CE 273
Markov Decision Processes

Lecture 22

Dynamic Discrete Choice Models

Previously on Markov Decision Processes

In such cases, we can write the state as (x_k, y_k) where x_k is affected by u_k and y_k is not. Let p_i represent the pmf of y_k. In such cases, the DP algorithm can be simplified as

$$\hat{J}_k(x_k) = \sum_{i=1}^{m} p_i J_k(x_k, i)$$

$$\hat{J}_k(x_k) = \sum_{i=1}^{m} p_i \min_{u_k \in U_k(x_k)} \mathbb{E}_{w_k} \left\{ g_k(x_k, u_k, w_k) + \hat{J}_{k+1}(f_k(x_k, u_k, w_k)) | y_k = i \right\}$$

In the case of Tetris, x_k is the board configuration and y_k is the shape of the block. There is no exogenous disturbance and the action uniquely determines the new state. Hence, we can write

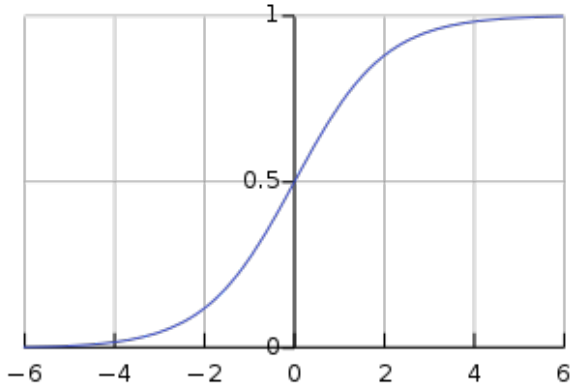
$$J_k(x_k) = \sum_{i=1}^{m} p_i \min_{u_k \in U_k(x_k)} \left\{ g_k(x_k, i, u_k) + J_{k+1}(f_k(x_k, i, u_k)) \right\}$$

f_k represents the new board position and g_k could be the number of rows cleared.

Previously on Markov Decision Processes

Specifically, we use any class of stochastic policies which guarantee that $\nabla_\theta \mu_\theta$ is differentiable and belongs to $(0, 1)$.

A standard way to encode such stochastic policies is to use a softmax policy. Imagine $f(i, u; \theta)$ denotes some kind of numerical preference/utility of u over all the other $u' \in U(i)$. The softmax policy is defined as



$$\mu_\theta(i, u) = \frac{\exp\left(f(i, u; \theta)\right)}{\sum_{u'} \exp\left(f(i, u; \theta)\right)}$$

This is identical to the multinomial logit model! We also explicilty parameterize $f(i, u; \theta)$ using a linear architecture

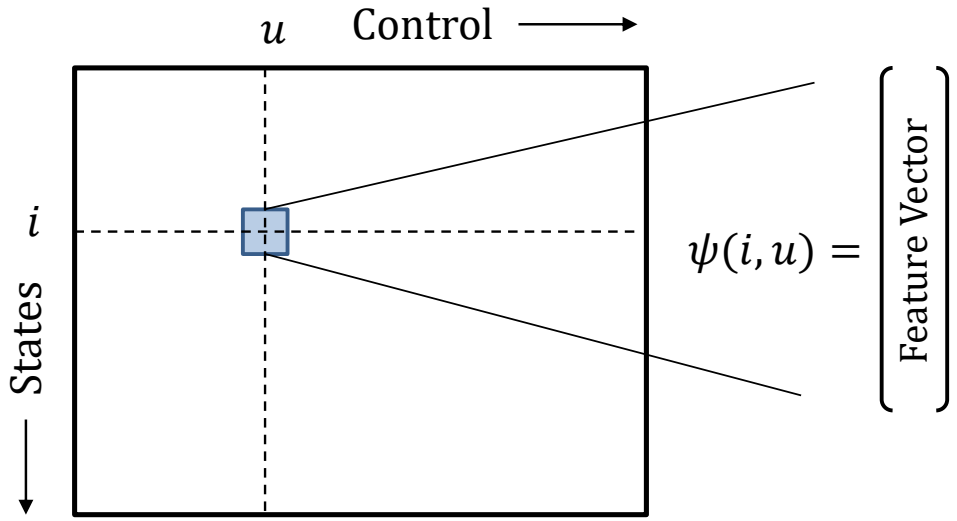
$$f(i, u; \theta) = \theta^T \psi(i, u)$$

where $\psi(i, u)$ is a feature vector for state-action pair (i, u).

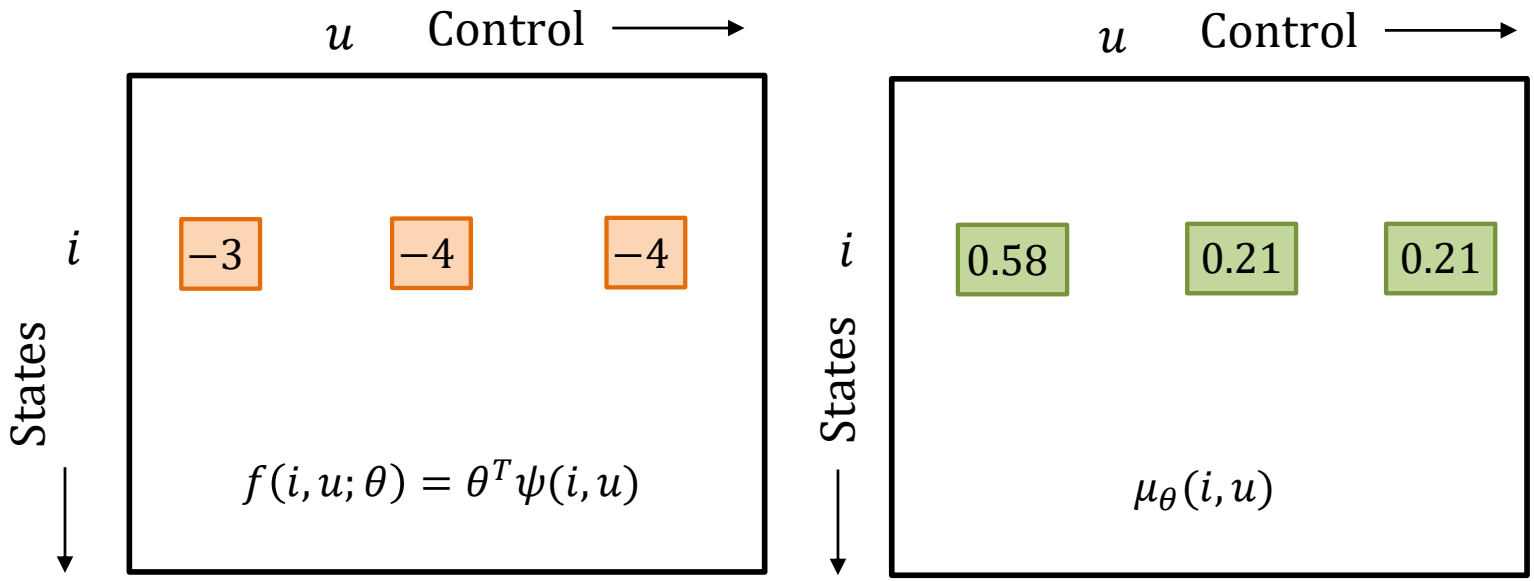
Previously on Markov Decision Processes

For instance a vector of features for Tetris could be

- Constant
- Number of lines cleared because of choosing u.
- Difference in column heights before and after taking control u in state i.



Previously on Markov Decision Processes



Notice that higher the value of f of an action, the odds of choosing it are greater.

$$\mu_\theta(i,u) = \frac{\exp\left(\theta^T \psi(i,u)\right)}{\sum_{u'} \exp\left(\theta^T \psi(i,u')\right)}$$

Previously on Markov Decision Processes

Let us revisit our original problem. We want to find θ to minimize

$$\min_{\theta \in \mathbb{R}^m} J_{\mu_\theta}(x_0)$$

Standard gradient descent approaches involve finding iterates

$$\theta^{k+1} = \theta^k - \eta_k \nabla_\theta J_{\mu_\theta}(x_0)$$

Expanding $J_{\mu_\theta}(x_0)$, in $\nabla_\theta J_{\mu_\theta}(x_0)$,

$$\nabla_\theta J_{\mu_\theta}(x_0) = \nabla_\theta \mathbb{E}_w \left\{ \sum_{k=0}^{\infty} g(x_k, \mu(x_k), w_k) \right\}$$

As before, we can think of replacing the expectation with simulated samples, but how do we deal with the ∇_θ operator that appears before the expectation?

Can we make it look like the expectation of something else? Yes.

Previously on Markov Decision Processes

The policy gradient theorem guarantees that

$$\nabla_\theta J_{\mu_\theta}(i) \propto \mathbb{E}_{j,v}\left\{ Q_{\mu_\theta}(j,v)\nabla_\theta \ln \mu_\theta(j,v) \right\}$$

Recall that for the softmax policy,

$$\nabla_\theta \ln \mu_\theta(j,v) = \psi(j,v) - \sum_{v'} \mu_\theta(j,v')\psi(j,v')$$

To simulate the above expectation, we can use a MC-like method in which we follow policy μ_θ for a full sample episode and for the state-action pair visited at step k, we calculate G_k (the sampled future cost) and $\nabla_\theta \ln_{\mu_\theta}(j_k, v_k)$ (using the above formula) and update θ

$$\theta^{k+1} = \theta^k - \eta_k G_k \nabla_\theta \ln \mu_\theta(j_k, v_k)$$

Lecture Outline

Introduction

So far, we have looked at sequential decision making probelms in which we had access to the problem data (MDP) or we inferred it from the environment (RL).

In the next two classes, we look at its inverse problem. We observe the actions taken by an agent and then try to uncover what was going in their mind when they took those actions.

That is, we assume that they are solving some MDP (knowingly or unknowingly) and we would like to uncover the input data so that their actions match with the optimal solutions.

This line of research appears in both behavioral economics as well as RL. Unfortunately, these two fields have been operating in closed communicating classes.

We will look at the econometric version in today's class and RL approaches in the next.

Introduction

Understanding Behavior

Before we step into the dynamic case, we will first discuss static models, in which decision makers choose among a finite set of alternatives in a single shot.

Each alternative is assumed to provide the decision maker some utility and the decision maker picks the choice with the highest utility.

In other words, the optimization model running in the decision makers mind is finding the max of a bunch of numbers.

Daniel McFadden, who won the 2000 Nobel in Economics, was instrumental in the development of discrete choice theory which is widely applied in transportation and marketing.

Introduction

Examples of Discrete Choice

Factors/features/attributes that can affect our choice:

- Cast
- Running time
- Genre
- IMDb rating
- Language
- Sequel or not

Introduction

Examples of Discrete Choice

Factors/features/attributes that can affect our choice:

- Price
- Speaker quality
- Operating System
- Battery Life
- Color
- Weight

Introduction

Examples of Discrete Choice





Factors/features/attributes that can affect our choice:

- Price
- Travel time
- Reliability
- Out-of-vehicle time
- Comfort
- Safety

Introduction

Applications

Choice models are very powerful since they can predict behavior of individuals when the choice attributes are altered. For example,

- What are the odds with which people will watch a film if the IMDb rating goes up by 0.5?
- How many products will I sell if I give a 10% discount on my product?
- What fraction of travelers might shift to buses if fuel prices increase by ₹5?

Lecture Outline

Discrete Choice Theory

Discrete Choice Theory

Introduction

Discrete choice theory is built on the assumption that decision makers calculate the utility from different alternatives and choose the one with maximum utility.

But an analyst may not have access to many attributes that individuals consider when choosing an alternative. For instance, consider the problem of selecting a mode.

Decision maker's world:

- Cost
- Time
- Reliability
- Safety

Utility of a mode is some function f(Cost, Time, Reliability, Safety)

Analyst's world:

- Cost
- Time

Utility of a mode is some parametric function u(Cost, Time; β) + random component

The random component captures the effect of all unobserved or latent attributes such as reliability and safety which are difficult to measure.

Discrete Choice Theory

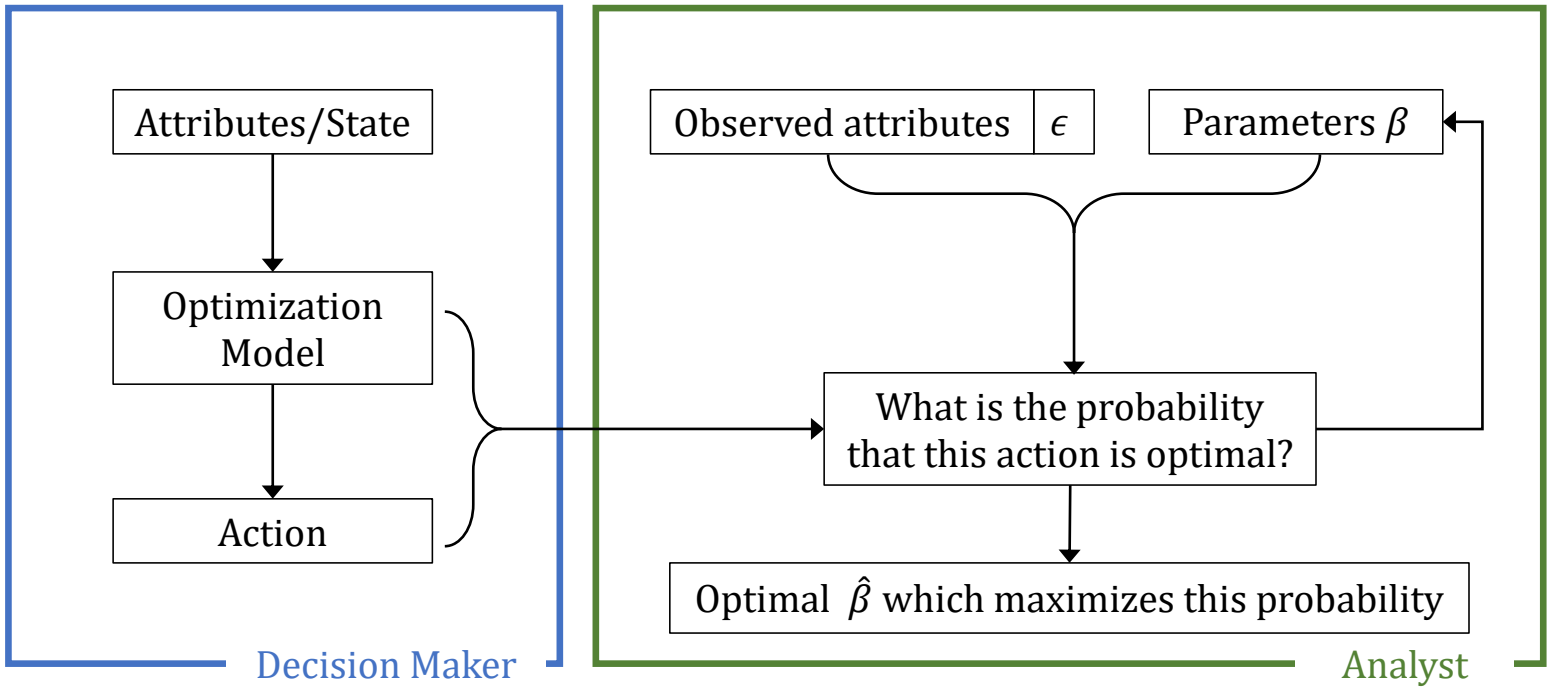
Introduction

The decision maker solves a deterministic problem of selecting the maximum utility of a finite number of alternatives.

The analyst on the other hand can only model the decision makers objective as random variable, and hence can say what is the probability of selecting a certain alternative.

The goal is to adjust β's such that the probability of selecting the observed alternative (i.e., the probability that the analyst was correct) is maximized or in other words.

The process of tweaking β's is also called maximum likelihood estimation. Estimating these parameters will help us understand the effects of different attributes on the choice probabilities.



Different optimization models maybe used in different contexts:

- Find the maximum of a bunch of numbers $\rightarrow$ Static choice models
- Knapsack problems $\rightarrow$ Multiple discrete-continuous choice models
- Markov Decision Process $\rightarrow$ Dynamic choice models

Discrete Choice Theory

Background

We will use slightly different notation for this lecture and let a denote actions and u represent deterministic component of utility. Suppose the set of alternatives is A.

The analyst assumes that the utility consists of

- An observable component $u(x, a)$ where $a \in A$ which depends on a vector of attributes/covariates/states x that is known to both the individual and the analyst.
- A latent component $\epsilon(a)$ which is known to the individual but not the analyst.

Individuals are assumed to select an alternative $a \in A$ which maximizes $u(x, a) + \epsilon(a)$.

Discrete Choice Theory

Example

Further, u is supposed to be linearly parameterized and we sometimes write $u(x, a; \beta)$ to point out this dependence.

For example, consider two choices of using a `Car` or taking the `Bus`. The state or attribute vector x can be the $\text{TT}_{\texttt{Car}}$ or $\text{TT}_{\texttt{Bus}}$. The utilities can be written as

$$u\big((\text{TT}_{\texttt{Car}}, \text{TT}_{\texttt{Bus}}), \texttt{Car}\big) = \beta_0 + \beta_1 \text{TT}_{\texttt{Car}} + \epsilon(\texttt{Car})$$
$$u\big((\text{TT}_{\texttt{Car}}, \text{TT}_{\texttt{Bus}}), \texttt{Bus}\big) = \beta_1 \text{TT}_{\texttt{Bus}} + \epsilon(\texttt{Bus})$$

Suppose the individual chose `Car`. Then the probability that the analyst got it right is

$$\begin{aligned}
&\mathbb{P}\Big[u\big((\text{TT}_{\texttt{Car}}, \text{TT}_{\texttt{Bus}}), \texttt{Car}\big) \geq u\big((\text{TT}_{\texttt{Car}}, \text{TT}_{\texttt{Bus}}), \texttt{Bus}\big)\Big] \\
&= \mathbb{P}\Big[\beta_0 + \beta_1 \text{TT}_{\texttt{Car}} + \epsilon(\texttt{Car}) \geq \beta_1 \text{TT}_{\texttt{Bus}} + \epsilon(\texttt{Bus})\Big] \\
&= \mathbb{P}\Big[\epsilon(\texttt{Car}) - \epsilon(\texttt{Bus}) \geq -\beta_0 - \beta_1 \text{TT}_{\texttt{Car}} + \beta_1 \text{TT}_{\texttt{Bus}}\Big]
\end{aligned}$$

There are multiple ways to proceed from here depending on the assumptions made on the error terms. The simplest option is to suppose that the ϵs are iid and type I extreme value or Gumbel distributed.

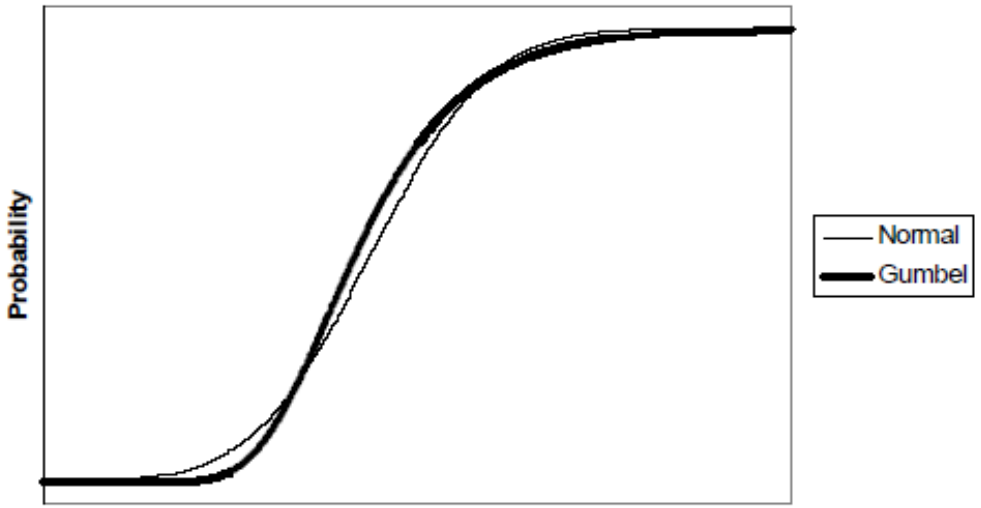
Discrete Choice Theory

Gumbel Distribution

The CDF and of a Gumbel random variable is

$$F(\epsilon) = e^{-e^{-\mu(\epsilon-\eta)}}$$

where μ is the scale parameter and η is the location parameter.



The reason for using Gumbel is that it is quite close to the normal distribution and the difference of Gumbel random variables follow logistic distribution, which has a closed form CDF!

Using the CDF of the logistic distribution, one can show that the probability of selecting Car is

$$\mathbb{P}[\text{Choosing } \texttt{Car}] = \frac{\exp u\big((\text{TT}_{\texttt{Car}}, \text{TT}_{\texttt{Bus}}), \texttt{Car}\big)}{\exp u\big((\text{TT}_{\texttt{Car}}, \text{TT}_{\texttt{Bus}}), \texttt{Car}\big) + \exp u\big((\text{TT}_{\texttt{Car}}, \text{TT}_{\texttt{Bus}}), \texttt{Bus}\big)}$$

In general, for more any finite number of alternatives the probability with which an individual will choose alternative a can be written as

$$\mathbb{P}[a|x] = \frac{\exp(u(x, a))}{\sum_{a' \in A} \exp(u(x, a'))}$$

These type of choice models are also called Multinomial Logit models (MNL). To explicitly represent the dependence on the parameter vector β, we can write the above notation as $\mathbb{P}[a|x; \beta]$.

Suppose an individual experiences a travel time of 20 minutes by `Car` and 30 minutes by `Bus`. Calculate the probabilities of selecting `Car` for

- $\beta_0 = 0.1, \beta_1 = -0.02$
- $\beta_0 = 0.2, \beta_1 = -0.01$

Discrete Choice Theory

Maximum Likelihood Estimation

In practice, we would have access to choices made by single individual in different contexts or multiple individuals from stated-preference surveys or revealed preferences (e.g., Amazon).

Thus, we can repeat this exercise for all observations $n = 1, \ldots, N$, to construct a likelihood function

$$\mathcal{L}(\beta) = \prod_{n=1}^{N} \mathbb{P}[a_n | x_n; \beta]$$

This is slightly modified by taking the logarithm on both sides to construct a log-likelihood function

$$\mathcal{LL}(\beta) = \sum_{n=1}^{N} \ln \mathbb{P}[a_n | x_n; \beta]$$

This objective is maximized to get β's that match the observed and predicted choices as closely as possible.

The above optimization problem is typically solved using a gradient ascent method. In general, depending on the assumptions on error terms, the objective can sometimes be non-concave and we are only guaranteed a local maxima.

In the earlier example, we did not have a constant in the utility of the bus. Having that would lead to what is called *identification issues*. This is very similar to linear independence of feature vectors.

There are a wide variety of statistical tests to ascertain if the estimates of coefficients are significant and if an alternate choice modeling framework gives better fit.

Discrete Choice Theory

Additional Reading

- Koppelman, F. S., & Bhat, C. (2006). A self instructing course in mode choice modeling: multinomial and nested logit models.
- Train, K. E. (2009). Discrete choice methods with simulation. Cambridge university press.

Dynamic Choice Models

Dynamic Choice Models

Introduction

The examples we saw so far are single shot models in which we select an option and the decision making process ends there.

However, there are several dynamic decisions humans have been making without the use of decision support tool.

Dynamic Choice Models

Examples

- When to replace or repair a machine
- When to retire/switch jobs/purse higher studies
- Buy a certain product now or wait for an upgraded version
- Build or sell property
- Select routes in a stochastic network
- Departure choices during a hurricane
- Sports (batting order, when to substitute?)

Dynamic Choice Models

Overview

In such instances, just as before, we can use the data to infer how decision makers weigh different attributes. This is equivalent to inferring the one-step costs or rewards.

However, since decisions are made over time, we can also uncover their beliefs (transition probabilities) of how the states might change.

For example, they may assume that the prices of a product would go up in the next time period according to some parametric function and these parameters are estimated using the same maximum likelihood estimation method.

Finally, we can also estimate their discount factors which tells us if they were myopic or forward looking!

Dynamic Choice Models

Notation

We will discuss dynamic models using finite horizon MDPs. Extensions to the infinite horizon case is straightforward.

Let time be divided into intervals $t = 1, 2, \ldots, T$. At time step t, the decision maker is assumed to choose an action a_t from A_t after observing a state vector (x_t, ϵ_t).

As before, the vector x_t is assumed to be observable to the analyst and the decision maker but ϵ_t is known to the decision maker only.

For notational ease, will use v and V to denote value functions and assume that the state variable is continuous.

Dynamic Choice Models

Conditional Value Functions

Just like static choice models, the expression for the probability can be written as follows

$$\mathbb{P}[a_t|x_t] = \mathbb{P}[v_t(x_t, a_t) + \epsilon_t(a_t) > v_t(x_t, a_t') + \epsilon_t(a_t') \, \forall \, a_t' \in A_t \backslash \{a_t\}]$$

where $v_t(x_t, a_t)$ is the **conditional value function**, which is a measure of the utility from choosing a_t at time t and behaving "optimally" thereafter.

If the errors are iid Gumbel, we can write

$$\mathbb{P}[a_t|x_t] = \frac{\exp\left(v_t(x_t, a_t)\right)}{\sum_{a_t' \in A_t} \exp\left(v_t(x_t, a_t')\right)}$$

The challenge is to compute v's. Earlier they just represented some linear parametric utility functions, but now it has to be calculated by solving an MDP.

Dynamic Choice Models

Bellman Equations

The decision maker selects an action after observing a state (x_t, ϵ_t) in time period t. Let u denote the one-step utility/reward.

Suppose that Π denotes the set of all admissible policies. Then, the objective of the decision maker is

$$\max_{\pi \in \Pi} \mathbb{E} \sum_{t=1}^{T} \alpha^{t-1} \Big\{ u\big(x_t, \pi_t(x_t, \epsilon_t)\big) + \epsilon_t\big(\pi_t(x_t, \epsilon_t)\big) \Big\}$$

where the expectation is taken over the distributions of x and ϵ.

The ϵ's are assumed to be iid over time periods with a probability density function $g(\epsilon_t)$ and the observable component of the state vector x is assumed Markovian with a probability density $f(x_{t+1}|x_t, \epsilon_t, a_t)$.

However, for the sake of tractability, it is assumed that $f(x_{t+1}|x_t, \epsilon_t, a_t) = f(x_{t+1}|x_t, a_t)$, which is commonly referred to as the **conditional independence** assumption. ϵ's are similar to uncontrollable state components.

Dynamic Choice Models

Bellman Equations

The decision maker uses the state (x_t, ϵ_t) to take an action at time t. Let $V_t(x_t, \epsilon_t)$ be the associated value function. The Bellman's optimality conditions for the individual are as follows

$$V_t(x_t, \epsilon_t) = \max_{a_t \in A_t} \left\{ u(x_t, a_t) + \epsilon_t(d_t) + \alpha \mathbb{E}[V_{t+1}(x_{t+1}, \epsilon_{t+1})] \right\}$$

where $\mathbb{E}[V_t(x_t, \epsilon_t)]$ is the expectation taken over the distributions of x and ϵ. That is,

$$\mathbb{E}[V_{t+1}(x_{t+1}, \epsilon_{t+1})] = \int \left(\int V_{t+1}(x_{t+1}, \epsilon_{t+1}) g(\epsilon_{t+1}) d\epsilon_{t+1} \right) f(x_{t+1}|x_t, a_t) dx_{t+1}$$

Dynamic Choice Models

Bellman Equations

Let the ex-ante value function denoted by $\bar{V}_t(x_t)$ be defined as follows:

$$\bar{V}_t(x_t) = \int V_t(x_t, \epsilon_t) g(\epsilon_t) d\epsilon_t$$

From earlier equations, we may write

$$V_t(x_t, \epsilon_t) = \max_{a_t \in A_t} \left\{ u(x_t, a_t) + \epsilon_t(a_t) + \alpha \int \bar{V}_{t+1}(x_{t+1}) f(x_{t+1}|x_t, a_t) dx_{t+1} \right\}$$

Multiplying both sides with density of ϵ and integrating,

$$\int V_t(x_t, \epsilon_t) g(\epsilon_t) d\epsilon_t =$$

$$\int \max_{a_t \in A_t} \left\{ u(x_t, a_t) + \epsilon_t(a_t) + \alpha \int \bar{V}_{t+1}(x_{t+1}) f(x_{t+1}|x_t, a_t) \right\} g(\epsilon_t) d\epsilon_t$$

$$\bar{V}_t(x_t) = \int \max_{a_t \in A_t} \left\{ u(x_t, a_t) + \epsilon_t(a_t) + \alpha \int \bar{V}_{t+1}(x_{t+1}) f(x_{t+1}|x_t, a_t) dx_{t+1} \right\} g(\epsilon_t) d\epsilon_t$$

Dynamic Choice Models

Bellman Equations

Recall that the conditional value function was the utility of choosing an action and behaving optimally thereafter. Hence, we can write,

$$v_t(x_t, a_t) = u(x_t, a_t) + \alpha \int \bar{V}_{t+1}(x_{t+1}) f(x_{t+1}|x_t, a_t) dx_{t+1}$$

Using the definition of the conditional value function,

$$\begin{aligned}\bar{V}_t(x_t) &= \int \max_{a_t \in A_t} \left\{ v_t(x_t, a_t) + \epsilon_t(a_t) \right\} g(\epsilon_t) d\epsilon_t \\ &= \mathbb{E}_\epsilon \left[\max_{a_t \in A_t} \left\{ v_t(x_t, a_t) + \epsilon_t(a_t) \right\} \right]\end{aligned}$$

If ϵs are assumed to be Gumbel distributed, the above expectation has a closed form solution which is given by

$$\bar{V}_t(x_t) = \gamma + \ln \sum_{a_t \in A_t} \exp\left(v_t(x_t, a_t)\right)$$

The conditional value functions can thus be solved using backward induction and the following sets of equations.

$$v_t(x_t, a_t) = u(x_t, a_t) + \alpha \int \bar{V}_{t+1}(x_{t+1}) f(x_{t+1}|x_t, a_t) dx_{t+1}$$

$$\bar{V}_t(x_t) = \gamma + \ln \sum_{a_t \in A_t} \exp\left(v_t(x_t, a_t)\right)$$

The one-step utilities are parameterized linearly using β's as done in static models, i.e., we can write $u(x_t, a_t)$ as $u(x_t, a_t; \beta)$.

Likewise, the transition beliefs are parametrized by λ. That is, we can write $f(x_{t+1}|x_t, a_t)$ as $f(x_{t+1}|x_t, a_t; \lambda)$.

Dynamic Choice Models

Maximum Likelihood Estimation

Finally, we can use the time-dependent choice probabilities to define a likelihood function and maximize it to obtain estimates of the discount factor and utility functions (and also the transition functions)

Recall,

$$\mathbb{P}[a_{nt}|x_{nt};\alpha,\beta,\lambda] = \frac{\exp\left(v_{nt}(x_{nt},a_{nt};\alpha,\beta,\lambda)\right)}{\sum_{a'_{nt}\in A_{nt}}\exp\left(v_{nt}(x_{nt},a'_{nt};\alpha,\beta,\lambda)\right)}$$

where n denotes an observation. The objective of the estimation procedure is to find $(\hat{\alpha},\hat{\beta},\hat{\lambda})$ which is a solution to

$$\max_{\alpha\in[0,1]} \mathcal{LL}(\alpha,\beta,\lambda) = \sum_{n=1}^{N}\sum_{t=1}^{T} \ln\left(\mathbb{P}[a_{nt}|x_{nt};\alpha,\beta,\lambda]\right)$$

Dynamic Choice Models

Estimation

In order to maximize the likelihood function, one could again use a gradient ascent algorithm or Newton's method with approximate Hessian (called BHHH algorithm).

The steps involved in estimation are as follows

- Start with initial values for the estimates
- Solve the dynamic program using backward induction to obtain the conditional value functions
- Compute the choice probabilities and the objective
- Compute the gradient vector using another backward pass (This can be quite challenging)
- Find a descent direction and step size and repeat the above steps.

Dynamic Choice Models

History

Dynamic discrete choice models were popularized by John Rust (who was a student of McFadden) in 1987.

Rust used 10 years of monthly data on bus mileage and engine condition that was meticulously collected by Harold Zurcher, a superindent of maintenance at the Madison Metropolitan Bus Company.

The sample consisted of 104 buses. The mileage since last replacement was used as the state variable and the one-step costs included cost of repair and replacement.

The transition probabilities were parameterized using an exponential distribution.

Additional Reading:

- Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. Econometrica: Journal of the Econometric Society, 999-1033.
- Rust, J. (1994). Structural estimation of Markov decision processes. Handbook of econometrics, 4, 3081-3143.

Your Moment of Zen