Aula 02

Modelos de Escolha Discreta com Dados Desagregados

Claudio R. Lucinda

FEA-RP/USP

Agenda

1. Modelos de Escolha Discreta
    - Modelo LOGIT Multinomial
    - Elasticidades e o Problema da IIA

Agenda

1. Modelos de Escolha Discreta
   - Modelo LOGIT Multinomial
   - Elasticidades e o Problema da IIA
2. Soluções para o Problema da IIA

Agenda

Agenda

Modelos de Escolha Discreta

- Agora, iremos discutir os modelos em que a escolha se dá sobre o "espaço de características"; os produtos derivam utilidade apenas na medida em que eles são agregados de características.
- Esta escolha no espaço de característica possui um elemento inerentemente idiosincrático; este lado idiosincrático é o que permite a estimação dos parâmetros.
- Inicialmente começaremos analisando o processo de estimação quando o analista possui dados sobre a escolha individual dos consumidores; depois discutiremos as situações em que apenas possuimos dados agregados.
- Uma bibliografia muito boa sobre esse assunto é o livro do Ken Train.

Modelos de Escolha Discreta

- O analista começará postulando uma função que relaciona estes dados observados com a escolha do consumidor, que chamaremos de $V(x_{nj}, s_{nj})$, sendo que x_{nj} representa as características observadas do produto e s_{nj} as características não observadas.
- Uma vez que alguns aspectos da utilidade do consumidor não são observados, em geral $V \neq U$, em que U é a "verdadeira" utilidade do consumidor. Desta forma, podemos fazer o seguinte ajuste:

$$U_{ij} = V_{ij} + \epsilon_{ij}$$

- Em que i denota o consumidor e j a alternativa. A idéia é que o termo ϵ_{ij} capture os aspectos do produto ou do indivíduo que não são observados pelo econometrista.
- Dada esta definição, as características deste termo dependem fundamentalmente de como o mesmo especifica V_{ij}.

Modelos de Escolha Discreta II

- No entanto, para que possamos estimar os componentes de V_{ij}, precisamos do termo ϵ_{ij}, e de uma distribuição conjunta para os ϵ_{ij} de todos os j. Denominando a distribuição conjunta de $\varepsilon =< \epsilon_{i1}, \epsilon_{i2}, \cdots, \epsilon_{iN} >$, temos:

$$
\begin{aligned}
P_{ij} &= Prob(U_{ij} > U_{ik}, \forall k \neq j) \\
&= Prob(V_{ij} + \epsilon_{ij} > V_{ik} + \epsilon_{ik}, \forall k \neq j) \\
&= Prob(V_{ij} - V_{ik} > \epsilon_{ik} - \epsilon_{ij}, \forall k \neq j) \\
&= Prob(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik}, \forall k \neq j)
\end{aligned}
$$

Modelos de Escolha Discreta III

- Esta última igualdade é uma distribuição acumulada, que nos diz a probabilidade que cada um dos termos aleatórios $\epsilon_{ik} - \epsilon_{ij}$ está abaixo das diferenças entre as avaliações observadas $V_{ij} - V_{ik}$. Podemos calcular este negócio, usando a distribuição conjunta dos ε, com a seguinte integral multidimensional:

$$P_{ij} = \int_{\varepsilon} I(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik}, \forall k \neq j) f(\varepsilon) d(\varepsilon)$$

- Em português, esta integral nos dá a área da distribuição conjunta de ε tal que as diferenças nos componentes idiosincráticos sejam menores do que as diferenças nos componentes determinísticos.
- Diferentes especificações de modelos de escolha discreta surgem em resposta a diferentes especificações da variável aleatória multidimensional ε. Por exemplo, se ε for uma distribuição $N(0, \Omega)$, isso nos dá o modelo probit multinomial.

Modelos LOGIT Multinomial:

- Se ε seguir uma distribuição de valores extremos I:

$$
\begin{aligned}
f(\epsilon_{ij}) &= e^{-\epsilon_{ij}} e^{-e^{-\epsilon_{ij}}} \\
F(\epsilon_{ij}) &= e^{e^{-\epsilon_{ij}}}
\end{aligned}
$$

- Temos o modelo LOGIT Multinomial. É importante notar que, para o caso dos modelos LOGIT, a integral multidimensional que fizemos anteriormente pode ser resolvida analiticamente.
- O primeiro passo para entendermos isso é uma regrinha que diz que as diferenças entre duas variáveis aleatórias que seguem esta distribuição valores extremos I têm distribuição logística:

$$
\begin{aligned}
\epsilon_{ij}^{*} &= \epsilon_{ik} - \epsilon_{ij} \\
F(\epsilon_{ij}^{*}) &= \frac{\epsilon_{ij}^{*}}{1 + \epsilon_{ij}^{*}}
\end{aligned}
$$

Modelo LOGIT Multinomial:

- A segunda parada é que os componentes idiosincráticos das utilidades são i.i.d.; mas antes, vamos reescrever a última das probabilidades antes da integral da seguinte forma:

$$P_{ij} = Prob(\epsilon_{ik} < \epsilon_{ij} + V_{ij} - V_{ik}, \forall k \neq j)$$

- Se o ϵ_{ij} é considerado como dado, esta função nos dá a função de distribuição acumulada para cada ϵ_{ik} avaliada em $\epsilon_{ij} + V_{ij} - V_{ik}$, o que, de acordo com a distribuição valores extremos I é igual a $\exp[-\exp[-(\epsilon_{ij} + V_{ij} - V_{ik})]]$. Como os elementos do vetor ε são independentes, isto significa que esta probabilidade conjunta – afinal de contas, vale para todos os elementos de ε exceto j – é igual a um produto das distribuições individuais:

$$P_{ij}|\epsilon_{ij} = \prod_{k \neq j} e^{-e^{-(\epsilon_{ij}+V_{ij}-V_{ik})}}$$

Modelo LOGIT Multinomial (II):

- Evidentemente, ϵ_{ij} não é dado, desta forma a probabilidade conjunta é a integral desta parada com respeito a todos os valores de ϵ_{ij}:

$$P_{ij} = \int_{\epsilon_{ij}=-\infty}^{\infty} \left(\prod_{k \neq j} e^{-e^{-(\epsilon_{ij}+V_{ij}-V_{ik})}} \right) e^{-\epsilon_{ij}} e^{-e^{-\epsilon_{ij}}} d\epsilon_{ij}$$

- Vamos cozinhar um pouco esta equação; lembrando que, para o produto j , $V_{ij} - V_{ij} = 0$, temos que a integral acima pode ser reconstruída da seguinte forma:

$$P_{ij} = \int_{\epsilon_{ij}=-\infty}^{\infty} \left(\prod_{k} e^{-e^{-(\epsilon_{ij}+V_{ij}-V_{ik})}} \right) e^{-\epsilon_{ij}} d\epsilon_{ij}$$

Modelo LOGIT Multinomial (III):

- Podemos transformar este produtório em soma, uma vez que as bases são iguais:

$$
\begin{aligned}
P_{ij} &= \int_{\epsilon_{ij}=-\infty}^{\infty} \exp\left(-\sum_k e^{-(\epsilon_{ij}+V_{ij}-V_{ik})}\right) e^{-\epsilon_{ij}} d\epsilon_{ij} \\
&= \int_{\epsilon_{ij}=-\infty}^{\infty} \exp\left(-e^{-\epsilon_{ij}} \sum_k e^{-(V_{ij}-V_{ik})}\right) e^{-\epsilon_{ij}} d\epsilon_{ij}
\end{aligned}
$$

- Redefinindo as variáveis de integração, tal que $e^{-\epsilon_{ij}} = t$, tal que $dt = -e^{-\epsilon_{ij}} d\epsilon_{ij}$. Note que, quando $\epsilon_{ij} \to \infty$, $t \to 0$, e quando $\epsilon_{ij} \to -\infty$, $t \to -\infty$, o que faz com que os limites de integração agora sejam 0 e ∞.

Modelo LOGIT Multinomial

- Usando este novo termo:

$$
\begin{aligned}
P_{ij} &= \int_{t=\infty}^{0} \exp\left(-t\sum_{k} e^{-(V_{ij}-V_{ik})}\right)(-dt) \\
&= \int_{t=0}^{\infty} \exp\left(-t\sum_{k} e^{-(V_{ij}-V_{ik})}\right) dt \\
&= \left.\frac{\exp\left(-t\sum_{k} e^{-(V_{ij}-V_{ik})}\right)}{\sum_{k} e^{-(V_{ij}-V_{ik})}}\right|_{0}^{\infty} \\
&= \frac{1}{\sum_{k} e^{-(V_{ij}-V_{ik})}} = \frac{e^{V_{ij}}}{\sum_{k} e^{V_{ik}}}
\end{aligned}
$$

- Podemos resumir os cuidados que temos na estimação dos modelos de escolha discreta em duas afirmações, "apenas diferenças de utilidade são importantes" e "a escala da utilidade é arbitrária".

A Escala da Utilidade é Arbitrária

- Se somarmos uma constante a cada um dos termos V_{ik}, a fórmula da probabilidade do slide anterior não se altera.
- Isso implica que os únicos parâmetros que podem ser estimados são aqueles que capturam diferenças entre as alternativas.
- Como fazer com variáveis que são constantes entre as alternativas:
  - Assumir diferentes coeficientes para cada alternativa
- Como só as diferenças de utilidade importam, na verdade o modelo de utilidade aleatória é expresso em termo de $J-1$ diferenças.

Apenas diferenças de utilidade são importantes

- Podemos notar que se multiplicarmos todos os termos V_{ik} por uma constante, a fórmula do slide anterior não se altera.
- Para lidar com isso, você precisa normalizar a escala dos termos erro, usualmente normalizando a variância dos ε
- No caso do Logit, a variância é $\frac{\pi^2}{6}$, ou aproximadamente 1.6. No Probit, a variância é 1.
- Por isso tem que tomar o cuidado em comparar os coeficientes do Probit e do Logit (os do logit são mais ou menos $\sqrt{1.6}$ o do Probit).
- Os coeficientes são $\frac{\beta}{\sigma}$, com σ sendo o DP do ε.

Estimação dos Modelos de Escolha Discreta:

- Em geral, os procedimentos de estimação do modelo Logit Multinomial está baseado no princípio da Máxima Verossimilhança. Inicialmente, vamos supor que a amostra seja aleatória, e que tenhamos dados sobre N tomadores de decisão. A probabilidade de um indivíduo i escolher a alternativa que ele efetivamente escolheu é igual a:

$$\prod_{j \in J} (P_{ij})^{y_{ij}}$$

- Em que $y_{ij} = 1$ se o indivíduo i escolheu o produto e $y_{ij} = 0$, caso contrário.

Estimação dos Modelos de Escolha Discreta (II):

- Supondo independência das escolhas dos indivíduos, a probabilidade de observação de uma amostra igual à que temos é:

$$L(\beta) = \prod_{i \in N} \prod_{j \in J} (P_{ij})^{y_{ij}}$$

- Em geral, os algoritmos numéricos maximizam o logaritmo desta probabilidade conjunta, o que dá:

$$\ln(L(\beta)) = LL(\beta) = \sum_{i \in N} \sum_{j \in J} y_{ij} \ln P_{ij}$$

Estimação dos Modelos de Escolha Discreta (III):

- Em geral, também podemos dar uma interpretação de GMM ao método de estimação utilizado da seguinte forma. O vetor de parâmetros que minimiza esta função deve atender à seguinte condição de primeira ordem:

$$\frac{\partial LL(\beta)}{\partial \beta} = 0$$

- Para facilitar, vamos supor que $V_{ij} = x_{ij}\beta$. Neste caso, temos:

$$\sum_{i \in N} \sum_{j \in J} (y_{ij} - P_{ij}) x_{ij} = 0$$

Efeitos Marginais

- Efeito Marginal sobre a Probabilidade de escolha do Produto j de uma alteração no atributo r do produto j (efeito marginal próprio):

$$\begin{aligned}\frac{\partial P_j}{\partial x_j^r} &= \frac{\partial(e^{V_{ij}}/\sum_{k\in J} e^{V_{ik}})}{\partial x_j^r} \\ &= \frac{\partial V_j}{\partial x_j^r} P_j(1-P_j)\end{aligned}$$

- Efeito Marginal sobre a Probabilidade de escolha do Produto j de uma alteração no atributo r de um produto $k \neq j$ (efeito marginal Cruzado):

$$\begin{aligned}\frac{\partial P_j}{\partial x_k^r} &= \frac{\partial(e^{V_j}/\sum_{c\in J} e^{V_{ic}})}{\partial x_k^r} \\ &= -\frac{\partial V_{ik}}{\partial x_k^r} P_j P_k\end{aligned}$$

Elasticidades

- Elasticidade Própria:

$$\epsilon = \frac{\partial P_j}{\partial x_j^r} \times \frac{x_j^r}{P_j} = \frac{\partial V_j}{\partial x_j^r}(1 - P_j)x_j^r$$

- Elasticidade Cruzada:

$$\epsilon_{ikr} = \frac{\partial P_j}{\partial x_k^r} \times \frac{x_k^r}{P_j} = -\frac{\partial V_{ik}}{\partial x_k^r} P_k x_k^r$$

- Esse último termo só depende de uma derivada parcial e de coisas relacionadas a k – e não a j
- Isso é chamada "Independência das Alternativas Irrelevantes"

Soluções para o Problema da IIA

1. Estipular "ex ante" que os termos ε sejam correlacionados – a saída do Logit Aninhado
2. Permitir variação aleatória entre os gostos, o que induz a correlação entre os ε – a saída do Random Coefficient Logit (ou Mixed Logit)
3. Algumas literaturas consideram interagir x_k^r com case specific variables pra resolver isso.

IIA – Problemas:

- Como vimos, um dos principais fatores que motivam a escolha do modelo Logit Multinomial para a a estimação de parâmetros estruturais do lado da demanda é a sua simplicidade computacional.
- No entanto, uma das principais críticas que podem ser levantadas a este modelo está relacionada com a hipótese de independência de alternativas irrelevantes (*Independence of Irrelevant Alternatives*-IIA).
- Para ilustrar este problema, vamos considerar um exemplo. Suponha que a probabilidade de escolha em modal de transporte seja de 65%, 15%, 10% e 10%, para carro sozinho, carro compartilhado, ônibus e metrô.
- Supondo que o serviço de metrô fosse melhorado ao ponto que a sua probabilidade de escolha subisse para 19%, as probabilidades de escolha iriam se alterar proporcionalmente.

Exemplo:

Tabela: Alterações nas Probabilidades de Escolha

Alternativa	Prob. Antes	Prob. Depois	Mudança Proporcional	Mudança Alg.
Carro	0,65	0,585	0,900	-0,065
Carro Comp.	0,15	0,135	0,900	-0,015
Ônibus	0,10	0,090	0,900	-0,010
Metrô	0,10	0,19	1,900	+0,90

Padrão de Substituição – Problemas:

- Como conseqüência disso, temos que a maior parte dos usuários novos do metrô viriam diretamente dos carros, simplesmente pelo fato que eles são maioria
  - Isto é o contrário da experiência, que diz que a maior parte dos usuários viriam dos ônibus e dos carros compartilhados.
- Esta limitação decorre do fato que os termos ϵ_{ij} são independentes entre si, hipótese utilizada na derivação das probabilidades de escolha do modelo.
- Podemos relaxar esta hipótese, o que dá margem a diferentes modelos.
- Iremos discutir uma alternativa que ainda preserva parte da facilidade computacional, o Modelo Logit Aninhado

Modelo Logit Aninhado

- A característica principal deste modelo é que agrupa as diferentes alternativas em grupos que são mais similares entre si do que aquelas que estão fora dos respectivos grupos.
- A derivação do modelo se baseia na premissa que algumas alternativas compartilham alguns componentes entre si nos seus termos representativos da parte aleatória da escolha do consumidor.
- Vamos manter o exemplo da tabela acima com respeito às alternativas e supor que Metrô e Ônibus sejam mais próximos entre si – pertencendo, portanto, ao mesmo ninho:

$$
\begin{aligned}
U_C &= V_C + \epsilon_C \\
U_{CC} &= V_{CC} + \epsilon_{CC} \\
U_O &= V_{TC} + V_O + \epsilon_{TC} + \epsilon_O \\
U_M &= V_{TC} + V_M + \epsilon_{TC} + \epsilon_M
\end{aligned}
$$

Modelo Logit Aninhado (II):

- Podemos racionalizar esta estrutura de escolha como uma escolha em dois estágios – ainda que, conceitualmente, a idéia de escolha em dois estágios aqui se baseie em elementos muito diferentes dos discutidos na parte de modelos neoclássicos de demanda.
- Podemos derivar as probabilidades de escolha das diferentes alternativas em duas partes. Inicialmente, condicional à escolha de Transporte Coletivo, temos as seguintes probabilidades de escolha:

$$
\begin{aligned}
P(O|TC) &= \frac{e^{V_O/\sigma}}{e^{V_O/\sigma} + e^{V_M/\sigma}} \\
P(M|TC) &= \frac{e^{V_M/\sigma}}{e^{V_O/\sigma} + e^{V_M/\sigma}}
\end{aligned}
$$

Modelo Logit Aninhado (III):

- No nível superior, temos três opções – carro, carro compartilhado ou transporte coletivo. A probabilidade de escolha neste nível superior é dada por:

$$
\begin{aligned}
P(C) &= \frac{e^{V_C}}{e^{V_C} + e^{V_{CC}} + e^{V_{TC}+\sigma I_{TC}}} \\
P(CC) &= \frac{e^{V_{CC}}}{e^{V_C} + e^{V_{CC}} + e^{V_{TC}+\sigma I_{TC}}} \\
P(TC) &= \frac{e^{V_{TC}+\sigma I_{TC}}}{e^{V_C} + e^{V_{CC}} + e^{V_{TC}+\sigma I_{TC}}}
\end{aligned}
$$

- A última fórmula merece uma análise mais pormenorizada. Podemos notar que o numerador desta função possui um termo diferente dos logit tradicional composto por dois termos.
- O primeiro deles é apenas a parte determinística da utilidade de se pegar transporte coletivo, enquanto o termo I_{TC}

Modelo Logit Aninhado (IV):

- Em especial, $I_{TC} = \log[\exp\left(\frac{V_O}{\sigma}\right) + \exp\left(\frac{V_M}{\sigma}\right)]$.
- O papel do parâmetro σ é mais importante, pois significa que os parâmetros dentro da função V de todas as alternativas dentro de um mesmo ninho são padronizadas por um mesmo valor.
- Uma vez que, por definição, σ é limitado entre zero e um, estes coeficientes são aumentados, aumentando assim a sensibilidade a alterações nos atributos da alternativa dentro do ninho.
- Podemos, a partir dos resultados acima, construir as probabilidades de escolha – não condicionais – das alternativas dentro do ninho "transporte coletivo":

$$\begin{aligned} P(O) &= P(O|TC) \times P(TC) \\ P(M) &= P(M|TC) \times P(TC) \end{aligned}$$

Logit Aninhado – O parâmetro σ

- Este parâmetro é uma função da correlação subjacente entre a parte aleatória da utilidade para os pares de alternativas dentro de um mesmo ninho, caracterizando o grau de substitutibilidade entre os parâmetros.
- Para manter a consistência com o processo de escolha colocado anteriormente, ele deve ser limitado entre zero e um.
  - Caso $\sigma = 0$, temos que existe perfeita correlação entre as alternativas dentro de um mesmo ninho, o que implica que a escolha deixa de ser estocástica e passa a ser determinística.
  - Caso $\sigma = 1$, temos que o modelo Logit Aninhado colapsa no modelo Logit Multinomial tradicional.

Logit Aninhado – Efeitos Marginais

- Vamos discutir como são calculados os efeitos marginais para o caso do Logit Aninhado.
- Para isto, temos que lembrar o cálculo mencionado na seção anterior, em que a probabilidade de escolha da alternativa j é o produto de dois termos:
  - A probabilidade de escolha do ninho K em que a alternativa se encontra, P_K
  - A probabilidade condicional de escolha de j, dado que o ninho escolhido foi o K – $P_{j|K}$:
  - Omitindo o subscrito i, temos que:

$$P_j = P_{j|K} P_K$$

Efeitos Marginais:

- Podemos calcular os efeitos marginais de uma alteração em um atributo de uma alternativa r qualquer com a seguinte fórmula:
$$\frac{\partial P_j}{\partial x_r} = \frac{\partial P_{j|K}}{\partial x_r} P_K + \frac{\partial P_K}{\partial x_r} P_{j|K}$$
- As elasticidades serão diferentes para o caso em que $j = r$, $j \neq r$, mas as alternativas estão no mesmo ninho, ou $j \neq r$ e as alternativas estão em ninhos diferentes.

Efeitos Marginais – Mesma Alternativa

- Supondo que não exista parte determinística na escolha de cada um dos ninhos, temos os seguintes efeitos marginais:

$$\frac{\partial P_j}{\partial x_r} = \frac{\partial V_r}{\partial x_r}\left[P_j^2 - \frac{P_j}{\sigma} + \frac{1-\sigma}{\sigma} P_j P_{j|K}\right]$$

- Caso $\sigma \to 1$, temos que este efeito marginal colapsa para o efeito marginal que vimos para o Logit Multinomial.

Efeitos Marginais – Alternativas no mesmo ninho

- Da mesma forma, caso os produtos sejam diferentes, mas pertencentes ao mesmo ninho, temos:

$$\frac{\partial P_j}{\partial x_r} = -\frac{\partial V_r}{\partial x_r}\left[P_j P_r + \frac{1-\sigma}{\sigma} P_r P_{j|K}\right]$$

- Esta também colapsa para o padrão de substituição visto anteriormente caso $\sigma \to 1$. Além disso, ficará igual ao efeito marginal para alternativas localizadas em ninhos diferentes.

Efeitos Marginais – Alternativas em ninhos diferentes

- Caso as alternativas pertençam a ninhos diferentes, temos a seguinte forma funcional para os efeitos marginais:

$$\frac{\partial P_j}{\partial x_r} = -\frac{\partial V_r}{\partial x_r} P_j P_r$$

- Neste caso, quando o Logit Aninhado colapsa para o logit multinomial, temos que o padrão de substituição se iguala com as alternativas dentro de um mesmo ninho.

Logit Aninhado – Estimação:

- O procedimento de estimação por Máxima Verossimilhança segue exatamente o procedimento que detalhamos no Logit Multinomial.
- Supondo independência das escolhas dos indivíduos, a probabilidade de observação de uma amostra igual à que temos é:

$$L(\beta) = \prod_{i \in N} \prod_{j \in J} (P_{ij})^{y_{ij}}$$

- Em geral, os algoritmos numéricos maximizam o logaritmo desta probabilidade conjunta, o que dá:

$$\ln(L(\beta)) = LL(\beta) = \sum_{i \in N} \sum_{j \in J} y_{ij} \ln P_{ij}$$

Mixed Logit

- As probabilidades de escolha no caso do modelo Mixed Logit são integrais definidas em termos de densidades – estipuladas pelo econometrista – dos parâmetros.
- Basicamente ela é uma média ponderada da probabilidade logit tradicional, ponderada pela densidade em diferentes pontos da distribuição dos coeficientes.

$$P_{ij} = \int \left(\frac{exp(\beta' \mathbf{X_{ij}})}{\sum_k exp(\beta' \mathbf{X_{ik}})} \right) f(\beta) d\beta$$

- Do ponto de vista da estimação, é a mesma coisa – desde que, a cada passo do otimizador, você recalcule esta integral.

Mixed Logit (II):

- Uma forma de ver porque o Mixed Logit é uma forma de se lidar com o problema da IIA é entendendo os coeficientes como componentes de erro que criam correlações entre as utilidades de diferentes alternativas, como em Train (2003).
- Imagine a seguinte especificação para a utilidade:

$$U_{ij} = \alpha' \mathbf{x_{ij}} + \mu_i' \mathbf{z_{ij}} + \varepsilon_{ij}$$

- Em que α é um vetor de coeficientes e μ é um vetor de termos aleatórios de média zero. Neste caso a parcela aleatória da utilidade da alternativa é dada por:

$$\nu_{ij} = +\mu_i' \mathbf{z_{ij}} + \varepsilon_{ij}$$

- E o componente μ_i induz correlação entre as alternativas

Train, Kenneth E. 2003. "Mixed logit." In *Discrete Choice Methods with Simulation*. . 1 ed., Chapter 6, 138–154. Cambridge, MA:Cambridge University Press.