Modelos de Escolha Discreta

Bibliografia: Greene, cap. 21

Claudio Lucinda

FEA/USP

Overview

Introdução

- Esta aula trata do truncamento e da censura e uma forma de truncamento chamada de problema de seleção de amostras
- O efeito de truncamento ocorre quando dados de amostra são extraídos de um subconjunto de uma população maior de interesse.
    - Por exemplo, estudos de renda baseados em rendas acima ou abaixo de alguma linha de pobreza podem ser de utilidade limitada para inferências sobre toda a população.
    - O truncamento é essencialmente uma característica da distribuição da qual os dados de amostra são extraídos.
- A censura é um problema mais comum em estudos recentes.
    - Para continuar o exemplo, suponha que, em vez de não serem observadas, todas as rendas abaixo da linha da pobreza sejam relatadas como se estivessem na linha da pobreza.
- A censura de um intervalo de valores da variável de interesse introduz uma distorção nos resultados estatísticos convencionais semelhante à do truncamento.
- Ao contrário do truncamento, no entanto, a censura é essencialmente um defeito nos dados de amostra. Presumivelmente, se não fossem censurados, os dados seriam uma amostra representativa da população de interesse.

Distribuições Truncadas

Uma distribuição truncada é a parte de uma distribuição não truncada que está acima ou abaixo de algum valor especificado. Este subconjunto é uma parte da distribuição total de renda que pode variar de zero a (essencialmente) infinito.

Theorem (Densidade de uma Variável Aleatória Truncada)

Se uma variável aleatória contínua x tem pdf $f(x)$ e a é uma constante, então

$$f(x \mid x > a) = \frac{f(x)}{\text{Prob}(x > a)}$$

A prova decorre da definição de probabilidade condicional e equivale apenas a dimensionar a densidade de modo que ela se integre a um na faixa acima de a. Observe que a distribuição truncada é uma distribuição condicional.

Normal Truncada

As aplicações mais recentes baseados em variáveis aleatórias contínuas usam a distribuição normal truncada. Se x tem uma distribuição normal com média μ e desvio padrão σ, então

$$\text{Prob}(x > a) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right) = 1 - \Phi(\alpha),$$

onde $\alpha = (a-\mu)/\sigma$ e $\Phi(\cdot)$ é a cdf normal padrão. A densidade da distribuição normal truncada é então

$$f(x \mid x > a) = \frac{f(x)}{1-\Phi(\alpha)} = \frac{\left(2\pi\sigma^2\right)^{-1/2} e^{-(x-\mu)^2/\left(2\sigma^2\right)}}{1-\Phi(\alpha)} = \frac{\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)}{1-\Phi(\alpha)}$$

onde $\phi(\cdot)$ é o pdf normal padrão.

Momentos de Distribuições Truncadas

Geralmente estamos interessados na média e na variância da variável aleatória truncada. Eles seriam obtidos pela fórmula geral:

$$E[x \mid x > a] = \int_a^{\infty} xf(x \mid x > a)dx$$

para a média e também para a variância.

1. Se o truncamento for de baixo, então a média da variável truncada é maior que a média da variável original. Se o truncamento for de cima, então a média da variável truncada é menor que a média da variável original. Isso é claramente visível na Figura 22.1.
2. O truncamento reduz a variância em comparação com a variância na distribuição não truncada.

Momentos da Normal Truncada

Theorem (Momentos da Normal Truncada)

Se $x \sim N\left[\mu, \sigma^2\right]$ e a é uma constante, então

$$E[x \mid \text{ truncamento }] = \mu + \sigma\lambda(\alpha),$$
$$\text{Var}[x \mid \text{ truncamento }] = \sigma^2[1 - \delta(\alpha)],$$

onde $\alpha = (a - \mu)/\sigma, \phi(\alpha)$ é a densidade normal padrão e

$$\begin{array}{ll} \lambda(\alpha) = \phi(\alpha)/[1 - \mathbf{\Phi}(\alpha)] & \text{se o truncamento for } x > a, \\ \lambda(\alpha) = -\phi(\alpha)/\mathbf{\Phi}(\alpha) & \text{se o truncamento for } x < a, \end{array}$$

e

$$\delta(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha]$$

Razão Inversa de Mills

Um resultado importante é

$$0 < \delta(\alpha) < 1 \text{ para todos os valores de } \alpha,$$

Um resultado que usaremos em vários pontos abaixo é $d\phi(\alpha)/d\alpha = -\alpha\phi(\alpha)$. A função $\lambda(\alpha)$ é chamada de razão inversa de Mills.

O Modelo de Regressão Truncada

No modelo dos exemplos anteriores, assumimos agora que

$$\mu_i = \mathbf{x}_i' \boldsymbol{\beta}$$

é a parte determinística do modelo de regressão clássico. Então

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i$$

onde

$$\varepsilon_i \mid \mathbf{x}_i \sim N\left[0, \sigma^2\right]$$

para que

$$y_i \mid \mathbf{x}_i \sim N\left[\mathbf{x}_i' \boldsymbol{\beta}, \sigma^2\right]$$

Estamos interessados na distribuição de y_i dado que y_i é maior que o ponto de truncamento a.

Modelo de Regressão Truncada II

A média condicional (aos x e à truncagem) é dada por

$$E\left[y_i \mid y_i > a\right] = \mathbf{x}_i'\boldsymbol{\beta} + \sigma \frac{\phi\left[\left(a - \mathbf{x}_i'\boldsymbol{\beta}\right)/\sigma\right]}{1 - \Phi\left[\left(a - \mathbf{x}_i'\boldsymbol{\beta}\right)/\sigma\right]}.$$

A média condicional é, portanto, uma função não linear de $a, \sigma, \mathbf{x}$ e $\boldsymbol{\beta}$. Os efeitos marginais deste modelo na subpopulação podem ser obtidos escrevendo

$$E\left[y_i \mid y_i > a\right] = \mathbf{x}_i'\boldsymbol{\beta} + \sigma\lambda\left(\alpha_i\right)$$

onde agora $\alpha_i = \left(a - \mathbf{x}_i'\boldsymbol{\beta}\right)/\sigma$. Por conveniência, seja $\lambda_i = \lambda\left(\alpha_i\right)$ e $\delta_i = \delta\left(\alpha_i\right)$. Então

$$\begin{aligned}
\frac{\partial E\left[y_i \mid y_i > a\right]}{\partial \mathbf{x}_i} &= \boldsymbol{\beta} + \sigma\left(d\lambda_i/d\alpha_i\right)\frac{\partial \alpha_i}{\partial \mathbf{x}_i} \\
&= \boldsymbol{\beta} + \sigma\left(\lambda_i^2 - \alpha_i\lambda_i\right)\left(-\boldsymbol{\beta}/\sigma\right) \qquad = \boldsymbol{\beta}\left(1 - \lambda_i^2 + \alpha_i\lambda_i\right) \\
&= \boldsymbol{\beta}\left(1 - \delta_i\right)
\end{aligned}$$

Consequências

- Para cada elemento de $\mathbf{x}_i$, o efeito marginal é menor que o coeficiente correspondente, e uma atenuação semelhante da variância.
- Se o efeito marginal do slide anterior ou o próprio coeficiente β é de interesse depende das inferências pretendidas do estudo.
- A primeira inclinação pode ser usar mínimos quadrados ordinários para estimar os parâmetros desse modelo de regressão. Para a subpopulação da qual os dados são extraídos, poderíamos escrever o modelo na forma

$$y_i \mid y_i > a = E\left[y_i \mid y_i > a\right] + u_i = \mathbf{x}_i'\beta + \sigma\lambda_i + u_i,$$

onde u_i é y_i menos sua expectativa condicional. Se estimarmos isso por regressão de mínimos quadrados ordinários de $\mathbf{y}$ em $\mathbf{X}$, então omitimos uma variável, o termo não linear λ_i.

Censura

Um problema muito comum em dados microeconômicos é a censura da variável dependente. Quando a variável dependente é censurada, os valores em um determinado intervalo são todos transformados em (ou relatados como) um único valor. Alguns exemplos que apareceram na literatura empírica são os seguintes:

1. Compras domésticas de bens duráveis [Tobin (1958)],
2. O número de casos extraconjugais [Fair (1977,1978)],
3. O número de horas trabalhadas por uma mulher na força de trabalho [Quester e Greene (1982)]
4. O número de prisões após a libertação da prisão [Witte (1980)],

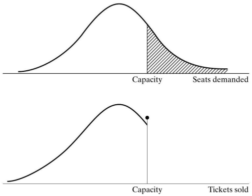
Distribuição Normal Censurada

- A teoria de distribuição relevante para uma variável censurada é semelhante àquela para uma truncada.
- Mais uma vez, começamos com a distribuição normal, pois muito do trabalho recebido foi baseado em uma suposição de normalidade. Também assumimos que o ponto de censura é zero, embora isso seja apenas uma normalização conveniente.
- Em uma distribuição truncada, apenas a parte da distribuição acima de $y = 0$ é relevante para nossos cálculos.
- Para fazer a distribuição integrar a um, nós a escalamos pela probabilidade de que uma observação na população não truncada caia no intervalo que nos interessa.
- Quando os dados são censurados, a distribuição que se aplica aos dados de amostra é uma mistura de distribuições discretas e contínuas.
- Para analisar esta distribuição, definimos uma nova variável aleatória y transformada da original, y^*, por

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

Normal Censurada II

- A distribuição que se aplica se $y^* \sim N\left[\mu, \sigma^2\right]$ é $\text{Prob}(y=0) = \text{Prob}\left(y^* \leq 0\right) = \Phi(-\mu/\sigma) = 1 - \Phi(\mu/\sigma)$, e se $y^* > 0$, então y tem a densidade de y^*.
- Esta distribuição é uma mistura de partes discretas e contínuas. A probabilidade total é um, conforme necessário, mas em vez de redimensionar a segunda parte, simplesmente atribuímos a probabilidade total na região censurada ao ponto de censura, neste caso, zero.



Momentos da Normal Censurada

Theorem (Momentos da Normal Censurada)

Se $y^* \sim N\left[\mu, \sigma^2\right]$ *e* $y = a$ *se* $y^* \leq a$ *ou então* $y = y^*$, *então*

$$E[y] = \Phi a + (1 - \Phi)(\mu + \sigma\lambda)$$

e

$$\text{Var}[y] = \sigma^2(1 - \Phi)\left[(1 - \delta) + (\alpha - \lambda)^2\Phi\right]$$

onde

$$\Phi[(a - \mu)/\sigma] = \Phi(\alpha) = \text{Prob}\,(y^* \leq a) = \Phi, \quad \lambda = \ phi/(1 - \Phi)$$

e

$$\delta = \lambda^2 - \lambda\alpha$$

Modelo TOBIT

A formulação geral é geralmente dada em termos de uma função de índice,

$$\begin{aligned} y_i^* &= \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \\ y_i &= 0 \quad \text{if } y_i^* \leq 0, \\ y_i &= y_i^* \quad \text{if } y_i^* > 0. \end{aligned}$$

para uma observação extraída aleatoriamente da população, que pode ou não ser censurada,

$$E\left[y_i \mid \mathbf{x}_i\right] = \Phi\left(\frac{\mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right)\left(\mathbf{x}_i'\boldsymbol{\beta} + \sigma\lambda_i\right),$$

onde

$$\lambda_i = \frac{\phi\left[\left(0 - \mathbf{x}_i'\boldsymbol{\beta}\right)/\sigma\right]}{1 - \Phi\left[\left(0 - \mathbf{x}_i'\boldsymbol{\beta}\right)/\sigma\right]} = \frac{\phi\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right)}{\Phi\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right)}.$$

TOBIT II

Theorem (Efeitos Marginais no Modelo de Regressão Censurada)

No modelo de regressão censurada com regressão latente $y^* = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$ *e variável dependente observada,* $y = a$ *se* $y^* \leq a, y = b$ *se* $y^* \geq b$*, e* $y = y^*$ *caso contrário, onde a e b constantes, seja* $f(\varepsilon)$ *e* $F(\varepsilon)$ *denotam a densidade e cdf de* ε*. Assuma que* ε *é uma variável aleatória contínua com média 0 e variância* σ^2*, e* $f(\varepsilon \mid \mathbf{x}) = f(\varepsilon)$*. Então*

$$\frac{\partial E[y \mid \mathbf{x}]}{\partial \mathbf{x}} = \boldsymbol{\beta} \times \text{Prob}\left[a < y^* < b\right]$$

Efeitos Marginais III

McDonald e Mofitt (1980) sugeriram uma decomposição útil de $\partial E\left[y_i \mid \mathbf{x}_i\right]/\partial \mathbf{x}_i$,

$$\frac{\partial E\left[y_i \mid \mathbf{x}_i\right]}{\partial \mathbf{x}_i}=\boldsymbol{\beta}\times\left\{\Phi_i\left[1-\lambda_i\left(\alpha_i+\lambda_i\right)\right]+\phi_i\left(\alpha_i+\lambda_i\right)\right\}$$

onde $\alpha_i=\mathbf{x}_i'\boldsymbol{\beta}$, $\Phi_i=\Phi\left(\alpha_i\right)$ e $\lambda_i=\phi_i/\Phi_i$. Tomando as duas partes separadamente, este resultado decompõe o vetor inclinação em

$$\frac{\partial E\left[y_i \mid \mathbf{x}_i\right]}{\partial \mathbf{x}_i}=\operatorname{Prob}\left[y_i>0\right]\frac{\partial E\left[y_i \mid \mathbf{x}_i, y_i>0\right]}{\partial \mathbf{x}_i}+E\left[y_i \mid \mathbf{x}_i, y_i>0\right]\frac{\partial \operatorname{Prob}\left[y_i>0\right]}{\partial \mathbf{x}_i}.$$

Assim, uma mudança em $\mathbf{x}_i$ tem dois efeitos: afeta a média condicional de y_i^* na parte positiva da distribuição e afeta a probabilidade de que a observação cairá nessa parte da distribuição.

TOBIT - Estimação

A estimação é por Máxima Verossimilhança A função log-verossimilhança é dada por:

$$\ln L = \sum_{y_i>0} -\frac{1}{2}\left[\log(2\pi) + \ln\sigma^2 + \frac{(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2}{\sigma^2}\right] + \sum_{y_i=0} \ln\left[1 - \Phi\left(\frac{\mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right)\right]$$

As duas partes correspondem à regressão clássica para as observações não-limite e as probabilidades relevantes para as observações limite, respectivamente. Com $\boldsymbol{\gamma} = \boldsymbol{\beta}/\sigma$ and $\theta = 1/\sigma$, a log-verossimilhança é

$$\ln L = \sum_{y_i>0} -\frac{1}{2}\left[\ln(2\pi) - \ln\theta^2 + (\theta y_i - \mathbf{x}_i'\gamma)^2\right] + \sum_{y_i=0} \ln\left[1 - \Phi\left(\mathbf{x}_i'\gamma\right)\right].$$

Seleção de Amostra - Truncagem Incidental

- Suponha que y e z tenham uma distribuição bivariada com correlação ρ.
- Estamos interessados na distribuição de y dado que z excede um determinado valor.
- A intuição sugere que se y e z estiverem positivamente correlacionados, então o truncamento de z deve empurrar a distribuição de y para a direita.
- Como antes, estamos interessados em (1) a forma da distribuição acidentalmente truncada e (2) a média e a variância da variável aleatória acidentalmente truncada.
- Vamos nos concentrar primeiro na distribuição normal bivariada.

A densidade conjunta truncada de y e z é

$$f(y, z \mid z > a) = \frac{f(y, z)}{\text{Prob}(z > a)}.$$

Para obter a densidade marginal truncada incidentalmente para y, integraríamos z para fora dessa expressão.

Momentos - Truncagem Incidental

Theorem (Momentos da Distribuição Normal Bivariada Incidentalmente Truncada)

Se y e z têm uma distribuição normal bivariada com médias μ_y e μ_z, desvios padrão σ_y e σ_z, e correlação ρ, então

$$E[y \mid z > a] = \mu_y + \rho\sigma_y\lambda(\alpha_z),$$
$$\text{Var}[y \mid z > a] = \sigma_y^2\left[1 - \rho^2\delta(\alpha_z)\right],$$

onde

$$\alpha_z = (a - \mu_z)/\sigma_z, \lambda(\alpha_z) = \phi(\alpha_z)/[1 - \Phi(\alpha_z)], \text{ e } \delta(\alpha_z) = \lambda(\alpha_z)[\lambda(\alpha_z) - \alpha_z]$$

Modelo de Seleção

Para colocar os exemplos anteriores em uma estrutura geral, deixe a equação que determina a seleção da amostra ser

$$z_i^* = \mathbf{w}'\boldsymbol{\gamma}_i + u_i$$

e deixe a equação de interesse ser

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$$

A regra é que y_i é observado apenas quando z_i^* é maior que zero. Suponha também que ε_i e u_i tenham uma distribuição normal bivariada com média zero e correlação ρ. Então podemos inseri-los no Teorema 22.5 para obter o modelo que se aplica às observações em nossa amostra:

$$\begin{aligned}
E\left[y_i \mid y_i \text{ is observed }\right] &= E\left[y_i \mid z_i^* > 0\right] \\
&= E\left[y_i \mid u_i > -\mathbf{w}'\boldsymbol{\gamma}_i\right] \\
&= \mathbf{x}_i'\boldsymbol{\beta} + E\left[\varepsilon_i \mid u_i > -\mathbf{w}'\boldsymbol{\gamma}_i\right] \\
&= \mathbf{x}_i'\boldsymbol{\beta} + \rho\sigma_\varepsilon\lambda_i\left(\alpha_u\right) \\
&= \mathbf{x}_i'\boldsymbol{\beta}_i + \beta_\lambda\lambda_i\left(\alpha_u\right),
\end{aligned}$$

Modelo de Seleção II

where $\alpha_u = -\mathbf{w}_i'\gamma/\sigma_u$ and $\lambda(\alpha_u) = \phi(\mathbf{w}_i'\gamma/\sigma_u)/\Phi(\mathbf{w}_i'\gamma/\sigma_u)$. So

$$\begin{aligned} y_i \mid z_i^* > 0 &= E[y_i \mid z_i^* > 0] + v_i \\ &= \mathbf{x}_i'\boldsymbol{\beta} + \beta_\lambda \lambda_i(\alpha_u) + v_i. \end{aligned}$$

- A regressão de OLS usando os dados para os quais temos informação produz estimativas inconsistentes de $\boldsymbol{\beta}$.
- Mais uma vez, podemos ver o problema como uma variável omitida.
- A regressão de mínimos quadrados de y em $\mathbf{x}$ e λ seria um estimador consistente, mas se λ for omitido, o erro de especificação de uma variável omitida será cometido.
- Finalmente, observe que a segunda parte do Teorema 22.5 implica que mesmo se λ_i fosse observado, os mínimos quadrados seriam ineficientes. A perturbação v_i é heterocedástica.

Modelo de Seleção III

Mecanismo de Seleção: $z_i^* = \mathbf{w}_i'\gamma + u_i, z_i = 1$ if $z_i^* > 0$ e 0 caso contrário;
Prob $(z_i = 1 \mid \mathbf{w}_i) = \mathbf{\Phi}(\mathbf{w}_i'\gamma)$ e
Prob $(z_i = 0 \mid \mathbf{w}_i) = 1 - \mathbf{\Phi}(\mathbf{w}_i'\gamma)$.
Modelo de Regressão: $\quad y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$ observado somente se $z_i = 1$,

$$(u_i, \varepsilon_i) \sim \text{ normal bivariada } [0, 0, 1, \sigma_\varepsilon, \rho].$$

Suponha que, como em muitos desses estudos, z_i e $\mathbf{w}_i$ sejam observados para uma amostra aleatória de indivíduos, mas y_i seja observado apenas quando $z_i = 1$. Este modelo é precisamente o que examinamos anteriormente, com

$$E[y_i \mid z_i = 1, \mathbf{x}_i, \mathbf{w}_i] = \mathbf{x}_i'\boldsymbol{\beta} + \rho\sigma_e\lambda(\mathbf{w}_i'\gamma)$$

Efeitos Marginais

- O efeito marginal dos regressores em y_i na amostra observada consiste em dois componentes.
- Existe o efeito direto na média de y_i, que é $\boldsymbol{\beta}$.
- Além disso, para uma determinada variável independente, se ela aparecer na probabilidade de que z_i^* seja positiva, ela influenciará y_i por meio de sua presença em λ_i .
- O efeito total das mudanças em um regressor que aparece em $\mathbf{x}_i$ e $\mathbf{w}_i$ em y é

$$\frac{\partial E\left[y_i \mid z_i^* > 0\right]}{\partial x_{ik}} = \beta_k - \gamma_k \left(\frac{\rho\sigma_\varepsilon}{\sigma_u}\right) \delta_i\left(\alpha_u\right),$$

onde

$$\delta_i = \lambda_i^2 - \alpha_i \lambda_i$$

Modelo de Seleção - Estimação em dois estágios

Os parâmetros do modelo de seleção amostral podem ser estimados por máxima verossimilhança. No entanto, o procedimento de estimativa em duas etapas de Heckman (1979) geralmente é usado. O método de Heckman é o seguinte.

1. Estime a equação probit por máxima verossimilhança para obter estimativas de γ. Para cada observação na amostra selecionada, calcule $\hat{\lambda}_i = \phi\left(\mathbf{w}_i'\hat{\gamma}\right)/\Phi\left(\mathbf{w}_i'\hat{\gamma}\right)$ e $\hat{\delta}_i = \hat{\lambda}_i\left(\hat{\lambda}_i - \mathbf{w}_i'\hat{\gamma}\right)$
2. Estimar $\boldsymbol{\beta}$ e $\beta_\lambda = \rho\sigma_e$ por regressão de mínimos quadrados de y em $\mathbf{x}$ e $\hat{\lambda}$.

É possível também construir estimadores consistentes dos parâmetros individuais ρ e σ_ε.