Máxima Verossimilhança - Parte II

Bibliografia: Greene

Claudio Lucinda

FEA/USP

Overview

ML Condicional

- Até o momento, fomos bastante soltos com o vetor $\mathbf{y}$, sem falar sobre o fato que na maior parte das vezes nós temos regressores neste modelo, e seria melhor que particionássemos este vetor em variáveis independentes e dependentes do modelo.
- Vamos representar $\mathbf{x}_i$ como sendo um conjunto de variáveis aleatórias e constantes que entram na distribuição condicional de y_i.
- Particionando a distribuição conjunta de y_i e $\mathbf{x}_i$ no produto da distribuição condicional e a marginal, podemos escrever a log-verossimilhança como sendo

$$\ln L(\alpha \mid \text{ data }) = \sum_{i=1}^{n} \ln f(y_i, \mathbf{x}_i \mid \alpha) = \sum_{i=1}^{n} \ln f(y_i \mid \mathbf{x}_i, \alpha) + \sum_{i=1}^{n} \ln g(\mathbf{x}_i \mid \alpha),$$

ML Condicional

- Vamos assumir que o processo gerador de $\mathbf{x}_i$ acontece externamente ao modelo de interesse.
  - Nem sempre isso é verdade - se tivéssemos endogeneidade, esse modelo teria que ser estudado também.
- Isso implica que os parâmetros do modelo $g\left(\mathbf{x}_i \mid \boldsymbol{\alpha}\right)$ não são os mesmos que aparecem em $f\left(y_i \mid \mathbf{x}_i, \boldsymbol{\alpha}\right)$.
- Portanto, podemos particionar $\boldsymbol{\alpha}$ em $[\boldsymbol{\theta}, \delta]$ e a Log-Verossimilhança fica sendo

$$\ln L(\boldsymbol{\theta}, \boldsymbol{\delta} \mid \mathbf{data}) = \sum_{i=1}^{n} \ln f\left(y_i, \mathbf{x}_i \mid \boldsymbol{\alpha}\right) = \sum_{i=1}^{n} \ln f\left(y_i \mid \mathbf{x}_i, \boldsymbol{\theta}\right) + \sum_{i=1}^{n} \ln g\left(\mathbf{x}_i \mid \boldsymbol{\delta}\right).$$

ML Condicional

Desde que $\boldsymbol{\theta}$ e δ não possuam elementos em comum nem restrições que os conectem, (tais como $\theta + \delta = 1$), podemos analisar as duas partes da Log-Verossimilhança separadamente e, a distribuição marginal de $\mathbf{x}_i$ não vai ser de muito interesse.
Os resultados assintóticos aqui precisam levar em conta a presença de $\mathbf{x}_i$ nas funções e derivadas de $\ln f\left(y_i \mid \mathbf{x}_i, \boldsymbol{\theta}\right)$.
Vamos supor dados razoavelmente comportados, tal que médias amostrais tais como

$$(1/n) \ln L(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{X}) = \frac{1}{n} \sum_{i=1}^{n} \ln f\left(y_i \mid \mathbf{x}_i, \boldsymbol{\theta}\right)$$

E o gradiente da função com respeito a θ irão convergir em probabilidade para suas esperanças populacionais.

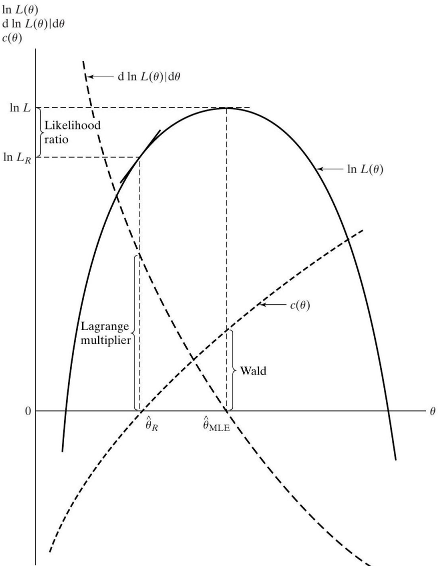
Condições

- Espaço de parâmetros. Espaço de parâmetros podem ter "vãos" e não convexidades que bagunçam os procedimentos. A função verossimilhança precisa ser uma função contínua de um espaço convexo de parâmetros (ainda que possa ser não limitado, como em $\sigma > 0$ no modelo de regressão.
- Identificação (numérica). Ou seja, a estimação precisa ser factível.
- Dados bem comportados. As leis dos grandes números tem que se aplicar às médias amostrais dos dados e alguma forma de Teorema Central do Limite precisa ser aplicável ao gradiente.

Os Três Testes Canônicos - no contexto de ML

- Teste de razão de verossimilhança. Se a restrição $c(\theta) = 0$ for válida, sua imposição não deve levar a uma grande redução na função de log-verossimilhança. Portanto, baseamos o teste na diferença, $\ln L_U - \ln L_R$, onde L_U é o valor da função de verossimilhança no valor irrestrito de θ e L_R é o valor da função de verossimilhança na estimativa restrita.
- Teste Wald. Se a restrição for válida, então $c\left(\hat{\theta}_{\text{MLE}}\right)$ deve ser próximo de zero, pois o MLE é consistente. Portanto, o teste é baseado em $c\left(\hat{\theta}_{\text{MLE}}\right)$. Rejeitamos a hipótese se esse valor for significativamente diferente de zero.
- Teste do multiplicador de Lagrange. Se a restrição for válida, o estimador restrito deve estar próximo do ponto que maximiza a verossimilhança logarítmica. Portanto, a inclinação da função de verossimilhança logarítmica deve ser próxima de zero no estimador restrito. O teste é baseado na inclinação da verossimilhança logarítmica no ponto em que a função é maximizada sujeita à restrição.

Os Três Testes Canônicos - no contexto de ML



Teste de Razão de Verossimilhança

O teste de Razão de Verossimilhança é

$$\lambda = \frac{\hat{L}_R}{\hat{L}_U}.$$

Por construção, esta função é limitada entre zero e um. As duas verossimilhanças são positivas, e $\hat{L}_R$ não pode ser maior que $\hat{L}_U$. Se λ é pequeno demais, temos motivos pra rejeitar as restrições.

Sob regularidade e sob H_0, a distribuição em grandes amostras $-2 \ln \lambda$ é qui-quadrada, com graus de liberdade iguais ao número de restrições impostas.

A hipótese nula é rejeitada se esse valor exceder o valor crítico apropriado das tabelas qui-quadrado.

Teste de Wald

- Uma deficiência prática do teste de razão de verossimilhança é que ele geralmente requer a estimativa de vetores de parâmetros restritos e irrestritos.
- Em modelos complexos, uma ou outra dessas estimativas pode ser muito difícil de calcular. Ou quando tanto a hipótese nula ou a alternativa são modelos restritos?
- Um dos testes para lidar com isso é o de Wald, baseado em um estimador que é assintoticamente distribuído normalmente.

$$\text{Se } \mathbf{x} \sim N_J[\boldsymbol{\mu}, \boldsymbol{\Sigma}], \text{ então } (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \sim \text{qui} - \text{quadrado } [J]$$

- No cenário de um teste de hipótese, sob a hipótese de que $E(\mathbf{x}) = \boldsymbol{\mu}$, a forma quadrática tem distribuição qui-quadrada.
- No entanto, se a hipótese de que $E(\mathbf{x}) = \boldsymbol{\mu}$ for falsa, então a forma quadrática que acabamos de fornecer será, em média, maior do que seria se a hipótese fosse verdadeira.

Teste de Wald para hipóteses

Seja $\hat{\boldsymbol{\theta}}$ o vetor de estimativas de parâmetros obtidas sem restrições. Nós hipotetizamos um conjunto de restrições

$$H_0 : \mathbf{c}(\boldsymbol{\theta}) = \mathbf{q}$$

A estatística de teste é

$$W = [\mathbf{c}(\hat{\boldsymbol{\theta}}) - \mathbf{q}]'(\text{ Asy. Var}[\mathbf{c}(\hat{\boldsymbol{\theta}}) - \mathbf{q}])^{-1}[\mathbf{c}(\hat{\boldsymbol{\theta}}) - \mathbf{q}]$$

Sob H_0, em grandes amostras, W tem uma distribuição qui-quadrada com graus de liberdade igual ao número de restrições [ou seja, o número de equações em $\mathbf{c}(\hat{\boldsymbol{\theta}}) - \mathbf{q} = \mathbf{0}$.

Estimando a Variância das Restrições

$$\text{Est. Asy. } \mathrm{Var}[\mathbf{c}(\hat{\boldsymbol{\theta}}) - \mathbf{q}] = \hat{\mathbf{C}} \text{ Est. Asy. } \mathrm{Var}[\hat{\boldsymbol{\theta}}]\hat{\mathbf{C}}',$$

$$\hat{\mathbf{C}} = \left[\frac{\partial \mathbf{c}(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}'}\right].$$

Para restrições lineares nos parâmetros:

$$H_0 : \mathbf{c}(\boldsymbol{\theta}) - \mathbf{q} = \mathbf{R}\boldsymbol{\theta} - \mathbf{q} = \mathbf{0},$$

$$\hat{\mathbf{C}} = \left[\frac{\partial \mathbf{c}(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}'}\right] = \mathbf{R}',$$

$$\text{Est. Asy. } \mathrm{Var}[\mathbf{c}(\hat{\boldsymbol{\theta}}) - \mathbf{q}] = \mathbf{R} \text{ Est. Asy. } \mathrm{Var}[\hat{\boldsymbol{\theta}}]\mathbf{R},$$

$$W = [\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{q}]' \left[\mathbf{R} \text{ Est. Asy. } \mathrm{Var}(\hat{\boldsymbol{\theta}})\mathbf{R}'\right]^{-1} [\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{q}]$$

Teste do Multiplicador de Lagrange

O terceiro procedimento de teste é o multiplicador de Lagrange (LM) ou teste de score eficiente (ou apenas score), a partir do modelo restrito ao invés do modelo irrestrito. Suponha que maximizamos a log-verossimilhança sujeita ao conjunto de restrições $\mathbf{c}(\boldsymbol{\theta}) - \mathbf{q} = \mathbf{0}$. Seja $\boldsymbol{\lambda}$ um vetor de multiplicadores de Lagrange e defina a função Lagrangeana:

$$\ln L^*(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}) + \lambda'(\mathbf{c}(\boldsymbol{\theta}) - \mathbf{q})$$

A solução do problema passa pelas CPO's:

$$\frac{\partial \ln L^*}{\partial \boldsymbol{\theta}} = \frac{\partial \ln L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \mathbf{C}'\lambda = \mathbf{0},$$
$$\frac{\partial \ln L^*}{\partial \boldsymbol{\lambda}} = \mathbf{c}(\boldsymbol{\theta}) - \mathbf{q} = \mathbf{0},$$

Teste do Multiplicador de Lagrange II:

No máximo restrito, as derivadas da função de log-verossimilhança são

$$\frac{\partial \ln L\left(\hat{\boldsymbol{\theta}}_R\right)}{\partial \hat{\boldsymbol{\theta}}_R} = -\hat{\mathbf{C}}'\hat{\boldsymbol{\lambda}} = \hat{\mathbf{g}}_R.$$

A estatística de teste do multiplicador de Lagrange é

$$\mathbf{LM} = \left(\frac{\partial \ln L\left(\hat{\boldsymbol{\theta}}_R\right)}{\partial \hat{\boldsymbol{\theta}}_R}\right)' \left[\mathbf{I}\left(\hat{\boldsymbol{\theta}}_R\right)\right]^{-1} \left(\frac{\partial \ln L\left(\hat{\boldsymbol{\theta}}_R\right)}{\partial \hat{\boldsymbol{\theta}}_R}\right).$$

Sob a hipótese nula, LM tem uma distribuição qui-quadrada limite com graus de liberdade iguais ao número de restrições.

Teste LM III:

Seja $\hat{\mathbf{g}}_{iR}$ denotando o i o termo no gradiente da função de log-verossimilhança. Então,

$$\hat{\mathbf{g}}_R = \sum_{i=1}^{n} \hat{\mathbf{g}}_{iR} = \hat{\mathbf{G}}'_R \mathbf{i},$$

Em que $\hat{\mathbf{G}}_R$ é a matriz $n \times K$ com i a décima linha igual a $\mathbf{g}'_{iR}$ e $\mathbf{i}$ é uma coluna de $1\,\mathrm{s}$. Se usarmos o estimador BHHH (produto externo de gradientes) em (17-18) para estimar o Hessiano, então

$$[\hat{\mathbf{I}}(\hat{\theta})]^{-1} = \left[\hat{\mathbf{G}}'_R \hat{\mathbf{G}}_R\right]^{-1}$$

e

$$\mathrm{LM} = \mathbf{i}'\hat{\mathbf{G}}_R \left[\hat{\mathbf{G}}'_R \hat{\mathbf{G}}_R\right]^{-1} \hat{\mathbf{G}}'_R \mathbf{i}$$