Aula 09

Simulation-Assisted Estimation

Claudio R. Lucinda

FEA/USP

Agenda

1. Monte Carlo Integration

Agenda

1. Monte Carlo Integration
2. Modelo de Escolha Discreta com Características Não-Observáveis
   - Estimando as Participações de Mercado
   - Calculando a Utilidade Média
   - Construindo as Condições de Momento

Monte Carlo Integration

- As integrais geralmente aparecem em estimadores econométricos em "forma aberta", ou seja, em uma forma para a qual não existe uma primitiva.
- Existem vários métodos numéricos disponíveis para aproximar integrais de forma aberta - as quadraturas de Gauss-Hermite e Gauss-Laguerre são duas.
- A técnica de integração de Monte Carlo pode frequentemente ser usada quando a integral está na forma

$$h(y) = \int_{w} g(y \mid w) f(w) dw = E_w[g(y \mid w)]$$

- onde $f(w)$ é a densidade de w ee w é uma variável aleatória que pode ser simulada.

Draws de Halton e Integração

- Se $w_1, w_2, \ldots, w_n$ são uma amostra aleatória de observações na variável aleatória w e $g(w)$ é uma função de w com média finita e variância, então pela lei dos grandes números [Teorema D.4 e o corolário em (D-5)],

$$\text{plim}\, \frac{1}{n} \sum_{i=1}^{n} g\left(w_i\right) = E[g(w)]$$

MC Integration

A integração de Monte Carlo é usada para avaliar a esperança

$$E[g(x)] = \int_x g(x) f(x) dx$$

onde $f(x)$ é a densidade da variável aleatória x e $g(x)$ é uma função suave. A aproximação de Monte Carlo é

$$E[\widehat{g(x)}] = \frac{1}{R} \sum_{r=1}^{R} g\left(x_r\right).$$

MC Integration II

- A convergência da aproximação para a esperança é baseada na lei dos grandes números - uma amostra aleatória de draws em $g(x)$ irá convergir em probabilidade para sua esperança.
- A abordagem padrão para integração baseada em simulação é usar sorteios aleatórios da distribuição especificada. a partir de um gerador de números aleatórios para produzir os sorteios de uma distribuição especificada.
- O componente central desta abordagem é extraído da distribuição uniforme contínua padrão, $U[0,1]$. Os sorteios de outras distribuições são obtidos a partir desses sorteios usando transformações. Em particular, para um sorteio da distribuição normal, onde u_i é um sorteio de $U[0,1]$, $v_i = \Phi^{-1}(u_i)$.

MC Integration III

- Dado que os sorteios iniciais satisfazem os pressupostos necessários, a questão central para fins de especificação da simulação é o número de sorteios.
- Um bom desempenho nesta conexão requer um número muito grande de empates.
- Os resultados diferem no número necessário em uma determinada aplicação, mas a conclusão geral é que, quando a simulação é feita dessa maneira, o número é grande (centenas ou milhares).
- Numerosos métodos foram concebidos para reduzir o número de extrações necessárias para obter uma aproximação satisfatória.
- Um desses métodos é introduzir alguma autocorrelação nos desenhos - uma pequena quantidade de correlação negativa entre os desenhos reduzirá a variância da simulação.
- Sorteios antitéticos, em que cada sorteio em uma sequência é incluído com sua imagem espelhada (w_i e $-w_i$ para empates normalmente distribuídos, w_i e $1 - w_i$ para uniforme, por exemplo) é um desses métodos.

Halton Draws

- Os procedimentos foram concebidos na literatura de análise numérica para obter sorteios "inteligentes" da distribuição uniforme, em vez de sorteios aleatórios.
- Uma literatura emergente documentou ganhos dramáticos de velocidade sem degradação no desempenho da simulação através do uso de um número menor de sorteios de Halton ou outros construídos, não aleatórios sequências em vez de um grande número de sorteios aleatórios.
- Esses procedimentos parecem reduzir muito o número de extrações necessárias para a estimativa (às vezes por um fator de 90 por cento ou mais) e reduzir o erro de simulação associado a um determinado número de draws.

Halton Draws II

Uma sequência de sorteios de Halton é gerada da seguinte forma: Seja r um número primo. Expanda a sequência de inteiros $g = 1, 2, \ldots$ em termos da base r como

$$g = \sum_{i=0}^{I} b_i r^i \text{ onde, por construção, } 0 \leq b_i \leq r - 1 \text{ e } r^I \leq g < r^{I+1}.$$

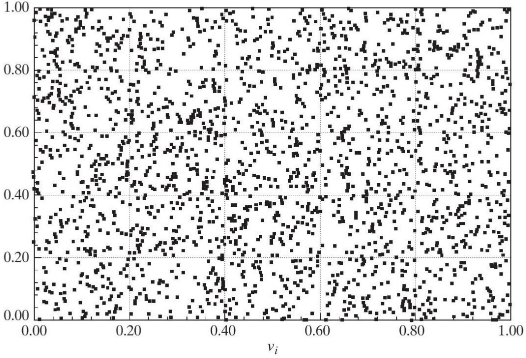
A sequência Halton de valores que corresponde a esta série é

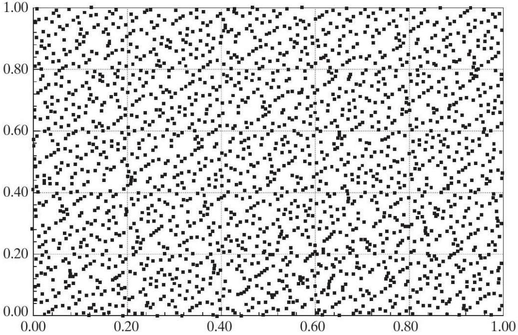
$$H(g) = \sum_{i=0}^{I} b_i r^{-i-1}$$

Por exemplo, usando a base 5 , o inteiro 37 tem $b_0 = 2, b_1 = 2$ e $b_3 = 1$. Então

$$H_5(37) = 2 \times 5^{-1} + 2 \times 5^{-2} + 1 \times 5^{-3} = 0,488$$

Comparação MC e Halton Draws





Escolha Discreta e Características Não Observáveis

- Vamos aqui discutir uma outra forma de driblar o problema decorrente da IIA.
- Neste caso, a sensibilidade da utilidade à atributos das alternativas depende de características individuais.
- A vantagem é que prescinde de imposições do analista acerca do padrão de substituição entre as alternativas.
- No Logit Aninhado, quem tem que escrever a árvore de escolhas é o analista.

Escolha Discreta...(II):

- Essa abordagem é apresentada no artigo clássico de BLP (1995)
- Uma vez que é baseada em uma microfundamentação bem robusta, permite que estendamos esta metodologia para um conjunto bem amplo de situações. Vamos então começar introduzindo o modelo microeconômico subjacente:

$$U_{ij} = \sum_k x_{jk}\beta_{ik} + \xi_j + \epsilon_{ij}$$

- O termo ξ_j captura as avaliações que os consumidores fazem das características não observaas da alternativa j.
- A diferença em relação ao que vimos na aula anterior é que voltamos a supor que a sensibilidade do consumidor i em relação ao atributo k do produto j pode ser diferente da mesma sensibilidade no caso de outro consumidor, ou seja $\beta_{ik} \neq \beta_{dk}, \forall i, d$.

Escolha Discreta...(III):

- A idéia aqui é modelar a heterogeneidade dos coeficientes dos indivíduos da seguinte forma:

$$\beta_{ik} = \bar{\beta}_k + \beta_k^o \mathbf{z}_i + \beta_k^u \mathbf{v}_i$$

- Ou seja, este coeficiente é composto por uma sensibilidade "média", $\bar{\beta}_k$, mais alguns efeitos:
    - As características individuais observadas possui sobre ele , que denotaremos $\mathbf{z}_i$
    - As características individuais **não observadas**, denominadas $\mathbf{v}_i$
- Substituindo esta definição dos coeficientes, temos que:

$$U_{ij} = \sum_k x_{jk}\bar{\beta}_k + \sum_k x_{jk}\beta_k^o \mathbf{z}_i + \sum_k x_{jk}\beta_k^u \mathbf{v}_i + \xi_j + \epsilon_{ij}$$

Escolha Discreta...(IV):

- Podemos reescrever de forma similar à da aula passada, separando esta especificação em uma parte correspondente á utilidade "média" da alternativa, e a parte aleatória:

$$
\begin{aligned}
U_{ij} &= V_j + \sum_k x_{jk}\beta_k^o \mathbf{z}_i + \sum_k x_{jk}\beta_k^u \mathbf{v}_i + \epsilon_{ij} \\
V_j &= \sum_k x_{jk}\bar{\beta}_k + \xi_j
\end{aligned}
$$

- Na verdade, não temos microdados, somente as participações de mercado e algumas características dos produtos.
- Como fazer neste caso?

Escolha Discreta...(V):

- Temos que a probabilidade de escolha de uma alternativa é, na verdade, a seguinte integral:

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

- O problema não é a integral de dentro – supondo uma distribuição valores extermos para ε, temos uma forma analítica – mas sim a integral de fora.
- Na maior parte dos casos, não teremos dados observados sobre as características dos consumidores que compraram cada uma das alternativas, de forma que apenas trabalharemos com características não observadas dos consumidores – ou seja, apenas os termos $\mathbf{v}_i$. A análise subseqüente será em três etapas.
  1. Especificar as participações de mercado em função dos coeficientes;
  2. Recuperar os sinais dos termos ξ_j a partir dos resultados da etapa anterior;
  3. Estimar os coeficientes por GMM

Estimando as Participações de Mercado

- A primeira parte é recuperar as participações de mercado em função da utilidade média, V_j, e dos coeficientes β_k^u. Uma vez que os termos $\mathbf{v}_i$ não são observados, iremos impor uma premissa sobre eles, que é a que eles seguem uma distribuição qualquer, enquanto que os termos ϵ_{ij} seguem a já tradicional distribuição de Valores Extremos I. Desta forma, temos que:

$$s_j(V_j, \beta) = \int \left(\frac{\exp[V_j + \sum_k x_{jk}\beta_k^u \mathbf{v}_i]}{1 + \sum_{q>0} \exp[V_q + \sum_k x_{qk}\beta_k^u \mathbf{v}_i]} \right) f(\mathbf{v}) d\mathbf{v}$$

- Para resolvermos esta integral, precisamos de uma forma funcional para $f(\mathbf{v})$.
- Geralmente, utiliza-se dados obtidos a partir de microdados – por exemplo, se a renda dos consumidores não é observada, usa-se distribuição da renda buscada na PNAD.

Calculando Integrais por Simulação

- Não existe uma forma analítica de resolução desta integral, de forma que teremos que usar ou métodos numéricos (por exemplo, método da quadratura) ou métodos de simulação para a obtenção do valor desta integral – e, consequentemente, o valor desta participação de mercado.
- Vamos falar um pouco sobre os métodos de simulação.
  - Em geral simulação consiste em sortear valores aleatórios de uma distribuição, calcular alguma coisa com cada um destas valores sorteados e depois tirar uma média destes cálculos. Em todos estes casos, o pesquisador quer calcular uma média da forma $\tilde{t} = \int t(\varepsilon)f(\varepsilon)d\varepsilon$.
- Vamos supor que você tenha sorteado *ns* valores desta distribuição $f(\mathbf{v})$. Vamos supor que $\mathbf{z} = \varnothing$.

Calculando Integrais por Simulação

- Com estes valores, podemos calcular a participação de mercado da alternativa j com a seguinte fórmula:

$$\hat{s}^{ns}(\delta_j, \beta) = \frac{1}{ns}\sum_{r=1}^{ns}\left[\frac{\exp[V_j + \sum_k x_{jk}\beta_k^u \mathbf{v}_{ir}]}{1 + \sum_{q>0}\exp[V_q + \sum_k x_{qk}\beta_k^u \mathbf{v}_{ir}]}\right]$$

- Ou seja, pegamos os valores de $\mathbf{v}_i$ associados com os sorteios de cada característica, e calculamos a fórmula do *share* com cada um deles. Depois disso, somamos e tiramos a média.
- Evidentemente, a utilização de métodos de simulação aumenta a imprecisão das estimativas, que pode ser reduzida quanto maior for o valor de ns.

Calculando a Utilidade Média

- O passo seguinte é calcular os efeitos das características dos produtos que não são observadas pelo analista. Para isso, precisamos obter estimativas do termo V_j.
- Tendo este negócio, podemos obter estimativas do ξ_j, do $\bar{\beta}_j$ e do β_k^u.
- Em primeiro lugar, no paper BLP, os autores notam que a seguinte relação:

$$V_j^h = V_j^{h-1} + \ln[s_j] - \ln[\hat{s}_j^{ns}]$$

- É um chamado *contraction mapping*, que possui um ponto fixo. Ou seja, se fizermos um processo sequencial, começando com um valor inicial para o V_j, e em cada iteração ajustando o valor de V_j no valor igual à diferença entre os logs da participação de mercado e a participação de mercado observada, acabaremos em um ponto fixo, em que não há alterações adicionais em V_j.

Calculando a Utilidade Média (II):

- Um bom valor inicial para fazer a recursão pode ser a participação de mercado obtida com a estimação de um modelo LOGIT multinomial.
- Na verdade, na programação é utilizada a seguinte versão não-linear do estimador:

$$\exp[V_j^h] = \exp[V_j^{h-1}] \times \frac{s_j}{\hat{s}_j^{ns}}$$

- Desta forma, obtemos uma estimativa para o termo V_j. A partir desta estimativa, podemos construir as condições de momento, lembrando que o termo ξ_j pode ser entendida como a diferença entre a utilidade média e os valores dos coeficientes médios.

Construindo as Condições de Momento

- Tendo este valor para V_j, podemos definir os erros da seguinte forma:

$$\xi_j = V_j - \sum_k \bar{\beta}_k x_{jk}$$

- O problema é que aqui temos que estimar os $\bar{\beta}_{\mathbf{j}}$. Isto é análogo a concentrar a função objetivo para ficar apenas em termos dos $\beta^{\mathbf{u}}$. Vamos fazer isso projetando a covariância entre os $\mathbf{x_j}$ e as variáveis instrumentais, $\mathbf{Z}$, na covariância entre o V_j e os instrumentos, ou seja:

$$\bar{\beta}_j = (\mathbf{K}(\mathbf{Z}^{\mathbf{T}}\mathbf{Z})\mathbf{K}^{\mathbf{T}})^{-1}\mathbf{K}(\mathbf{Z}^{\mathbf{T}}\mathbf{Z})\mathbf{L}$$

Construindo as Condições de Momento (II)

- Supondo que tenhamos variáveis exógenas e instrumentos para os preços, podemos construir as condições de momento da seguinte forma:

$$m(\theta) = \Xi \mathbf{Z}$$

- Em que Ξ é o empilhamento dos ξ_j. Com isto, podemos construir uma função objetivo da seguinte forma:

$$q = \Xi^T \mathbf{Z} \mathbf{\Phi}^{-1} \mathbf{Z}^T \Xi$$

- Em que Φ representa uma estimativa da matriz variância-covariância dos momentos das equações, ou $\Phi = E(\mathbf{Z}^T \Xi \Xi^T \mathbf{Z})$.

Iterando até Convergir

- Neste ponto, perdemos de vista os coeficientes β_k^u. Para isto, precisamos ter em mente que as etapas anteriores são repetidas todas ao longo de cada iteração. Na verdade, uma vez que, para os valores de β_k^u temos forma analítica para os coeficientes $\bar{\beta}_k$, podemos fazer o seguinte; em cada iteração apenas usar métodos numéricos para obter o mínimo valor da função critério na iteração. Ou seja, em cada iteração temos o seguinte roteiro:
    1. Dados os valores novos dos parâmetros β_k^u, podemos recalcular todo o processo e calcular a função objetivo;
    2. Pelo algoritmo numérico, temos novas sugestões para os valores β_k^u.