Modelos de Escolha Discreta

Bibliografia: Greene, cap. 21

Claudio Lucinda

FEA/USP

Overview

Discrete Choice

- Os modelos para explicar uma variável dependente binária (0/1) normalmente surgem em dois contextos.
- Com dados sobre a variável de interesse e um conjunto de covariadas, o analista está interessado em especificar uma relação entre a primeira (discreta) e a última (contínua).
- Em outros casos, o modelo de escolha binária surge no contexto de um modelo no qual a natureza dos dados observados dita o tratamento especial de um modelo de escolha binária.
- Por exemplo, num modelo de procura de bilhetes para eventos desportivos, em que a variável de interesse seja o número de bilhetes, pode acontecer que a observação consista apenas em saber se o recinto desportivo está lotado (demanda superior ou igual a capacidade então $Y = 1$) ou não ($Y = 0$).
- Geralmente, os modelos e técnicas usados em ambos os casos são os mesmos. No entanto, é útil examinar ambos.

Discrete Choice

- O entrevistado trabalha ou procura trabalho ($Y = 1$) ou não ($Y = 0$) no período em que nossa pesquisa é realizada.
- Acreditamos que um conjunto de fatores, como idade, estado civil, escolaridade e histórico profissional, reunidos em um vetor $\mathbf{x}$ explicam a decisão, de modo que

$$\text{Prob}(Y = 1 \mid \mathbf{x}) = F(\mathbf{x}, \beta) \tag{1}$$

$$\text{Prob}(Y = 0 \mid \mathbf{x}) = 1 - F(\mathbf{x}, \beta). \tag{2}$$

O conjunto de parâmetros β reflete o impacto das mudanças em $\mathbf{x}$ na probabilidade. Por exemplo, entre os fatores que podem nos interessar está o efeito marginal do estado civil na probabilidade de participação na força de trabalho. O problema neste ponto é criar um modelo adequado para o lado direito da equação.

Linear Probability Models

Uma possibilidade é manter a conhecida regressão linear,

$$F(\mathbf{x}, \boldsymbol{\beta}) = \mathbf{x}'\boldsymbol{\beta}$$

Como $E[y \mid \mathbf{x}] = F(\mathbf{x}, \boldsymbol{\beta})$, podemos construir o modelo de regressão,

$$y = E[y \mid \mathbf{x}] + (y - E[y \mid \mathbf{x}]) = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$$

O modelo de probabilidade linear tem uma série de deficiências. Uma pequena complicação surge porque ε é heterocedástico de uma forma que depende de $\boldsymbol{\beta}$. Como $\mathbf{x}'\boldsymbol{\beta} + \varepsilon$ deve ser igual a 0 ou 1, ε é igual a $-\mathbf{x}'\boldsymbol{\beta}$ ou $1 - \mathbf{x}'\boldsymbol{\beta}$, com probabilidades $1 - F$ e F, respectivamente. Assim, você pode facilmente mostrar que

$$\text{Var}[\varepsilon \mid \mathbf{x}] = \mathbf{x}'\boldsymbol{\beta}\left(1 - \mathbf{x}'\boldsymbol{\beta}\right)$$

Uma falha mais séria é que, sem algum ajuste ad hoc com o termo erro, não podemos ter certeza de que as previsões desse modelo realmente parecerão probabilidades.

Modelos de Escolha Discreta e link functions

Nosso requisito, então, é um modelo que produzirá previsões consistentes com a teoria subjacente em (21-1). Para um determinado vetor de regressores, esperaríamos

$$\lim_{\mathbf{x}'\beta \to +\infty} \text{Prob}(Y = 1 \mid \mathbf{x}) = 1$$

$$\lim_{\mathbf{x}'\beta \to -\infty} \text{Prob}(Y = 1 \mid \mathbf{x}) = 0.$$

A princípio, qualquer distribuição de probabilidade contínua adequada definida sobre a linha real será suficiente. A distribuição normal tem sido utilizada em muitas análises, dando origem ao modelo probit,

$$\text{Prob}(Y = 1 \mid \mathbf{x}) = \int_{-\infty}^{\mathbf{x}'\beta} \phi(t)dt = \Phi\left(\mathbf{x}'\beta\right).$$

A função $\Phi(\cdot)$ é uma notação comumente usada para a distribuição normal padrão.

Modelos de Escolha Discreta

Em parte devido à sua conveniência matemática, a distribuição logística,

$$\text{Prob}(Y = 1 \mid \mathbf{x}) = \frac{e^{\mathbf{x}'\beta}}{1 + e^{\mathbf{x}'\beta}} = \Lambda\left(\mathbf{x}'\beta\right)$$

também tem sido usado em muitas aplicações. Usaremos a notação $\boldsymbol{\Lambda}\,(.)$ para indicar a função de distribuição cumulativa logística, também é chamado de modelo logit por motivos que discutiremos na próxima seção. Ambas as distribuições têm a conhecida forma de sino das distribuições simétricas. Outros modelos que não assumem simetria, como o modelo Weibull

$$\text{Prob}(Y = 1 \mid \mathbf{x}) = \exp\left[-\exp\left(\mathbf{x}'\beta\right)\right]$$

e modelo log log complementar (CLOGLOG),

$$\text{Prob}(Y = 1 \mid \mathbf{x}) = 1 - \exp\left[\exp\left(-\mathbf{x}'\beta\right)\right]$$

Mas o Probit e o Logit são os mais comumente utilizados

Efeitos Marginais

O modelo de probabilidade é uma regressão:

$$E[y \mid \mathbf{x}] = 0\left[1 - F\left(\mathbf{x}'\boldsymbol{\beta}\right)\right] + 1\left[F\left(\mathbf{x}'\boldsymbol{\beta}\right)\right] = F\left(\mathbf{x}'\boldsymbol{\beta}\right).$$

Qualquer que seja a distribuição utilizada, é importante observar que os parâmetros do modelo, como os de qualquer modelo de regressão não linear, não são necessariamente os efeitos marginais que estamos acostumados a analisar. Em geral,

$$\frac{\partial E[y \mid \mathbf{x}]}{\partial \mathbf{x}} = \left\{\frac{dF\left(\mathbf{x}'\boldsymbol{\beta}\right)}{d\left(\mathbf{x}'\boldsymbol{\beta}\right)}\right\}\boldsymbol{\beta} = f\left(\mathbf{x}'\boldsymbol{\beta}\right)\boldsymbol{\beta},$$

Efeitos Marginais – Probit e Logit

Para a distribuição normal, esse resultado é

$$\frac{\partial E[y \mid \mathbf{x}]}{\partial \mathbf{x}} = \phi\left(\mathbf{x}'\boldsymbol{\beta}\right)\boldsymbol{\beta}$$

onde $\phi(t)$ é a densidade normal padrão. Para a distribuição logística,

$$\frac{d\Lambda\left(\mathbf{x}'\boldsymbol{\beta}\right)}{d\left(\mathbf{x}'\boldsymbol{\beta}\right)} = \frac{e^{\mathbf{x}'\boldsymbol{\beta}}}{\left(1+e^{\mathbf{x}'\boldsymbol{\beta}}\right)^2} = \Lambda\left(\mathbf{x}'\boldsymbol{\beta}\right)\left[1-\Lambda\left(\mathbf{x}'\boldsymbol{\beta}\right)\right]$$

Assim, no modelo logit,

$$\frac{\partial E[y \mid \mathbf{x}]}{\partial \mathbf{x}} = \Lambda\left(\mathbf{x}'\boldsymbol{\beta}\right)\left[1-\Lambda\left(\mathbf{x}'\boldsymbol{\beta}\right)\right]\boldsymbol{\beta}$$

Efeitos Marginais

- Para calcular os efeitos marginais, pode-se avaliar as expressões nas médias amostrais dos dados ou avaliar os efeitos marginais em cada observação e usar a média amostral dos efeitos marginais individuais.
- As funções são contínuas com primeiras derivadas contínuas, então o Teorema D.12 (o teorema de Slutsky) e assumindo que os dados são "bem comportados" uma lei de grandes números (Teoremas D.4 e D.5) se aplica;
- Em grandes amostras, eles darão a mesma resposta. Mas isso não ocorre em amostras pequenas ou moderadas.
- A prática atual favorece a média dos efeitos marginais individuais quando é possível fazê-lo.
- Efeito Marginal de uma dummy:

$$\text{Ef. marginal} = \text{Prob}\left[Y = 1 \mid \bar{\mathbf{x}}_{(d)}, d = 1\right] - \text{Prob}\left[Y = 1 \mid \bar{\mathbf{x}}_{(d)}, d = 0\right]$$

Index Models e Variáveis Latentes

- Vemos o resultado de uma escolha discreta como um reflexo de uma regressão subjacente.
- Como um exemplo frequentemente citado, considere a decisão de fazer uma grande compra.
- A teoria afirma que o consumidor faz um cálculo de custo-benefício na margem com base nas utilidades obtidas ao fazer a compra e ao não fazer a compra e ao usar o dinheiro para outra coisa.
- Modelamos a diferença entre benefício e custo como uma variável não observada y^* tal que

$$y^* = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$$

Index Models e Variáveis Latentes

- Assumimos que ε tem média zero e possui uma logística padronizada com variância (conhecida) $\pi^2/3$ ou uma distribuição normal padrão com variância um.
- Não observamos o benefício líquido da compra, apenas se ela é feita ou não.
- Portanto, nossa observação é:

$$
\begin{aligned}
y &= 1 \quad \text{se } y^* > 0, \\
y &= 0 \quad \text{se } y^* \leq 0.
\end{aligned}
$$

Nessa formulação, $\mathbf{x}'\beta$ é chamada de index function.

Index Models e Variáveis Latentes- Identificação

- Primeiro, a suposição de variância conhecida de ε é uma normalização inocente.
- Suponha que a variância de ε seja escalada por um parâmetro irrestrito σ^2.
- A regressão latente será $y^* = \mathbf{x}'\boldsymbol{\beta} + \sigma\varepsilon$.
- Mas, $(y^*/\sigma) = \mathbf{x}'(\boldsymbol{\beta}/\boldsymbol{\sigma}) + \varepsilon$ é o mesmo modelo com os mesmos dados.
- Isso significa que σ não pode ser numericamente identificado pelos dados.

Identificação - II

- Em segundo lugar, a suposição de zero para o limite também é inocente somente se o modelo contiver uma constante
- Seja a o suposto limiar diferente de zero e α seja um termo constante desconhecido e, por enquanto, $\mathbf{x}$ e $\boldsymbol{\beta}$ contêm o resto do índice não incluindo o termo constante. Então, a probabilidade de que y seja igual a um é

$$\text{Prob}\left(y^* > a \mid \mathbf{x}\right) = \text{Prob}\left(\alpha + \mathbf{x}'\boldsymbol{\beta} + \varepsilon > a \mid \mathbf{x}\right) = \text{Prob}\left[(\alpha - a) + \mathbf{x}'\boldsymbol{\beta} + \varepsilon > 0 \mid \mathbf{x}\right]$$

Como α é desconhecido, a diferença $(\alpha - a)$ permanece um parâmetro desconhecido. Com as duas normalizações,

$$\text{Prob}\left(y^* > 0 \mid \mathbf{x}\right) = \text{Prob}\left(\varepsilon > -\mathbf{x}'\boldsymbol{\beta} \mid \mathbf{x}\right).$$

Estimação e Inferência em Modelos de Escolha Discreta

- Com exceção do modelo de probabilidade linear, a estimativa de modelos de escolha binária geralmente é baseada no método de máxima verossimilhança.
- Cada observação é tratada como um único sorteio de uma distribuição de Bernoulli (binomial com um sorteio).
- O modelo com probabilidade de sucesso $F\left(\mathbf{x}'\beta\right)$ e observações independentes leva à probabilidade conjunta, ou função de verossimilhança,

$$\text{Prob}\left(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n \mid \mathbf{X}\right) = \prod_{y_i=0}\left[1 - F\left(\mathbf{x}_i'\beta\right)\right]\prod_{y_i=1} F\left(\mathbf{x}_i'\beta\right)$$

Em que $\mathbf{X}$ representa $[\mathbf{x}_i]_{i=1,\ldots,n}$. A log verossimilhança é

$$\ln L = \sum_{i=1}^{n}\left\{y_i \ln F\left(\mathbf{x}_i'\beta\right) + (1 - y_i)\ln\left[1 - F\left(\mathbf{x}_i'\beta\right)\right]\right\}$$

Scores da ML

As derivadas da função verossimilhança são:

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{n} \left[\frac{y_i f_i}{F_i} + (1 - y_i) \frac{-f_i}{(1 - F_i)} \right] \mathbf{x}_i = \mathbf{0}$$

Em que f_i é a densidade, $dF_i / d\left(\mathbf{x}_i'\beta\right)$. Para o logit, isso fica:

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} (y_i - \Lambda_i)\, \mathbf{x}_i = \mathbf{0}.$$

Observe que se $\mathbf{x}_i$ contém um termo constante, as condições de primeira ordem implicam que a média das probabilidades previstas deve ser igual à proporção de uns na amostra. Essa implicação também tem alguma semelhança com as equações normais de mínimos quadrados se considerarmos o termo $y_i - \Lambda_i$ como um resíduo.

Scores da ML II

$$\ln L = \sum_{y_i=0} \ln\left[1 - \Phi\left(\mathbf{x}_i'\beta\right)\right] + \sum_{y_i=1} \ln \Phi\left(\mathbf{x}_i'\beta\right)$$

Os scores de L são

$$\frac{\partial \ln L}{\partial \beta} = \sum_{y_i=0} \frac{-\phi_i}{1-\Phi_i}\mathbf{x}_i + \sum_{y_i=1} \frac{\phi_i}{\Phi_i}\mathbf{x}_i = \sum_{y_i=0} \lambda_i^0 \mathbf{x}_i + \sum_{y_i=1} \lambda_i^1 \mathbf{x}_i .$$

Ou – derivação para o leitor interessado:

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{n} \left[\frac{q_i \phi\left(q_i \mathbf{x}_i'\beta\right)}{\Phi\left(q_i \mathbf{x}_i'\beta\right)}\right] \mathbf{x}_i = \sum_{i=1}^{n} \lambda_i \mathbf{x}_i = \mathbf{0}.$$

where $q_i = 2y_i - 1$

Hessianos

Logit:
The actual second derivatives for the logit model are quite simple:

$$\mathbf{H} = \frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = -\sum_i \Lambda_i \left(1 - \Lambda_i\right) \mathbf{x}_i \mathbf{x}_i'.$$

Para o probit

$$\mathbf{H} = \frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \sum_{i=1}^{n} -\lambda_i \left(\lambda_i + \mathbf{x}_i' \boldsymbol{\beta}\right) \mathbf{x}_i \mathbf{x}_i'.$$

Efeitos Marginais e Elasticidades

- As probabilidades previstas, $F\left(\mathbf{x}'\hat{\beta}\right) = \hat{F}$ e os efeitos marginais estimados $f\left(\mathbf{x}'\hat{\beta}\right) \times \hat{\beta} = \hat{f}\hat{\beta}$ são funções não lineares das estimativas dos parâmetros.
- Para calcular os erros padrão, podemos usar a abordagem de aproximação linear (método delta). Para as probabilidades previstas,

$$\text{Asy. Var}[\hat{F}] = [\partial\hat{F}/\partial\hat{\beta}]'\mathbf{V}[\partial\hat{F}/\partial\hat{\beta}]$$

em que

$$\mathbf{V} = \text{Asy. Var}[\hat{\beta}]$$

Efeitos Marginais e Elasticidades II

A matriz de covariância assintótica estimada de $\hat{\beta}$ pode ser qualquer uma das descritas anteriormente. Seja $z = \mathbf{x}'\hat{\beta}$. Então o vetor de derivadas é

$$[\partial \hat{F}/\partial \hat{\beta}] = [d\hat{F}/dz][\partial z/\partial \hat{\beta}] = \hat{f}\mathbf{x}$$

Reorganizando, temos

$$\text{Asy. Var}[\hat{F}] = \hat{f}^2 \mathbf{x}'\mathbf{V}\mathbf{x}$$

No caso de um x binário, isso fica

$$\text{Asy. Var}[\Delta\hat{F}] = [\partial \Delta\hat{F}/\partial \hat{\beta}]'\mathbf{V}[\partial \Delta\hat{F}/\partial \hat{\beta}]$$

em que

$$[\partial \Delta\hat{F}/\partial \hat{\beta}] = \hat{f}_1 \begin{pmatrix} \overline{\mathbf{x}}_{(d)} \\ 1 \end{pmatrix} - \hat{f}_0 \begin{pmatrix} \overline{\mathbf{x}}_{(d)} \\ 0 \end{pmatrix}.$$

Efeitos Marginais - Probit e Logit

No modelo Probit, $df/dz = -z\phi$, então

$$\text{Asy. Var}[\hat{\gamma}] = \phi^2 \left[\mathbf{I} - (\boldsymbol{\beta}'\mathbf{x})\, \boldsymbol{\beta}\mathbf{x}'\right] \mathbf{V} \left[\mathbf{I} - (\boldsymbol{\beta}'\mathbf{x})\, \boldsymbol{\beta}\mathbf{x}'\right]'.$$

Para o Logit, $\hat{f} = \hat{\Lambda}(1 - \hat{\Lambda})$, então

$$\frac{d\hat{f}}{dz} = (1 - 2\hat{\Lambda}) \left(\frac{d\hat{\Lambda}}{dz}\right) = (1 - 2\hat{\Lambda})\hat{\Lambda}(1 - \hat{\Lambda}).$$

Reorganizando, obtemos

$$\text{Asy. Var}[\hat{\gamma}] = [\Lambda(1 - \Lambda)]^2 \left[\mathbf{I} + (1 - 2\Lambda)\boldsymbol{\beta}\mathbf{x}'\right] \mathbf{V} \left[\mathbf{I} + (1 - 2\Lambda)\mathbf{x}\boldsymbol{\beta}'\right].$$

Modelos de Escolha Discreta Multivariados

Uma extensão natural do modelo probit seria permitir mais de uma equação, com perturbações correlacionadas, no mesmo espírito do modelo de regressão aparentemente não relacionado. A especificação geral para um modelo de duas equações seria:

$$
\begin{aligned}
y_1^* &= \mathbf{x}_1'\beta_1 + \varepsilon_1, \quad y_1 = 1 \quad \text{if } y_1^* > 0, 0 \text{ otherwise} \;, \\
y_2^* &= \mathbf{x}_2'\beta_2 + \varepsilon_2, \quad y_2 = 1 \quad \text{if } y_2^* > 0, 0 \text{ otherwise}, \\
E\left[\varepsilon_1 \mid \mathbf{x}_1, \mathbf{x}_2\right] &= E\left[\varepsilon_2 \mid \mathbf{x}_1, \mathbf{x}_2\right] = 0, \\
\operatorname{Var}\left[\varepsilon_1 \mid \mathbf{x}_1, \mathbf{x}_2\right] &= \operatorname{Var}\left[\varepsilon_2 \mid \mathbf{x}_1, \mathbf{x}_2\right] = 1, \\
\operatorname{Cov}\left[\varepsilon_1, \varepsilon_2 \mid \mathbf{x}_1, \mathbf{x}_2\right] &= \rho.
\end{aligned}
$$

MVDC - Estimação

Isso vai ser por MV A CDF Normal bivariada é

$$\text{Prob}\left(X_1 < x_1, X_2 < x_2\right) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} \phi_2\left(z_1, z_2, \rho\right) dz_1 dz_2$$

Representada por $\Phi_2\left(x_1, x_2, \rho\right)$. A PDF é

$$\phi_2\left(x_1, x_2, \rho\right) = \frac{e^{-(1/2)\left(x_1^2 + x_2^2 - 2\rho x_1 x_2\right)/\left(1-\rho^2\right)}}{2\pi\left(1-\rho^2\right)^{1/2}}.49$$

Seja $q_{i1} = 2y_{i1} - 1$ and $q_{i2} = 2y_{i2} - 1$. Portanto, $q_{ij} = 1$ se $y_{ij} = 1$ e -1 se $y_{ij} = 0$ para $j = 1$ e 2 . Definindo

$$z_{ij} = \mathbf{x}_{ij}' \beta_j \quad \text{e} \quad w_{ij} = q_{ij} z_{ij}, \quad j = 1, 2,$$

MVDC - Estimação II

As probabilidades que entram na Log-Verossimilhança são:

$$\text{Prob}\left(Y_1 = y_{i1}, Y_2 = y_{i2} \mid \mathbf{x}_1, \mathbf{x}_2\right) = \Phi_2\left(w_{i1}, w_{i2}, \rho_{t^*}\right)$$

Em que

$$\rho_{i^*} = q_{i1} q_{i2} \rho$$

A Log-Verossimilhança fica sendo

$$\log L = \sum_{i=1}^{n} \ln \Phi_2\left(w_{i1}, w_{i2}, \rho_{i^*}\right)$$

E os Scores

$$\frac{\partial \ln L}{\partial \beta_j} = \sum_{i=1}^{n} \left(\frac{q_{ij} g_{ij}}{\Phi_2}\right) \mathbf{x}_{ij}, \quad j = 1, 2,$$

$$\frac{\partial \ln L}{\partial \rho} = \sum_{i=1}^{n} \frac{q_{i1} q_{i2} \phi_2}{\Phi_2}, g_{i1} = \phi\left(w_{i1}\right) \Phi\left[\frac{w_{i2} - \rho_{i^*} w_{i1}}{\sqrt{1 - \rho_{i^*}^2}}\right]$$