因果推論中之維度縮減問題

中央研究院統計科學研究所　黃名鉞

一、平均處理效果

在因果推論問題裡，研究者主要探討的是一個母體經過不同的處理之後是否會有不同的反應。令$Y(0)$與$Y(1)$分別爲在$T=0$與$T=1$兩種處理之下的反應變數，則平均處理效果(Average treatment effect)定義爲

$$\delta = E\{T(1)-T(0)\}.$$

一般來說，研究者可以利用隨機對照試驗所得到的實驗型資料來估計δ。但在許多實際的例子中，研究者往往只能搜集到觀測型資料。在此資料結構之下，由於$(Y(0),Y(1))$與T並不獨立，因此直接比較$T=0$與$T=1$兩群反應變數的平均往往無法得到δ的無偏或一致性估計。針對此種資料，我們可以使用 Neyman-Rubin 模型[1]來刻劃$(Y(0),Y(1))$與T之間的相依性，其假設存在一群混淆因子X，使得$(Y(0),Y(1))$與T在給定X之下獨立。在此模型之下，我們可發現

$$E[E\{Y(T)|T=k,X\}]$$

$$=E[E\{Y(k)|T=1,X\}]$$

$$=E[E\{Y(k)|X\}]=E\{Y(k)\},$$

當中$k=0,1$。換言之，研究者可藉由控制對照混淆因子的不同數值來得到$E\{Y(T)|T=1,X=x\}-E\{Y(T)|T=0,X=x\}$的估計，再對混淆因子的數值取加權平均，即可得到$\delta$的無偏或一致性估計。

二、充分維度縮減

針對$E\{Y(T)|T=1,X=x\}$的估計，過去常用的方法爲引入適當的參數化模型。但在許多實際應用中，研究者往往沒有先驗的資訊幫助建模，或者沒有適當的方式來診斷模式的正確性。爲避免錯誤的模型假設，另一種方式爲使用無母數平滑法來直接估計此條件期望。雖然無母數法不需要特定的模型假設，但是當混淆因子的個數p過多時，此平滑估計法往往會因維數災難(curse of dimensionality)而在有限樣本表現得非常不穩定。因此，[2]提出了使用傾向評分匹配(propensity score matching)來估計此條件期望。傾向評分最大的好處是它是一維的，但[3]發現根據此評分所得到δ的估計並不會是有效估計。因此在此研究中，我們試圖尋找其它低維度的評分，以得到相對穩定並且依然爲有效的估計式。令span(B_0)爲滿足

$$T \perp X|B_0^T X, Y(0) \perp X|B_0^T X, Y(1) \perp X|B_0^T X$$

之最小線性子空間，當中B爲一$p \times d_0$之係數矩陣，則$B_0^T X$即爲我們提出的匹配評分。我們稱此法爲聯合充分維度縮減。當處理變數非二元時，此法亦可推廣爲尋找滿足

$$T \perp X|B_0^T X, Y(t) \perp X|B_0^T X \ \forall t \in \mathrm{supp}(T)$$

之最小線性子空間span(B_0)。值得注意的是d_0亦爲待估之參數，因此此法可視爲一巢狀多指標模型選擇問題。特別當$d_0=p$時，此模型約化爲完全非參數化迴歸模型，因此我們不需特定的模型假設，而是藉由資料來搜尋正確的模型。此外，我們也證明了利用$B_0^T X$做匹配所得到δ的估計依然是有效估計（在完全非參數化迴歸模型之下）。

三、估計準則

針對二元的處理變數，我們提出使用以下準則來估計B_0：

$$\text{cv}(d,B,h)=\frac{1}{n}\sum_{i=1}^{n}\left[\left\{T_i-\hat{\pi}^{-i}(B^TX_i)\right\}^2(1-\bar{T})+\sum_{k=0}^{1}T_i^k(1-T_i)^{1-k}\int\left\{1(Y_i\le y)-\hat{F}_k^{-i}(y|B^TX_i)\right\}^2d\hat{F}_Y(y)\right],$$

當中$\bar{T}=n^{-1}\sum_{i=1}^{n}T_i$, $\hat{F}_Y(y)=n^{-1}\sum_{i=1}^{n}1(Y_i\le y)$，

$$\hat{\pi}(B^Tx)=\frac{\sum_{j=1}^{n}T_jK_h\left(B^TX_j-B^Tx\right)}{\sum_{j=1}^{n}K_h\left(B^TX_j-B^Tx\right)}\text{ ，}$$

$$\hat{F}_k(y|B^Tx)=\frac{\sum_{j=1}^{n}T_j^k\left(1-T_j\right)^{1-k}1\left(Y_j\le y\right)K_h\left(B^TX_j-B^Tx\right)}{\sum_{j=1}^{n}T_j^k\left(1-T_j\right)^{1-k}K_h\left(B^TX_j-B^Tx\right)}\text{ ，}$$

以及上標$-i$代表將第i筆資料去除所得之估計式。使$\text{cv}(d,B,h)$最小的$(\hat{d},\hat{B},\hat{h})$即爲我們所提之估計式。此準則之特色爲同時選取正確的維度以及相對應的子空間基底，與過去常用之充分維度縮減將兩者分開處理的手法相當不同。實務上也因只須計算單一準則，故有較好之計算效率。

當T爲連續處理變數時，因$T=t$可能只包含一筆資料或沒有資料，故此準則無法直接套用。爲解決這個問題，我們利用$(T,Y)\perp X|B_0^TX$之性質，改提出以下準則

$$\text{cv}(d,B,h)=\frac{1}{n}\sum_{i=1}^{n}\int\left\{1(T_i\le t,Y_i\le y)-\hat{F}^{-i}(t,y|B^TX_i)\right\}^2d\hat{F}_{T,Y}(t,y),$$

當中$\hat{F}_{T,Y}(t,y)=n^{-1}\sum_{i=1}^{n}1(T_i\le t,Y_i\le y)$**以及**

$$\hat{F}(t,y|B^Tx)=\frac{\sum_{j=1}^{n}1\left(T_j\le t,Y_j\le y\right)K_h\left(B^TX_j-B^Tx\right)}{\sum_{j=1}^{n}K_h\left(B^TX_j-B^Tx\right)}.$$

此準則爲[4]推廣至多變量反應變數之情況。

四、數值模擬

我們考慮一個二元處理變數的例子。令$X=(X_1,\dots,X_{10})$爲互相獨立之標準常態分配混淆因子，傾向評分設定爲$\pi(X)=e^{X_1+X_2}/(1+e^{X_1+X_2})$，反應變數則分別爲$Y(1)=1+2X_1-X_2+\varepsilon_1$以及$Y(0)=X_1-X_2+\varepsilon_0$，當中$\varepsilon_1,\varepsilon_0$爲標準差 0.1 之獨立常態誤差。由此可知待估計之平均處理效果爲$\delta=1$。下表爲利用不同的匹配評分以及 1000 次模擬所得到的δ估計之平均及標準差。我們發現$\hat{B}^TX$可以得出較好的δ估計，特別是與無降維的原始評分X比較時，所提之估計式可得到較小的標準差。另外，我們在樣本數 100,200,400 之下分別計算δ估計標準差與效率界限(efficiency bound)之比值爲2.614,2.180,1.555。由此可驗證當樣本數足夠大時，所提之估計式確爲有效估計。

	$n=100$		$n=200$	
score	mean	s.d.	mean	s.d.
$\hat{B}^TX$	1.045	0.2726	1.018	0.1607
X	1.005	0.3345	1.090	0.2526
$\hat{B}_\pi^TX$	1.019	0.3796	1.027	0.2827
$\hat{\pi}$	0.685	0.1976	0.683	0.1416

參考文獻

[1] D. B. Rubin, Estimating causal effects of treatments in randomized and nonrandomized studies. J. Educ. Psychol. **66** (1974), 688-701.

[2] P. R. Rosenbaum and D. B. Rubin, The central role of the propensity score in observational studies for causal effects. Biometrika **70** (1983), 41-55.

[3] J. Hahn, On the role of the propensity score in efficient semiparametric estimation of average treatment effects. Econometrica **66** (1998), 315-331.

[4] M.-Y. Huang and C.-T. Chiang, An effective semiparametric estimation approach for the sufficient dimension reduction model. J. Am. Statist. Assoc. **112**, 1296-1310.