Introduction to Dynamic Discrete Choice

Robert A. Miller

Tilburg University

November 2023

- The lecture material for this course is based on 28 sessions found at:
  - http://comlabgames.com/structuraleconometrics/
- The data for problems in dynamic discrete choice typically comprise a sample of individuals or firms with records on some of their:
  - background characteristics
  - choices
  - outcomes from those choices.
- Suppose our model generated the data.
- What are the challenges to estimation and testing?
  1. The choices and outcomes of economic models are typically nonlinear in the underlying parameters of the model we wish to estimate.
  2. The data variables on background, choices and outcomes might be an incomplete description about what is relevant to the model.

A Dynamic Discrete Choice Model

Choices

- 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:

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

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

- At an abstract level assuming that choices are mutually exclusive is innocuous, because two combinations of choices sharing some features but not others can be interpreted as two different choices.
- For example in a female labor supply and fertility model, suppose:

$$j \in \{(\text{work, no birth}), (\text{work, birth}), (\text{no work, no birth}), (\text{no work, birth})\}$$

A Dynamic Discrete Choice Model

Information and states

- Suppose that actions taken at time t can potentially depend on the state $z_t \in Z$.
- For Z finite denote by $f_{jt}(z_{t+1}|z_t)$, the probability of z_{t+1} occurring in period $t+1$ when action j is taken at time t.
- For example in the example above, suppose $z_t = (w_t, k_t)$ where:
  - $k_t \in \{0, 1, \ldots\}$ are the number of births before t.
  - $w_t \equiv d_{1,t-1} + d_{2,t-1}$ is her wage in period t.
- Thus $w_t = 1$ if the female worked in period $t-1$, and $w_t = 0$ otherwise.
- With up to 5 offspring, 3 levels of experience, the number of states including age (say 50 years) is 750.
- Adding in 4 levels of education (less than high school, high school, some college and college graduate) and 3 racial categories, increases this number to 9000.

A Dynamic Discrete Choice Model

Large but sparse matrices

- When Z is finite there is a $Z \times Z$ transition matrix for each (j, t).
- In many applications the matrices are sparse.
- In the example above they have $9,000^2 = 81$ million cells.
- However households can only increase the number of kids one at time.
- They can only increase or decrease their work experience by one unit at most.
- Hence there are at most six cells they can move from (w_t, k_t):

$$\left\{ \begin{array}{l} (w_t, k_t), (w_t, k_t + 1), (w_t + 1, k_t), \\ (w_t + 1, k_t + 1), (w_t - 1, k_t), (w_t - 1, k_t + 1) \end{array} \right\}$$

- Therefore a transition matrix has at most $54,000$ nonzero elements, and all the nonzero elements are one.
- Given a deterministic sequence of actions sequentially taken over S periods, we can form the S period transition matrix by producting the one period transitions.

A Dynamic Discrete Choice Model

More on information and states

- If Z is a Euclidean space $f_{jt}(z_{t+1}|z_t)$ is the probability (density function) of z_{t+1} occurring in period $t+1$ when j is picked at time t.
- With almost identical notation we could model $z_t \in Z_t$ and in this way generalize from states of the world to histories, or information known at t, or t-measurable events.
- For example in a health application we might define $z_t \equiv \{h_s\}_{s=1}^{t-1}$ as a medical record with $h_s \in \{\text{healthy at } s, \text{ sick at } s\}$.

A Dynamic Discrete Choice Model

Preferences and expected utility

- The individual's current period payoff from choosing j at time t is determined by z_t, which is revealed to the individual at the beginning of the period t.
- 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 $t \in \{1, 2, \ldots, T\}$ and each state $z_t \in Z$ the individual's expected utility is:

$$E\left\{\sum_{t=1}^{T}\sum_{j=1}^{J}\beta^{t-1}d_{jt}u_{jt}(z_t)\,|z_1\right\}$$

  where $\beta \in (0,1)$ is the subjective discount factor, and at each period t the expectation is taken over $z_2, \ldots, z_T$.
- Formally, β is redundant if u is subscripted by t; we typically include a geometric discount factor so that infinite sums of utility are bounded, and the optimization problem is well posed.

- Write the optimal decision at period t as a decision rule denoted by $d_t^o(z_t)$ formed from its elements $d_{jt}^o(z_t)$.
- Let $V_t(z_t)$ denote the value function in period t, conditional on behaving according to the optimal decision rule:

$$V_t(z_t) \equiv E\left[\sum_{\tau=t}^{T}\sum_{j=1}^{J}\beta^{\tau-t}d_{j\tau}^o(z_\tau)\,u_{j\tau}(z_\tau)\,|z_t\right]$$

- In terms of period $t+1$:

$$\beta V_{t+1}(z_{t+1}) \equiv \beta E\left\{\sum_{\tau=t+1}^{T}\sum_{j=1}^{J}\beta^{\tau-t-1}d_{j\tau}^o(z_\tau)\,u_{j\tau}(z_\tau)\,|z_{t+1}\right\}$$

Characterizing the Solution

Recursive Representation

- Appealing to Bellman's (1958) principle we obtain, when Z is finite:

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

- A similar expression holds when Z is Euclidean using an integral.

Characterizing the Solution

Optimization

- To compute the optimum for T finite, we first solve a static problem in the last period to obtain $d_T^o(z_T)$ for all $z_T \in Z$.
- Applying backwards induction $i \in \{1, \ldots, J\}$ is chosen to maximize:

$$u_{it}(z_t) + E\left\{\sum_{\tau=t+1}^{T} \sum_{j=1}^{J} \beta^{\tau-t-1} d_{j\tau}^o(z_\tau)\, u_{j\tau}(z_\tau)\, |z_t, d_{it} = 1\right\}$$

- In the stationary infinite horizon case we assume $u_{jt}(z) \equiv u_j(z)$ and that $u_j(z) < \infty$ for all (j, z).
- Consequently expected utility each period is bounded and the contraction mapping theorem applies, proving $d_t^o(z) \to d^o(z)$ for large T.

Inference

Estimating a model when all heterogeneity is observed

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

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

- By definition:

$$d_{jt}^o(z_t) \equiv I\{v_{jt}(z_t) \geq v_{kt}(z_t) \forall\ k\}$$

- Suppose we observe the states z_{nt} and decisions $d_{nt} \equiv (d_{n1t}, \ldots, d_{nJt})$ of individuals $n \in \{1, \ldots, N\}$ over time periods $t \in \{1, \ldots, T\}$.
- Could we use such data to infer the primitives of the model:
  1. A consistent estimator of $f_{jt}(z_{t+1}|z_t)$ can be obtained from the proportion of observations in the (t, j, z_t) cell transitioning to z_{t+1}.
  2. There are $(J-1)\sum_{n=1}^{N} I\{z_{nt} = z_t\}$ inequalities relating pairs of mappings $v_{jt}(z_t)$ and $v_{kt}(z_t)$ for each observation on d_{nt} at (t, z_t).
  3. Can we recursively derive the values of $u_{jt}(z_t)$ from the $v_{jt}(z_t)$ values?

Inference

Why unobserved heterogeneity is introduced into data analysis

- Note that if two people in the data set with the same (t, z_t) made different decisions, say j and k, then $v_{jt}(z_t) = v_{kt}(z_t)$.
- There are two potential problems with taking this approach:
  1. In a large data set it is easy to imagine that for every choice $j \in \{1, \ldots, J\}$ and every (t, z_t) at least one sampled person n sets $d_{njt} = 1$. If so, we would infer the population was indifferent between all the choices. Hence the model would lack empirical content because no behavior can be ruled out.
  2. This approach does not make use of the information that some choices are more likely than others. The sample proportions taking different choices at (t, z_t) might vary, some choices being observed often, others infrequently.
- So treating all heterogeneity as observed, and trying to predict the decisions of individuals, is not a promising approach to analyzing data.

Inference

Unobserved heterogeneity

- A more modest objective is to predict the *probability distribution of choices* margined over the *unobserved heterogeneity*.
- This essentially obliterates differences between macroeconomics and microeconomics.
- We now assume the states can be partitioned into those which are observed, x_t, and those that are not, ϵ_t.
- Define $z_t \equiv (x_t, \epsilon_t)$ and the current payoff from taking action j at t given (x_t, ϵ_t) by $u^*_{jt}(x_t) + \epsilon_{jt}$.
- We might interpret $u^*_{jt}(x_t)$ as $E\left[u_{jt}(z_t)\,|x_t\right]$ when only the j^{th} option is offered (so there is no choice).
- To satisfy a transversality condition, assume $\left\{u^*_{jt}(x)\right\}_{t=1}^{T}$ is a bounded sequence for each $(j, x) \in \{1, \ldots, J\} \times \{1, \ldots, X\}$, and so is:

$$\left\{\int \max\left\{|\epsilon_{1t}|, \ldots, |\epsilon_{Jt}|\right\} g_t\left(\epsilon_t|x_t\right) d\epsilon_t\right\}_{t=1}^{T}$$

- Denote the mixed probability (density) of the pair $(x_{t+1}, \epsilon_{t+1})$, conditional on (x_t, ϵ_t) and the optimal action is j, as:

$$H_{jt}\left(x_{t+1}, \epsilon_{t+1} \left| x_t, \epsilon_t \right.\right) \equiv d_{jt}^{o}\left(x_t, \epsilon_t\right) f_{jt}\left(x_{t+1}, \epsilon_{t+1} \left| x_t, \epsilon_t \right.\right)$$

- The probability of $\{d_1, x_2, \ldots, d_{T-1}, x_T, d_T\}$ given x_1 is:

$$\Pr\left\{d_1, x_2, \ldots, d_{T-1}, x_T, d_T \left| x_1 \right.\right\} =$$
$$\int_{\epsilon_T} \cdots \int_{\epsilon_1} \left[\begin{array}{c} g\left(\epsilon_1 \left| x_1 \right.\right) \sum_{j=1}^{J} d_{jT} d_{jT}^{o}\left(x_T, \epsilon_T\right) \times \\ \prod_{t=1}^{T-1} \sum_{j=1}^{J} d_{jt} H_{jt}\left(x_{t+1}, \epsilon_{t+1} \left| x_t, \epsilon_t \right.\right) \end{array} \right] d\epsilon_1 \ldots d\epsilon_T$$

where $g\left(\epsilon_1 \left| x_1 \right.\right)$ is the density of ϵ_1 conditional on x_1.

Inference

Maximum Likelihood Estimation

- Suppose the data consist of N independent and identically distributed draws from the string of random variables $(X_1, D_1, \ldots, X_T, D_T)$.
- Observation $n \in \{1, \ldots, N\}$ is given by $\left\{x_1^{(n)}, d_1^{(n)}, \ldots, x_T^{(n)}, d_T^{(n)}\right\}$.
- Let $\theta \in \Theta$ uniquely index a specification of $u_{jt}(z_t)$, $f_{jt}\left(z_{t+1} | z_t\right)$ and β.
- Conditional on $x_1^{(n)}$, suppose some $\theta_0 \in \Theta$ generated $\left\{d_1^{(n)}, x_2^{(n)}, \ldots, d_T^{(n)}\right\}_{n=1}^{N}$ for all $n \in \{1, 2, \ldots\}$.
- The *maximum likelihood* (ML) estimator selects $\theta \in \Theta$ to maximize the joint probability of the observed occurrences conditional on the initial conditions:

$$\theta_{ML} \equiv \underset{\theta \in \Theta}{\arg\max} \left\{ N^{-1} \sum_{n=1}^{N} \log\left(\Pr\left\{ d_1^{(n)}, x_2^{(n)}, \ldots, x_T^{(n)}, d_T^{(n)} \middle| x_1^{(n)}; \theta \right\} \right) \right\}$$

- This model is point identified if and only if (iff) θ_0 is the unique solution when $\theta \in \Theta$ is chosen to maximize:

$$\int_{x_1^{(n)}} \log \left(\Pr \left\{ d_1^{(n)}, x_2^{(n)}, \ldots, x_T^{(n)}, d_T^{(n)} \left| x_1^{(n)}; \theta \right. \right\} \right) dF \left(x_1^{(n)} \right)$$

- If the model is point identified, θ_{ML} is $\sqrt{N}$ consistent, asymptotically normal, and asymptotically efficient:
  1. a model is *point identified* if no other model in the Θ set of models has the same *data generating process*.
  2. an estimator of an identified model is *consistent* if it converges to θ_0 in some probabilistic sense as N increases without bound.
  3. the *rate of convergence*, $1/2$ in this case, is the greatest α leaving the limit of $N^{\alpha} (\theta_{ML} - \theta_0)$ bounded in some probabilistic sense.
  4. asymptotic normality means the *limiting distribution* (again as N increases without bound), of $\sqrt{N} (\theta_{ML} - \theta_0)$ is normal.
  5. *asymptotic efficiency* refers to the lowest asymptotic variance of all consistent estimators with the same rate of convergence.

Separable Transitions in the Observed Variables

A simplification

- The multiple integration is computationally demanding.
- We could assume that for all (t, j, x_t, ϵ_t) the transition of the observed variables does not depend on the unobserved variables:

$$F_{jt}\left(x_{t+1} \left| x_t, \epsilon_t \right. ; \theta\right) = F_{jt}\left(x_{t+1} \left| x_t \right. ; \theta\right)$$

- Note $F_{jt}\left(x_{t+1} \left| x_t \right.\right)$ is identified for each (t, j) from the transitions, so there is no conceptual reason for parameterizing this distribution.
- The ML estimator maximizes the same criterion function but $H_{jt}\left(x_{n,t+1}, \epsilon_{t+1} \left| x_{nt}, \epsilon_t \right. ; \theta\right)$ simplifies to:

$$\begin{aligned} & H_{jt}\left(x_{t+1}, \epsilon_{t+1} \left| x_t, \epsilon_t \right. ; \theta\right) \equiv \\ & d_{jt}^{o}\left(x_t, \epsilon_t; \theta\right) f_{jt}\left(x_{t+1} \left| x_t \right. ; \theta\right) f_{j.t+1}\left(\epsilon_{t+1} \left| x_{t+1}, x_t, \epsilon_t \right. ; \theta\right) \end{aligned}$$

,

- Instead of jointly estimating the parameters, we could use a two stage estimator to reduce computation costs:
  1. Estimate $F_{jt}\left(x_{t+1}\left|x_t\right.;\theta\right)$ with a cell estimator, a parametric function, or a nonparametric estimator, with $\widehat{F}_{jt}\left(x_{t+1}\left|x_t\right.;\theta\right)$.
  2. Define:

$$\widehat{H}_{jt}\left(x_{t+1},\epsilon_{t+1}\left|x_t,\epsilon_t\right.;\theta\right)\equiv$$
$$d^o_{jt}\left(x_t,\epsilon_t;\theta\right)\widehat{f}_{jt}\left(x_{t+1}\left|x_t\right.;\theta\right)f_{j.t+1}\left(\epsilon_{t+1}\left|x_{t+1},x_t,\epsilon_t\right.;\theta\right)$$

  3. Choose θ to maximize the product over n of:

$$\int_{\epsilon_T}\ldots\int_{\epsilon_1}\left[\begin{array}{c} g\left(\epsilon_1\left|x_1\right.\right)\sum_{j=1}^{J}d_{jT}d^o_{jT}\left(x_T,\epsilon_T\right)\times \\ \prod_{t=1}^{T-1}\sum_{j=1}^{J}d_{jt}\widehat{H}_{jt}\left(x_{t+1},\epsilon_{t+1}\left|x_t,\epsilon_t\right.\right)\end{array}\right]d\epsilon_1\ldots d\epsilon_T$$

  4. Correct standard errors induced at the first stage of estimation.

Conditional Independence

Conditional independence defined

- Separable transitions do not, however, free us from:
  1. the curse of multiple integration.
  2. numerical optimization to obtain the value function.
- Suppose we assume in addition that ϵ_{t+1}, conditional on x_{t+1}, is independent of x_t (plausible) and ϵ_t (questionable).
- Conditional independence embodies both assumptions:

$$
\begin{aligned}
F_{jt}\left(x_{t+1}\left|x_t, \epsilon_t\right.\right) &= F_{jt}\left(x_{t+1}\left|x_t\right.;\theta\right) \\
F_{j,t+1}\left(\epsilon_{t+1}\left|x_{t+1}, x_t, \epsilon_t\right.\right) &= G_{t+1}\left(\epsilon_{t+1}\left|x_{t+1}\right.;\theta\right)
\end{aligned}
$$

- Conditional independence implies:

$$
F_{jt}\left(x_{t+1}, \epsilon_{t+1}\left|x_t, \epsilon_t\right.\right) = F_{jt}\left(x_{t+1}\left|x_t\right.;\theta\right) G_{t+1}\left(\epsilon_{t+1}\left|x_{t+1}\right.;\theta\right)
$$

Conditional Independence

Simplifying expressions within the likelihood

- Conditional independence implies:

$$
\begin{aligned}
& \sum_{j=1}^{J} d_{njT} d_{jT}^{o}\left(x_{nT}, \epsilon_{T}; \theta\right) g_{1}\left(\epsilon_{1} \left| x_{n1}\right. ; \theta\right) \\
& \times \prod_{t=1}^{T-1} H_{t}\left(x_{t+1}, \epsilon_{t+1} \left| x_{t}, \epsilon_{t}\right. ; \theta\right) \\
= & \sum_{j=1}^{J} d_{nTj} d_{jT}^{o}\left(x_{nT}, \epsilon_{T}; \theta\right) g_{1}\left(\epsilon_{1} \left| x_{n1}\right. ; \theta\right) \\
& \times \prod_{t=1}^{T-1} \sum_{j=1}^{J} \left[d_{jt} d_{jt}^{o}\left(x_{t}, \epsilon_{t}; \theta\right) f_{jt}\left(x_{t+1} \left| x_{t}\right. ; \theta\right) g_{t+1}\left(\epsilon_{t+1} \left| x_{t+1}\right. ; \theta\right)\right] \\
= & \prod_{t=1}^{T-1} \sum_{j=1}^{J} d_{jt} f_{jt}\left(x_{t+1} \left| x_{t}\right. ; \theta\right) \\
& \times \prod_{t=1}^{T} \sum_{j=1}^{J} d_{jt} d_{jt}^{o}\left(x_{t}, \epsilon_{t}; \theta\right) g_{t}\left(\epsilon_{t} \left| x_{t}\right. ; \theta\right)
\end{aligned}
$$

Conditional Independence

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

$$\int_{\epsilon_T \ldots \epsilon_1} \left[\begin{array}{l} g_1 \left(\epsilon_1 \left| x_{n1} \right. ; \theta \right) \sum_{j=1}^{J} d_{njT} d_{jT}^{o} \left(x_{nT}, \epsilon_T ; \theta \right) \times \\ \prod_{t=1}^{T-1} \sum_{j=1}^{J} d_{njt} H_{jt} \left(x_{n,t+1}, \epsilon_{t+1} \left| x_{nt}, \epsilon_t \right. ; \theta \right) \end{array} \right] d\epsilon_1 \ldots d\epsilon_T$$

$$= \int_{\epsilon_T \ldots \epsilon_1} \left[\begin{array}{l} \prod_{t=1}^{T-1} \sum_{j=1}^{J} d_{njt} f_{jt} \left(x_{n,t+1} \left| x_{nt} \right. \right) \times \\ \prod_{t=1}^{T} \sum_{j=1}^{J} d_{njt} d_{jt}^{o} \left(x_{nt}, \epsilon_t ; \theta \right) g_t \left(\epsilon_t \left| x_{nt} \right. ; \theta \right) \end{array} \right] d\epsilon_1 \ldots d\epsilon_T$$

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

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

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:

$$\begin{aligned} p_{tj}(x_t) &\equiv \int_{\epsilon_t} d^o_{tj}(x_t, \epsilon_t)\, g_t(\epsilon_t | x_t)\, d\epsilon_t \\ &= E\left[d^o_{tj}(x_t, \epsilon_t) | x_t\right] \\ &= \int_{\epsilon_t} \prod_{k=1}^{J} I\{v_{tk}(x_t, \epsilon_t) \leq v_{tj}(x_t, \epsilon_t)\}\, g_t(\epsilon_t | x_t)\, d\epsilon_t \end{aligned}$$

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

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

Conditional Independence

Conditional value function

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

$$
\begin{aligned}
v_{jt}(x_t, \epsilon_t) &= u_{jt}(x_t, \epsilon_t) \\
&\quad + \beta \int_{\epsilon} \int_{x_{t+1}} \left\{ \begin{array}{l} V_{t+1}(x_{t+1}, \epsilon) \times \\ f_{tj}\left(x_{t+1} \left| x_t \right.\right) g_{t+1}\left(\epsilon \left| x_{t+1} \right.\right) dx_{t+1} d\epsilon \end{array} \right\}
\end{aligned}
$$

- Given conditional independence, define the *conditional value function* as:

$$
v^*_{tj}(x_t) \equiv u^*_{tj}(x_t) + \beta \int_{\epsilon} \int_{x_{t+1}} \left\{ \begin{array}{l} V_{t+1}(x_{t+1}, \epsilon) \times \\ f_{tj}\left(x_{t+1} \left| x_t \right.\right) g_{t+1}\left(\epsilon \left| x_{t+1} \right.\right) dx_{t+1} d\epsilon \end{array} \right\}
$$

Conditional Independence

Conditional choice probabilities

- Similarly define $p_{jt}(x_t)$, the *conditional choice probability* (CCP) for each (j,t), by integrating over $(\epsilon^*_{t1},\ldots,\epsilon^*_{tJ})$ in the regions where $d^o_{jt}(x_{nt},\epsilon_t)=1$, namely:

$$\epsilon_{tk}-\epsilon_{tj}\leq v^*_{tj}(x_t)-v^*_{tk}(x_t)$$

hold for all $k\in\{1,\ldots,J\}$:

$$\begin{aligned}
p_{jt}(x_t) &= \int_{\epsilon_t}\prod_{k=1}^{J}\mathbf{1}\left\{v_{tk}(x_{nt},\epsilon_t)\leq v_{tj}(x_{nt},\epsilon_t)\right\}g_t\left(\epsilon_t\left|x_t\right.\right)d\epsilon_t \\
&= \int_{\epsilon_t}\prod_{k=1}^{J}I\left\{\epsilon^*_{tk}-\epsilon^*_{tj}\leq v^*_{tj}(x_{nt})-v^*_{tk}(x_{nt})\right\}g^*_t\left(\epsilon^*_t\left|x_t\right.\right)d\epsilon^*_t
\end{aligned}$$

Conditional Independence

Connection with static models

- 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 $v^*_{jT}(x_T) \equiv u^*_{jT}(x_T)$.
- For example to the probability of observing the J^{th} choice is:

$$p_{JT}(x_T) \equiv \int_{-\infty}^{\epsilon_{JT} + u^*_{JT}(x_T) - u^*_{1T}(x_T)} \cdots \int_{-\infty}^{\epsilon_{JT} + u^*_{JT}(x_T) - u^*_{J-1,T}(x_T)} \int_{-\infty}^{\infty} g_T\left(\epsilon^*_T \,|x_T\right) d\epsilon^*_T$$

- The main difference between a estimating a static discrete choice model using ML versus its dynamic analogue satisfying conditional independence using ML is that parameterizations of $v^*_{jt}(x_t)$ based on $u^*_{jt}(x_t)$ do not have a closed form, but must be computed numerically.
- For example if ϵ_{jt} is Type 1 Extreme Value (T1EV), then we would replace "$u^*_{jt}(x_t)$" with "$v^*_{jt}(x_t)$" in a logit.