Estimação do Impacto de Múltiplos Tratamentos e do Efeito Dosagem: Escore de Propensão Generalizado

Fabio Veras Soares
Centro Internacional de Pobreza (PNUD/IPEA)

Vimos nas aulas anteriores que o “escore de propensão” é utilizado para reduzir o viés de seleção na estimação de relações causais, quando a seleção (para participar do tratamento) é baseada em variáveis observáveis.

O viés seria derivado do fato dos grupos de tratamento e controle diferirem sistematicamente em termos de suas características observáveis.

Todos os procedimentos que vimos anteriormente relacionados ao escore de propensão se baseavam na idéia de uma variável de tratamento binária: tratados versus controles.

O tratamento era então representado por uma variável *dummy*.

Entretanto há casos em que o tratamento não pode ser representado por uma variável *dummy*. Nestes casos o parâmetro de interesse pode ser o impacto:

a) da dosagem do tratamento: montante de transferência, anos de estudo, quantidade de cigarros consumidos por dia, números de anos em que a pessoa fumou cigarros, valor do prêmio da loteria, numero de dias de treinamento profissional, tipos de treinamento profissional.

b) de diferentes tipos de tratamento: transferência de renda e/ou suplemento nutricional; computador e/ou professor adicional; diferentes tipos de grupo de controle (externalidade), diferentes tipos de treinamento profissional;

Imai and Dyk (2003)

Escore de Propensão:

Escore de Balanceamento: $T \perp X \mid P(X)$

- O tratamento é independente das variáveis observáveis, se controlamos pelo escore de propensão (condicional ao escore de propensão)

- Hipótese da Independência Condicional (CIA - unconfoundedness):

$(Y_1, Y_0) \perp T \mid P(X)$

- As técnicas de *matching* (empareamento), ponderação pelo inverso da probabilidade, e de subclassificação podem ser usadas em conjunto com o escore de propensão para diminuir o impacto do viés de seleção.

Vantagens do escore de propensão:

- Reduz o problema da dimensionalidade;
- Melhor do que se o escore fosse conhecido;

É possível generalizar o "escore de propensão" para avaliar situações onde o tratamento não é representado por uma variável *dummy*?

Sim. Através de uma função de propensão ou GPS (Escore de Propensão Generalizado)

Modelo Rubin: Tratamento Generalizado
Define-se um conjunto de resultados potenciais: $\Upsilon = (Y_i(t^P), t^P \in \tau$ para i=1,...,n,
Onde τ é um conjunto de possíveis valores de tratamento e $Y_i(t^p)$ é uma variável aleatória que relaciona um tratamento particular (t^p) a um resultado potencial.

Hipótese 1: SUTVA

- Ausência de equilíbrio geral e de externalidades:

$p\{Y_i(t_i^p) \mid T_j^A = t_j^p, X_i\} = p\{Y_i(t_i^p) \mid, X_i\}$ para qualquer $i \neq j$ e qualquer t_i^p, $t_j^p \in \tau$

Hipótese 2: CIA (Ignorabilidade da alocação do tratamento):

- A distribuição do tratamento não depende do resultado potencial dados os observáveis.

$p\{T^A \mid Y(t^p), X\} = p\{T^A \mid X\}$ para todo $X \in \chi$ *e conjuntos mensuráveis de* $A \subset \tau$.

Efeito médio causal:

$$E\{Y(t_1^p) - Y(t_2^p) \mid X\} = E\{Y(t_1^p) \mid T^A = t_1^p, X\} - E\{Y(t_2^p) \mid T^A = t_2^p, X\}$$

- Problema: dimensionalidade de X

A Função de Propensão

A função de propensão é a probabilidade condicional do tratamento -- seja ele binário, contínuo ou multivariado – dada as covariáveis observadas:

$e_{\psi}(T^{A} \mid X) = p_{\psi}\{T^{A} \mid X_{i}\}$ onde ψ parametriza a função.

- A função de propensão, $e_{\psi},$ é desconhecida e sua distribuição condicional $p_{\psi}\{T^{A} \mid X_{i}\}$ deve ser modelada de modo a estimar os parâmetros desconhecidos, ψ (máximo verossimilhança).

<u>Hipótese 3</u>: Função de Propensão é unicamente parametrizada

Esta hipótese implica que a função de propensão depende de X apenas através de $\theta_{\psi}(X)$. Ou seja, a função propensão é efetivamente "sintetizada" no parâmetro θ que tipicamente tem uma dimensão menor que X.

Exemplo com uma variável de tratamento contínua:

Pode-se modelar a distribuição do tratamento (contínuo) dado um vetor de covariáveis (X) como:

$$T^{A} \mid X \sim N(X^{\mathrm{T}}\beta, \sigma^{2})$$

$$e\ (T^{A} \mid \theta_{\psi}(X)), \psi(\beta, \sigma^{2}), \theta_{\psi}(X) = X^{\mathrm{T}}\beta$$

Dado ψ, a função de propensão é completamente determinada pelo escalar θ. Portanto, *matching* ou subclassificação de função de propensão podem ser realizados através de *matching* ou subclassificação de θ, independentemente da dimensão de X.

Teorema 1: Função de propensão com um escore de balanceamento

$$p\{T^A \mid e(.\mid X)\} = p\{T^A \mid X, e(.\mid X)\} = p\{T^A \mid X\}$$

Na prática este teorema pode ser checado através da regressão de cada uma das covariáveis (X) contra a variável de tratamento T^A, controlando-se pela função de propensão. Se o coeficiente estimado não for significativo, então a função de propensão foi bem especificada e balanceou a distribuição.

Teorema 2: CIA dada a função de propensão (unconfoundedness)

$p\{Y(t^p) \mid T^A, e(.\mid X)\} = p\{Y(t^p) \mid e(.\mid X)\}$ para qualquer $t^p \in \tau$

Dada a função de propensão, o resultado potencial, $Y(t^P)$, e o tratamento, T^A , são independentes.

Da teoria para a Prática

$p\{Y(t^p) \mid e(. \mid X)\}$ pode ter sua média calculada ao longo da distribuição da função de propensão de modo a obter a distribuição do resultado do tratamento, como uma função de valores específicos do tratamento (tipo ou intensidade).

Os autores advogam o seguinte método para calcular o impacto do tratamento a partir da integração de $p\{Y(t^p) \mid e(. \mid X)\}$ sobre a distribuicao da função de propensão:

1º passo: Modelar $p_{\psi}\{T^A \mid X_i\}$ de como a estimar os parâmetros de interesse $\hat{\psi}$.

2º passo: Calcular $\hat{\theta} = \theta_{\hat{\psi}}(X)$ para cada observação e subclassificar observações com valores similares de $\hat{\theta}$ em um numero parcimonioso de subclasses de tamanho parecidos.

3º passo: Dentro de cada subclasse modelar: $p\{Y(t^p) \mid T^A = t^p\}$ e estimar o coeficiente da regressão de $Y(t^P)$ contra (t^P). Na pratica, um ajuste adicional pela função de propensão pode ser requerido:

$p\{Y(t^p) \mid T^A = t^p, \hat{\theta}\}$, e talvez por algum X também.

4º passo: Calcular o efeito causal médio como a media ponderada do efeito de cada uma das subclasses.

- *Matching* parece ser uma alternativa de mais difícil implementação neste caso.

- Hirano e Imbens (2003) não usam subclassificação. Eles utilizam uma regressão saturada na função de propensão e na variável de tratamento.

Há muita vantagem em usar este procedimento ao invés de usar uma regressão simples que meça diretamente o impacto do tratamento sobre o indicador de interesse?

O que dizem os experimentos de Monte Carlo?

- dosagem
- múltiplo tratamento

a) a função de propensão é mal especificada de propósito

b) o resultado de uma regressão de Y em T^A e X é reportado para fins de comparação.

<u>Conclusão</u>: o modelo com a função de propensão <u>com subclassificação</u> é menos sensível à má especificação do modelo.

Exemplos:

1) Efeitos de fumar sobre gastos médicos

a) Variável de tratamento (interesse): carteira por dia x anos de fumo

Modelo de duas partes:

T=0 (logístico)

T>0 (regressão línea)

b) Variável de interesse: carteira por dia (freq) e anos de fumo (duração)

Modelo de duas partes para cada uma das variáveis de tratamento

2) Efeito da educação sobre a renda:

Angrist e Krueger (1995):

Balancear a variável instrumental.