Introduction Structural Estimation of Markov Decision Processes

Miguel Alcobendas

Yahoo!

February 18, 2018

Dynamic Models: Applications

- Fancy word in machine learning $--->$ "Reinforcement Learning"
- Interactions between ads and content from user's perspective
- Model search behavior (allocation of a sequence of ads)
- Repeated clicks on the same search page (title, sitelinks,...)
- Counterfactuals: Welfare Analysis
- Advertiser behavior: bidding, budgeting, exit

Structural Estimation of Markov Decision Processes:

- MDP provides a framework for modelling sequential decision making under uncertainty

Structural Estimation of Markov Decision Processes:

- MDP provides a framework for modelling sequential decision making under uncertainty
- Two type of variables over period $t = 0, 1, ..., T$:
  - State variables: $s_t = (x_t, \epsilon)$. x_t observed by the econometrician part and ϵ_t observed only by the agent
  - Control variables d_t: Discrete decision process (DDP) vs Continuous Decision Process (CDP)

Structural Estimation of Markov Decision Processes:

- MDP provides a framework for modelling sequential decision making under uncertainty
- Two type of variables over period $t = 0, 1, ..., T$:
  - State variables: $s_t = (x_t, \epsilon)$. x_t observed by the econometrician part and ϵ_t observed only by the agent
  - Control variables d_t: Discrete decision process (DDP) vs Continuous Decision Process (CDP)
- Agent represented by a set of primitives (u, p, β)
  - $u(s_t, d_t)$ represents the agent's preferences at time t
  - $p(s_{t+1}|s_t, d_t)$ is a Markov transition probability representing the agent's belief about uncertain future states
  - β utility discount factor

Structural Estimation of Markov Decision Processes:

- Rational agents behaving according to an optimal decision rule $d_t = \delta(s_t)$ that solves

$$V_0^T(s) = max_\delta E_\delta\{\sum_{t=0}^{T} \beta^t u(s_t, d_t)|s_0 = s\}$$

where E_δ expectation wrt the stochastic process $\{s_t, d_t\}$ induced by the decision rule δ

- The Markov Decision Process can be solved using Dynamic Programing
- In periods $t = 0, 1, ..., T$ the value V_t functions are recursively defined by

$$V_t(s_t) = max_{d_t \in D_t(s_t)}\{u_t(s_t, d_t) + \beta \int V_{t+1}[s_{t+1,\delta_{t+1}(s_{t+1})}]p_{t+1}(ds_{t+1}|s_t, d_t)\}$$

and the policy function δ_t solves the previous equation

Structural Estimation of Markov Decision Processes:

- Additive Separability Assumption (AS)

$$u(s,d) = u(x,d) + \epsilon(d)$$

- Conditional Independence Assumption (CI)
  The transition density of the controlled Markov process $\{x_t, \epsilon_t\}$ factors as

$$p(dx_{t+1}, d\epsilon_{t+1}|x_t, \epsilon_t, d_t) = q(d\epsilon_{t+1}|x_{t+1})\pi(dx_{t+1}|x_t, d_t)$$

  comments:

  1. x_{t+1} is a sufficient statistic for ϵ_{t+1}. Dependence between ϵ_t and ϵ_{t+1} is transmitted through observed x_{t+1}
  2. x_{t+1} depends on x_t not on ϵ_t

- Under AS and CI assumptions, the Bellman's equation has the form

$$v(x,\epsilon) = u(x,d) + \beta \int max_{d' \in D(y)}[v(y,d') + \epsilon(d')]q(d\epsilon|y)\pi(y|x,d)$$

Structural Estimation of Markov Decision Processes:

- If $\{s_t, d_t\}$ is a DDP satisfying AS, CI and other regularity conditions, then the controlled process $\{x_t, \epsilon_t\}$ is Markovian with transition probability

$$Pr\{dx_{t+1}, d_{t+1}|x_t, d_t\} = P(d_{t+1}|x_{t+1})\pi(dx_{t+1}|x_t, d_t)$$

- Given panel data $\{x_t^a, d_t^a\}$ on observed states and decisions of a collection of agents, the max likelihood estimator $\hat{\theta}^f$ is

$$\hat{\theta}^f = argmax_\theta L^f(\theta) = \prod_{a=1}^{A}\prod_{t=1}^{T_a} P(d_t^a|x_t^a, \theta)\pi(x_t^a|x_{t-1}^a, d_{t-1}^a, \theta)$$

- In practice, we use a two step model

$$\hat{\theta}_1^p = argmax_{\theta_1} L_1^p(\theta_1) = \prod_{a=1}^{A}\prod_{t=1}^{T_a} \pi(x_t^a|x_{t-1}^a, d_{t-1}^a, \theta)$$

$$\hat{\theta}_2^p = argmax_{\theta_2} L_2^p(\hat{\theta}_1^p, \theta_2) = \prod_{a=1}^{A}\prod_{t=1}^{T_a} P(d_t^a|x_t^a, \hat{\theta}_1^p, \theta_2)$$

Structural Estimation of Markov Decision Processes:

- Assumptions:
  - $q(d\epsilon|x)$ is a multivariate extreme-value distribution. Then

$$q(d\epsilon|x) = \prod_{d \in D(x)} exp\{-\epsilon(d) + \gamma\} exp[-exp\{-\epsilon(d) + \gamma\}]$$

  - $P(d|x)$ follows a multinomial logit formula
- Then

$$v_t(x, d) = u_t(x, d) + \beta \int log[\sum_{d' \in D(y)} exp[v_{t+1}(y, d')]]\pi_t(dy|x, d)$$

and

$$P(d|x) = \frac{exp[v(x, d)]}{\sum_{l \in D(x)} exp[v(x, l)]}$$

Structural Estimation of Markov Decision Processes:

- In finite-horizon Markov Decision Processes MDP, the value functions are computed using backward recursion
- Step 1:

$$v_T(x,d) = u_T(x,d)$$

- Step 2:

$$v_t(x,d) = u_t(x,d) + \beta \int log[\sum_{d' \in D(y)} exp[v_{t+1}(y,d')]]\pi_t(dy|x,d)$$