swissQuant
Mastering Complexity

Mathematical Challenge September 2019

Uncertain Decision Making in Dynamic Optimization Models

References

[1] M. L. Puterman. Markov decision processes: Discrete stochastic dynamic programming. Wiley, 2005.

[2] N. Bäuerle and U. Rieder. Markov decision processes with application to finance. Springer, 2011.

[3] W. Zucchini and I. MacDonald. Hidden markov models for time series - an introduction using r. CRC Press, 2009.

[4] R. M. Alexander. Optima for animals. Princeton University Press, 1996.

[5] U. H. Thygesen, L. Sommer, K. Evans, and T. A. Patterson. Dynamic optimal foraging theory explains vertical migrations of bigeye tuna. *Ecology*, 2016.

Description

Motivation

Stochastic controlled models [1] are characterized by the fact that the random transitions between the states of the system depend upon a control variable, which can be chosen by a decision maker. Controlled models are usually coupled with optimal control theory, where the control variables are chosen in order to optimize a given value function.
The standard theory for dynamic decision making can be extended by including uncertainty in the decision making process. The main motivation is that it allows an external observer to make inference on observed behaviour of natural systems which are assumed to behave optimally. Since observed behaviour typically is not in perfect agreement with model predictions, which might be due to modelling limitations, a statistical analysis requires a model of deviations from optimal behaviour.

Technical Details

For the sake of exposition consider the time and state discrete case. The state of the system is described by a random process $(X_t)_{t\in\mathbb{N}}$, where the transitions between two consecutive states are determined by the controls $(U_t)_{t\in\mathbb{N}}$. As the states are often unknown, measurements $(Y_t)_{t\in\mathbb{N}}$ are introduced. The dynamic is described via a hidden Markov process coupled with a control problem:

- $\delta(x) = \mathbb{P}(X_0 = x)$
- $\gamma_u(i,j) = \mathbb{P}(X_{t+1} = j | X_t = i, U_t = u) \quad \forall t$
- $q_t(x,u) = \mathbb{P}(U_t = u | X_t = x)$
- $p(x,y) = \mathbb{P}(Y_t = y | X_t = x) \quad \forall t$

More specifically, δ and Γ_u depend on the experiment design, p defines the accuracy of the measures, while the policy $\pi = (q_0, \ldots, q_{T-1})$ characterizes the optimality (or rather the near-optimality) of the decision maker. Defining the value function V as the expected final payoff

$$V_t(x) = \mathbb{E}_\pi[V_T(X_T)|X_t = x] \tag{1}$$

where the final condition $V_T(x)$ depends on the experiment design, the state-control transition can be expressed

- for optimal control as:

$$q_t(x,u) = q_t^*(x,u) = \begin{cases} 1 & \text{if } \; u = \text{argmax}\{\tilde{q}_t(x,a), a \in \mathbf{U}\} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

- for near-optimal control as:

$$q_t(x,u) = q_t^{(\beta)}(x,u) = \frac{exp(\beta\,\tilde{q}_t(x,u))}{\sum_{a\in\mathbf{U}} exp(\beta\,\tilde{q}_t(x,a))} \quad \text{with} \quad \beta \in [0, +\infty) \tag{3}$$

where

$$\tilde{q}_t(x,u) = \sum_j \gamma_u(x,j) V_{t+1}(j) \tag{4}$$

In order to be able to solve the optimization problem, i.e. to show that q^* is optimal, it is necessary to re-write the value function V in the backwards formulation [2]:

$$V_t(x) = \sum_u q_t(x,u)\tilde{q}_t(x,u) \tag{5}$$

with

$$\tilde{q}_t(x,u) = \sum_j \gamma_u(x,j) V_{t+1}(j) \tag{6}$$

In fact, using (2), (5) and (6), one can write

$$\begin{aligned} V_t^*(x) &= \sum_{u \in \mathbf{U}} q_t^*(x,u)\,\tilde{q}_t(x,u) \\ &= 1 \cdot \max_{u \in \mathbf{U}} \tilde{q}_t(x,u) \\ &= \sum_{u \in \mathbf{U}} q_(x,u) \max_{u \in \mathbf{U}} \tilde{q}_t(x,u) \\ &\geq \sum_{u \in \mathbf{U}} q_t(x,u)\,\tilde{q}_t(x,u) = V_t(x)\ . \end{aligned}$$

The likelihood of the observed data [3], which are both the decision maker's actions $u^{T-1} = (u_0, \ldots, u_{T-1})$ and the measurments $y^T = (y_0, \ldots, y_T)$, is given by

$$L_T = \delta P(y_0) Q_0(u_0) \Gamma_{u_0} P(y_1) Q_1(u_1) \ldots \Gamma_{u_{T-2}} P(y_{T-1}) Q_{T-1}(u_{T-1}) \Gamma_{u_{T-1}} P(y_T) 1'_m \tag{7}$$

where

- $Q_t(u) = diag((q_t(x,u)_x))$,
- $P(y) = diag((p(x,y)_x))$.

When long time series are available, it is quite common to face numerical underflow problems in the calculation of the likelihood. Since (7) expresses the likelihood as a product of matrices, not of scalars, it is not straightforward how to compute the log-likelihood.
Moreover, since the computation of value function can only be performed backward in time, the classical algorithms, which run forwards, should be modified to run the two computations backwards, reducing memory needs and consequently computational time.
The following quantities are needed for the log-likelihood backward computation:

$$\alpha_t = \begin{cases} \Gamma_{u_{T-1}} P(y_T) 1'_m & \text{if } t = 0 \\ \Gamma_{u_{T-(t+1)}} P(y_{T-t}) Q_{T-t}(u_{T-t}) \alpha_{t-1} & \text{if } t \in \{1, \ldots, T-1\} \\ P(y_0) Q_0(u_0) \alpha_{T-1} & \text{if } t = T \end{cases} \tag{8}$$

$$\omega_t = \begin{cases} \max(\alpha_t) & \text{if } t \in \{0, \ldots, T-1\} \\ \delta \alpha_T & \text{if } t = T \end{cases} \tag{9}$$

$$\phi_t = \frac{\alpha_t}{\omega_t} \tag{10}$$

Combining quantities (8)-(10), the likelihood function can be expressed as the product of scalars as expressed in (11), which allows to directly run the computation backwards:

$$L_T = \omega_T = \omega_1 \cdot \prod_{t=2}^{T} \frac{\omega_t}{\omega_{t-1}} \tag{11}$$

In fact, at each iteration of the backward algorithm only few information need to be stored to compute the following step of the backward filter.

Questions

- **Q1:** *The general purpose of introducing near-optimality in dynamic programming is the generalization of the optimal theory. Show that* (3) *extends* (2)*, highlighting the role of* β *in defining the optimality of the decisions.*
- **Q2:** *How would you extend the discrete case presented to the continuous framework, i.e. in the context of the Hamilton-Jacobi-Bellman's equation?*
- **Q3:** *The theory presented was first applied to an ecological case study, where an investigation of optimal and near-optimal foraging theory was performed (see [4], [5]). What applications to the financial sector would be possible?*

We look forward to your opinions and insights.

Best Quant Regards,

swissQuant Group Leadership Team