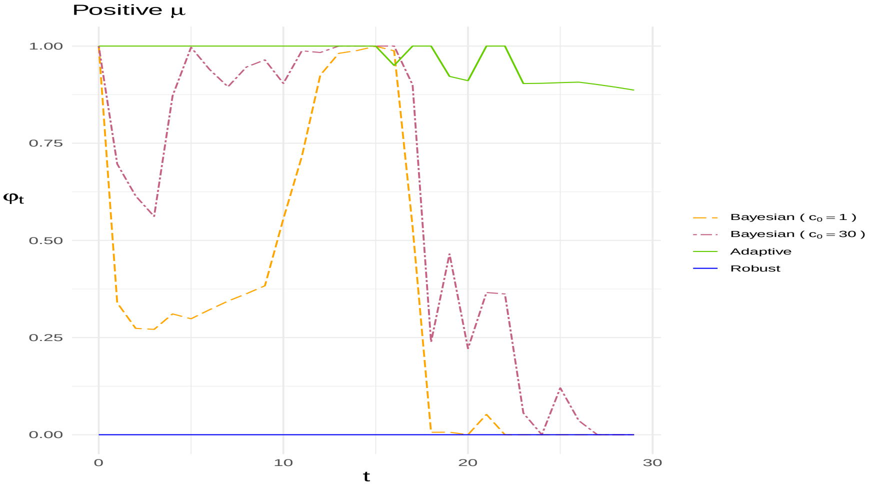

Figure: *Path of adaptive Bayesian optimal strategy in comparison to strong robust, and adaptive with* $t_0 = 100$, $\eta = 1.5$, $\hat{\mu}_0 = 4.615 \times 10^{-3}$, $\hat{\sigma}_0 = 5.609 \times 10^{-2}$.

Structure of Optimal Control

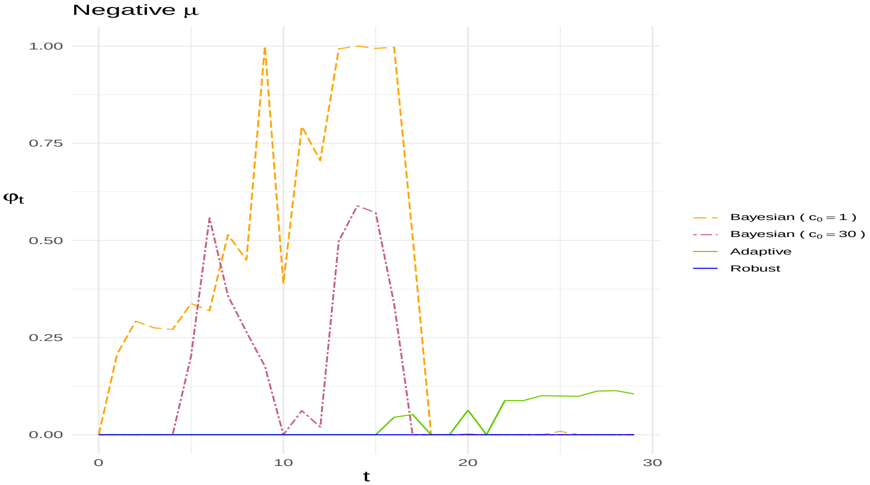

Figure: *Path of adaptive Bayesian optimal strategy in comparison to strong robust, and adaptive with* $t_0 = 100$, $\eta = 1.5$, $\hat{\mu}_0 = -3.987 \times 10^{-3}$, $\hat{\sigma}_0 = 6.288 \times 10^{-2}$.

Comparison of Performance on Out-of-Sample Paths

	$\hat{\mu}_0 = 4.615 \times 10^{-3}$, $\hat{\sigma}_0 = 5.609 \times 10^{-2}$				
	BA			SR	AR
	$c_0 = 1$	$c_0 = 10$	$c_0 = 30$		
mean(V)	1.8037	1.8036	1.8034	1.8020	1.8026
var(V)	4.295e-4	5.421e-4	6.653e-4	4.917e-14	1.162e-3
$q_{0.30}(V)$	1.7919	1.7891	1.7849	1.8020	1.7841
$q_{0.90}(V)$	1.8352	1.8405	1.8485	1.8020	1.8483
max(V)	1.8721	1.8720	1.8656	1.8020	1.8783
min(V)	1.7711	1.7536	1.7647	1.8020	1.7123

Table: *Mean, variance, 30%-quantile, 90%-quantile, maximum, and minimum of the out-of-sample terminal utility for the BA, SR and AR methods; Case 1.*

Comparison of Performance on Out-of-Sample Paths

	$\hat{\mu}_0 = -3.987 \times 10^{-3}$, $\hat{\sigma}_0 = 6.288 \times 10^{-2}$				
	BA			SR	AR
	$c_0 = 1$	$c_0 = 10$	$c_0 = 30$		
mean(V)	1.8043	1.8041	1.8038	1.8020	1.8016
var(V)	4.092e-4	3.356e-4	1.768e-4	4.917e-14	2.020e-5
$q_{0.30}(V)$	1.7940	1.7959	1.7981	1.8020	1.8006
$q_{0.90}(V)$	1.8362	1.8334	1.8256	1.8020	1.8050
max(V)	1.8702	1.8639	1.8590	1.8020	1.8146
min(V)	1.7594	1.7574	1.7801	1.8020	1.7715

Table: *Mean, variance, 30%-quantile, 90%-quantile, maximum, and minimum of the out-of-sample terminal utility for the BA, SR and AR methods; Case 2.*

Conclusions

- We also tried data with trend: $Z_{15k}, \ldots, Z_{15k+14} \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Z_{15k+15}, \ldots, Z_{15k+29} \sim \mathcal{N}(\mu_2, \sigma_2^2)$, where k is even, and BA performs even better comparatively.

- For statistically sound confidence levels, the confidence region $\boldsymbol{\Theta}(\hat{\mu}, \hat{\sigma}^2)$ used by SR is usually too large and it leads to conservative optimal strategies: investing in the banking account all the time.

- AD is too sensitive to the initial estimates $(\hat{\mu}_0, \hat{\sigma}_0^2)$ as it produces very different strategies even on the same future paths.

- The nonparametric Bayesian approach is more robust to model misspecification and error in estimation.

- The parameter c_0 can serve as a tuning parameter for the purpose of risk management and it balances the strategies between being aggressive and conservative. A relatively large c_0 can prevent the learning from overfitting during early stages.

Thank You !

The end of the talk ...
but not of the story