Dynamic Programming and Machine Learning

Chris Rogers
Statistical Laboratory, University of Cambridge

November 28, 2020

Controlled Markov processes.

Controlled Markov process $(X_0, X_1, \ldots, X_T)$ taking values in $\mathcal{X}$:if $X_t = x$, action $a_t = a$, then

$$P(\ X_{t+1} \in dy \mid X_t = x, a_t = a\ ) = p_t(y|x, a)\, dy.$$

Objective is

$$\min E\Big[\ \sum_{s=0}^{T-1} c_s(X_s, a_s) + C(X_T)\ \Big].$$

Bellman equation (Bellman [1966]).

If

$$V_t(x) = \min E\left[\sum_{s=t}^{T-1} c_s(X_s, a_s) + C(X_T) \,\middle|\, X_t = x \right],$$

then for $t = 0, \ldots, T-1$

$$\begin{aligned} V_t(x) &= \min_a \left[c_t(x,a) + \int V_{t+1}(y)\, p_t(y|x,a)\, dy \right] \\ V_T(x) &= C(x). \end{aligned}$$

Solve by backward recursion. Seems like a complete solution, but

- explicit examples rare;
- numerical solution struggles if dimension of $\mathcal{X}$ is large
- *really* want optimal policy!

Examples.

Suppose $X_t \in \mathcal{X} = (0, \infty)^N$, set $\xi_t^i \equiv \log(X_t^i / X_{t-1}^i)$ and suppose that $\xi_t \sim N(\mu, \Sigma) \quad \forall t$, independent.

1. American min put. Given X_0, Σ, r and $\mu^i = r - \frac{1}{2}\Sigma_{ii}$. We have $a_t \in \{0, 1\}$, $a_t = 0$ means 'continue', $a_t = 1$ means 'stop'. If $\tau = \inf\{t : a_t = 1\}$, then reward is

$$E\left[\, e^{-r\tau}(K - \min_i X_\tau^i)^+ \,\right] \equiv E[Z_\tau].$$

Here, $K > 0$ given. Even for $N = 1$, no closed-form solution. Numerics struggle beyond $N = 3$.

2. Optimal investment. Enter day t holding θ_t^i units of asset i, so overall wealth is $w_t = \theta_t \cdot X_t$. Choose to hold θ_{t+1} through to day $t+1$, so

$$z_t = (\theta_t - \theta_{t+1}) \cdot X_t$$

is consumption on day t. Aim is to

$$\max E\Big[\sum_{t=0}^{T-1} u_t(z_t) + U(w_T) \Big].$$

State is (X_t, w_t), evolution depends on the controls.

3. Parameter uncertainty. If μ is unknown, could do Bayesian story, so now state is $(X_t, w_t, \hat{\mu}_t)$. Gauss-Wishart prior on (μ, Σ^{-1}) if Σ also unknown.

What else could we do?

European min-put for $N = 20$

$$E\left[\ e^{-rT}(K - \min_i X_T^i)^+\ \right] \equiv E[Z_T]$$

could be evaluated by *simulation*; American min-put?

Theorem (Rogers [2002], Haugh and Kogan [2004]) If $\mathcal{M}_0$ is the space of martingales vanishing at $t = 0$ then

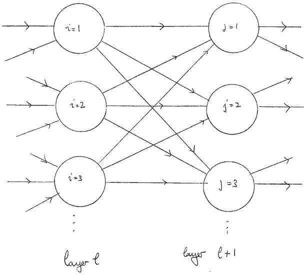
$$\sup_\tau E[Z_\tau] = \min_{M \in \mathcal{M}_0} E\left[\ \sup_{0 \le t \le T} (Z_t - M_t)\ \right].$$

A big step forward:

- gives upper bounds ;
- opens way to simulation.

But how to find 'good' martingales M? Andersen and Broadie [2004] used a good τ to generate a good M.

Machine learning.



Nodes at layer ℓ receive inputs $Y_{\ell-1} = (y^1_{\ell-1}, \ldots, y^n_{\ell-1})$ from layer $\ell - 1$, and turn these into outputs

$$y^j_\ell = \varphi\Big(\sum_i w_{ji}(\ell)\, y^i_{\ell-1} + b_j(\ell) \Big)$$

where φ is non-linear, constants w, b to be learned.

We have a training set of many pairs (Y_0, Y_{end}) and want a function f of the above form such that $f(Y_0^j) = Y_{end}^j$, $j = 1, \ldots, N_{train}$.

Notice $f(Y) = f(Y; \{w_{ji}(\ell), b_j(\ell)\})$. Choose a suitable loss function L and

$$\min_{w,b} \sum_{j=1}^{N_{train}} L(f(Y_0^j, Y_{end}^j)).$$

Optimal stopping problem? (Han and E [2016])

We seek a function $\pi : \{0, 1, \dots, T-1\} \times \mathcal{X} \to \{0, 1\}$, where

$$\begin{aligned} \pi(t, x) &= 1 \quad \text{if we want to stop at time } t \text{ in state } x \\ &= 0 \quad \text{else.} \end{aligned}$$

This is a function of the path so far, and determines a stopping time τ. We take Y_0^j to be a simulated path of X, and set

$$f(Y_0^j) = K - Z_\tau.$$

Finally, define $L(x) = x$. That's all.

REMARKS. (i) In practice, $\pi : \{0, 1, \dots, T-1\} \times \mathcal{X} \to (0, 1)$

(ii) Becker et al. [2020] combine with dual methods to get bounds within 1% in dimension up to 500.

(iii) Number of time steps relatively small ($\sim 10 - 15$).

What now?

Dual approach to more general dynamic programming problems?

Rogers [2007]. Suppose that $(\varepsilon_t)_{t=1}^T$ are IID $U(0,1)$ and

$$X_{t+1} = F(t, X_t, a_t, \varepsilon_{t+1}), \qquad t = 0, \ldots, T-1.$$

Then

$$V_0(X_0) = \max_{(h_t)} E\left[\inf_{\mathbf{a}} \left\{ \sum_{t=0}^{T-1} \left\{ c_t(X_s, a_s) - \Delta M_{t+1} \right\} + C(X_T) \right\} \right]$$

where

$$\Delta M_{t+1} = h_{t+1}(X_{t+1}) - Ph_{t+1}(X_t, a_t).$$

ML technology should work for this too!

References

L. Andersen and M. Broadie. Primal-dual simulation algorithm for pricing multidimensional American options. *Management Science*, 50(9):1222–1234, 2004.

S. Becker, P. Cheridito, and A. Jentzen. Pricing and hedging American-style options with deep learning. *Journal of Risk and Financial Management*, 13(7):158, 2020.

R. Bellman. Dynamic Programming. *Science*, 153(3731):34–37, 1966.

J. Han and W. E. Deep learning approximation for stochastic control problems. *arXiv preprint arXiv:1611.07422*, 2016.

M. B. Haugh and L. Kogan. Pricing American options: a duality approach. *Operations Research*, 52(2):258–270, 2004.

L. C. G. Rogers. Monte Carlo valuation of American options. *Mathematical Finance*, 12(3):271–286, 2002.

L. C. G. Rogers. Pathwise stochastic optimal control. *SIAM Journal on Control and Optimization*, 46(3):1116–1132, 2007.