Additive Model とゆかいな仲間たち

Mona ちゃん
ケロロ大学タママ大学ペコポン研究所

January 6, 2025

1. 加法モデル (Additive Model)
2. 加法部分線形モデル (Additive Partially Linear Model)
3. セミパラメトリックなめらか係数モデル (Semiparametric Smooth Coefficient Model)

1. 加法モデル (Additive Model)
2. 加法部分線形モデル (Additive Partially Linear Model)
3. セミパラメトリックなめらか係数モデル (Semiparametric Smooth Coefficient Model)

説明変数ベクトルを分解して足し合わせる

- 説明変数ベクトル $Z_i = (Z_{1i}, ..., Z_{qi})$ を各要素にばらす。
- c_0 をスカラーパラメータ、$g_l(.), l = 1, .., q$ を未知のなめらかな関数として以下のようなモデルを考える。

$$Y_i = c_0 + g_1(Z_{1i}) + ... + g_q(Z_{qi}) + u_i$$

このようなモデルは**加法モデル (Additive Model)** と呼ばれる。

ノンパラじゃダメ?

- 加法モデルは「限りなくノンパラに近いセミパラ」である。
  cf. 村上龍『限りなく透明に近いブルー』(1976.7, 講談社)
- それならもうノンパラ回帰でいいのでは?

$$Y_i = g(Z_i) + u_i$$

- しかし、説明変数 Z_i の次元が大きい場合、ノンパラ回帰だと「次元の呪い」が起きてしまう...
- 加法モデルでは、説明変数ベクトルを要素に分解してそれぞれを未知関数に入れてやることで、各ノンパラ構造に含まれる説明変数の次元を抑えている。

識別のための仮定

- 任意の c および $l \neq m$ に対して
$$[g_l(z_l) + c] + [g_m(z_m) - c] = g_l(z_l) + g_m(z_m)$$
が成立するため、このままだと $g_1, ..., g_q$ たちを識別することはできない。
- そこで、識別のために $l = 1, ..., q$ に対して
$$E[g_l(Z_l)] = 0$$
という仮定をおく。この仮定の帰結として
$$E(Y_i) = c_0$$
が成り立つ。

α 番目の説明変数を孤立させる

- $Z_{\underline{\alpha}i} \coloneqq (Z_{1i}, \ldots, Z_{\alpha-1,i}, Z_{\alpha+1,i}, \ldots, Z_{qi})$ ($Z_{\alpha i}$ だけ除外)
- $G_{\underline{\alpha}(z_{\underline{\alpha}})} \coloneqq g_1(z_i) + \ldots + g_{\alpha-1}(z_{\alpha-1}) + g_{\alpha+1}(z_{\alpha+1}) + \ldots + g_q(z_q)$
- 加法モデルは以下のように書ける:

$$Y_i = c_0 + g_\alpha(Z_{\alpha i}) + G_{\underline{\alpha}}(Z_{\underline{\alpha}i}) + u_i$$

条件付き期待値 $E[.|Z_\alpha = z_\alpha, Z_{\underline{\alpha j}}]$ を取る

- $\xi(z_\alpha, Z_{\underline{\alpha j}}) := E(Y_j|Z_\alpha = z_\alpha, Z_{\underline{\alpha j}})$ ($Z_{\underline{\alpha}}$ は確率変数のまま!)
- 加法モデルの両辺の条件付き期待値 $E[.|Z_\alpha = z_\alpha, Z_{\underline{\alpha j}}]$ をとると

$$\xi(z_\alpha, Z_{\underline{\alpha j}}) = c_0 + g_\alpha(z_\alpha) + G_{\underline{\alpha}}(Z_{\underline{\alpha j}})$$

$Z_{\underline{\alpha}j}$ に関して Marginal Integration

- $m_\alpha(z_\alpha) := E[\xi_\alpha(z_\alpha, Z_{\underline{\alpha}j})]$
- $Z_{\underline{\alpha}j}$ に関して周辺期待値をとると、識別のための仮定から

$$E[G_{\underline{\alpha}}(Z_{\underline{\alpha}j})] = 0$$

なので

$$m_\alpha(z_\alpha) = c_0 + g_\alpha(z_\alpha)$$

である (Z のうち Z_α しか残っていないことに注意!)。

Marginal Integration Method

- したがって

$$g_\alpha(z_\alpha) = m_\alpha(z_\alpha) - c_0$$

ここで

$$c_0 = E(Y_i) = E[E[\xi(Z_{\alpha i}, Z_{\underline{\alpha} i})|Z_{\alpha i}]] = E[m_\alpha(Z_{\alpha i})]$$

に注意すると

$$g_\alpha(z_\alpha) = m_\alpha(z_\alpha) - E[m_\alpha(Z_{\alpha i})]$$

これを利用して $g_\alpha(z_\alpha)$ を推定しよう。

$m_\alpha(z_\alpha)$ の推定

- まず、$E[Y_j|Z_{\alpha j} = z_\alpha, Z_{\underline{\alpha}j}]$ を Local Linear に推定する。
- $\tilde{\alpha}_\alpha(z_\alpha, Z_{\underline{\alpha}j})$ を以下の最小化問題の解の a として定義する。

$$\min_{a,b} \sum_{l=1}^{n} [y_l - a - (Z_{\alpha l} - z_\alpha)b]^2 k_{h_\alpha}(Z_{\alpha l} - z_\alpha) K_{h_{\underline{\alpha}}}(Z_{\underline{\alpha}l} - Z_{\underline{\alpha}j})$$

- この $\tilde{\alpha}_\alpha(z_\alpha, Z_{\underline{\alpha}j})$ を用いると、$m_\alpha(z_\alpha) = E[E[Y_j|Z_{\alpha j} = z_\alpha, Z_{\underline{\alpha}j}]]$ は以下のように推定することができる。

$$\tilde{m}_\alpha(z_\alpha) = \frac{1}{n} \sum_{j=1}^{n} \tilde{\alpha}_\alpha(z_\alpha, Z_{\underline{\alpha}j})$$

$E[m_\alpha(Z_\alpha)]$ の推定

- $E[m_\alpha(Z_\alpha)]$ も $\tilde{m}_\alpha(z_\alpha)$ を用いて

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{m}_\alpha(Z_{\alpha i})$$

  により推定することができる。

- 以上より、$g_\alpha(z_\alpha)$ の推定量は以下のようになる。

$$\tilde{g}_\alpha(z_\alpha) = \tilde{m}_\alpha(z_\alpha) - \frac{1}{n}\sum_{i=1}^{n}\tilde{m}_\alpha(Z_{\alpha i})$$

計算コストを下げよう

- Marginal Integration では $i, j = 1, ..., n$ に対して $\tilde{\alpha}_\alpha(Z_{\underline{\alpha}i}, Z_{\underline{\alpha}j})$ を求める必要があり、計算コストが高い...
- そこで、心機一転 Y の Z_α へのノンパラ回帰を考える:

$$E(Y|Z_\alpha = z_\alpha) = c_0 + g_\alpha(z_\alpha) + \underbrace{\sum_{s \neq \alpha} E[g_s(Z_s)|Z_\alpha = z_\alpha]}_{\text{邪魔!!!}}$$

操作変数を投入

- 以下の条件を満たす操作変数 $w_\alpha(Z)$ を考える

$$E[w_\alpha(Z)|Z_\alpha = z_\alpha] = 1$$

$$E[w_\alpha(Z)g_s(Z_s)|Z_\alpha = z_\alpha] = 0 \text{ for } s \neq \alpha$$

- $Y_i = c_0 + g_\alpha(Z_{\alpha i}) + G_{\underline{\alpha}}(Z_{\underline{\alpha} i}) + u_i$ の両辺に $w_\alpha(Z_i)$ をかけて、条件付き期待値 $E[.|Z_\alpha = z_\alpha]$ をとると

$$E[w_\alpha(Z_i)Y_i|Z_{\alpha i} = z_\alpha] = c_0 + g_\alpha(z_\alpha)$$

$c_0 + g_\alpha(z_\alpha)$ が推定できそう!

どんな操作変数が適任?

- 以下のような操作変数がよさそう

$$w_\alpha(Z) = \frac{f_\alpha(Z_\alpha) f_{\underline{\alpha}}(Z_{\underline{\alpha}})}{f(Z_\alpha, Z_{\underline{\alpha}})}$$

- 任意の確率変数ξに関して

$$E[w_\alpha(Z)\xi | Z_\alpha = z_\alpha] = \int \xi f_{\underline{\alpha}}(z_{\underline{\alpha}}) dz_{\underline{\alpha}}$$

これは確率変数ξの$Z_{\underline{\alpha}}$に関するMarginal Integration!
(この事実はもう少し後で活躍)

ありがとう $w_\alpha(Z)$

- 操作変数

$$w_\alpha(Z) = \frac{f_\alpha(Z_\alpha) f_{\underline{\alpha}}(Z_{\underline{\alpha}})}{f(Z_\alpha, Z_{\underline{\alpha}})}$$

に関して

$$E[w_\alpha(Z_i)Y_i | Z_{\alpha i} = z_\alpha] = c_0 + g_\alpha(z_\alpha)$$

が成り立つので、

$$E[w_\alpha(Z_i)Y_i | Z_{\alpha i} = z_\alpha]$$

をノンパラメトリック推定すればよい。

$c_0 + g_\alpha(z_\alpha)$ の推定

$$
\begin{aligned}
\tilde{\gamma}(z_\alpha) &= \frac{1}{nh_o}\sum_{j=1}^{n}\frac{k(\frac{Z_{\alpha j}-z_\alpha}{h_o})}{\hat{f}_\alpha(z_\alpha)}\frac{\hat{f}_\alpha(z_\alpha)\hat{f}_{\underline{\alpha}}(Z_{\underline{\alpha}j})}{\hat{f}(z_\alpha, Z_{\underline{\alpha}j})}Y_j \\
&= \frac{1}{nh_o}\sum_{j=1}^{n}k\left(\frac{Z_{\alpha j}-z_\alpha}{h_o}\right)\frac{\hat{f}_{\underline{\alpha}}(Z_{\underline{\alpha}j})}{\hat{f}(z_\alpha, Z_{\underline{\alpha}j})}Y_j \\
&= \frac{1}{n^2h_o^q}\sum_{j=1}^{n}\sum_{l=1}^{n}\frac{k(\frac{Z_\alpha - z_\alpha}{h_o})K_{\underline{\alpha}}(\frac{Z_{\underline{\alpha}l}-Z_{\underline{\alpha}j}}{h_o})}{\hat{f}(z_\alpha, Z_{\underline{\alpha}j})}Y_j
\end{aligned}
$$

$g_\alpha(z_\alpha)$ の推定

- $\tilde{\gamma}(z_\alpha)$ によって $c_0 + g_\alpha(z_\alpha)$ を推定できることがわかった。
- ここで識別のために課した条件 $E[g_\alpha(Z_\alpha)] = 0$ を思い出すと c_0 は $\tilde{\gamma}(z_{\alpha j})$ の標本平均によって推定できそうである。
  (Y_i の標本平均じゃダメ?)
- 従って
$$\tilde{g}_\alpha(z_\alpha) = \tilde{\gamma}(z_\alpha) - \frac{1}{n}\sum_{j=1}^{n} \tilde{\gamma}_\alpha(Z_{\alpha j})$$
  により $g_\alpha(z_\alpha)$ を推定できる。
- 短所: 漸近分散が大きくなってしまう。

Oracle推定量

- c_0 と $g_s(z_s)(s \neq \alpha)$ が分かっているとする。
- $Y_{\alpha i}^{oracle} := Y_i - c_0 - \sum_{s \neq \alpha} g_s(Z_{si})$
- $Y_{\alpha i}^{oracle}$ を $Z_{\alpha i}$ に Local Constant にノンパラ回帰すると $g_\alpha(z_\alpha)$ の (infeasible な)**Oracle 推定量**が得られる:

$$\tilde{g}_\alpha^{oracle}(z_\alpha) = \frac{\sum_{j=1}^{n} k(\frac{Z_{\alpha j} - z_\alpha}{h_\alpha}) Y_{\alpha j}^{oracle}}{\sum_{j=1}^{n} k(\frac{Z_{\alpha j} - z_\alpha}{h_\alpha})}$$

バンド幅は $\delta_1 > 0$ に対して $h_\alpha = \delta_1 n^{-\frac{1}{5}}$ とするとよい。

- Local Linear にやってもよい。

infeasibleをfeasibleに

- しかし実際には $g_s(z_s)(s \neq \alpha)$ はわからない。
- そこで、c_0 と $g_s(z_s)(s \neq \alpha)$ をそれぞれの推定量で代用する。
- c_0 は Y_i の標本平均でO.K.
- $c_0 + g_s(z_s)$ は以下の推定量で一致推定できる。

$$\tilde{\gamma}(z_s) = \frac{1}{nh_\alpha} \sum_{j=1}^{n} k\left(\frac{Z_{sj} - z_s}{h_\alpha}\right) \frac{\hat{f}_{\underline{s}}(Z_{\underline{s}j})}{\hat{f}(Z_{sj}, Z_{\underline{s}j})} Y_j$$

feasibleなOracle推定量

- まず $Y_{\alpha i}^{oracle}$ を推定

$$Y_{\alpha i}^{2step} = Y_i - \sum_{s \neq \alpha} \tilde{\gamma}(Z_{si}) + (q-1)\bar{Y}$$

- $Y_{\alpha i}^{oracle}$ を $Y_{\alpha i}^{2step}$ で代用すれば feasible な **Oracle 推定量**が得られる:

$$\hat{g}_{\alpha}^{oracle}(z_\alpha) = \frac{\sum_{j=1}^{n} k(\frac{Z_{\alpha j}-z_\alpha}{h_\alpha})Y_{\alpha j}^{2step}}{\sum_{j=1}^{n} k(\frac{Z_{\alpha j}-z_\alpha}{h_\alpha})}$$

- (漸近的) 効率性の観点で、feasible な Oracle 推定量は infeasible なものに引けをとらない!

もう1つの推定方法: Backfitting

- 加法モデルの両辺の条件付き期待値 $E[.|Z_{\alpha i} = z_\alpha]$ をとり、整理すると、各 $\alpha = 1, ..., q$ に対して

$$g_\alpha(z_\alpha) = E(Y_i|Z_{\alpha i} = z_\alpha) - c_0 - \sum_{s \neq \alpha} E[g_s(Z_{si})|Z_{\alpha i} = z_\alpha]$$

  が成り立つ。
- 左辺に $g_\alpha(z_\alpha)$、右辺に $g_s(z_s)(s \neq \alpha)$ があるので、逐次的に推定するのがよさそう...?

Backfittingによる逐次的推定

$l=0$ 各 $\alpha=1,...,q$ に対して初期推定量 $\hat{g}_\alpha^{[0]}(z_\alpha)$ を決める
(例えば $\hat{g}_\alpha^{[0]}(z_\alpha)=0$)。さらに $\bar{Y}$ で c_0 を推定。

$l=1,2,\ldots$ $(\hat{g}_1^{[l-1]}(z_1),...,\hat{g}_q^{[l-1]}(z_q))$ を所与として、新しい推定量
$\hat{g}_\alpha^{[l]}(z_\alpha)$ を $\alpha=1,2,\ldots$ の順で以下のように設定

$$\hat{g}_\alpha^{[l]}(z_\alpha)=\hat{E}[Y_i|Z_{\alpha i}=z_\alpha]-\hat{c}_0-\sum_{s=1}^{\alpha-1}\hat{E}[\hat{g}_s^{[l]}(Z_{si})|Z_{\alpha i}=z_\alpha]$$
$$-\sum_{s=\alpha+1}^{q}\hat{E}[\hat{g}_s^{[l-1]}(Z_{si})|Z_{\alpha i}=z_\alpha]$$

- ただし、確率変数 A_i に対して $\hat{E}[A_i|Z_{\alpha i}=z_\alpha]$ は $E[A_i|Z_{\alpha i}=z_\alpha]$ のカーネル推定量 (Local Constant/Linear でも可)

リンク関数に通してみる

- 加法モデルをリンク関数 $G(.)$ に通してみる:

$$Y_i = G\left[c_0 + \sum_{\alpha=1}^{q} g_\alpha(z_\alpha)\right] + u_i$$

- ここでリンク関数 $G(.)$ は既知(例えば logit 関数)
  さらに G の逆関数も求まっているとする ($M := G^{-1}$)
- 目標:$m(z) = E(Y|Z = z)$ を推定したい。

条件付き期待値を取って逆関数 G^{-1} に通す

- 両辺の条件付き期待値 $E[.|Z=z]$ をとると

$$m(z) = G\left[c_0 + \sum_{\alpha=1}^{q} g_\alpha(z_\alpha)\right]$$

両辺をリンク関数 G の逆関数に通すと

$$M[m(z)] = c_0 + \sum_{\alpha=1}^{q} g_\alpha(z_\alpha)$$

Marginal Integration 再び

- $M[m(Z)]$ を $Z_{\underline{\alpha}}$ に関して Marginal Integration すると

$$\int M[m(Z_\alpha, z_{\underline{\alpha}})] f_{\underline{\alpha}}(z_{\underline{\alpha}}) dz_{\underline{\alpha}} = c_0 + g_\alpha(Z_\alpha)$$

  この最左辺を $\phi_\alpha(Z_\alpha)$ として定義すると

$$\phi_\alpha(z_\alpha) = c_0 + g_\alpha(z_\alpha)$$

- $m(z)$ を推定するために c_0 と $g_\alpha(z_\alpha)$ が必要。そのためには $\phi_\alpha(z_\alpha)$ が必要。そのためには $m(z)$ が必要... どうする?

とりあえず $m(z)$ をざっくり推定

- 悩んでいても仕方ないので、とりあえず $m(z) = E(Y|Z=z)$ をざっくり推定する

$$\hat{m}(z_\alpha, z_{\underline{\alpha}}) = \frac{\sum_i k_{h_\alpha}(z_\alpha, Z_{\alpha i}) W_{h_{\underline{\alpha}}}(z_{\underline{\alpha}}, Z_{\underline{\alpha i}}) Y_i}{\sum_i k_{h_\alpha}(z_\alpha, Z_{\alpha i}) W_{h_{\underline{\alpha}}}(z_{\underline{\alpha}}, Z_{\underline{\alpha i}})}$$

- k_{h_α} は2次カーネル、
$W_{h_{\underline{\alpha}}}$ は(特に $q > 2$ のとき) 高次カーネルがよい(らしい)。

$\phi_\alpha(z_\alpha)$ の推定

- ざっくりとした推定 $\hat{m}(z)$ を用いると $\phi_\alpha(z_\alpha)$ は以下のように推定することができる。

$$\tilde{\phi}_\alpha(z_\alpha) = \frac{1}{n}\sum_{i=1}^{n} M[\hat{m}(z_\alpha, Z_{\underline{\alpha} i})]$$

- c_0 は以下のように推定できる (今回は $\bar{Y}$ じゃダメ)。

$$\tilde{c}_0 = \frac{1}{nq}\sum_{\alpha=1}^{q}\sum_{i=1}^{n} \tilde{\phi}_\alpha(Z_{\alpha i})$$

$g_\alpha(z_\alpha)$ そして $m(z)$ の推定

- 推定量 $\tilde{c}_0, \tilde{\phi}_\alpha(z_\alpha)$ を用いると、$g_\alpha(z_\alpha)$ は以下のように推定できる。

$$\tilde{g}_\alpha(z_\alpha) = \tilde{\phi}_\alpha(z_\alpha) - \tilde{c}_0$$

- $\tilde{g}_\alpha(z_\alpha)(\alpha = 1, ..., q)$ を用いると (最終的な)$m(z)$ の推定量 $\hat{m}(z)$ が得られる:

$$G\left[\tilde{c}_0 + \sum_{\alpha=1}^{q} \tilde{g}_\alpha(z_\alpha)\right]$$

1. 加法モデル (Additive Model)
2. 加法部分線形モデル (Additive Partially Linear Model)
3. セミパラメトリックなめらか係数モデル (Semiparametric Smooth Coefficient Model)

加法モデルの限界

- 加法モデルでは、最初に説明変数をばらしてから関数に通すことにより、各ノンパラ構造に入っている説明変数の次元を 1 次元に抑え、「次元の呪い」を回避している。
- しかし、この最初の「ばらす」工程があだとなって説明変数の間での「交差効果」は無視されてしまう。

加法部分線形モデル

- 説明変数ベクトル $Z_i = (Z_{1i}, ..., Z_{qi})$ を各要素にばらす。
- β_0 をスカラーパラメータ、β を p 次元パラメータ、$g_l(.), l = 1, .., q$ を未知のなめらかな関数として以下のようなモデルを考える。

$$Y_i = \beta_0 + X_i'\beta + g_1(Z_{1i}) + ... + g_q(Z_{qi}) + u_i$$

- X_i は $(Z_{1i}, ..., Z_{qi})$ たちの「交差項」を含む。
- このようなモデルは**加法部分線形モデル (Additive Partially Linear Model)** という。

加法部分線形モデルのココが凄い

- 加法モデルと同様に、各 $g_l(.), l = 1, ..., q$ に入る説明変数の次元は 1 次元であるため、「次元の呪い」を回避できる。
- さらに、$X_i'\beta$ により $(Z_{1i}, ..., Z_{qi})$ の「交差効果」を表現することも可能。

Robinson のマネ

- まず β を推定しよう。
- 加法部分線形モデルの両辺の条件付き期待値 $E[.|Z_i]$ を取ると

$$E[Y_i|Z_i] = \beta_0 + E[X_i|Z_i]'\beta + g_1(Z_{1i}) + ... + g_q(Z_{qi})$$

元のモデル式との差をとると

$$(Y_i - E[Y_i|Z_i]) = (X_i - E[X_i|Z_i])'\beta + u_i$$

ノンパラパートが消えた!

β の推定量

- Robinson の推定方法に習い $Y_i - E[Y_i|Z_i]$ を $X_i - E[X_i|Z_i]$ に線形回帰すれば、β の infeasible な推定量 $\tilde{\beta}_{\mathrm{inf}}$ が得られる。
- $E[Y_i|Z_i], E[X_i|Z_i]$ をそれぞれの (ノンパラ) 推定量 $\hat{E}[Y_i|Z_i], \hat{E}[X_i|Z_i]$ で置き換えれば、feasible な推定量 $\tilde{\beta}$ が得られる:

$$\tilde{\beta} = \left\{ \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{E}[X_i|Z_i])(X_i - \hat{E}[X_i|Z_i])' \right\}^{-1} \times \left\{ \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{E}[X_i|Z_i])(Y_i - \hat{E}[Y_i|Z_i]) \right\}$$

feasible はいいけど本当に invertible?

- β の feasible な推定量 $\tilde{\beta}$ の一部

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{E}[X_i|Z_i])(X_i - \hat{E}[X_i|Z_i])' \right\}^{-1}$$

inverse の中身は、$E[(X_i - E[X_i|Z_i])(X_i - E[X_i|Z_i])']$ に収束する。
この収束先は本当に逆行列を持つ?

「交差項」に一工夫が必要

- 「交差項」と聞いて多くの人が思い浮かべるであろう $Z_l Z_m (l \neq m)$ はどうだろうか?$X_i = Z_{li} Z_{mi}$ とすると

$$X_i - E[X_i | Z_i] = Z_{li} Z_{mi} - E[Z_{li} Z_{mi} | Z_i] = 0$$

  となるため、逆行列が存在しない!
- 「交差項」に一工夫が必要。

こんな「交差項」ならO.K.

- $E[(X_i - E[X_i|Z_i])(X_i - E[X_i|Z_i])']$ が正定値行列であるとき、X は Z の「非確率的関数でない」という。
- X が Z の「非確率的関数でない」のであれば、$E[(X_i - E[X_i|Z_i])(X_i - E[X_i|Z_i])']$ が逆行列を持つため問題ナシ。

ノンパラパート $g_l(.), l = 1, .., q$ の推定

- β の $\sqrt{n}$ 一致推定値 $\tilde{\beta}$ を用いると加法部分線形モデルは以下のように書ける。

$$Y_i - X_i'\tilde{\beta} = \beta_0 + \sum_{\alpha=1}^{q} g_\alpha(Z_{\alpha i}) + \varepsilon_i$$

$$\varepsilon_i = u_i + X_i'(\beta - \tilde{\beta})$$

- $\tilde{\beta}$ は $\sqrt{n}$ 一致性を持つため、真の値への収束レートは一般のノンパラ推定量よりも速い。したがって、β をその $\sqrt{n}$ 一致推定量 $\tilde{\beta}$ で置き換えても問題ない。
- あとは加法モデルと同様にして $g_l(.), l = 1, .., q$ を推定すればよい。

やっぱり計算コストがネック

- 加法モデルと同様に、加法部分線形モデルでも計算コストがネックとなる。
- 計算コストをなんとかして抑えることができないだろうか?

初心 (Robinson) に帰ろう

- Robinson の推定方法の核となるアイデアは、「ノンパラパート $g(Z)$ を変化させることなくパラメトリックパートを変化させる」である。
- 「ノンパラパート $g(Z)$ を変化させることなくパラメトリックパートを変化させる」ような"作用"は他にないだろうか?

操作変数$w_\alpha(Z_i)$が再び活躍

- 確率変数ξ_iに対して$\xi_{\alpha i}$を以下のように定義する。

$$\xi_{\alpha i} := E[\xi_i w_\alpha(Z_i)|Z_{\alpha i}] = \int E[\xi_i|Z_i = z] f_{\underline{\alpha}}(z_{\underline{\alpha}}) dz_{\underline{\alpha}}$$

ただし$w_\alpha(Z_i)$は操作変数:

$$w_\alpha(z) = \frac{f_\alpha(z_\alpha) f_{\underline{\alpha}}(z_{\underline{\alpha}})}{f(z_\alpha, z_{\underline{\alpha}})}$$

- ここで$\xi_{\alpha i}$は確率変数$\xi_{\alpha i}$の$Z_{\underline{\alpha}}$に関する Marginal Integration となっていることに注意。

加法関数空間への射影

- $\{g_1(Z_{1i}), ..., g_q(Z_{qi})\}$ が張る空間を加法関数空間と呼ぶ。
- 確率変数 ξ_i の加法関数空間への射影 $E_A(\xi_i)$ を以下のように定義する。

$$E_A(\xi_i) := \sum_{\alpha=1}^{q} \xi_{\alpha i}$$

- 各 α について $E_A[g_\alpha(Z_{\alpha i})] = g_\alpha(Z_{\alpha i})$ が成り立ち、より一般に

$$\Xi_i = a_1 g_1(Z_{1i}) + ... + a_q g_q(Z_{qi})$$

の形で表される確率変数 Ξ_i について $E_A(\Xi_i) = \Xi_i$ が成り立つ。
射影してもノンパラパートは変化しない!

射影してノンパラパートとさよなら

- 加法部分線形モデルの両辺を加法関数空間に射影すると

$$E_A(Y_i) = E_A(X_i)'\beta + \sum_{s=1}^{q} g(Z_{si})$$

元のモデル式との差をとると

$$Y_i - E_A(Y_i) = (X_i - E_A(X_i))'\beta + u_i$$

ノンパラパートが消えた!

$E_A(Y_i), E_A(X_i)$ の推定

- β の (feasible な) 推定量を得るためには $E_A(Y_i), E_A(X_i)$ を推定する必要がある。
- $\xi = Y_i, X_i$ として

$$\hat{\xi}_{\alpha i} = \frac{1}{nh_\alpha}\sum_{j=1}^{n} k\left(\frac{Z_{\alpha j} - Z_{\alpha i}}{h_\alpha}\right)\frac{\hat{f}_{\underline{\alpha}}(Z_{\underline{\alpha j}})}{\hat{f}(Z_{\alpha j}, Z_{\underline{\alpha j}})}$$

  を計算すると $Y_{\alpha i}, X_{\alpha i}$ の推定量が得られる。
- $E_A(Y_i), E_A(X_i)$ は以下のように推定できる。

$$\hat{E}_A(Y_i) = \sum_{\alpha=1}^{q}\hat{Y}_{\alpha i}, \hat{E}_A(X_i) = \sum_{\alpha=1}^{q}\hat{X}_{\alpha i}$$

feasible な β の推定方法と $g_\alpha(z_\alpha)$ の推定

- $E_A(Y_i), E_A(X_i)$ をそれぞれの推定量 $\hat{E}_A(Y_i), \hat{E}_A(X_i)$ で置き換えた

$$Y_i - \hat{E}_A(Y_i) = (X_i - \hat{E}_A(X_i))'\beta + u_i$$

で線形回帰を行えば、β の推定量 $\hat{\beta}$ が得られる。
- ノンパラパート $g_\alpha(z_\alpha)$ も

$$Y_i - X_i'\hat{\beta} = \beta_0 + \sum_{\alpha=1}^{q} g_\alpha(Z_{\alpha i}) + \varepsilon_i$$

から推定することができる。
- $g_l(.), l = 1, ..., q$ を求めることなく β を推定できるため、β にだけ興味があるなら計算コストが抑えられる。

1. 加法モデル (Additive Model)
2. 加法部分線形モデル (Additive Partially Linear Model)
3. セミパラメトリックなめらか係数モデル (Semiparametric Smooth Coefficient Model)

部分線形モデルを拡張

- あけましてノメでとうございます!
- 部分線形モデルを拡張した以下のようなモデルを考える。

$$Y_i = X_i' \beta(Z_i) + u_i$$

  X_i は定数項を含む p 次元説明変数ベクトル。
  $\beta(.)$ は q 次元説明変数ベクトル Z_i の関数。
  (パラメトリックな回帰係数 β に説明変数 Z がノンパラメトリックに寄生!?)
- このようなモデルは**セミパラメトリックなめらか係数モデル (Semiparametric Smooth Coefficient Model)** という。

Local Constantに推定

- モデル式の両辺に左から X_i をかけ、条件付き期待値 $E(.|Z_i = z)$ をとると、
$$E(X_iY_i|z) = E(X_iX_i'|z)\beta(z)$$
よって
$$\beta(z) = [E(X_iX_i'|z)]^{-1}E(X_iY_i|z)$$
- したがって以下のような $\beta(z)$ の Local Constant な推定量が考えられる。

$$\hat{\beta}(z) = \left[\sum_{j=1}^{n} X_jX_j'K\left(\frac{Z_j - z}{h}\right)\right]^{-1}\sum_{j=1}^{n} X_jY_jK\left(\frac{Z_j - z}{h}\right)$$

Local Constant推定量の解釈

- Z をスカラー確率変数とし、カーネル関数 $K(.)$ を一様カーネルとすると、Local Constant 推定量は以下のようになる。

$$\hat{\beta}(z) = \left[\sum_{|Z_j - z| \leq h} X_j X_j' \right]^{-1} \sum_{|Z_j - z| \leq h} X_j Y_j$$

  これは、対応する Z_j が z に近い値をとるような観測 (X_j, Y_j) のみを用いて行った、Y_j から X_j への線形回帰!
- 他のカーネル関数を用いた場合についても同様の解釈が可能。

漸近理論のための仮定(の解釈)

- $\hat{\beta}(z)$ が一致性・漸近正規性を持つためには、$n \to \infty$ としたときに、$h \to 0$ と $nh \to \infty$ が成り立たなければならない。
- 今、Z_i の定義域が $O(h^{-1})$ 個の幅 h の区間に分割されていることに注意すると、各区間に入っている観測の個数は $O(nh)$ 個。
- つまり、$nh \to \infty$ はバンド幅 h が 0 に収束していったとしても、その区間内には十分な大きさの観測が含まれていることを要求している。

Local Linearにも推定してみる

- Z_i はスカラー確率変数とする。
- 以下のような最小化問題

$$\min_{\{a_s(z),b_s(z)\}_{s=1}^{p}} \sum_{i=1}^{n} \left[Y_i - \sum_{s=1}^{p} \{a_s + b_s(Z_i - z)\} X_{is} \right]^2 K_h(Z_i, z)$$

の解 $\{(\hat{a}_s(z), \hat{b}_s(z)\}_{s=1}^{p}$ のうち、$\hat{a}_s(z)$ が $\hat{\beta}_s(z)$ に対応 $(s = 1, ..., p)$。

- 回帰係数の偏微分 $\frac{\partial \beta(z)}{\partial z}$ が知りたいかも

回帰係数とその偏微分を同時に推定

- $\delta(z) := (a_1(z), ..., a_p(z), b_1(z), ..., b_p(z))'$(推定したい)
- $\mathcal{X}$:i 行目が $(X_i', X_i'(Z_i - z))$ の $n \times 2p$ 行列
- $\mathcal{K} := diag\{K_h(Z_1 - z), ..., K_h(Z_n - z)\}$
- $\mathcal{Y} := (Y_1, ..., Y_n)$
- $\delta(z)$ の推定量は WLS の形で得られる:

$$\hat{\delta}(z) = (\mathcal{X}'\mathcal{K}\mathcal{X})^{-1}\mathcal{X}'\mathcal{K}\mathcal{Y}$$

- $\hat{a}_s(z)$ が $\beta_s(z)$ の、$\hat{b}_s(z)$ が $\frac{\partial \beta_s(z)}{\partial z}$ の推定量となっている。

$\beta(.)$ ってホントにノンパラメトリック?

- ここまで、β がノンパラメトリックに推定することを考えてきた。
- 難しく考えすぎ?
- もしかしたら $\beta(z)$ は z に依存しないかも:$\beta(z) = \beta_0$
- パラメトリックに表現できるかも:
  $\beta(z) = \beta_0(z, \gamma)$ ($\beta_0(.)$ は既知の関数、γ は有限次元パラメータ)

推定量$\hat{\gamma}$を元に検定

- $\hat{\gamma}$をγの$\sqrt{n}$-consistentな推定量とする。
- Local Constant 推定量$\hat{\beta}(z)$をパラメトリックな推定量$\hat{\beta}_0(z) := \beta_0(z, \hat{\gamma})$で代用できないか検定する。
- ノンパラメトリックな推定量$\hat{\beta}(z)$とパラメトリックな推定量$\hat{\beta}_0(z)$のズレを以下のI_nにより評価:

$$I_n := \int \left\{ [\hat{\beta}(z) - \hat{\beta}_0(z)]' A_n [\hat{\beta}(z) - \hat{\beta}_0(z)] \right\} dz$$

A_nは正定値行列。

ランダムな分母問題に対処

- Local Constant な推定量 $\hat{\beta}(z)$ はランダムな分母

$$D_n(z) := \frac{1}{n}\sum_{j=1}^{n} X_j X_j' K\left(\frac{Z_j - z}{h}\right)$$

を持つ。

- このままだとランダムな分母問題が起きてしまうので、$A_n = D_n(z)' D_n(z)$ としてしまおう。

ランダムな分母問題を解決1

- $A_n = D_n(z)'D_n(z)$ とした場合、I_n は以下のようになる。

$$
\begin{aligned}
I_n =& \int \{D_n(z)[\hat{\beta}(z) - \hat{\beta}_0(z)]\}' D_n(z)[\hat{\beta}(z) - \hat{\beta}_0(z)] dz \\
=& \frac{1}{n^2} \int \sum_i \{X_i Y_i K_{h,Z_i,z} - X_i X_i' \hat{\beta}_0(z) K_{h,Z_i,z}\}' \\
& \times \sum_j \{X_j Y_j K_{h,Z_j,z} - X_j X_j' \hat{\beta}_0(z) K_{h,Z_j,z}\} dz \\
=& \frac{1}{n^2} \sum_i \sum_j \int \{X_i [Y_i - X_i' \hat{\beta}_0(z)] K_{h,Z_i,z}\}' \\
& \times \{X_j [Y_j - X_j' \hat{\beta}_0(z)] K_{h,Z_j,z}\} dz
\end{aligned}
$$

ランダムな分母問題を解決2

- $\hat{u}_i \coloneqq Y_i - X_i'\hat{\beta}_0(Z_i)$

$$\bar{k}(v) \coloneqq \int k(u)k(v-u)du$$

$$\bar{K}_{h,Z_i,Z_j} \coloneqq \prod_{s=1}^{q} \frac{1}{h_s}\bar{k}\left(\frac{Z_{is}-Z_{js}}{h_s}\right)$$

- すると I_n は以下のように近似できる:

$$I_n \approx \frac{1}{n^2}\sum_i \sum_j X_i'X_j\hat{u}_i\hat{u}_j\bar{K}_{h,Z_i,Z_j}$$

ちょっとラクできるかも

- たたみこみ積分は 2 次カーネル K_2 で代用可能
- $i = j$ の項は無視できる
  (帰無仮説の下で検定統計量が 0 に近い値を取りやすくなるため
  むしろ無視した方がよい?)

検定統計量とその分布

- 以上より、以下のような検定統計量を用いるのがよさそう

$$\hat{I}_n := \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i} X_i' X_j \hat{u}_i \hat{u}_j K_2$$

- 実際、帰無仮説の下で以下が成り立つ(らしい)

$$\frac{n(h_1 ... h_q)^{\frac{1}{2}} \hat{I}_n}{\sqrt{\hat{\sigma}_0^2}} \xrightarrow{d} N(0,1)$$

$$\hat{\sigma}_0 := \frac{2h_1 ... h_q}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}(X_i' X_j)^2 \hat{u}_i^2 \hat{u}_j^2 K_2^2$$

部分線形モデルをさらに拡張

- 部分線形モデルをさらに拡張した以下のようなモデルを考える。

$$Y_i = W_i'\gamma + X_i'\beta(Z_i) + u_i$$

  説明変数が (W_i, X_i, Z_i) に 3 分割されていることに注意
- このようなモデルを**部分線形なめらかな係数モデル (Partially Linear Smooth Coefficient Model)** という。
- $\beta(.)$ をなめらか係数関数 (smooth coefficient function) という。

γをわかったフリ

- とりあえずγは既知(そんなことはない)としてギロンを進める。
- $Y_i - W_i'\gamma$を新たな被説明変数として以下のようなモデルを考える。

$$Y_i - W_i'\gamma = X_i'\beta(Z_i) + u_i$$

通常のなめらか係数モデルに帰着!

Local Constantに推定

- $\beta(z)$ の Local Constant 推定量は以下のようになる:

$$\hat{\beta}(z) = \left[\sum_j X_j X_j' K_{h,Z_j,z}\right]^{-1} \sum_j X_j [Y_j - W_j'\gamma] K_{h,Z_j,z}$$

推定量 $\hat{\beta}(z)$ を分解

- $\hat{\beta}(z)$ は以下のように分解できる:

$$\hat{\beta}(z) = \hat{\beta}_y(z) - \hat{\beta}_w(z)'\gamma$$

$$\hat{\beta}_y(z) := \left[\sum_j X_j X_j' K_{h,Z_j,z}\right]^{-1} \sum_j X_j Y_j K_{h,Z_j,z}$$

$$\hat{\beta}_w(z) := \left[\sum_j X_j X_j' K_{h,Z_j,z}\right]^{-1} \sum_j X_j W_j K_{h,Z_j,z}$$

(お目当ての?)γの推定へ

- 分解した推定量 $\hat{\beta}(Z_i) = \hat{\beta}_y(Z_i) - \hat{\beta}_w(Z_i)'\gamma$ を用いるとモデルは以下のように書き換えることができる:

$$Y_i - X_i'\hat{\beta}_y(Z_i) = (W_i - X_i'\hat{\beta}_w(Z_i)')\gamma + v_i$$

- $Y_i - X_i'\hat{\beta}_y(Z_i)$ の $W_i - X_i'\hat{\beta}_w(Z_i)'$ への回帰と見ると、OLS により γ が推定できる。