

Logit

(rascunho de notas de aula)

(Train, 2009: caps. 1, 2, 3 e 4)

Victor Gomes

Universidade de Brasilia

Preliminares

O problema principal relacionado à necessidade de simulação numérica é a capacidade limitada dos computadores lidarem com integrais. Ver a discussão no livro.

Nos modelos de escolha discreta se tenta entender os efeitos causais. *Alguns destes fatores que causam as escolhas são observados e outros não.*

- x fatores observados;
- ε fatores não-observados;

- $y = h(x, \varepsilon)$ a escolha dos agentes é relacionada aos fatores por meio de uma função (processo comportamental).

Como temos um fator não observado, ε , então a escolha dos agentes não é determinística.

- $f(\varepsilon)$ função de densidade de ε , que é considerado aleatório.

A probabilidade de que um agente escolha um resultado do conjunto de possibilidades é simplesmente a probabilidade de que os fatores não observados são tais que o processo comportamental resulta na escolha:

$$P(y | x) = \text{Prob}(\varepsilon \text{ s.a. } h(x, \varepsilon) = y)$$

Defina a função indicadora $I[h(x, \varepsilon) = y]$: ela assume 1 quando for verdade e 0 caso contrário. Então a probabilidade de que o agente escolha y é o valor esperado da função indicadora sobre todos os valores possíveis dos fatores não observados:

$$\begin{aligned} P(y | x) &= \text{Prob}(I[h(x, \varepsilon) = y] = 1) = \\ &= \int I[h(x, \varepsilon) = y] f(\varepsilon) d\varepsilon \quad (1) \end{aligned}$$

Para calcular a probabilidade, a integral deve ser solucionada.

Expressão forma-fechada completa

Suponha um modelo logit binário. Na compra ou não de um produto a pessoa obtém algum benefício (positivo ou negativo).

A parte da função utilidade observada pelo economista é $\beta'x$: x é o vetor de variáveis e β o vetor de parâmetros. A parte não observada, ε , da função utilidade é aditiva: $U = \beta'x + \varepsilon$. Assuma que ε é distribuído logisticamente. Então a densidade é:

$$f(\varepsilon) = e^{-\varepsilon}/(1 + e^{-\varepsilon})^2$$

e distribuição cumulativa é:

$$F(\varepsilon) = 1/(1 + e^{-\varepsilon})$$

A pessoa toma uma ação se a utilidade é positiva (benefício).

A probabilidade de ação é dada por:

$$P = \int I[\beta'x + \varepsilon > 0]f(\varepsilon)d\varepsilon$$

$$\begin{aligned} &= \int I[\varepsilon > -\beta'x] f(\varepsilon) d\varepsilon \\ &= \int_{\varepsilon=-\beta'x}^{\infty} f(\varepsilon) d\varepsilon \\ &= 1 - F(-\beta'x) = 1 - \frac{1}{1 + e^{\beta'x}} \\ P &= \frac{e^{\beta'x}}{1 + e^{\beta'x}} \end{aligned} \tag{2}$$

Simulação completa

A simulação é aplicável a qualquer especificação de h ou f . A simulação se apoia no fato de que integração sobre uma densidade é uma forma de calcular média.

Considere a integral $\bar{t} = \int t(\varepsilon)f(\varepsilon)d\varepsilon$, tal que $t(\varepsilon)$ é uma estatística baseada em ε que tem densidade $f(\varepsilon)$.

Esta integral \bar{t} é o valor esperado de t sobre todos os valores possíveis de ε . Esta média pode ser aproximada em uma forma intuitiva. Tome muito sorteios de ε da distribuição f , calcule $t(\varepsilon)$ para cada sorteio (draw) e tome a média dos resultados. A média simulada é um estimador não-viesado da média verdadeira. A média verdadeira é aproximada a média que os sorteios $\rightarrow \infty$.

Este conceito de simulação é a base de todos os métodos de simulação. Como dado pela equação (1), a probabilidade de um resultado particular é uma média da indicadora $I(\cdot)$ sobre todos os valores possíveis de ε . A probabilidade, quando expressa dessa forma, pode ser simulada diretamente usando a seguinte receita:

1. Tome um sorteio de ε de $f(\varepsilon)$. Chame este sorteio de ε^1 , onde o superescrito representa aqui a primeiro sorteio.
2. Determine o quanto $h(x, \varepsilon^1) = y$ com este valor de ε . Então crie $I^1 = 1$; caso contrário faça $I^1 = 0$.
3. Repita os passos 1 a 2 muitas vezes, para um total de R sorteios. O indicador para cada sorteio é denominado I^r para $r = 1, 2, \dots, R$.

4. Calcule a média dos I^r 's. A média é a probabilidade simulada:

$$\tilde{P}(y | x) = \frac{1}{R} \sum_{r=1}^R I^r$$

Esta é a proporção das vezes que os sorteios dos fatores não-observados, quando combinados com as variáveis observadas x resultam no resultado y .

Simulação parcial, forma-fechada parcial

Suponha que os termos aleatórios possam ser decompostos em duas partes: ε_1 e ε_2 . Faça a densidade conjunta ser $f(\varepsilon) = f(\varepsilon_1, \varepsilon_2)$. A densidade conjunta pode ser expressa como o produto de uma densidade marginal com e uma condicional: $f(\varepsilon_1, \varepsilon_2) =$

$f(\varepsilon_2 | \varepsilon_1)f(\varepsilon_1)$. Com esta decomposição a probabilidade na equação (1) pode ser expressa como:

$$= \int_{\varepsilon_1} \left[\int_{\varepsilon_2} I[h(x, \varepsilon_1, \varepsilon_2) = y] f(\varepsilon_2 | \varepsilon_1) d\varepsilon_2 \right] f(\varepsilon_1) d\varepsilon_1 \quad (3)$$

Suponha que existe uma forma-fechada para a integral no colchetes. Denomine esta fórmula de $g(\varepsilon_1) \equiv \int_{\varepsilon_2} I[h(x, \varepsilon_1, \varepsilon_2) = y] f(\varepsilon_2 | \varepsilon_1) d\varepsilon_2$ que é condicional ao valor de ε_1 . A probabilidade então passa a ser:

$$P(y | x) = \int_{\varepsilon_1} g(\varepsilon_1) f(\varepsilon_1) d\varepsilon_1 \quad (4)$$

Se a solução de forma fechada não existe para para esta integral, então ela é aproximada por simulação. Observe que isto é apenas

a média de g sobre a densidade marginal de ε_1 . A probabilidade é simulada tomando draws de $f(\varepsilon_1)$, calculando $g(\varepsilon_1)$ para cada draw, e tomando a média dos resultados.

Este procedimento é chamado de particionamento conveniente do erro. A integral sobre ε_2 dado ε_1 é calculada exatamente, enquanto a integral sobre ε_1 é simulada.

Existem claras vantagens desta abordagem sobre a simulação completa. Integrais analíticas são ambas mais precisas e fáceis de calcular do que as integrais simuladas.

Modelos de escolha discreta

Conjunto de escolhas

1. alternativas mutuamente exclusivas (escolha de apenas uma alternativa);
2. o conjunto de escolha deve ser exaustivo (mas se pode incluir a “não-compra”);
3. o número de alternativas deve ser finito (contagem).

Probabilidades de escolha: Random Utility Models

Quem escolhe, n , possui J alternativas.

A utilidade obtida da alternativa j é U_{nj} , $j = 1, \dots, J$. Esta utilidade é conhecida por quem toma a decisão, mas não pelo economista.

A alternativa i que possui maior utilidade é a escolhida: se e somente se $U_{ni} > U_{nj}$, $\forall j \neq i$.

O economista não observa a utilidade do tomador de decisão. A função utilidade *representativa* é $V_{nj} = V(x_{nj}, s_n) \forall j$. x_{nj} são os atributos das escolhas e s_n são atributos de quem decide.

Uma vez que não se pode observar a utilidade do tomador de decisão temos, a utilidade pode ser decomposta em: $U_{nj} =$

$V_{nj} + \varepsilon_{nj}$. ε_{nj} captura os fatores que afetam a utilidade mas não estão em V_{nj} .

A função de densidade conjunta do vetor $\varepsilon'_{nj} = \{\varepsilon_{n1}, \dots, \varepsilon_{nJ}\}$ é representada por $f(\varepsilon_n)$. Esta densidade permite escrever as probabilidades de escolha da alternativa i :

$$P_{ni} = \text{Prob}(U_{ni} > U_{nj}, \forall j \neq i) \quad (5)$$

$$= \text{Prob}(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}, \forall j \neq i) \quad (6)$$

$$= \text{Prob}(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj}, \forall j \neq i) \quad (7)$$

Esta probabilidade é uma distribuição cumulativa: probabilidade de que cada termo aleatório $\varepsilon_{nj} - \varepsilon_{ni}$ esteja abaixo da quantidade observada $V_{ni} - V_{nj}$. Usando a densidade $f(\varepsilon_n)$, esta probabilidade cumulativa pode ser escrita como:

$$P_{ni} = \int_{\varepsilon} I(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj}, \forall j \neq i) f(\varepsilon_n) d\varepsilon_n \quad (8)$$

tal que $I(\cdot)$ é uma função indicadora. Assume 1 quando o que está entre parênteses é verdadeiro e 0 caso contrário. A integral assume forma-fechada para alguns casos.

- Logit: utilidade é distribuição de valor extremo iid;
- Nested logit: GEV;
- Probit: normal multivariada;
- Mixed logit: um parte que é especificada e outra parte de ε é valor extremo iid.

Identificação

O nível absoluto de utilidade não importa (ver (8)). Apenas os parâmetros que são estimados são aqueles que capturam diferenças entre as alternativas.

- É razoável especificar constante específica: $V_{nj} = x'_{nj} + k_j$, $\forall k_j$, tal que k_j é uma constante associada a alternativa j . Ela captura o efeito médio sobre a utilidade. o procedimento padrão é normalizar uma constante como zero: interpretação de deslocamento em relação a constante da alternativa normalizada.
- Variáveis sociodemográficas. Atributos dos “indivíduos” são especificados para criar diferenças entre as alternativas. Um

exemplo de modelo é o seguinte:

$$U_b = \alpha T_b + \beta M_b + \theta_b Y + k_b + \varepsilon_b$$

$$U_c = \alpha T_c + \beta M_c + \varepsilon_c$$

tal que $\theta_b = \theta_b^0 - \theta_c^0$ é interpretado como o efeito diferencial de renda Y entre as alternativas b e c . O valor de θ_b pode ser positivo ou negativo.

- Termo de erro independentes. A integral (8) pode ter a dimensão reduzida para $(J-1)$ utilizando a diferença de termos de erro $\tilde{\varepsilon}_{nji} \equiv \varepsilon_{nj} - \varepsilon_{ni}$, i.e. a diferença em erros das alternativas i e j . Então $\tilde{\varepsilon}_{ni} = \{\tilde{\varepsilon}_{n1i}, \dots, \tilde{\varepsilon}_{nJi}\}$ é um vetor de dimensão $(J-1)$. Escreva $g(\cdot)$ como a densidade destas diferenças dos erros ($\tilde{\varepsilon}_{ni}$). A integral de dimensão $(J-1)$ é:

$$P_{ni} = \int I(\tilde{\varepsilon}_{nji} < V_{ni} - V_{nj}, \forall j \neq i) g(\tilde{\varepsilon}_{ni}) d\tilde{\varepsilon}_{ni} \quad (9)$$

Normalização e estrutura dos erros

Adicionando uma constante à utilidade de todas as alternativas não muda a escolha. O mesmo acontece quando se multiplica cada alternativa por uma constante. Equivalência: $U_{nj}^0 = V_{nj} + \varepsilon_{nj} \forall j$ é equivalente a $U_{nj}^1 = \lambda V_{nj} + \varepsilon_{nj} \forall j$ para qualquer $\lambda > 0$. É preciso normalizar a escala da utilidade.

Solução: normalizar a variância dos termos de erro. Quando a utilidade é multiplicada por $\lambda > 0$, a variância de cada ε_{nj} muda em λ^2 : $\text{Var}(\lambda\varepsilon_{nj}) = \lambda^2\text{Var}(\varepsilon_{nj})$. Portanto, normalizando a variância do termo de erro equivale a normalizar a escala da utilidade.

Normalização com erro iid. (Normalização direta) Muito importante para a interpretação.* Considere $U_{nj}^0 = x'_{nj}\beta + \varepsilon_{nj}^0$, tal

*Cuidado ao se comparar modelos.

que a variância no termo de erro é $\text{Var}(\varepsilon_{nj}^0) = \sigma^2$. Suponha que a escala seja normalizada para 1. O modelo original passa a ser: $U_{nj}^1 = x'_{nj}(\beta/\sigma) + \varepsilon_{nj}^1$ com $\text{Var}(\varepsilon_{nj}^1) = 1$. Os coeficientes originais são divididos pelo SD da parte não-observada da utilidade. O novo coeficiente reflete o efeito observado das variáveis relativo ao desvio-padrão dos fatores não-observados.

No modelo logit a normalização padrão é $\pi^2/6$, que é em torno de $\sqrt{1.6}$. I.e. $U_{nj}^1 = x'_{nj}(\beta/\sigma)\sqrt{1.6} + \varepsilon_{nj}$ com $\text{Var}(\varepsilon_{nj}) = 1.6$.

Normalização com erros heterocedásticos. As vezes a variância do erro pode ser diferente para diferentes segmentos da população – a variância é diferente em cada segmento. Estratégia: normalização geral da utilidade pela variância de UM segmento,

e então estimar a variância para cada segmento relativo ao “normalizador.”

Normalização com erros correlacionados. Caso complexo. Suponha a matriz de covariância para o caso de 4 alternativas.

$$\Omega = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \cdot & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \cdot & \cdot & \sigma_{33} & \sigma_{34} \\ \cdot & \cdot & \cdot & \sigma_{44} \end{pmatrix} \quad (10)$$

Uma solução comum é normalizar os termos de erro com respeito a uma variância. Com erros iid: $\sigma_{jj} = \sigma_{ii}$ e $\sigma_{ij} = 0$ para $i \neq j$. Normalizando para o primeiro termo, temos:

$$\Omega = \begin{pmatrix} k & \omega_{ab} & \omega_{ac} \\ \cdot & \omega_{bb} & \omega_{bc} \\ \cdot & \cdot & \omega_{cc} \end{pmatrix} \quad (11)$$

tal que k é a diferença para o primeiro termo de erro. Por exemplo, o parâmetro ω_{bb} é a variância da diferença entre os erros da 1a e 3a alternativa relativa a variância da diferença entre os erros da 1a e 2a alternativas. Uma complicação a mais aqui é que a diferença dos termos de erro inclui a covância entre as alternativas.[†]

Agregação

Resultado agregado tem que ser ponderado pela participação de cada grupo n : enumeração da amostra. Considere uma escolha discreta com probabilidade P_{ni} que o tomador de decisão n irá escolher a alternativa i de um conjunto de alternativas. Suponha uma amostra de N tomadores de decisão, $n = 1, \dots, N$, sorteada

[†]E.g. $\text{Var}(\tilde{\varepsilon}_{n21}) = \text{Var}(\varepsilon_{n2} - \varepsilon_{n1}) = \text{Var}(\varepsilon_{n2}) + \text{Var}(\varepsilon_{n1}) - 2\text{Cov}(\varepsilon_{n2}, \varepsilon_{n1})$.

da população. Cada tomador de decisão n tem um peso associado, w_n . Uma estimativa consistente do número total de tomadores de decisão na população que escolhem a alternativa i é denominado de \hat{N}_i , que é simplesmente a soma ponderada das probabilidades individuais:

$$\hat{N}_i = \sum_n w_n P_{ni} \quad (12)$$

A probabilidade média é \hat{N}_i/N . Derivadas médias e elasticidades são obtidas de forma similar.

Segmentação. Quando o número de variáveis explicativas é pequeno e assumem poucos valores é possível estimar os resultados agregados sem utilizar a amostra de tomadores de decisão. Suponha 8 segmentos: 4 níveis educacionais \times 2 gêneros. Possuindo dados do número de pessoas em cada segmento, o resultado agregado pode ser estimado pelo cálculo da probabilidade

em cada segmento, não sobre indivíduos em cada segmento. O número de estimado de pessoas que escolheram alternativa i é

$$\hat{N}_i = \sum_{s=1}^8 w_s P_{si} \quad (13)$$

P_{si} é a probabilidade de que um tomador de decisão no segmento s escolheu alternativa i e w_s é o número de pessoas em cada segmento s .

Previsão

Similar ao procedimento de agregação. Mas os pesos são precisam ser ajustados para refletir mudanças antecipadas ao longo do tempo. Com enumeração amostral, a amostra é ajustada tal que parece que a amostra é selecionada no futuro. A amostra

pode ser ajustada pela (i) mudança do valor das variáveis associadas com cada amostra dos indivíduos e/ou (ii) mudando os pesos associados com cada tomador de decisão para refletir mudanças ao longo do tempo desses indivíduos na população.

Modelo Logit

Modelo Logit

Probabilidades das escolhas

O modelo logit é obtido assumindo que cada ε_{nj} é distribuído como valor extremo iid (ou valor extremo tipo I / duplo exponencial). A densidade para cada componente não-observável da utilidade é:

$$f(\varepsilon_{nj}) = e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}} \quad (14)$$

e a distribuição cumulativa é

$$F(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}} \quad (15)$$

A variância desta distribuição é $\pi^2/6$. Assumindo que a variância é $\pi^2/6$ estamos implicitamente normalizando a escala da utilidade.

A diferença entre duas variáveis valor extremo é distribuída como logística. I.e., se ε_{nj} e ε_{ni} são valor extremo iid, então $\varepsilon_{nji}^* = \varepsilon_{nj} - \varepsilon_{ni}$ segue a distribuição logística:

$$F(\varepsilon_{nji}^*) = \frac{e^{\varepsilon_{nji}^*}}{1 + e^{\varepsilon_{nji}^*}} \quad (16)$$

A distribuição de valor extremo possui caudas um pouco mais grossas do que uma normal (pode suportar mais casos extremos). A distinção entre as duas não costuma ser perceptível quando aplicadas aos dados. A hipótese ou característica crucial da valor extremo é que os erros são independentes uns dos outros. A independência significa que a porção não-observada da utilidade de uma alternativa não é relacionado com o termo não-observado de outra alternativa. Essa hipótese não é tão

restritiva e pode ser interpretada como resultado natural de um modelo bem especificado.*

A probabilidade das escolhas é uma reorganização de (5):

$$P_{ni} = \text{Prob}(\varepsilon_{nj} < \varepsilon_{ni} + V_{ni} - V_{nj}, \forall j \neq i)$$

Se esse é considerado como dado, esta expressão é a distribuição cumulativa para cada ε_{nj} solucionado em $\varepsilon_{ni} + V_{ni} - V_{nj}$. Em termos da distribuição valor extremo será: $\exp(-\exp(-(\varepsilon_{ni} + V_{ni} - V_{nj})))$. Como por hipótese a distribuição cumulativa sobre todo $j \neq i$ é o produto das distribuições cumulativa individuais:

$$P_{ni} | \varepsilon_{ni} = \prod_{j \neq i} e^{-e^{-(\varepsilon_{ni} + V_{ni} - V_{nj})}} \quad (17)$$

*Sob a hipótese de independência, o termo de erro de uma alternativa não adiciona informação ao pesquisador sobre o termo de erro de outra alternativa. Explicado de outra forma, se especifica V_{nj} tal que o restante da função utilidade (não-observado) se essencialmente um “white noise.”

Como ε_{ni} não é dado, a probabilidade de escolha é a integral de $P_{ni} | \varepsilon_{ni}$ sobre todos os valores de ε_{ni} ponderada pela densidade (14):

$$P_{ni} = \int \left(\prod_{j \neq i} e^{-e^{-(\varepsilon_{ni} + V_{ni} - V_{nj})}} \right) e^{-\varepsilon_{ni}} e^{-e^{-\varepsilon_{ni}}} d\varepsilon_{ni} \quad (18)$$

A manipulação desta equação e fazendo $V_{nj} = \beta' x_{nj}$ resulta na expressão de forma-fechada para a *probabilidade de escolha logit*:

$$P_{ni} = \frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}} \quad (19)$$

Propriedades das probabilidades logit:

1. $0 < P_{ni} < 1$. A probabilidade de uma alternativa nunca é exatamente 0 e somente será 1 com uma alternativa.

2. $\sum_{i=1}^J P_{ni} = 1.$

3. Relação da probabilidade logit com a utilidade representativa é *sigmoid*. Relação em formato de S. Mudanças no regressores tem pouco impacto quando a probabilidade é elevada ou baixa. Os regressores tem impacto quando a probabilidade está mais próxima de 0.5.

Por exemplo, no caso de uma escolha binária a probabilidade pode ser representada em uma forma mais suscinta:

$$P_1 = \frac{1}{1 + e^{\beta' x_{nj} - \beta' x_{n1}}} \quad (20)$$

Parâmetro de escala

Em geral a utilidade pode ser expressada como: $U_{nj}^* = V_{nj} + \varepsilon_{nj}^*$, onde a variância do erro é $\sigma^2 \times (\pi^2/6)$. Utilidade normalizada: $U_{nj} = V_{nj}/\sigma + \varepsilon_{nj}$, tal que $\varepsilon_{nj} = \varepsilon_{nj}^*/\sigma$. O termo não observado tem variância $\pi^2/6$: $\text{Var}(\varepsilon_{nj}) = \text{Var}(\varepsilon_{nj}^*/\sigma) = (1/\sigma^2)\text{Var}(\varepsilon_{nj}^*) = (1/\sigma^2) \cdot \sigma^2 \cdot \pi^2/6 = \pi^2/6$. Então a probabilidade de escolha é:

$$P_{ni} = \frac{e^{(\beta^*/\sigma)x_{ni}}}{\sum_j e^{(\beta^*/\sigma)x_{nj}}} \quad (21)$$

Cada coeficiente é re-escalado por $1/\sigma$. Observe que apenas a razão (β^*/σ) pode ser estimada. Usualmente o modelo é expresso na forma re-escalada: $\beta = (\beta^*/\sigma)$.

O parâmetro de escala não afeta a razão dos coeficientes. Disposição à pagar, valor do tempo, e outras medidas de taxas marginais de substituição não são afetadas pelo parâmetro de escala. Apenas a interpretação da magnitude é afetada.

Efeitos Marginais

- O interesse é a determinação dos efeitos marginais de uma mudança em um regressor sobre a probabilidade condicional que $y = 1$.
- Para o modelo geral de probabilidade, o efeito marginal do j -ésimo regressor é:

$$\frac{\partial \text{Prob}[y_i = 1 \mid \mathbf{x}_i]}{\partial x_{ij}} = F'(\mathbf{x}_i' \boldsymbol{\beta}) \beta_j,$$

- Dado um modelo específico existem diversas formas de se calcular o efeito marginal.

- É indicado se utilizar $N^{-1} \sum_i F'(\mathbf{x}'_i \hat{\boldsymbol{\beta}}) \hat{\beta}_j$, a média amostral dos efeitos marginais.
- Para o modelo logit, podemos utilizar a média amostral de y . Então $F(\mathbf{x}'\boldsymbol{\beta}) = \bar{y}$ e $F'(\mathbf{x}'\boldsymbol{\beta}) = F'(F^{-1}(\bar{y}))$. Isto implica que o efeito marginal será:

$$\bar{y}(1 - \bar{y})\hat{\beta}_j$$

- Dicas práticas:
 - Vários estudos apresentam apenas o coeficientes de regressão.

- Os modelos padrão de resultado binário são modelos de um único índice, portanto a razão de dois coeficientes de regressores é igual a razão de efeitos marginais.
- O sinal do coeficiente fornece o sinal do efeito marginal.
- Os coeficientes podem ser usados para obter um limite superior dos efeitos marginais.

Poder e limite

IIA - O modelo logit apresenta certo padrão de substituição. É resumido na independência de alternativas irrelevantes (IIA).

Para qualquer 2 alternativas i e k , a razão das probabilidades logit é:

$$\begin{aligned}\frac{P_{ni}}{P_{nk}} &= \frac{e^{V_{ni}} / \sum_j e^{V_{nj}}}{e^{V_{nk}} / \sum_j e^{V_{nj}}} = \\ &= \frac{e^{V_{ni}}}{e^{V_{nk}}} = e^{V_{ni} - V_{nk}}\end{aligned}\tag{22}$$

Essa razão não depende de outras alternativas que não sejam i e k . Em muitas aplicações esta propriedade é muito útil e realista, em outras não.

Inadequação: caso do ônibus azul e vermelho (Chipman e Debreu).[†] É o pior cenário para o modelo logit.

[†]Baseado nas notas de aula de Aviv Nevo e Ariel Pakes.

- Consumidores escolhem entre dirigir um carro para o trabalho ou tomar um ônibus (vermelho):
 - trabalhar em casa não é opção
 - decisão do quanto trabalhar não depende do transporte
- Metade dos consumidores escolhe o carro e metade o ônibus
- Artificialmente introduza uma nova opção: um ônibus azul
 - consumidores não enxergam cores
 - sem mudança de preço ou serviço

- Na realidade metade continua escolhendo carro e a outra metade se divide entre ônibus vermelho e azul. Também não há mudança no bem-estar dos consumidores.
- Suponha que usamos um modelo logit simples para analisar o bem-estar gerado pela introdução do ônibus azul:

$$u_{ijt} = V_{jt} + \varepsilon_{ijt}$$

opção	$t = 0$		$t = 1$	
	observado	previsto	observado	previsto
	u	V_{j0}	u	V_{j1}
carro	0.5	0		
ônib. vermelho	0.5	0		
ônib. azul				
bem-estar	ln(2)			

Agora introduzimos o ônibus azul. Como os consumidores não enxergam cor, a utilidade observada V é a mesma para o ônibus azul ou vermelho. Se isto é verdade temos:

opção	$t = 0$		$t = 1$			
	observado		previsto		observado	
	u	V_{j0}	u	V_{j1}	u	V_{j1}
carro	0.5	0	0.33	0	0.5	0
ônib. vermelho	0.5	0	0.33	0	0.25	$\ln 0.5$
ônib. azul			0.33	0	0.25	$\ln 0.5$
bem-estar	$\ln(2)$		$\ln(3)$		$\ln(2)$	

Substituição proporcional

A mesma questão pode ser expressada em termos de elasticidades-cruzadas da probabilidade logit. Vamos considerar mudar um

atributo da alternativa j . Queremos saber o efeito desta mudança sobre as probabilidade de todas as demais alternativas. A elasticidade-cruzada é:

$$E_{iz_{nj}} = -\beta_z z_{nj} P_{nj} \quad (23)$$

onde z_{nj} é o atributo da alternativa j e β_z é o seu coeficiente.

Essa elasticidade-cruzada é a mesma para todo i : i não entra na fórmula. A melhora nos atributos de uma alternativa reduz as probabilidades de todas as outras alternativas na mesma proporção.

Este padrão de substituição é uma manifestação da propriedade IIA.

Dados em painel

Pode ser incluído no modelo logit análise de painel: efeitos fixos. A utilidade que o indivíduo n obtém da alternativa j no período t é:

$$U_{njt} = V_{njt} + \varepsilon_{njt} \quad (24)$$

Se ε_{njt} é distribuída como valor extremo, independente sobre n , j e t , então as probabilidades de escolha são:

$$P_{nit} = \frac{e^{V_{nit}}}{\sum_j e^{V_{ejt}}} \quad (25)$$

Neste caso, cada situação de escolha de cada indivíduo é uma observação separada.

Aspecto dinâmico, por exemplo, de formação de hábito de consumo pode ser capturado com a seguinte estrutura:

$$V_{njt} = \alpha y_{nj(t-1)} + \beta x_{njt}$$

tal que $y_{njt} = 1$ quando n escolhe a alternativa j em t . Com $\alpha > 0$, a utilidade da alternativa j no período corrente é maior se a alternativa j foi escolhida no período anterior.

Função não-linear

Em modelos estruturais é comum a necessidade de estimativa de funções não-lineares. Um exemplo discutido por Train é o caso da escolha entre bens e lazer. Assuma que cada preferência do trabalhador possa ser representada por uma utilidade Cobb-Douglas:

$$U = (1 - \beta) \ln G + \beta \ln L$$

O parâmetro β reflete a preferência relativa por bens e lazer, com níveis elevados de β implicando maior preferência por lazer em relação ao trabalho. Cada trabalhador possui um salário por hora fixo w . O trabalhador escolhe o número de horas de trabalho que maximize U sujeito a dotação de horas e a restrição orçamentária.

Quando é adicionado a escolha do meio de transporte ao problema temos um problema condicional da escolha do carro. O trabalhador maximiza U sujeito a restrição de horas para trabalhar, lazer e transporte e a restrição orçamentária (incluindo o custo do transporte). A utilidade que Train e McFadden (1978) usam no BART project é:

$$U_j = -\alpha \left(c_j/w^\beta + w^{1-\beta}t_j \right) \quad (26)$$

aqui $j =$ carro e ônibus. O custo do transporte (viagem) é dividido por w^β e o tempo de viagem é multiplicado por $w^{1-\beta}$. O parâmetro que representa a preferência por bens e lazer entra na utilidade de forma não-linear.

Excedente do consumidor (CS)

Definição: utilidade (em moeda) que a pessoa recebe na situação de escolha.

Defina:

$$CS_n = \frac{1}{\alpha_n} \max_j (U_{nj}) \quad (27)$$

tal que α_n é a utilidade marginal da renda: $dU_n/dY_n = \alpha_n$, dado que Y_n é a renda da pessoa n . Sabemos que o pesquisador não observa U_{nj} . Na notação logit o CS é:

$$E(CS_n) = \frac{1}{\alpha_n} E[\max_j (V_{nj} + \varepsilon_{nj})] \quad (28)$$

com a expectativa sendo todos os valores possíveis de ε_{nj} . Como se assume ε_{nj} valor extremo iid e a utilidade é linear na renda, a

equação acima pode ser escrita como:

$$E(CS_n) = \frac{1}{\alpha_n} \ln \left(\sum_{j=1}^J e^{V_{nj}} \right) + C, \quad (29)$$

tal que C é uma constante desconhecida que representa o fato de que o nível absoluto de utilidade não pode ser medido (em termos práticos C pode ser ignorada).

O total do excedente do consumidor é calculado na população é calculado como soma ponderada de $E(CS_n)$ sobre uma amostra de tomadores de decisão. Os pesos refletem o número de pessoas na população que lidam com as mesmas utilidades representativas do modelo amostral.

A mudança no excedente do consumidor que resulta de mudanças na alternativas e/ou do conjunto de escolha é calculado

a partir de (29) (é o fundamental para análise de efeitos de política). A mudança é calculada pela diferença do CS pós (1) do pré alteração (0):

$$\Delta E(CS_n) = \frac{1}{\alpha_n} \left[\ln \left(\sum_{j=1}^{J^1} e^{V_{nj}^1} \right) - \ln \left(\sum_{j=1}^{J^0} e^{V_{nj}^0} \right) \right] \quad (30)$$

Para calcular CS é preciso ter estimativa da utilidade marginal da renda α_n . Usualmente uma variável de custo ou preço entra na utilidade (e se toma o valor absoluto como α_n). (Em geral se assume que a utilidade marginal da renda é independente da renda).

Derivadas e elasticidade

Quanto muda a probabilidade logit com respeito a alteração em algum fator observado na função utilidade. A mudança na probabilidade de escolha da alternativa i dado mudança no fator observado z_{ni} (mantendo a utilidade representativa das demais alternativas constante):

$$\begin{aligned} \frac{\partial P_{ni}}{\partial z_{ni}} &= \frac{\partial \left(e^{V_{ni}} / \sum_j e^{V_{nj}} \right)}{\partial z_{ni}} \\ &= \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} \frac{\partial V_{ni}}{\partial z_{ni}} - \frac{e^{V_{ni}}}{\left(\sum_j e^{V_{nj}} \right)^2} e^{V_{ni}} \frac{\partial V_{ni}}{\partial z_{ni}} = \end{aligned}$$

$$\frac{\partial P_{ni}}{\partial z_{ni}} = \frac{\partial V_{ni}}{\partial z_{ni}} P_{ni} (1 - P_{ni}) \quad (31)$$

Se a utilidade representativa é linear em z_{ni} com coeficiente β_z , a derivada se torna $\beta_z P_{ni}(1 - P_{ni})$.

O quanto muda a probabilidade quando muda um atributo de outra alternativa? Quanto a probabilidade muda quando muda z_{nj} ?

$$\begin{aligned} \frac{\partial P_{ni}}{\partial z_{nj}} &= \frac{\partial \left(e^{V_{ni}} / \sum_j e^{V_{nk}} \right)}{\partial z_{nj}} \\ &= - \frac{e^{V_{ni}}}{\left(\sum_j e^{V_{nk}} \right)^2} e^{V_{nj}} \frac{\partial V_{nj}}{\partial z_{nj}} = \end{aligned}$$

$$\frac{\partial P_{ni}}{\partial z_{nj}} = - \frac{\partial V_{nj}}{\partial z_{nj}} P_{ni} P_{nj} \quad (32)$$

Quando V_{nj} é linear em z_{nj} com coeficiente β_z , então esta derivada cruzada é $-\beta_z P_{ni} P_{nj}$.

Elasticidades. A elasticidade de P_{ni} com respeito a z_{ni} , uma variável entrando na utilidade da alternativa i , é:

$$E_{iz_{ni}} = \frac{\partial P_{ni}}{\partial z_{ni}} \frac{z_{ni}}{P_{ni}} \quad (33)$$

$$= \frac{\partial V_{ni}}{\partial z_{ni}} P_{ni} (1 - P_{ni}) \frac{z_{ni}}{P_{ni}}$$

$$E_{iz_{ni}} = \frac{\partial V_{ni}}{\partial z_{ni}} (1 - P_{ni}) z_{ni} \quad (34)$$

Se a utilidade representativa é linear em z_{ni} com coeficiente β_z , então $E_{iz_{ni}} = \beta_z (1 - P_{ni}) z_{ni}$.

A elasticidade-cruzada de P_{ni} com respeito a uma variável entrando na alternativa j é:

$$E_{iz_{nj}} = \frac{\partial P_{ni} z_{nj}}{\partial z_{nj} P_{ni}} \quad (35)$$

$$E_{iz_{nj}} = -\frac{\partial V_{nj}}{\partial z_{ni}} (1 - P_{nj}) z_{nj} \quad (36)$$

No caso da utilidade linear a elasticidade cruzada é $E_{iz_{nj}} = -\beta_z(1 - P_{nj})z_{nj}$.

Estimação

Amostra exógena. Suponha uma amostra de N indivíduos. Como a probabilidade logit é forma-fechada o ML tradicional pode ser aplicado. A probabilidade da pessoa n escolher a alternativa que observada é:

$$\prod_i (P_{ni})^{y_{ni}} \quad (37)$$

onde $y_{ni} = 1$ se a pessoa n escolher i e zero caso contrário. Note que uma vez que $y_{ni} = 0$ para todas as alternativas não escolhidas e P_{ni} é elevado a zero, então esse termo é a probabilidade de escolha de i .

Assumindo que cada escolha é independente, a probabilidade de

cada indivíduo na amostra escolher i é:

$$L(\beta) = \prod_{n=1}^N \prod_i (P_{ni})^{y_{ni}} \quad (38)$$

onde β é um vetor contendo os parâmetros do modelo. A função log-verossimilhança é:

$$L(\beta) = \sum_{n=1}^N \sum_i y_{ni} \ln P_{ni} \quad (39)$$

e o estimador do valor para β maximiza esta função.

Amostras baseadas na escolha. Em algumas situações a seleção da amostra pode ou deve possuir viés sobre alguma escolha. Isto pode dar origem a base de dados híbridas: parte com amostragem aleatória e outra parte sem. Se o pesquisador está usando

uma amostra baseada nas escolhas e incluiu uma constante específica às alternativas, então o modelo logit estimado como se a amostra fosse exógena produz estimativas consistentes com a excessão das constantes. Além disso estas constantes são viesadas por um fator conhecido e podem ser ajustadas para serem consistentes.

Em particular, a expectativa da constante estimada para a alternativa j , denominada $\hat{\alpha}_j$ é relacionada com a constante verdadeira α_j^* por

$$E(\hat{\alpha}_j) = \alpha_j^* - \ln(A_j/S_j) \quad (40)$$

onde A_j é a participação dos tomadores de decisão na população que escolhe a alternativa j , e S_j é a participação na *amostra baseada na escolha* de quem escolheu a alternativa j .

Testes

Métrica de ajuste: razão de verossimilhança (LR). LR mede o quão bom é o modelo estimado em comparação a um modelo em que todos os parâmetros são zero. A estatística LR é definida como:

$$\rho = 1 - \frac{LL(\hat{\beta})}{LL(0)} \quad (41)$$

tal que $LL(\hat{\beta})$ é o valor da função log-verossimilhança com os parâmetros estimados e $LL(0)$ é a mesma com função assumindo 0 para todos os parâmetros. A LR vai de zero a um. Zero quando quando os parâmetros estimados não melhores do que $LL(0)$. Um quando os parâmetros estimados preveem perfeitamente as escolhas observadas.

Assim como em regressões se usa estatísticas t . Para hipóteses mais complexas se pode usar estatísticas LR específicas (ver Hayashi, cap. 7).

GEV

Nested Logit

Um modelo nested logit é adequado quando o conjunto de alternativas que o indivíduo enfrenta pode ser dividida em subconjuntos, chamados de ninhos (*nests*). Para este caso se tem as seguintes propriedades:

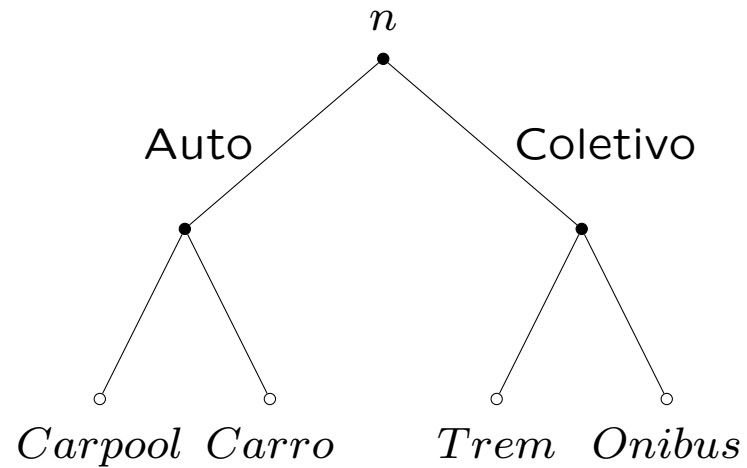
1. Para duas alternativas que estão no mesmo nest, a razão de probabilidades é independente dos atributos ou existência de todas as demais alternativas. IIA vale dentro de cada ninho.
2. Para duas alternativas em *nests diferentes*, a razão de probabilidade pode depender dos atributos de outras alternativas

nos dois nests. IIA não vale em geral para alternativas em nests diferentes.

Exemplo. Suponha um conjunto de alternativas para o trabalhador: dirigir sozinho para o trabalho, carona (carpooling), ônibus ou trem. Se qualquer alternativa for removida, as probabilidades de outras alternativas deveria aumentar. A questão relevante sobre as probabilidades é: qual a proporção de aumento das probabilidades quando uma alternativa é removida?

Probabilidade ($\Delta\%$): com a alternativa removida					
Alternativas	Original	Carro	Carpool	Ônibus	Trem
Carro	.40		.45 (12.5)	.52 (30)	.48 (20)
Carpool	.10	.20 (100)		.13 (30)	.12 (20)
Ônibus	.30	.48 (60)	.33 (10)		.40 (33)
Trem	.20	.32 (60)	.22 (10)	.35 (70)	

Observe que as probabilidades para ônibus e trem aumentam na mesma proporção quando uma das outras alternativas é removida. IIA vale entre estas duas alternativas. Portanto, chame este ninho de 'coletivo.' Similarmente, a probabilidade de carro e carpool aumentam na mesma proporção quando uma das alternativas anteriores for removida. IIA vale entre estas duas alternativas. Forme um nest com estas duas alternativas: 'auto.'



Uma forma de mostrar os ninhos é com o diagrama de árvore. Em cada tronco a propriedade IIA vale. I.e. no segmento 'auto' vale a propriedade IIA, mas não entre escolhas de nests diferentes.

Probabilidades

Faça o conjunto de alternativas j ser dividido em K subconjuntos não sobrepostos representados por B_1, B_2, \dots, B_k e denomine de *nests* (ninhas). A utilidade do indivíduo n obtida da alternativa j no nest B_k é representada por

$$U_{nj} = V_{nj} + \varepsilon_{nj},$$

tal que V_{nj} é observado pelo pesquisador enquanto ε_{nj} é o termo não observado.

Distribuição

A distribuição cumulativa do termo não observado é:

$$F(\varepsilon_{nj}) = \exp \left\{ - \sum_{k=1}^K \left[\sum_{j \in B_k} \exp \left(-\varepsilon_{nj} / \lambda_k \right) \right]^{\lambda_k} \right\}. \quad (42)$$

Esta distribuição é um tipo de GEV. Aqui a distribuição marginal de cada ε_{nj} é valor extremo univariada, mas com os ε_{nj} correlacionados com os ninhos. Para qualquer duas alternativas j e m no ninho B_k , ε_{nj} é correlacionado com ε_{nm} . Para qualquer duas alternativas em nests diferentes, a parte não-observada da utilidade é ainda não-correlacionada: $\text{Cov}(\varepsilon_{nj}, \varepsilon_{nm}) = 0$ para qualquer $j \in B_k$ e $m \in B_l$ com $l \neq k$.

Aqui λ_k mede o grau de independência da utilidade não-observada entre as alternativas no ninho k . Um alto valor de λ_k significa maior independência e menor correlação. A estatística $(1 - \lambda_k)$ é a medida de correlação – se λ_k aumenta então se tem menos correlação (a estatística cai).* Se $\lambda_k = 1$ então se tem completa

*Na verdade, esta correlação é complexa, mas esta estatística é um bom indicador.

independência entre as alternativas e o nested logit se torna o logit tradicional.

Probabilidades

Dada a distribuição da utilidade não-observada, a probabilidade de escolha da alternativa $i \in B_k$ é:

$$P_{ni} = \frac{e^{V_{ni}/\lambda_k} \left(\sum_{j \in B_k} e^{V_{nj}/\lambda_k} \right)^{\lambda_k - 1}}{\sum_{t=1}^K \left(\sum_{j \in B_t} e^{V_{nj}/\lambda_t} \right)^{\lambda_t}} \quad (43)$$

Podemos usar esta fórmula para mostrar que a IIA vale em cada subconjunto de alternativas mas não entre subconjuntos. Considere as alternativas $i \in B_k$ e $m \in B_l$. Dado que o denominador

de (43) é o mesmo para todas as alternativas a razão de probabilidades é a razão dos *numeradores*:

$$\frac{P_{ni}}{P_{nm}} = \frac{e^{V_{ni}/\lambda_k} \left(\sum_{j \in B_k} e^{V_{nj}/\lambda_k} \right)^{\lambda_k - 1}}{e^{V_{nm}/\lambda_l} \left(\sum_{j \in B_l} e^{V_{nj}/\lambda_l} \right)^{\lambda_l - 1}} \quad (44)$$

Se $k = l$ (i.e. i e m estão no mesmo ninho) então os fatores entre parênteses se cancelam e temos:

$$\frac{P_{ni}}{P_{nm}} = \frac{e^{V_{ni}/\lambda_k}}{e^{V_{nm}/\lambda_l}} \quad (45)$$

Esta razão é independente para todas as outras alternativas. Para $k \neq l$ (i.e. i e m estão em diferentes nests), os fatores entre parênteses não se cancelam. Neste caso a razão das probabilidades dependem dos atributos de todas as alternativas nos ninhos que contém i e m . Observe que as probabilidades apenas

dependem dos ninhos que contém i e m (“independência dos nests irrelevantes - IIN”).

O valor de λ_k deve ser fixo entre indivíduos e situações de escolha. O valor de λ_k deve estar em um intervalo particular para o modelo ser consistente com o comportamento maximizador de utilidade. Se $0 \leq \lambda_k \leq 1$ para qualquer k , então o modelo é consistente com o comportamento maximizador. Para $\lambda_k > 1$ o modelo é consistente para certo intervalo de variáveis explicativas mas não todos.[†]

Decomposição em dois logits

[†]Valor negativo não consistente, mas implica redução de utilidade com melhoria de atributos.

Forma tradicional de representação do nested logit. A utilidade pode ser decomposta em duas partes: (1) uma parte denominada de W que é constante para todas as alternativas em um ninho e (2) uma parte denominada Y que varia sobre as alternativas em um nest. A utilidade é escrita como:

$$U_{nj} = W_{nk} + Y_{nj} + \varepsilon_{nj} \quad (46)$$

para $j \in B_k$, tal que:

W_{nk} depende apenas das variáveis que descrevem o ninho k . Estas variáveis diferem entre os nests mas não nas alternativas em cada ninho.

Y_{nj} depende das variáveis que descrevem a alternativa j . Estas variáveis mudam nas alternativas em cada ninho k .

Esta decomposição é geral, pois para qualquer W_{nk} , $Y_{nj} \equiv V_{nj} - W_{nk}$.

Esta decomposição permite escrever as probabilidades nested logit como o produto de duas probabilidades logit. A escolha da alternativa $i \in B_k$ é:

$$P_{ni} = P_{ni|B_k} P_{nB_k}, \quad (47)$$

onde $P_{ni|B_k}$ é a probabilidade condicional de escolha da alternativa i dado que a alternativa do nest B_k é feita e P_{nB_k} é a probabilidade marginal de escolher uma alternativa no nest B_k .

As probabilidade marginal e condicional são:

$$P_{nB_k} = \frac{e^{W_{nk} + \lambda_k I_{nk}}}{\sum_{l=1}^K e^{W_{nl} + \lambda_l I_{nl}}} \quad (48)$$

$$P_{ni|B_k} = \frac{e^{Y_{ni}/\lambda_k}}{\sum_{j \in B_k} e^{Y_{nj}/\lambda_k}} \quad (49)$$

tal que

$$I_{nk} = \ln \sum_{j \in B_k} e^{Y_{nj}/\lambda_k}$$

(I_{nk} conecta os o modelo logit inferior com o superior).

A probabilidade de escolher o nest B_k depende da utilidade esperada que a pessoa recebe daquele nest. Esta utilidade esperada inclui (i) a utilidade recebida sem importar a escolha a ser feita (W_{nk}), (ii) mais a utilidade adicional que se recebe pela escolha da melhor alternativa dentro do nest escolhido ($\lambda_k I_{nk}$).

Paired Combinatorial Logit

Cada parte de alternativas é considerado como um nest. Dado que cada alternativa forma um par com cada outra alternativa, então cada alternativa é membro de $J - 1$ nests. Um parâmetro λ_{ij} indica o grau de independência entre as alternativas i e j .

A probabilidade de escolha do modelo PCL é:

$$P_{ni} = \frac{\sum_{j \neq i} e^{V_{ni}/\lambda_{ij}} \left(e^{V_{ni}/\lambda_{ij}} + e^{V_{nj}/\lambda_{ij}} \right)^{\lambda_{ij}-1}}{\sum_{k=1}^{J-1} \sum_{l=k+1}^J \left(e^{V_{nk}/\lambda_{kl}} + e^{V_{nl}/\lambda_{kl}} \right)^{\lambda_{kl}}} \quad (50)$$

A soma no numerador é sobre todos os $J - 1$ nests que a alternativa i faz parte. Para cada um destes nests, o termo sendo

adicionado o mesmo como o numerador da probabilidade nested logit. Então o PCL é como o nested logit exceto pelo fato de que ele permite que i esteja em mais de um ninho. O denominador também assume a forma como no nested logit.

Nested logit generalizado

Ninhos de alternativas são denominados de B_1, B_2, \dots, B_K . Cada alternativa pode ter um membro em mais de um ninho. Uma alternativa é alocada entre os nests, e estando mais alocada em alguns ninhos do que em outros. Um parâmetro de alocação α_{jk} reflete o quanto cada alternativa j é membro do nest k : tal que $\alpha_{jk} \geq 0 \forall j, k$. Assumindo que $\sum_k \alpha_{jk} = 1 \forall j$, isso descreve a proporção de distribuição da alternativa entre os nests.

A probabilidade de escolha da alternativa i é:

$$P_{ni} = \frac{\sum_k \left(\alpha_{ik} e^{V_{ni}} \right)^{1/\lambda_k} \left(\sum_{j \in B_k} \left(\alpha_{jk} e^{V_{nj}} \right)^{1/\lambda_k} \right)^{\lambda_k - 1}}{\sum_{l=1}^K \left(\sum_{j \in B_l} \left(\alpha_{jl} e^{V_{nj}} \right)^{1/\lambda_l} \right)^{\lambda_l}} \quad (51)$$

A fórmula é similar ao nested logit, exceto pelo fato de que numerador é a soma sobre todos os nests que contém a alternativa i , com os pesos aplicados a estes nests. A probabilidade também pode ser escrita na forma de “dois logits.”

$$P_{ni} = \sum_k P_{ni|B_k} P_{nk} \quad (52)$$

tal que a probabilidade do nest k é:

$$P_{nk} = \frac{\left(\sum_{j \in B_k} (\alpha_{jk} e^{V_{nj}})^{1/\lambda_k} \right)^{\lambda_k}}{\sum_{l=1}^K \left(\sum_{j \in B_k} (\alpha_{jl} e^{V_{nj}})^{1/\lambda_l} \right)^{\lambda_l}}$$

e a alternativa da alternativa i dado o nest k é

$$P_{ni|B_k} = \frac{(\alpha_{ik} e^{V_{ni}})^{1/\lambda_k}}{\sum_{j \in B_k} (\alpha_{jk} e^{V_{nj}})^{1/\lambda_k}}$$

Efeito parcial

- Para x_k contínua podemos escrever o efeito parcial:

$$\frac{\partial P(y = j | \mathbf{x})}{\partial x_k} = P(y = j | \mathbf{x}) \left\{ \beta_{jk} - \left[\sum_{h=1}^J \beta_{hk} \exp(\mathbf{x}\beta_h) \right] / g(\mathbf{x}, \beta) \right\}$$

onde β_{hk} é o k -ésimo elemento de β_h e $g(\mathbf{x}, \beta) = 1 + \sum_{h=1}^J \exp(\mathbf{x}\beta_h)$.