Conditional Independence

Robert A. Miller

Structural Econometrics

September 2017

A Recapitulation

A dynamic discrete choice model

- Each period $t \in \{1, 2, \ldots, T\}$ for $T \leq \infty$, an individual chooses among J mutually exclusive actions.
- Let d_{jt} equal one if action $j \in \{1, \ldots, J\}$ is taken at time t and zero otherwise:

$$d_{jt} \in \{0, 1\}$$

$$\sum_{j=1}^{J} d_{jt} = 1$$

- Suppose that actions taken at time t can potentially depend on the state $z_t \in Z$.
- The current period payoff at time t from taking action j is $u_{jt}(z_t)$.
- Given choices $(d_{1t}, \ldots, d_{Jt})$ in each period $t \in \{1, 2, \ldots, T\}$ the individual's expected utility is:

$$E\left\{\sum_{t=1}^{T}\sum_{j=1}^{J}\beta^{t-1} d_{jt} u_{jt}(z_t)\right\}$$

A Recapitulation

Value function and optimization

- Writing the optimal decision rule as $d_t^o(z) \equiv (d_{1t}^o(z_t), \ldots, d_{Jt}^o(z_t))$, and denoting the value function by $V_t(z_t)$, we obtained:

$$
\begin{aligned}
V_t(z_t) &= \sum_{t=1}^{T} \sum_{j=1}^{J} d_{jt}^o u_{jt}(z_t) \\
&= \sum_{j=1}^{J} d_{jt}^o \left[u_{jt}(z_t) + \beta \sum_{z_{t+1}=1}^{Z} V_{t+1}(z_{t+1}) f_{jt}(z_{t+1} | z_t) \right]
\end{aligned}
$$

- Let $v_{jt}(z_t)$ denote the flow payoff of action j plus the expected future utility of behaving optimally from period $t+1$ on:

$$
v_{jt}(z_t) \equiv u_{jt}(z_t) + \beta \sum_{z_{t+1}=1}^{Z} V_{t+1}(z_{t+1}) f_{jt}(z_{t+1} | z_t)
$$

- Bellman's principle implies:

$$
d_{jt}^o(z_t) \equiv \prod_{k=1}^{K} I\{v_{jt}(z_t) \geq v_{kt}(z_t)\}
$$

A Recapitulation

Estimation

- Partitioning the states $z_t \equiv (x_t, \epsilon_t)$ into those which are observed, x_t, and those that are not, ϵ_t, indexing a given specification of $u_{jt}(z_t)$, $f_{jt}(z_{t+1} | z_t)$ and β by $\theta \in \Theta$, we showed the maximum likelihood estimator, $\theta_{ML} \in \Theta$ selects θ to maximize the joint probability of the observed occurrences:

$$\prod_{n=1}^{N} \int_{\epsilon_T} \ldots \int_{\epsilon_1} \left[\begin{array}{l} \sum_{j=1}^{J} I\{d_{njT} = 1\} d_{jT}^{o}(x_{nT}, \epsilon_T) \times \\ \prod_{t=1}^{T-1} H_{nt}(x_{n,t+1}, \epsilon_{t+1} | x_{nt}, \epsilon_t) g(\epsilon_1 | x_{n1}) \end{array} \right] d\epsilon_1 \ldots d\epsilon_T$$

where:

$$H_{nt}(x_{n,t+1}, \epsilon_{t+1} | x_{nt}, \epsilon_t) \equiv \sum_{j=1}^{J} I\{d_{njt} = 1\} d_{jt}^{o}(x_{nt}, \epsilon_t) f_{jt}(x_{n,t+1}, \epsilon_{t+1} | x_{nt}, \epsilon_t)$$

is the probability density of the pair $(x_{n,t+1}, \epsilon_{t+1})$ conditional on (x_{nt}, ϵ_t) when choices are optimal for θ, and $d_{njt} = 1$.

A Recapitulation

A computational challenge

- What are the computational challenges to enlarging the state space?
    1. Computing the value function;
    2. Solving for equilibrium in a multiplayer setting;
    3. Integrating over unobserved heterogeneity.
- These challenges have led researchers to compromises on several dimensions:
    1. Shrink large data set or use a small data set;
    2. Keep the dimension of the state space small;
    3. Assume all choices and outcomes are observed;
    4. Model unobserved states as a matter of computational convenience;
    5. Consider only one side of market to finesse equilibrium issues;
    6. Adopt parameterizations based on convenient functional forms.

Separable Transitions in the Observed Variables

A simplification

- We could assume that for all (j, t, x_t, ϵ_t) the transition of the observed variables does not depend on the unobserved variables:

$$f_{jt}(x_{t+1} | x_t, \epsilon_t) = f_{jt}(x_{t+1} | x_t)$$

- Since x_{t+1} conveys all the information of x_t for the purposes of forming probability distributions at $t+1$:

$$\begin{aligned} f_{jt}(x_{t+1}, \epsilon_{t+1} | x_t, \epsilon_t) &\equiv g_{t+1}(\epsilon_{t+1} | x_{t+1}, x_t, \epsilon_t) f_{jt}(x_{t+1} | x_t, \epsilon_t) \\ &\equiv g_{t+1}(\epsilon_{t+1} | x_{t+1}, \epsilon_t) f_{jt}(x_{t+1} | x_t) \end{aligned}$$

- The ML estimator maximizes the same criterion function but $H_{nt}(x_{n,t+1}, \epsilon_{t+1} | x_{nt}, \epsilon_t)$ simplifies to:

$$H_{nt}(x_{n,t+1}, \epsilon_{t+1} | x_{nt}, \epsilon_t) = \sum_{j=1}^{J} I\{d_{njt} = 1\} d_{jt}^{o}(x_{nt}, \epsilon_t) g_{t+1}(\epsilon_{t+1} | x_{n,t+1}, \epsilon_t) f_{jt}(x_{n,t+1} | x_{nt})$$

- Note $f_{jt}\left(x_{t+1}\left|x_t\right.\right)$ is identifed for each (j,t) from the transitions.
- Instead of jointly estimating the parameters, we could use a two stage estimator to reduce computation costs:
  1. Estimate $f_{jt}\left(x_{t+1}\left|x_t\right.\right)$ with a cell estimator (for x finite), a nonparametric estimator, or a parametric function;
  2. Define:

$$\widehat{H}_{nt}\left(x_{n,t+1},\epsilon_{t+1}\left|x_{nt},\epsilon_t;\theta\right.\right)\equiv$$
$$\sum_{j=1}^{J} I\left\{d_{njt}=1\right\} d_{jt}^{o}\left(x_{nt},\epsilon_t;\theta\right) g_{t+1}\left(\epsilon_{t+1}\left|x_{n,t+1},\epsilon_t;\theta\right.\right)\widehat{f}_{jt}\left(x_{n,t+1}\left|x_{r}\right.\right.$$

$$\widehat{H}_{nt}\left(x_{n,t+1},\epsilon_{t+1}\left|x_{nt},\epsilon_t;\theta\right.\right)\equiv \sum_{j=1}^{J}\left[\begin{array}{l} I\left\{d_{njt}=1\right\} d_{jt}^{o}\left(x_{nt},\epsilon_t;\theta\right) \\ \times g_{t+1}\left(\epsilon_{t+1}\left|x_{n,t+1},\epsilon_t;\theta\right.\right)\widehat{f}_{jt}\left(x_{n,t+1}\left|x_{nt}\right.\right)\end{array}\right]$$

  3. Select the remaining (preference) parameters to maximize:

$$\prod^{N}\int\prod^{T}\widehat{H}_{nt}\left(x_{n,t+1},\epsilon_{t+1}\left|x_{nt},\epsilon_t\right.;\theta\right) g_1\left(\epsilon_1\left|x_{n1}\right.;\theta\right) d\epsilon$$

- Separable transitions do not, however, free us from:
  1. the curse of multiple integration;
  2. numerically optimization to obtain the value function.
- Suppose in addition, that conditional on x_t the unobserved variable ϵ_{t+1} is are independent of ϵ_t.
- Conditional independence embodies both assumptions:

$$\begin{aligned} f_{jt}\left(x_{t+1}\left|x_t, \epsilon_t\right.\right) &= f_{jt}\left(x_{t+1}\left|x_t\right.\right) \\ g_{t+1}\left(\epsilon_{t+1}\left|x_{t+1}, \epsilon_t\right.\right) &= g_{t+1}\left(\epsilon_{t+1}\left|x_{t+1}\right.\right) \end{aligned}$$

  It implies:

$$f_{jt}\left(x_{t+1}, \epsilon_{t+1}\left|x_t, \epsilon_t\right.\right) = f_{jt}\left(x_{t+1}\left|x_t\right.\right) g_{t+1}\left(\epsilon_{t+1}\left|x_{t+1}\right.\right)$$

- Note that the model in Assignment 1 does not satisfy conditional independence, because posterior beliefs are unobserved state variables governed by a controlled markov process.

Conditional Independence

Simplifying expressions within the likelihood

- Conditional independence simplifies $H_{nt}\left(x_{n,t+1},\epsilon_{t+1}\left|x_{nt},\epsilon_{t}\right.\right)$ to:

$$H_{nt}\left(x_{n,t+1},\epsilon_{t+1}\left|x_{nt},\epsilon_{t}\right.\right)=$$
$$\sum_{j=1}^{J} I\left\{d_{njt}=1\right\} d_{jt}^{o}\left(x_{nt},\epsilon_{t}\right) g_{t+1}\left(\epsilon_{t+1}\left|x_{n,t+1}\right.\right) f_{jt}\left(x_{n,t+1}\left|x_{nt}\right.\right)$$

- Also note that:

$$\prod_{t=1}^{T}\left\{\sum_{j=1}^{J} I\left\{d_{njt}=1\right\} d_{jt}^{o}\left(x_{nt},\epsilon_{t}\right) f_{jt}\left(x_{n,t+1}\left|x_{nt}\right.\right)\right\}=$$
$$\prod_{t=1}^{T}\left\{\sum_{j=1}^{J} I\left\{d_{njt}=1\right\} f_{jt}\left(x_{n,t+1}\left|x_{nt}\right.\right)\right\}\times\prod_{t=1}^{T}\left\{\sum_{j=1}^{J} I\left\{d_{njt}=1\right\} d_{jt}^{o}\left(x \right.\right.$$

$$\prod_{t=1}^{T}\left\{\sum_{j=1}^{J} I\left\{d_{njt}=1\right\} d_{jt}^{o}\left(x_{nt},\epsilon_{t}\right) f_{jt}\left(x_{n,t+1}\left|x_{nt}\right.\right)\right\}$$
$$=\prod_{t=1}^{T}\left\{\sum_{j=1}^{J} I\left\{d_{njt}=1\right\} f_{jt}\left(x_{n,t+1}\left|x_{nt}\right.\right)\right\}$$
$$\times\prod_{t=1}^{T}\left\{\sum_{j=1}^{J} I\left\{d_{njt}=1\right\} d_{jt}^{o}\left(x_{nt},\epsilon_{t}\right)\right\}$$

- Hence the contribution of $n \in \{1, \ldots, N\}$ to the likelihood is the product of:

$$\prod_{t=1}^{T-1} \sum_{j=1}^{J} I\{d_{njt} = 1\} f_{jt}(x_{n,t+1} | x_{nt})$$

and:

$$\int_{\epsilon_T} \ldots \int_{\epsilon_1} \prod_{t=1}^{T-1} \sum_{j=1}^{J} I\{d_{njt} = 1\} d_{jt}^{o}(x_{nt}, \epsilon_t) g_{t+1}(\epsilon_{t+1} | x_{n,t+1}) g_1(\epsilon_1 | x_{n1}) d\epsilon_1 \ldots$$

- The second expression simplifies to:

$$\prod_{t=1}^{T} \left[\sum_{j=1}^{J} I\{d_{njt} = 1\} \int_{\epsilon_t} d_{jt}^{o}(x_{nt}, \epsilon_t) g_t(\epsilon_t | x_{nt}) d\epsilon_t \right]$$

Conditional Independence

Conditional choice probabilities defined

- Under conditional independence, we define for each (t, x_t) the conditional choice probability (CCP) for action j as:

$$\begin{aligned} p_{jt}(x_t) &\equiv \int_{\epsilon_t} d_{jt}^o(x_{nt}, \epsilon_t)\, g_t(\epsilon_t | x_{nt})\, d\epsilon_t \\ &= E\left[d_{jt}^o(x_t, \epsilon_t) | x_t\right] \\ &= \int_{\epsilon_t} \prod_{k=1}^{J} I\{v_{kt}(x_{nt}, \epsilon_t) \leq v_{jt}(x_{nt}, \epsilon_t)\}\, g_t(\epsilon_t | x_{nt})\, d\epsilon_t \end{aligned}$$

- Using this notation, the likelihood can now be compactly expressed as:

$$\sum_{n=1}^{N} \sum_{t=1}^{T-1} \sum_{j=1}^{J} I\{d_{njt} = 1\} \ln\left[f_{jt}(x_{n,t+1} | x_{nt})\right] + \sum_{n=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{J} I\{d_{njt} = 1\} \ln p_{jt}(x_t)$$

Conditional Independence

Reformulating the primitives

- Conditional independence implies that $v_{jt}(x_t, \epsilon_t)$ only depends on ϵ_t through $u_{jt}(x_t, \epsilon_t)$ because:

$$
\begin{aligned}
& v_{jt}(x_t, \epsilon_t) \equiv \\
& u_{jt}(x_t, \epsilon_t) + \beta \int_{\epsilon_t+1} \sum_{x_{t+1}=1}^{X} V_{t+1}(x_{t+1}, \epsilon_{t+1}) f_{jt}\left(x_{t+1} \left| x_t\right.\right) g_{t+1}\left(\epsilon_{t+1} \left| x_{t+}\right.\right.
\end{aligned}
$$

- Without further loss of generality we now define:

$$
u_{jt}(x_t, \epsilon_t) \equiv E\left[u_{jt}(x_t, \epsilon_t) \left| x_t\right.\right] + \epsilon^*_{jt} \equiv u^*_{jt}(x_t) + \epsilon^*_{jt}
$$

- In this way we redefine the primitives by the preferences $u^*_{jt}(x_t)$, the observed variables transitions $f_{jt}\left(x_{t+1} | x_t\right)$, and the distribution of unobserved variables $g^*_t\left(\epsilon^*_t | x_t\right)$ where $\epsilon^*_t \equiv \left(\epsilon^*_{1t}, \ldots, \epsilon^*_{Jt}\right)$.

Conditional Independence

Conditional value functions defined

- Given conditional independence, define the conditional valuation function as:

$$v_{jt}^{*}(x_t) \equiv u_{jt}^{*}(x_t) + \beta \int_{\epsilon_{t+1}} \sum_{x_{t+1}=1}^{X} V_{t+1}^{*}(x_{t+1}, \epsilon_{t+1}^{*}) f_{jt}(x_{t+1} | x_t) g_{t+1}^{*}(\epsilon_{t+1}^{*} | x_{t+1})$$

- Thus $p_{jt}(x)$ is found by integrating over $(\epsilon_{1t}, \ldots, \epsilon_{Jt})$ in the regions:

$$\epsilon_{kt}^{*} - \epsilon_{jt}^{*} \leq v_{jt}^{*}(x_t) - v_{kt}^{*}(x_t)$$

hold for all $k \in \{1, \ldots, J\}$. That is $p_{jt}(x_t)$ can be rewritten:

$$\int_{\epsilon_t} \prod_{k=1}^{J} I\{v_{kt}(x_{nt}, \epsilon_t) \leq v_{jt}(x_{nt}, \epsilon_t)\} g_t(\epsilon_t | x_t) d\epsilon_t$$

$$= \int_{\epsilon_t} \prod_{k=1}^{J} I\{\epsilon_{kt}^{*} - \epsilon_{jt}^{*} \leq v_{jt}^{*}(x_{nt}) - v_{kt}^{*}(x_{nt})\} g_t^{*}(\epsilon_t^{*} | x_t) d\epsilon_t^{*}$$

- Suppose we only had data on the last period T, and wished to estimate the preferences determining choices in T.
- By definition this is a static problem in which $v_{jT}(z_T) \equiv u_{jT}(z_T)$.
- For example to the probability of observing the J^{th} choice is:

$$p_{JT}(z_T) \equiv \int_{-\infty}^{\epsilon_{JT}+u_{JT}(z_T)-u_{1T}(z_T)} \dots \int_{-\infty}^{\epsilon_{JT}+u_{JT}(z_T)-u_{J-1,T}(z_T)} \int_{-\infty}^{\infty} g_T(\epsilon_T | x_T) \, d\epsilon_T$$

- The only essential difference between a estimating a static discrete choice model using ML and a estimating a dynamic model satisfying conditional independence using ML is that parametrizations of $v_{jt}(x_t)$ based on $u_{jt}(x_t)$ do not have a closed form, but must be computed numerically.