

npregress intro - ノンパラメトリック回帰の概要 【 評価版 】

ノンパラメトリック回帰は共変量を与えられたときのアウトカム変数の平均値をその関数形に関する前提を置かずにモデル化します。これによってノンパラメトリック回帰は関数形の指定に関する誤りに対して頑健な推定手法を提供するものとなります。npregress は系列回帰 (series regression) とカーネル回帰 (kernel regression) という 2 種類のノンパラメトリック回帰推定法を実装しています。

ノンパラメトリック系列回帰推定法の場合、アウトカム変数は共変量の関数として回帰されます。共変量の関数は基底関数 (basis function) と呼ばれます。基底とは平滑な関数を自在に良く近似できる項 (terms) の集合のことを言います。基底関数にはこれらの項のサブセットが含まれます。npregress series によって用いられる基底としては多項式、スプライン、B-スプラインがあります。

ノンパラメトリックカーネル回帰推定法の場合、アウトカム変数の加重平均が計算されます。その場合の重みはカーネルと呼ばれる関数です。npregress kernel は局所的線形 (local-linear) と局所的定数 (local-constant) という 2 種類の回帰推定法を実装しています。

アウトカム変数の平均値を系列回帰によって近似しようがカーネル回帰によって近似しようが、いずれの場合にもその推定結果は関数形状に関する仮定に対してロバストな (頑健な) ものとなります。しかしそれには代償 — 数多くの観測データと多くの計算量 — が伴います。

1. ノンパラメトリック回帰
2. ノンパラメトリック系列回帰
 - 2.1 Runge の現象
 - 2.2 スプラインと B-スプライン
3. ノンパラメトリックカーネル回帰
4. ノンパラメトリックな手法の限界

1. ノンパラメトリック回帰

ノンパラメトリック回帰は共変量が与えられたときのアウトカム変数の平均値がどのような関数形に従うかについて不確かな場合に使用されます。例えば線形回帰を用いて推定を行う場合、アウトカム変数の平均値の関数形としては共変量の線形結合であることを仮定します。パラメトリック回帰にしろノンパラメトリック回帰にしろ、それらは共変量の種々の値に対する平均値を推定します。今、図1に示されるような擬似的なデータについて考えてみましょう。 x のすべての値に対するアウトカム変数の平均値を表す曲線が散布図上に示されています。



図1

図1に示されたデータの平均値は x に関し線形ではないため、線形回帰を行ったとするなら、それはアウトカム変数に対する共変量 x の効果を正確にグラフ化したものとは言えなくなります。実際、このデータに対し線形回帰をフィットさせると図2のようなプロットがもたらされることになります。

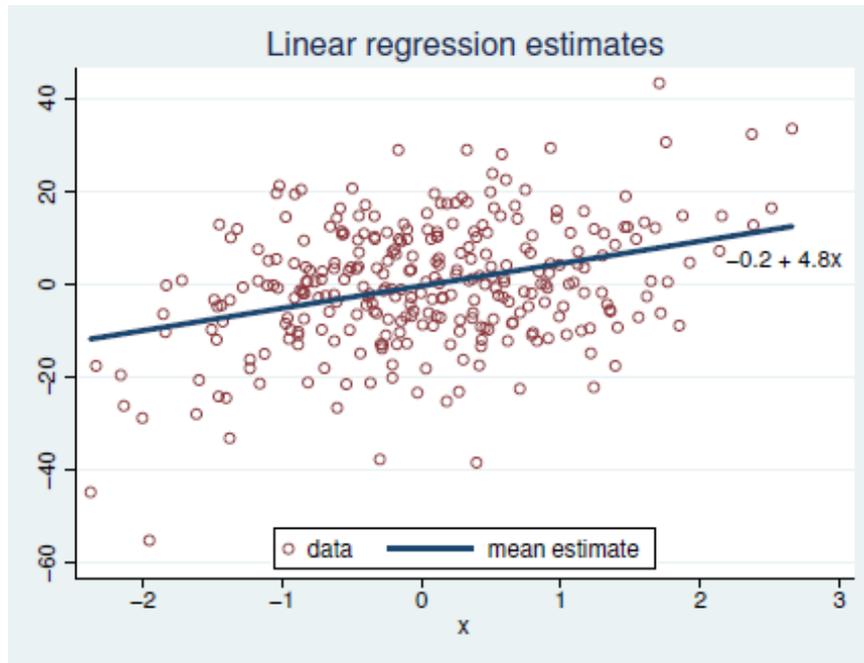


図 2

x が変化したときの応答変数予測値の変化は正の値で一定であるわけですが、真の平均値は非線形な形で変化しています。平均値の関数形に関する仮定が正しくなかった場合、得られる推定値は不整合なものとなります。これに対し `nptest` を用いてモデルをフィットさせ推定結果をグラフ化すると図 3 のようなプロットが得られます。

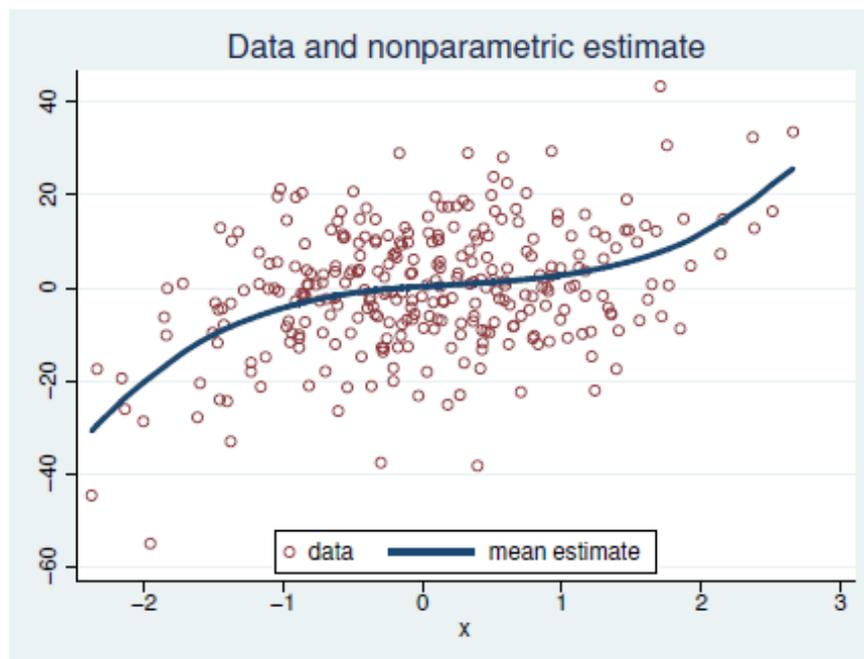


図 3

nmpregress はアウトカム変数と共変量間の関係を正しく表現できています。ノンパラメトリック回帰による推定結果は真の関数が十分滑らかなものである限り整合性が取れたものとなります。線形回帰の前提が正しかった場合であってもノンパラメトリック回帰は整合性のある結果をもたらしますが、効率はより低いものとなります。

ノンパラメトリック回帰は関数形の指定に関する誤りに対して頑健な推定手法を提供するものですが、より多くの観測データと計算時間を必要とするという代償を伴うものでもあります。この代償コストは共変量の数と共に増加します — 次元の呪い (curse of dimensionality) —。

2. ノンパラメトリック系列回帰

評価版では割愛しています。

2.1 Runge の現象

評価版では割愛しています。

2.2 スプラインと B-スプライン

評価版では割愛しています。

3. ノンパラメトリックカーネル回帰

評価版では割愛しています。

4. ノンパラメトリックな手法の限界

評価版では割愛しています。

