
• •



جامعة قاصدي مرباح ورقلة



كلية الرياضيات و علوم المادة

قسم الرياضيات

ماستر

مسار: رياضيات وإعلام آلي

فرع: رياضيات

تخصص: احصاء واحتمالات

من إعداد الطالب : مؤذن احمد

الموضوع

التقدير التكراري لدالة الربط للنموذج الشبه المعلي

تناقش يوم 2018/06/05 من طرف لجنة المناقشة :

مشرفا	جامعة قاصدي مرباح ورقلة	استاذ	بهدي عيسى
مناقشا	جامعة قاصدي مرباح ورقلة	استاذ	عقون رشيد
رئيسا	جامعة قاصدي مرباح ورقلة	استاذ	مزايبة محمد الهادي



شكر و عرفان

الحمد لله الذي علمنا ما لم نكن نعلم ، و نشكرك اللهم ما أعطيتنا من النعم .
أولا وقبل كل شيء نشكر الله عزوجل الذي وفقني في إنهاء هذا العمل المتواضع ،
كما أتقدم بخالص عبارات الشكر و العرفان والتقدير إلى كل من ساعدوني في إنجاز هذا
البحث المتواضع و نخص بالذكر الأستاذ المشرف: (بهدي عيسى) حفظه الله
الذي لم يبخل علي بعلمه و نصائحه و توجهاته القيمة للإتمام هذا العمل و أتمنى له المزيد من
النجاح و التوفيق في حياته العملية و العلمية
وأسأهل بتوجيه أعمق عبارات الشكر و العرفان الى جميع أساتذتي الكرام الذين أشرفوا على
تكويني طيلة مشواري الجامعي
ولا يفوتني بتوجيه الشكر والتقدير الى أساتذة اللجنة المناقشة : الاستاذ عقون رشيد و
الاستاذ مزايبة محمد الهادي و اتمني لهما و لجميع اساتذة كلية الطب و الرياضيات و عمال
الادارة و كل الطلبة التوفيق و النجاح



إهداء

بسم الله الرحمن الرحيم

و الصلاة و السلام على أشرف المرسلين سيدنا محمد صلى الله عليه و سلم تسليما كثيرا
أهدي هذا العمل المتواضع إلى امي التي سهرت الليالي في تربيتي و التي اختارت لي ماهو
احسن لي في حياتي و هو العلم و التي طالما انتظرت هذه اللحظة -التخرج - و التي كانت
بدعواتها الصالحة في كل صلاتها تتمنى لي كل التوفيق و النجاح و التسهيل ,

حفظها الله و بارك في ايام عمرها

و الى ابي الذي ضحى بنفسه من اجلي و تحمل مشاق الحياة من أجل أن أتعلم و أصل إلى
ما واصلت إليه , حفظه الله و بارك الله في ايام عمره و جزاه الله كل خير
دون أن أنسى من تقاسمت معهم حلاوة الحياة و مرارتها و من كانوا لي عوناً على غدر
الزمان و الدنيا الى اخوتي ** عبد الله . * . فاطمة . * . عبد الجليل . * . عبد الوهاب و

اسرته الصغيرة . * . و عبد الكريم و اسرته الصغيرة **

الى كل عائلة مؤدب و براكو

و الى كل زملائي و اصدقائي اللذين كانوا معي في الحياة الدراسية و الى صديقي , * ,
دحمان سعيد , * , و ابو القاسم سنكيي , * , شباشي مسعود , هم نعم الاصدقاء و هبهم لي

الله اتمني من الله ان يفتح عليهم ابواب الخير و على جميع المسلمين

الفهرس

3	1 استعراض المراجع
4	1.1 استعراض المراجع
4	1.1.1 التقدير التكراري للنماذج شبه المطبة
4	2.1.1 تقدير متكرر في بعض نماذج التشوه
5	3.1.1 التقديرات الوظيفية التكرارية وتطبيقاتها للتنبؤ
5	4.1.1 تقدير تكراري في الانحدار المعاكس بإشرايح (SIR)
6	2.1 اخلاصة
7	2 تعاريف، طريقة SIR، مقدرات
8	1.2 نماذج الانحدار
8	1.1.2 مفهوم الانحدار
8	2.1.2 نموذج المطبي
8	3.1.2 نموذج اللامطبي
9	4.1.2 النموذج شبه المطبي
9	2.2 نموذج شبه المطبي والانحدار العكسي
11	3.2 التكرار العكسي بالشريحة طريقة SIR
11	1.3.2 خوارزمية عامة لطريقة SIR
12	4.2 تقدير كثافة الاحتمال
12	1.4.2 الافتراض وبناء المقدر
13	2.4.2 التقدير باستخدام طريقة التواء
15	5.2 تقدير دالة الانحدار
16	1.5.2 مقدر Nadaraya-Watson
19	2.5.2 تقدير Nadaraya-Watson التكراري
23	3 النهج المتكرر نتائج SIR المقاربة و محاكاة
24	1.3 التقدير التكراري للمطبة
24	1.1.3 طريقة الانحدار العكسي المتقطع SIR
30	2.3 نتائج مقارب
31	3.3 بعض نتائج المحاكاة



33	نتائج المقاربة	4.3
33	1.4.3 نموذج ومقدرات	
34	2.4.3 القرضيات	
35	النظريات	5.3
35	1.5.3 قانون الأعداد الكبرى	
36	2.5.3 نظرية الحدود المركبة	
36	6.3 بعض نتائج المحاكاة	

مقدمة

طرق الانحدار مفيدة لنمذجة الربط بين متغير مفسر y و متغير تفسيري x وهي تستخدم في كثير من مجالات العلوم التطبيقية مثل الاقتصاد و الهندسة الزراعية و العلوم البيئية . في الادب الاحصائي هناك فئتان رئيسيتان من نماذج الانحدار, نموذج معلمي و نموذج غير معلمي و تتميز هذه النماذج بمزايا و عيوب, وقد نجد نموذج شبه معلمي الذي يربط بين النموذجين المعلمي و الغير معلمي, يتم التعريف بالنموذج المعلمي واللا معلمي و النموذج الشبه المعلمي في الجزء الثاني و تم تقديم المقدرات اللامعلمية للكثافة و دالة الربط. نعطي أولاً بناء مقدر نواة كلاسيكية (مقدرة Parzen-Rosenblatt) لكثافة احتمالية بعد ذلك نقدم المقدر التكراري لكثافة الاحتمال قدها Wolverson- فاغر ، ثم من Deheuvels و تقديم بعض نتائج التقارب من هذا المقدر, ثم ننتقل إلى تقدير Nadaraya-Watson التكراري وكذلك إلى التقدير العكسي ل Révész وأخيراً نذكر الخصائص التقاربية الرئيسية لهؤلاء المقدرين.

في الجزء الثالث تم تقديم الطريقة التي قدمها Li¹ (1991) التي تدعي بالانحدار العكسي للشرائح التكرارية (SIR) حيث يقوم بتقدير الجزء المعلمي θ من نموذج الشبه المعلمي دون الحاجة إلى تقدير دالة الربط f . بعد الحصول على تقدير θ ، نقترح إجراء تكراري لتقدير دالة الربط f بتقدير Nadaraya-Watson التكراري ثم نقترح طريقة للجمع بين التقديرات المتكررة لدالة الربط f من قبل مقدر Nadaraya-Watson التكرارية و تقدير المعلمة θ عبر مقدر SIR التكرارية. ثم نؤسس قانون الأعداد الكبيرة وكذلك نظرية الحد المركزي.²

الاشكالية

كيف يتم تقدير دالة الربط في نموذج شبه المعلمي بواسطة المقدر التكراري ؟

من هذه الاشكالية نستخلص الاسئلة التالية

- 1- ما هي الطريقة لتقدير الجزء المعلمي θ من نموذج شبه المعلمي ؟
- 2- كيف يتم تقدير دالة الربط $\hat{f}_n(x)$ بواسطة مقدر Nadaraya-Watson ؟

¹Li, K. C. .(1991) Sliced inverse regression for dimension reduction, with discussion. Journal of the American Statistical Association, 86, 316-342

²المقدمة من اطروحة ملخصة Thi Mong Ngoc NGUYEN

الفرضيات

1- تسمى طريقة SIR التكرارية المسماة بالانحدار الانعكاسي المقسم او المتقطع التكراري, حيث تقوم هذه الطريقة بتقدير الجزء المعلمي θ من النموذج شبه المعلمي دون الحاجة إلى تقدير دالة الربط f

2- يتم تقدير دالة الربط $\hat{f}_n(x)$ بتقدير مقدر Nadaraya-Watson انطلاقا من مقدر النواة التكرارية ل Parzen-Rosenblatt (دالة الربط f يمكن تقديره عن طريق الطريقة اللامعلمية).

الهدف

الهدف من هذه المذكرة هو وضع طريقة لتقدير دالة الربط f للنموذج شبه المعلمي $y = f(\theta x, \varepsilon)$ و تطبيق طريقة الانحدار العكسي المتقطع التكراري SIR التي قدمها Li (1991) لتقدير المعلمة θ و مقارنة بعض مقدرات ودراسة تقاربها .

الفصل الأول

استعراض المراجع

قائمة المحتويات

4	1.1 استعراض المراجع
6	2.1 الخلاصة



1.1 استعراض المراجع

1.1.1 التقدير التكراري للنماذج شبه المعلمة

• المؤلف : Thi Mong Ngoc Nguyen

• عنوان المرجع : التقدير التكراري للنماذج شبه المعلمة

Estimation réursive pour les modèles semi-paramétriques

• مفاتيح البحث : تقدير تكراري ، نموذج شبه المعلمي ، طريقة الانحدار.

• الملخص: اهتم الباحث في الأطروحة بدراسة النموذج الشبه المعلمي، $y = f(x\theta \varepsilon)$ و مشاكل تقدير المعلمات و دالة الربط بالطرق التكرارية، حسب طريقة Li(1991) التي تسمى بالانحدار العكسي المتقطع SIR و اهتم ايضا بالمقدرات التكرارية لدالة الربط دالة Nadaraya Watson التكرارية ثم اقترح طريقة للجمع بين التقدير المتكرر لدالة الربط f من قبل مقدر Nadaraya Watson التكراري وتقدير المعلمات ب SIR ثم أسس قانون الأعداد الكبيرة loi des grands nombres وكذلك نظرية الحد المركزي. théorème de limite centrale.

2.1.1 تقدير متكرر في بعض نماذج التشوه

• المؤلف : Philippe Fraysse

• عنوان المرجع : تقدير متكرر في بعض نماذج التشوه

Estimation réursive dans certains modèles de déformation

• مفاتيح البحث : تقدير تكراري ، نماذج شبه بارامترية ، خوارزميات عشوائية ، مقدرات النواة ، مارتينجال.

• الملخص: قام الباحث في هذه الأطروحة بدراسة بعض نماذج التشوه شبه المعلمي

$$y = f(X_n - \theta) + \varepsilon_n$$

وهدفه هو اقتراح طرق متكررة ، مستمدة من خوارزميات العشوائية اعتمادا على خوارزميات the Kiefer-Woflowitz، the Robbins-Monro لتقدير معالم هذه النماذج، و قام بتقديم طرق النوة لتقدير دالة الكثافة أو الانحدار ثم ركز على اثنين من أكثر مقدرات النواة شيوعاً ، وهما مقدر بارزين روزينبلات ومقدر Nadaraya-Watson ، من خلال تقديم النسخ المتكررة من هذين المقدرين.

قام بتقدير شبه تكراري لمعلمة الترجمة ودالة الانحدار لنموذج الترجمة. والحالة التي تكون فيها دالة الربط دورية.

3.1.1 تقديرات الدوال التكرارية وتطبيقاتها للتنبؤ

• المؤلف : Aboubacar Amiri

• عنوان المرجع : تقديرات الدوال المتكررة وتطبيقاتها للتنبؤ
Estimateurs fonctionnels récurrente et leurs applications à la prévision

- : مفاتيح البحث تقدير تكراري ، نموذج شبه المعلي ، طريقة الانحدار
- الملخص: في هذه الأطروحة ،اهتم الباحث بأساليب تقدير النواة التكرارية من الشكل

$$r_n^l = \frac{Q_n^l}{f_n^l} \quad Q_n^l = \frac{1}{\sum_{i=1}^n h_i^{(1-l)d}} \sum_{i=1}^n \frac{m(Y_i)}{h_i^{ld}} K \left(\frac{x - X_i}{h_i} \right)$$

وتطبيقاتها للتنبؤ حيث $l = 0$ او $l = 1$ و قام بتقديم معايير مقارنة تعتمد على التحيز التقاربي ، والتباين ، وانحطاً التريبي ، ثم حدد عائلة من مقدري الكثافة، عائلة من التقديرات المتكررة في عكس دالة الانحدار ثم درس خصائصه التقارب وفقاً للمعلمة، و استخدم النتائج التي تم الحصول عليها لتقدير الانحدار لبناء التنبؤ غير المعلي لكل نواة.

4.1.1 تقدير تكراري في الانحدار المعاكس بشرائح (SIR)

• المؤلف : Thi Mong Ngoc Nguyen et Jérôme Saracco

• عنوان المرجع : تقدير تكراري في الانحدار المعاكس بشرائح (SIR)
Estimation récurrente en régression inverse par)tranches sliced inverse regression)

- مفاتيح البحث : تقدير تكراري ، نموذج شبه بارامتري ، انحدار العكسي شرائح (SIR)
- الملخص: ركز الباحث على طريقة SIR (شرائح الانحدار العكسي، التي يمكن أن تترجم من قبل شرائح الانحدار عكسية) لتقدير θ معلمة في نموذج شبه المعلي $y = f(x\theta, \varepsilon)$ دون الحاجة إلى تقدير الدالة f أو لتحديد قانون الخطأ ε حيث قام بدراستها وفق $H = 2$ شرائح، اقترح في هذه المقالة صيغة متكررة لهذا المقدر. و قام بتقديم خصائص المقاربة لهذا المقدر (sûre et normalité asymptotique) (convergence presque



2.1 الخلاصة

باستخدام مفاتيح البحث ، نستنتج أن هناك تشابهاً في البيانات الأساسية لكل باحث ، على الرغم من الاختلاف في العناوين التي تمت دراستها ، التي هي التقدير التكراري للنماذج شبه المعلمية للمؤلف Nguyen Thi Mong Ngoc ، والتقدير المتكراري في نماذج التشوه من تأليف الكاتب Fraysse Philippe ، والمقدرات التكرارية وتطبيقاتها للتنبؤ من قبل المؤلف Amiri Aboubacar ، و تقدير تكراري في الانحدار المعاكس بشرائح (SIR) للكاتبان Th Mong Ngoc Nguyen et Jérôme Saracco ، تقتصر هذه المذكرات او البحوث كلها في حيز المفاتيح التالية ، التقدير التكراري ، نموذج شبه المعلمي ، طريقة الانحدار ، و التي تنسجم مع موضوع البحث .

الفصل الثاني

تعريف, طريقة SIR , مقدرات

قائمة المحتويات

8 نماذج الانحدار	1.2
9 نموذج شبه المعلي والانحدار العكسي	2.2
11 التكرار العكسي بالشريحة طريقة SIR	3.2
12 تقدير كثافة الاحتمال	4.2
15 تقدير دالة الانحدار	5.2

1.2 نماذج الانحدار

تعتبر أساليب الانحدار مفيدة في نمذجة الربط بين متغير مفسر y ومتغير تفسيري x . وهي تستخدم في العديد من مجالات العلوم التطبيقية مثل الاقتصاد ، أو الهندسة الزراعية ، أو الإحصاء الحيوي ، أو القياسات الحيوية ، أو العلوم البيئية ، توجد فئتان رئيسيتان من نماذج الانحدار هما نماذج معلمية ونماذج غير معلمية يتميز كلا النوعين من النماذج بمزايا وعيوب محددة، حيث يوجد نموذج شبه معلمية يجمع بين النموذج المعلمي و اللامعلمي .

1.1.2 مفهوم الانحدار

يقصد بالانحدار Regression تقدير العلاقة بين متغيرات أحدها تابع ومتغير أو أكثر من المتغيرات المستقلة في صورة جبرية باستخدام البيانات المتوفرة عن المتغيرات بصياغة معادلة أو منحني يفسر هذه العلاقة. وتعرف المعادلة المفسرة للعلاقة بمعادلة الانحدار.

2.1.2 نموذج المعلمي

بالنسبة لنموذج المعلمي ، فإن دالة الربط بين متغير مفسر $y \in \mathbb{R}$ ومتغير تفسيري $x \in \mathbb{R}^p$ ، يعتمد على عدد محدد من المعلمات لتقدير. ويصاغ النموذج على الشكل التالي :

$$y = f_{\theta}(x) + \varepsilon$$

دالة الربط f_{θ} تنتمي إلى مجموعة من الدوال ذات معلمات θ متجه للمعلمات الحقيقية و ε هو مصطلح خطأ عشوائي في هذا النموذج ، الهدف هو تقدير المعلمة θ ، وتكون تقنيات تقدير المعلمات بالطرق التالية (أقصى احتمال maximum de vraisemblance المربعات الصغرى moindres carrés ...) فعالة عندما يتم تحديد دالة الربط المرتبطة بالمعلمة f_{θ} بشكل صحيح و تسمح هذه النماذج بتفسير تأثير المتغير التفسيري على المتغير المراد تفسيره.

3.1.2 نموذج اللامعلمي

تظهر النماذج اللامعلمية كبديل يوفر الاستمرارية المرغوبة في النمذجة متغير مفسر y مرتبط بالمتغير التفسيري x بواسطة دالة الربط غير معروفة يجب تقديرها:

$$y = f(x) + \varepsilon$$

الموضوع المشترك للتراجع الوظيفي لتقدير دالة الربط $f(x)$ هو فكرة التجانس المحلي وتعتمد جودتها على وجود بيانات كافية بقيمة x , عندما يكون المتغير التفسيري أحادي البعد، تتوفر العديد من تقنيات التقدير: طريقة النواة، وخطوط التنعيم lissage ولكن عندما يصبح حجم x كبيراً، فإن عدد الملاحظات اللازمة للتنعيم المحلي ينمو بشكل كبير مع هذا البعد.

4.1.2 النموذج الشبه المعلمي

النموذج الشبه المعلمي الذي اقترحه Li و Duan (1991). المتغير مفسر y هي قيمة في \mathbb{R} والمتغير التفسيري x الذي ينتمي الى \mathbb{R}^p وهذا النموذج يكتب في الشكل التالي:

$$y = f(\theta x, \varepsilon)$$

حيث:

- المعلمة θ هي متجهة غير معروفة في \mathbb{R}^p ,
- الضجيج ε bruit le عبارة عن خطأ عشوائي غير مستقل عن x لا توجد افتراضات حول توزيع ε ,
- الدالة f هي دالة ذات قيمة في \mathbb{R} غير معروفة.

تم تطوير نماذج شبه المعلمية للجمع بين مزايا الأساليب المعلمية وغير المعلمية، وهي القدرة على تفسير النماذج المعلمية ومرونة النماذج اللامعلمية. في مثل هذه النماذج، يعتمد المتغير مفسر y عادةً على x من خلال عدد محدود من المعلمات الاقليدية $\theta'_1 \dots \theta'_k$ و دالة الربط f .

2.2 نموذج شبه المعلمي والانحدار العكسي

في نموذج شبه المعلمي من الانحدار المقترح من قبل لي¹ (1991):

$$y = f(\theta'_1 x, \dots, \theta'_k x) \quad (1.2)$$

- y هو متغير أحادي للمفسر
- يعتبر x متغيراً عشوائياً تفسيري من البعد p بتوقع $\mu = \mathbb{E}$ و مصفوفة التباين $\Sigma = \mathbb{V}$ معرفة موجبة.
- $\theta'_1 \dots \theta'_k$ هي متجهات حقيقية ذات البعد p ، غير معروفة ومستقلة خطياً، $k < p$

¹ Li, K. C. (1991) Sliced inverse regression for dimension reduction, with discussion. Journal of the American Statistical Association, 86, 316-342

- ε عبارة عن خطأ عشوائي غير مستقل عن x ، ولا يوجد افتراض حول توزيع ε
- f هي الدالة المعرفة على \mathbb{R} غير معروفة .

وبما أن قيمة K أقل من p ، فإن هذا النموذج هو أيضاً نموذج يسمح بتخفيض البعد الفضائي للمتغيرات التفسيرية.

تسمى المصطلحات $(\theta'_k x)$ مؤشرات $k = 1 \dots K$ ، بالنسبة إلى $K > 1$

تعريف 1

نوظف الفضاء E على EDR ("تخفيض البعد التناسبي") . بالإضافة إلى ذلك ، أي مجموعة خطية من (θ_k) تنتمي إلى E ، وتسمى اتجاه EDR . الهدف الرئيسي الذي وضعناه في هذا النموذج هو تقدير الجزء المعلمي ، أي تقدير قاعدة فضاء EDR . E يمكن بعد ذلك الحصول على الجزء الوظيفي لدالة الربط f بالطرق الالعملية المعتادة.

كتابة نموذج آخر نفرض وجود علاقة بين x و z حيث :

$$z = \Sigma^{-1/2}(x - \mu)$$

يمكننا بعد ذلك إعادة كتابة النموذج (2.1) في النموذج:

$$y = g(\eta'_1 z, \dots, \eta'_k z, \varepsilon), \quad (2.2)$$

حيث $E^s = Vect(N)$ و $N = [\eta_1, \dots, \eta_k]$ لدينا ، $\eta_k = \Sigma^{1/2}\theta_k$ ، $k = 1, \dots, K$

تعريف 2 يُسمى E^s مساحة EDR القياسية ، ويطلق أي اتجاه يخص E^s على اتجاه EDR "تخفيض البعد التناسبي" القياسي.

ملاحظه 1 نقوم بمطابقة الدالة g مع الدالة f بالقرب من الوسيطات K الأولى ، نجد

$$\begin{aligned} g(\eta'_1 z, \dots, \eta'_k z, \varepsilon) &= g(\eta'_1 \Sigma^{-1/2}(x - \mu), \dots, \eta'_k \Sigma^{-1/2}(x - \mu), \varepsilon) \\ &= g(\theta'_1 x - \theta'_1 \mu, \dots, \theta'_k x - \theta'_k \mu, \varepsilon) \end{aligned}$$

3.2 التكرار العكسي بالشريحة طريقة SIR

1.3.2 خوارزمية عامة لطريقة SIR

في الممارسة العملية ، لدينا عينة $\{(x_i, y_i), i = 1 \dots n\}$ من المتغيرات العشوائية المستقلة ذات توزيعات متماثلة من النموذج المرجعي, نحن نريد أن نقدر قاعدة من الفضاء EDR ثم نقوم بتقدير المصفوفة Γ_T^S او Γ_T^S نقوم باختيار الدالة $T(y)$ بحيث يسهل تقدير هذه المصفوفة, وبالتالي ، فإن التكرار العكسي في الشرائح يأخذ دالة T تقديرا ل y في الشرائح H المميزة: $s_1 \dots s_H$, يسمى هذا التحول بالتقسيم. في هذا الإطار ، تم كتابة التوقعات الشرطية $\mathbb{E}[z|T(y)]$ كالتالي :

$$m_h^s = \mathbb{E}[z|y \in s_h] \quad h = 1 \dots H$$

نلاحظ ان $p_h = Pr(y \in s_h)$ والمصفوفة Γ_T^S تكون في الشكل التالي :

$$\Gamma_T^S = \sum_{h=1}^H m_h^s m_h^{s'}$$

و منه خوارزمية SIR تكون على النحو التالي:

• المرحلة 1 توحيد x_i في z_i عن طريق التحول

$$z_i = \hat{\Sigma}^{-1/2}(x_i - \bar{x}), \quad i = 1 \dots n$$

حيث \bar{x} و $\hat{\Sigma}$ هما مصفوفة التباين و التغير التجريبية والتوقع التجريبي لعينة x_i : $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ و

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$$

• المرحلة 2 تقسيم دعائم y_i الى شرائح H , $s_1 \dots s_H$, وحساب النسبة التجريبية من y_i تنتمي الى شريحة s_h :

$$\hat{p}_h = \frac{1}{n} \sum_{i=1}^n I_{[y_i \in s_i]}, \quad h = 1 \dots H$$

• المرحلة 3 حساب التوقعات التجريبية لكل z_i

$$\hat{m}_h^s = \frac{1}{n \hat{p}_h} \sum_{i=1}^n z_i I_{[y_i \in s_i]}, \quad h = 1 \dots H$$

• المرحلة 4 بناء تقدير لمصفوفة التباين و التغير الخاصة بمتوسطات الشريحة:

$$\hat{\Gamma}_T^S = \sum_{h=1}^H \hat{p}_h \hat{m}_h^s \hat{m}_h^{s'}$$

حيث نستخرج القيم الذاتية لهذه المصفوفة

نلاحظ \hat{v}_k^s حيث $k = 1 \dots K$ و القيم K هي القيم الذاتية الاكبر من القيم الذاتية للمصفوفة $\hat{\Gamma}_T^S$

• المرحلة 5 حساب قيم K المباشر EDR المقدرة:

$$\hat{v}_k = \hat{\Sigma}^{-1/2} \hat{v}_k^s \quad k = 1 \dots K$$

4.2 تقدير كثافة الاحتمال

نفترض أن لدينا تسلسل (X_n) للمتغيرات العشوائية المستقلة و نفس قانون الكثافة لدالة الربط f

1.4.2 الافتراض وبناء المقدر

لدينا الدالة F المرتبطة بالتوزيع f ، ثم بالقانون القوي للأعداد الكبيرة loi forte des grands nombres ، فإن المقدار $\hat{F}_n(x)$ ، لكل $x \in \mathbb{R}$ ،

$$\hat{F}_n(x) = \frac{1}{n} \sum_{k=1}^n I_{\{X_n \leq x\}} \quad (3.2)$$

تتقارب بالتأكيد ، إلى $F(x)$ علاوة على ذلك ، ان

$$F(x) = \int_{-\infty}^x f(t) dt$$

لأن $h > 0$ صغير بما فيه الكفاية ، ومن الفرق التفاضلي يعطي بالشكل التالي

$$\frac{F(x+h) - F(x-h)}{2h}$$

ومنه تقدير $\hat{f}_n(x)$ يكتب في الشكل :

$$\begin{aligned} \hat{f}_n(x) &= \frac{\hat{F}_n(x+h) - \hat{F}_n(x-h)}{2h} \\ &= \frac{1}{2h} \sum_{k=1}^n I_{\{x-h < X_n < x+h\}} \end{aligned}$$

هو مقدر ثابت من دالة الربط $f(x)$.

يسمى المقدر $\hat{f}_n(x)$ بمقدّر Rosenblatt. يمكن كتابتها على النحو التالي:

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K_u \left(\frac{X_i - x}{h_n} \right) \quad (4.2)$$

$K_u(\cdot)$ النواة الموحدة Unifforme.

و منه مقدر Rosenblatt هو مقدر Parzen-Rosenblatt الذي يكتب في الشكل التالي :

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right)$$

بالإضافة إلى ذلك ، تُسمى المعلمة h نافذة التنعيم lissage للمقدّر، في كل ما يلي ، سننظر في نافذة تنعيم h بناءً على n و ملاحظة h_n وسنفترض أن h_n يميل إلى 0 عندما يميل n إلى اللانهاية، لاحظ أن \hat{f}_n هي كثافة احتمالية.

2.4.2 التقدير باستخدام طريقة النواة

تقدير Parzen-Rosenblatt

لتكن (X_n) سلسلة من المتغيرات العشوائية المستقلة التي تتبع قانون كثافة الاحتمالات g . نريد تقدير g من عينة (X_1, X_2, \dots, X_n) .

ولتكن h_n متتالية موجبة تماماً حيث $h_n \xrightarrow{n \rightarrow \infty} 0$ و $nh_n \xrightarrow{n \rightarrow \infty} \infty$ دالة موجبة ومحدودة تسمى النواة حيث :

$$\int_{\mathbb{R}} k(x) dx = 1, \int_{\mathbb{R}} x k(x) dx = 0 \text{ et } \int_{\mathbb{R}} k^2(x) dx = \mu^2$$

يمكننا تقدير g من مقدر Parzen-Rosenblatt

$$\hat{g}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right)$$

لتكن

$$G_n = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i \leq x\}}$$

و منه G_n تُعطى بواسطة:

$$\begin{aligned}\hat{g}_n(x) &= \frac{1}{h_n} \left(G_n\left(x + \frac{h_n}{2}\right) - G_n\left(x - \frac{h_n}{2}\right) \right) \\ &= \frac{1}{nh_n} \sum_{i=1}^n \mathbb{I}_{\left\{x - \frac{h_n}{2} \leq X_i \leq x + \frac{h_n}{2}\right\}} \\ &= \frac{1}{nh_n} \sum_{i=1}^n \mathbb{I}_{\left\{\frac{1}{2} \leq \frac{X_i - x}{h_n} \leq \frac{1}{2}\right\}} \\ &= \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right)\end{aligned}$$

حيث $K(\cdot)$ هو النواة الوحدة. Uniforme.

في ظل الشروط المعطاة على h_n و K ، أظهر Rosenblatt (1971) أن $\tilde{g}_n(x)$ يتقارب في المتوسط التربيعي إلى $g(x)$ حيث:

$$\text{Var}(\hat{g}_n(x)) \sim \frac{\tau^2}{nh_n} g(x) \text{ pour } n \rightarrow \infty$$

وبالنظر إلى الحالة المتعددة للمتغير من خلال قيم $p \in \mathbb{N}$ ، و النواة k هي كثافة على \mathbb{R}^p ، $\forall x \in \mathbb{R}^p$

$$\hat{g}_n(x) = \frac{1}{nh_n^p} \sum_{i=1}^n K\left(\frac{X_i - x}{h_i}\right)$$

تقدير النواة التكرارية

يتم إعطاء تقدير النواة التكرارية من خلال العلاقة التراجعية التالية:

$$\hat{g}_n(x) = (1 - \gamma_n)g_{n-1}(x) + \frac{\gamma_n}{h_n} \sum_{i=1}^n K\left(\frac{x - X_n}{h_n}\right)$$

h_n و γ_n هما متاليتين موجبتان تؤلان إلى 0.

مقدر Wolverson-Wagner، وهو يتوافق مع الخيار $\gamma_n = 1/n$ ويكتب:

$$\hat{g}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_i} K\left(\frac{X_i - x}{h_i}\right)$$

تقدير Deheuvels

هناك غيرها من مقدرات النواة لتقدير كثافة الاحتمال. يمكننا أن نقبس بشكل ملحوظ مقدر Deheuvels لكل $n \geq 1$ و $x \in \mathbb{R}$ لدينا ²

$$\tilde{g}_n(x) = \frac{1}{S_n} \sum_{i=1}^n K \left(\frac{X_i - x}{h_i} \right)$$

$$S_n = \sum_{i=1}^n h_i$$

أظهر Deheuvels أن مقدره $\tilde{g}_n(x)$ يتقارب في المتوسط التربيعي إلى $g(x)$ بالإضافة إلى تقاربه شبه المؤكد وتقارب شبه طبيعي. *normalité asymptotique*.

نظريه 1 اذا كانت f قابلة للتفاضل و تفاضلها محدود و النافذة المختارة h_n حيث $h_n = n^{-\alpha}$ و تكون $0 < \alpha < 1$ اذا لكل $x \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} \tilde{g}_n(x) = g(x) \quad p,s$$

اذا كانت $1 < \alpha < 1/5$

$$\sqrt{nh_n} (\tilde{g}_n(x) - g(x)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mu^2 g(x))$$

اختيار النواة والنافذة

يمكن إجراء هذه الخيارات فقط من خلال استخدام بعض النوافذ h وتبين أن اختيار النواة ليس له تأثير كبير إذا تم اختياره في فئة معقولة. من ناحية أخرى، يعد اختيار نافذة h_n أمراً حاسماً. بشكل عام، يتم الحصول على h_n من خلال تقنيات التحقق المتقاطع. ³

5.2 تقدير دالة الانحدار

نفترض أن لدينا أزواجاً عشوائية (x_n, y_n) مرتبطة بعلاقة:

$$y_n = f(x_n) + \varepsilon_n$$

² Paul Deheuvels : Sur l'estimation séquentielle de la densité. C. R. Acad. Sci. Paris Sér. A-B, 276:A1119-A1121, 1973.

³ تقدر كثافة الاحتمال من اطروحة Fraysse Philippe 36-41

تسلسل المتغيرات العشوائية (ε_n) هو استمرار للمتغيرات العشوائية المستقلة المتمركزة و تتبع نفس القانون وحيث دالة الانحدار f غير معروفة ، ثم نحدد مقدر f كجموع مرجح لسلسلة الملاحظات y_n

$$\hat{f}_n(x) = \sum_{k=1}^n W_{n.k}(x)y_k \quad (2.6.1)$$

الأوزان $W_{nk}(x)$ تعتمد فقط على nk و x و $(x_1 \dots x_n)$ حيث $x \in \mathbb{R}$

$$\sum_{k=1}^n W_{n.k}(x) = 1$$

1.5.2 مقدر Nadaraya-Watson

أما بالنسبة لتقدير معظم كثافة احتمال للانحدار الالعملية ، فهي \hat{f}_n مقدر ل Nadaraya-Watson⁵⁴ المعروف لجميع $n \geq 1$ ومن أجل $x \in \mathbb{R}$:

$$\hat{f}_n(x) = \begin{cases} \frac{\sum_{k=1}^n K\left(\frac{X_k-x}{h_n}\right)Y_k}{\sum_{k=1}^n K\left(\frac{X_k-x}{h_n}\right)} & \text{si } \sum_{k=1}^n K\left(\frac{X_k-x}{h_n}\right) \neq 0 \\ 0 & \text{sinon} \end{cases}$$

يتوافق المقدر Nadaraya-Watson مع اختيار

$$W_{n.k} = \frac{K\left(\frac{X_k-x}{h_n}\right)}{\sum_{k=1}^n K\left(\frac{X_k-x}{h_n}\right)} \mathbb{I}_{\{\sum_{k=1}^n K\left(\frac{X_k-x}{h_n}\right) \neq 0\}}$$

في المعادلة (2.6.1). نلاحظ ان قيمة X متغيراً عشوائياً من نفس القانون ل (X_n) و Y متغير عشوائي لقانوني (Y_n) ونفترض أن الزوج (X, Y) له كثافة $g(x, y)$ بالنسبة لقياس Lebesgue على \mathbb{R}^2 ، ثم يتم إعطاء الكثافة الهامشية X ل densité marginale ، لكل $x \in \mathbb{R}$ ، بواسطة

$$g(x) = \int_{\mathbb{R}} g(x, y)dy > 0$$

⁴E. A. Nadaraja : On a regression estimate. Teor. Verojatnost. i Primenen., -9:157

⁵Geoffrey S. Watson : Smooth regression analysis. Sankhya Ser. A, ,26:359-372 ,1964

يمكننا بعد ذلك كتابته على الشكل التالي ، لكل $x \in \mathbb{R}$ حيث $g(x) \neq 0$ نجد

$$f(x) = \mathbb{E}[Y|X = x] = \frac{\int_{\mathbb{R}} yg(x, y)dy}{\int_{\mathbb{R}} g(x, y)dy} = \frac{\int_{\mathbb{R}} yg(x, y)dy}{g(x)}$$

الاتساق و طبيعة التقارب

من أجل الحصول على التقارب الخاص لمقدر Nadaraya-Watson ، نستخدم مرة أخرى تحليل التباين المتحيز الذي يحدث عند تقدير كثافة الاحتمال ، في اطروحة Collomb.⁶ بخصوص الحالة الطبيعية المقاربة لمقدار Nadaraya-Watson ، فإن النتائج الأولى ترجع إلى Schuster⁷ .
ومنه نتحصل على مقدر لNW، عن طريق التحيز التباين التالي :

$$\mathbb{E} \left[(\hat{f}_n(x) - f(x))^2 \right] = \mathbb{V} \left(\hat{f}_n(x) \right) + \left(\mathbb{E}[\hat{f}_n(x)] - f(x) \right)^2$$

حساب التباين

$$\sigma^2 = \mathbb{V}(Y|X = x) = \frac{1}{g(x)} \int y^2 g_{X,Y}(x, y) dy - (f(x))^2.$$

قضيه 1 من المفترض أن $\mathbb{E}[Y^2] < \infty$ في كل نقطة $x, g(x), f(x), \sigma^2(x)$ حيث $g(x) > 0$ لدينا :

$$\mathbb{V} \left(\hat{f}_n(x) \right) = \frac{1}{nh_n} \times \left\{ \frac{\sigma^2(x)}{g(x)} \int_{\mathbb{R}} K^2(u) du \right\} (1 - o(1))$$

هذا المصطلح $o(1)$ يؤل الى الصفر الى $n \rightarrow \infty$

حساب التحيز

يكون المقدر NW في شكل حاصل عشوائي ، لذلك يتم استخدام التقريب التالي بشكل عام كمصطلح مركزي:

$$\tilde{\mathbb{E}}[\hat{f}_n(x)] = \frac{\mathbb{E}[\hat{r}_n(x)]}{\mathbb{E}[\hat{g}_n(x)]}$$

⁶ Gérard Collomb : Quelques propriétés de la méthode du noyau pour l'estimation non paramétrique de la régression en un point fixé. C. R. Acad. Sci. Paris Sér. A-B, 285(4):A289-A292, 1977

⁷ Eugene F. Schuster : Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. Ann. Math. Statist., 43:84-88, 1972

هذه الصيغة أسهل في التعامل معها وتسمح على وجه الخصوص بخط الانحراف

$$\tilde{d}_n = \hat{f}_n(x) - \tilde{\mathbb{E}}[\hat{f}_n(x)]$$

من أجل تبرير اختيار المصطلح الوسطي ، أظهر Nadaraya (1989) أنه عندما يحد Y و $nh_n \rightarrow \infty$

$$\mathbb{E}[\hat{f}_n(x)] = \tilde{\mathbb{E}}[\hat{f}_n(x)] + O((nh_n)^{-1})$$

عندما $nh_n^2 \rightarrow \infty$ و $\mathbb{E}(Y^2) < \infty$

$$\mathbb{E}[\hat{f}_n(x)] = \tilde{\mathbb{E}}[\hat{f}_n(x)] + O((n^{1/2}h_n)^{-1})$$

قضيه 2 نفترض ان f و g من الصنف $C^q(\mathbb{R})$ و النواة K رتبها p حيث :

$$\int_{\mathbb{R}} \mathbf{K}(z) dz = 1, \int_{\mathbb{R}} z^j \mathbf{K}(z) dz = 0, j = 1 \dots q-1, \int_{\mathbb{R}} z^q \mathbf{K}(z) dz < \infty$$

بعد ذلك نجد $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$

$$\mathbb{E}[\hat{f}_n(x)] - f(x) = \frac{h_n^q}{2} \left(\left\{ f^{(q)}(x) + 2f^{(q-1)}(x) \frac{g^{(q-1)}(x)}{g(x)} \right\} \int_{\mathbb{R}} z^q \mathbf{K}(z) dz \right) (1 + O(1)).$$

طبيعية شبه مقارب

نظريه 2 نفترض أن Y يكون بعدها اكبر من 2 و تُفترض الدالتين f و g قابلتين لتفاضل مرتين ومستمرتين في الفضاء او المجال \mathbb{E} عند كل نقطة x من التباين σ^2 حيث $g(x) > 0$ لدينا :

$$\sqrt{nh_n} \left(\hat{f}_n(x) - f(x) \right) \xrightarrow{\mathcal{L}} \mathcal{N}(B(x), v^2(x))$$

و

$$v^2(x) = \frac{\tau^2 \sigma^2(x)}{g(x)}$$

و

$$B(x) = \left\{ f''(x) + 2f'(x) \frac{g'(x)}{g(x)} \right\} \times \int_{\mathbb{R}} u^2 \mathbf{K}(u) du$$

.

نظريه 3 لتكن $0 < \alpha < 1$ و $h_n = n^{-\alpha}$, و نفرض ان f محدودة وقابلة لتفاضل مرتين و نفرض x معرف بالكثافة g المحدودة والقابلة للتفاضل مرتين و النواة K محققة

$$\int_{-\infty}^{+\infty} k^2(x)d(x) = \mu^2 < +\infty, \quad \frac{1}{2} \int_{-\infty}^{+\infty} x^2 k(x)d(x) = \nu^2 < +\infty$$

إذا كان الضجيج bruit le مربع متكامل والتباين $\sigma^2 > 0$ اذا لكل قيمة $x \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} \hat{f}_n(x) = f(x) \quad p,s$$

اذا كانت $1/5 < \alpha < 1$ اذا لكل قيمة $x \in \mathbb{R}$ حيث $g(x) \neq 0$ نستنتج

$$\sqrt{nh_n} \left(\hat{f}_n(x) - f(x) \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\sigma^2 \mu^2}{g(x)} \right)$$

2.5.2 تقدير Nadaraya-Watson التكراري

كما هو الحال بالنسبة لمقدار Parzen-Rosenblatt, إذا كانت الأزواج (X_n, Y_n) متتابعة, فن الأفضل استخدام نسخة متكررة من مقدر Nadaraya-Watson⁸. هذا المقدر التكراري غير محدد, لكل $n > 1$ ولجميع $x \in \mathbb{R}$

$$\hat{f}_n(x) = \begin{cases} \frac{\sum_{k=1}^n \frac{1}{h_k} K\left(\frac{X_k - x}{h_n}\right) Y_k}{\sum_{k=1}^n \frac{1}{h_k} K\left(\frac{X_k - x}{h_n}\right)} & \text{si } \sum_{k=1}^n \frac{1}{h_k} K\left(\frac{X_k - x}{h_n}\right) \neq 0 \\ 0 & \text{si } \sum_{k=1}^n \frac{1}{h_k} K\left(\frac{X_k - x}{h_n}\right) = 0 \end{cases}$$

نظريه 4 لتكن $0 < \alpha < 1$ و $h_n = n^{-\alpha}$ نفرض ان f محدودة وقابلة للتفاضل مرتين و تفاضلها محدود ايضا و نفرض ان X معرفة بالكثافة g و تكون g محدودة وقابلة لتفاضل مرتين و تفاضلها محدود و النواة k تحقق:

$$\int_{-\infty}^{+\infty} k^2(x)d(x) = \mu^2 < +\infty, \quad \text{et} \quad \frac{1}{2} \int_{-\infty}^{+\infty} x^2 k(x)d(x) = \nu^2 < +\infty$$

إذا كان الضجيج bruit le مربع متكامل والتباين $\sigma^2 > 0$ اذن لكل $x \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} \hat{f}_n(x) = f(x) \quad p,s$$

اذا كانت $1/3 < \alpha < 1$ و (ε) معرف بالترتيب اكبر من 2 فان لكل $x \in \mathbb{R}$ حيث تكون $g(x) \neq 0$

$$\sqrt{nh_n} \left(\hat{f}_n(x) - f(x) \right) \xrightarrow{\mathcal{L}} N \left(0, \frac{\sigma^2 \mu^2}{g(x)(1 + \alpha)} \right)$$

⁸E. A. Nadaraja : On a regression estimate. Teor. Verоятnost. i Primenen., -9:157 ,159 .1964

تقدير Révész

يوجد مقدر آخر متكرر لدالة الانحدار f الذي يدعي بمقدر Révész⁹ لقد تم استخدام هذا المقدر في القليل من الطروحات على الرغم من السلوك المقارب المفيد. ومع ذلك، يمكننا أن نذكر أطروحة Slaoui و al et Mokkadem¹⁰ اللذان قاما بدراسة مقدر Révész في أطروحتهم، يتم التعريف بهذا المقدر ب

مهما يكن $x \in \mathbb{R}$

$$\check{f}_n(x) = \check{f}_{n-1}(x) + \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{X_i - x}{h_i}\right) (Y_n - \check{f}_{n-1}(x))$$

درس Revész الخصائص المقاربة لهذا المقدر للنافذة $h_n = n^{-\alpha}$ وتحصل على النتائج التالية :¹¹

نظريه 5 مهما تكن $1/2 < \alpha < 1$ ولكل $x \in \mathbb{R}$ حيث $2g(x) > 1 - \alpha$ لدينا :

$$\lim_{n \rightarrow \infty} \hat{f}_n(x) = f(x) \quad p,s$$

و

$$\sqrt{nh_n} \left(\check{f}_n(x) - f(x) \right) \xrightarrow{\mathcal{L}} N \left(0, \frac{\sigma^2 \mu^2 g(x)}{2g(x) - (1 - \alpha)} \right)$$

مقارنة بين المقدرين ، (NW) Nadaraya-Watson ، (NWR) التكراري (NWR) و (R) Révész

تمت مقارنة بين NW ، NWR و R عن طريق التجربة التي أجريت على الحيوان المائي (المحار) حيث تم تثبيته بالمستقبلات لتقييم استمراريته ونوعية المياه التي يعيش فيها بتوفير الوسط الحيوي لهذا الحيوان كالهواء و الطعام. الهدف هو وصف نشاط الصمام وقياس السلوك الحركي لهذه الرخويات وفقا لمعلبات وسلوكيات اللانمطية المرتبطة بتلوث المياه من خلال عوامل الطحالب أو الأنشطة البشرية. استخدام Valvometry كنظام مراقبة بيولوجية لجودة البيئة و مراقبة التأثيرات المحتملة للتأثيرات الجديدة ، من خلال اقتراح الاحتياطات غير الحدودية ومقارنة ثلاث وظائف غير بارامترية أو متكررة و تم هذا العمل بشكل روتيني للحصول على البيانات

لدينا عينة تتكون من أزواج مستقلة من المتغيرات العشوائية $(T_1 Y_1) \dots (T_n Y_n)$ وننظر في نماذج الطبقات اللامعملية في $i = 1 \dots n$

$$Y_i = f(T_i) + \varepsilon$$

⁹P. Révész : How to apply the method of stochastic approximation in the nonparametric estimation of a regression function. Math. Operationsforsch. Statist. Ser. Statist., 119-126:(1)8 .1977

¹⁰Abdelkader Mokkadem, Mariane Pelletier et Yousri Slaoui : Revisiting Révész's stochastic approximation method for the estimation of a regression function. ALEA Lat. Am. J. Probab. Math. Stat., 6:63-114 .2009

¹¹تقدير دالة الانحدار, من اطروحة Frayse Philippe 41-44

ومن ثم تحصلو على وظائف السيرة الذاتية للمكونات الثلاثة بعد الحصول على البيانات من المختبر وتم تصميمها ومعالجتها إحصائياً، يتم اختيار هذه المعلمة كحل وسط بين الاختلاف والتحيز التقديري ، أخيراً ، نحصل على التدفقات التالية للتحقق من صحة التقدير. وتم استخدام كمييار أسلوب التحقق المتقاطع والذي يتكون من تقليل دالة h

$$CV(h) = \sum_{i=1}^n (y_i - \hat{m}_{(-i)}(T_i; h))^2$$

$\hat{m}_{(-i)}(T_i; h)$ يعني المقدّر (NW) ، NWR أو R لدالة الربط عند النقطة T_i تحسب على عينة خاصة من الثنائية $(T_i; Y_i)$

نحدد عرض النوافذ باستخدام طريقة التحقق من الصحة (انظر شكل 1). ثم اخترنا النواة الغوسية لمصادر NW و NWR بالنسبة للمقدّر R ، توصي مقالة Révész (1977) باختيار النواة المنتظمة. الشكل 1 يمثل المعيار $CV(h) = CV(\alpha)$ من خلال طريقة التحقق من صحة التجربة ، نحن نرى بوضوح أن وظائف السيرة الذاتية التي تم الحصول عليها عددياً جميعها محدثة لمقدراتنا الثلاثة. النوافذ المقدره هي $h = n^{-\alpha_{cv}}$

α_{cv} لكل مقدر يساوي على التوالي $n^{-0.47}$ ل NW و $n^{-0.50}$ ل NWR و $n^{-0.99}$ ل R مع هذه الخيارات من النوافذ ، نحصل على توافق جيد جداً لنماذج الانحدار مع البيانات (انظر شكل 2).

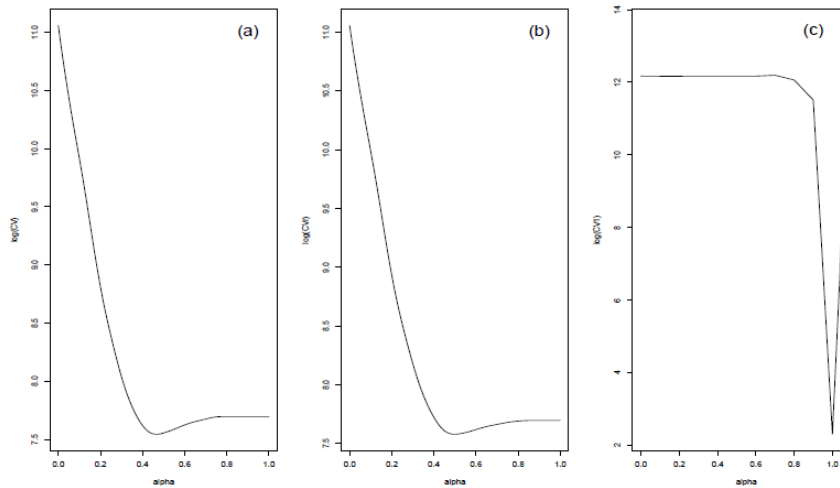
من أجل المقارنة بسهولة بين المقدرات الثلاثة ، نقدم الشكل 2 الذي يمثل بقايا المقدر. نحن نرى بوضوح أن مقيم Rezez أفضل بكثير من الآخرين .

مقارنة التقديرات الثلاثة من حيث معيار التباين ، حيث نجد ان :

$$\mathbb{V}(R) < \mathbb{V}(NWR) < \mathbb{V}(NW) (p < 0.05)$$

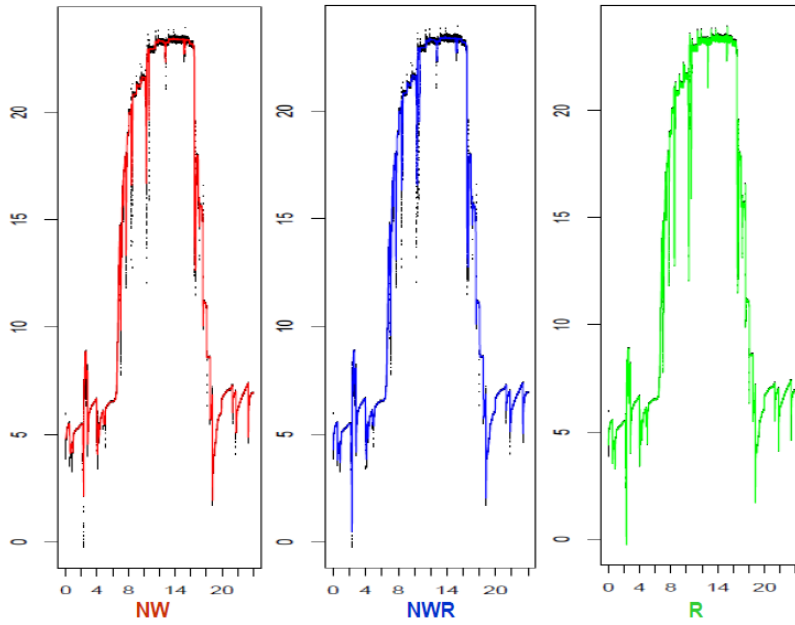
وبالتالي ، تشير هذه النتيجة إلى أنه يجب علينا بدلاً من ذلك استخدام مقيم متكرر للنوع NWR أو R بمعنى الاختلاف المقارب الأدنى. ومع ذلك ، فإن وقت الحساب لمقدار R أكبر.

12

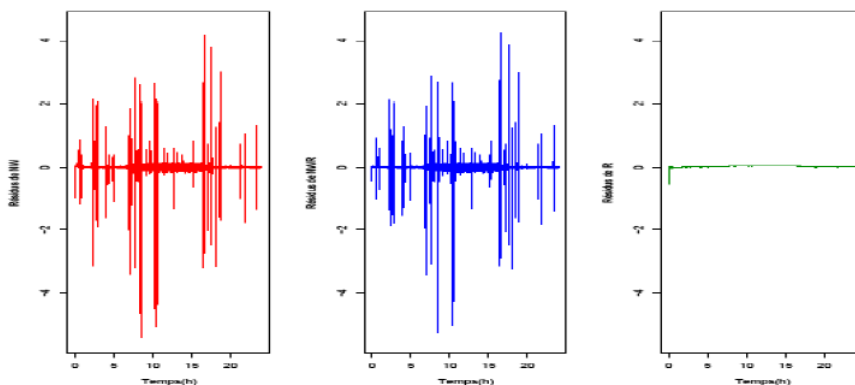


(a) : $\alpha_{cv} = 0 : 47NW$, (b) : $\alpha_{cv} = 0 : 5 NWR$, (c) : $\alpha_{cv} = 0 : 99R$

الشكل 1: تقدير أجمام النوافذ .



الشكل 2: الكفاءة اليومية المقدمة للقييمات الثلاثة



الشكل 3. اخطاء المقدرات

الفصل الثالث

النهج المتكرر نتائج, SIR المقاربة و محاكاة

قائمة المحتويات

24	التقدير التكراري للمعلمة θ	1.3
30	نتائج مقارب	2.3
31	بعض نتائج المحاكاة	3.3
33	نتائج المقاربة	4.3
35	النظريات	5.3
36	بعض نتائج المحاكاة	6.3

النهج المتكرر لطريقة SIR

وقد وضعت نماذج شبه المعلمي لدمج مزايا المعلمي وغير المعلمي ، وهي القدرة على تفسير النماذج المعلمية ومرورنة النماذج غير المعلمية. وتعتمد هذه النماذج عموماً على معلمة البعد المحدودة ، المشار إليها θ ، وكذلك دالة الربط f المراد تقديرها.

في هذا الفصل ، تم دراسة نموذج بنموذج الشبه المعلمة الذي اقترحه دوان ولي (1991). المتغير مفسر y هي قيمة في \mathbb{R} والمتغير تفسيري x الذي ينتمي الى \mathbb{R}^p وهذا النموذج يكتب في الشكل التالي :

$$y = f(\theta x, \varepsilon) \quad (3.1)$$

حيث :

- المعلمة θ هو الناقل غير معروف في \mathbb{R}^p .
- الضجيج ε bruit le عبارة عن خطأ عشوائي غير مستقل عن x لا توجد افتراضات حول توزيع ε .
- دالة الربط f هي دالة ذات قيمة في \mathbb{R} غير معروفة.

1.3 التقدير التكراري للمعلمة θ

نقدم أولاً باختصار طريقة SIR ونذكر تعريف مقدر اتجاهات EDR (لأي عدد من الشرائح) كحل لمشكلة مع القيم الذاتية للمصفوفة، ثم نحدد شكل تحليلي لهذا المقدر عندما $H = 2$. ثم نقترح في الخطوة الأولى شكل تكراري لمقدر المصفوفة . ثم نعطي نموذجاً متكرراً لمقدر اتجاه EDR عند $H = 2$.

1.1.3 طريقة الانحدار العكسي المتقطع SIR

الشريحة العكسية أو الانحدار SIR هي طريقة الانحدار شبه المعلمي على أساس حجة هندسية. تم تقديمه بواسطة Li¹ و Duan و Li² . تعتمد هذه الطريقة على خاصية دالة الانحدار العكسي (أي بدلاً من التراجع عن y على x ، فهي خصائص انحدار x على y التي سيتم استخدامها). وتمثل ميزة هذا الدور في عكس ذلك ، وابعاد هذه المشكلة قد تنخفض للحصول على النتيجة المطلوبة لدينا في الواقع مشكلات في البعد

¹Li, K. C. .(1991) Sliced inverse regression for dimension reduction, with dicussion. Journal of the American Statistical Association, ,86 .316-342

²Duan, N. and Li, K. C. .(1991) Slicing regression : a link-free regression method. The Annals of Statistics, ,19 .505-530

1 ، وهو الانحدار العكسي مما يجعل من الممكن التراجع عن كل إحداثي x على y و للحصول على نتيجة نقوم بإضافة افتراض احتمالي لتوزيع المتغير التفسيري x :

(H) المتغير التفسيري x له توزيع احتمالية غير متدنية بحيث ، لكل $b \in \mathbb{R}^p$ و التوقع الشرطي $\mathbb{E}[b'x|\theta'x]$ هو خطي في $\theta'x$ ،

هذه الفرضية ، وتسمى أيضا شرط الخطي ، يتم التحقق منها عند المتغير التفسيري x يتبع التوزيع الاهليلجي ، خاصة عندما يكون توزيع x متعدد الأعمدة.

نضع $\mu = \mathbb{E}[x]$ و $\Sigma = \mathbb{V}(x)$ ، نعطي الآن نظرية التوصيف لاتجاه θ الذي وضعه Li³

نظريه 6 في إطار النموذج (3,1) وتحت الفرضية (H) ، تركّز منحنى الانحدار العكسي $Y \rightarrow \mathbb{E}[x|\mathbb{T}(y)]$ ينتمي إلى الفضاء الفرعي الخطي ل \mathbb{R}^p مولدة من قبل θ و \mathbb{T} هو تحويل *monotone* ل y (والذي يتوافق مع "slicing" المحدد لاحقا).

النتيجة المباشرة لهذه النظرية هي أن مصفوفة-التباين والتغاير لهذا المنحنى $\Gamma = \mathbb{V}(\mathbb{E}[x|\mathbb{T}(y)])$ يتدهور في أي اتجاه Σ -متعامد ل θ و بالتالي فإن المتجه الصحيح $\tilde{\theta}$ المرتبطة والغير معدومة لقيم الذاتية للمصفوفة $\Sigma^{-1}\Gamma$ هو متحد (colinéaire) ل θ للقيمة الذاتية $\tilde{\theta}$ وهذا هو اتجاه EDR

ملاحظه 2 كجزء من تعميم النموذج (3,1) إلى نموذج مع عدة مؤشرات ، $\theta'_1 x, \dots, \theta'_d x$ حيث $d < p$ حسب لي (1991) القيم الذاتية $\tilde{\theta}'_1, \dots, \tilde{\theta}'_k$ ترتبط بقيم كبرى غير معدومة من المصفوفة $\Sigma^{-1}\Gamma$ التي تنتمي إلى الفضاء الجزئي الخطي المولدة للقيم $\theta'_1, \dots, \theta'_b$. نقول انها تنتمي إلى الفضاء EDR

لتطبيق العملية ، لدينا عينة من الملاحظات $L\{(x_i, y_i), i = 1 \dots N\}$ سوف نذكر الآن طريقة التقدير (غير التكرارية) كما قدمها Duan و Li .

نرمز ل \bar{x} و Σ_N المتوسط والتباين التجريبي للعينة $\{x_i, i = 1 \dots N\}$ على الترتيب حيث :

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{et} \quad \Sigma_N = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}_N)(x_i - \bar{x}_N)'$$

³Li, K. C. .(1991) Sliced inverse regression for dimension reduction, with discussion. Journal of the American Statistical Association, 86 .316-342

لتقدير المصفوفة Γ نختار التحويل \mathbb{T} بحيث يمكن تقدير هذه المصفوفة بسهولة. تم اقتراح خيار معين من \mathbb{T} من قبل Duan و Li : "التقسيم" الذي هو تقديرية ل y استناداً إلى تقسيم y إلى شرائح H المميزة حيث s_1, \dots, s_H , نرمل $p_h = \mathbb{P}(y \in s_h)$ و $m_h = \mathbb{E}[x|y \in s_h]$ و $h = 1 \dots H$ يتم كتابة المصفوفة Γ ببساطة شديدة في النموذج:

$$\Gamma = \sum_{h=1}^H P_h(m_h - \mu)(m_h - \mu)' \quad (3.2)$$

لا يشكل تقدير الكميات المختلفة المتضمنة في المصفوفة Γ أي صعوبة: فهو يكفي لاستبدال اللحظات التجريبية باللحظات النظرية. ثم نحصل على مقدر Γ_N حيث

$$\Gamma_N = \sum_{h=1}^H p_{h.N}(m_{h.N} - \tilde{x}_N)(m_{h.N} - \tilde{x}_N)'$$

و

$$p_{h.N} = \frac{1}{N} \sum_{i=1}^N \mathbb{I}[y_i \in s_h] = \frac{N_h}{N} \text{ et } m_{h.N} = \frac{1}{N_h} \sum_{i \in s_h} x_i$$

هي نسبة تجريبية من y_i لسقوط في شريحة s_h ، والمعيار التجريبي من X_I بحيث ينتمي y_i إلى شريحة s_h ثم نحدد القيم الذاتية الصحيحة $\tilde{\theta}_N$ المرتبطة بالقيمة ذاتية من المصفوفة $\Sigma_N^{-1} \Gamma_N$. نقول ان $\tilde{\theta}_N$ هو مقدر لاتجاه EDR.

تتوفر نتائج المقاربة لمراجع SIR غير المتكرر لاتجاه θ دون أن تكون شاملة ، يمكن الإشارة إلى أن سرعة التقارب لاتجاه EDR المقدرة نحو الاتجاه الصحيح هو $n^{1/2}$ وأن الحالة الطبيعية المقاربة للقيم الذاتية للاهتمام (جهاز عرض نظيف على مساحة EDR المقدرة للقيم الذاتية و القيم الذاتية الرئيسية ، أي مقدرات اتجاهات EDR) تم الحصول عليها من قبل Saracco ⁴ .

دراسة الحالة التي يسمح فيها $H = 2$ بالحصول على تعبير تحليلي للمقدر $\tilde{\theta}_N$

سنركز على الحالة الخاصة التي ننظر فيها إلى شريحتين فقط تمت الإشارة إليهما s_1 و s_2 السبب الأساسي لهذا الاختيار هو أنه يمكننا بسهولة الحصول على تعبير تحليلي لمقدر اتجاه θ في الواقع ، عندما $H = 2$ ، يتم كتابة

⁴Saracco, J. , (1997) An asymptotic theory for Sliced Inverse Regression. Communications in Statistics - Theory and methods, 26, 2141-2171

المصفوفة Γ :

$$\Gamma = p_1 z_1 z_1' + p_2 z_2 z_2' \quad (3.3)$$

$$z_h = m_h - \mu \quad h = 1, 2$$

يمكننا بعد ذلك إظهار القيمة الذاتية الغير معدومة λ للمصفوفة $\Sigma^{-1}\Gamma$ و القيمة الذاتية $\tilde{\theta}$ تم كتابة على الشكل التالي:

$$\lambda = \frac{p_1}{p_2} z_1' \Sigma^{-1} z_1 \quad \text{et} \quad \tilde{\theta} = \Sigma^{-1}(z_1 - z_2) \quad (3.4)$$

يمكننا بالتالي استخلاص من العلاقات (3,4) مقدرات λ_N و $\tilde{\theta}_N$ لكل λ و $\tilde{\theta}$

$$\lambda_N = \frac{p_{1.N}}{p_{2.N}} z_{1.N}' \Sigma_N^{-1} z_{1.N} \quad (3.5)$$

$$\tilde{\theta}_N = \Sigma_N^{-1}(z_{1.N} - z_{2.N})$$

$$z_{h.N} = m_{h.N} - \tilde{x}_N \quad \text{حيث} \quad h = 1, 2$$

في القسم الفرعي التالي ، سنقترح أولاً كتابة متكررة لمقدر مصفوفة d'intérêt

التقدير التكراري للمصفوفة $\Sigma^{-1}\Gamma$

عند النظر إلى الشكل غير التكراري للمقدرات ، لدينا عينة من الملاحظات $L = \{(x_i, y_i), i = 1 \dots N\}$ للمتغير العشوائي (x, y) المستقلة وموزعة بشكل متساوٍ للنموذج (3,1) ، بالنسبة للنموذج التكراري للمقدرات ، تنقسم هذه العينة إلى قسمين: العينة الفرعية $\{(x_i, y_i), i = 1 \dots N - 1\}$ والملاحظة (x_N, y_N)

سنقدم أشكال متكررة من Σ و \bar{x}_N و Σ_N^{-1} ثم Γ_N بعد ذلك سنصف الشكل التكراري لمصفوفة d'intérêt $\Sigma_N^{-1}\Gamma_N$

الشكل المتكرر لكل من Σ_N و \bar{x}_N و Σ_N^{-1}

الشكل التكراري للمتوسط التجريبي \bar{x}_N ل N من ملاحظات x_1, \dots, x_N تكتب في الشكل التالي :

$$\tilde{x}_N = \frac{1}{N} \sum_{i=1}^N x_i = \frac{N-1}{N} \tilde{x}_{N-1} + \frac{1}{N} x_N = \tilde{x}_{N-1} + \frac{1}{N} \Phi_N \quad (3.6)$$

$$\Phi_N = x_N - \tilde{x}_{N-1} \text{ حيث}$$

الشكل التكراري لمصفوفة الفروق التباينية التجريبية Σ_N تعطى في الشكل :

$$\begin{aligned} \Sigma_N &= \frac{1}{N} \sum_{i=1}^N (x_i - \tilde{x}_N)(x_i - \tilde{x}_N)' \\ &= \frac{N-1}{N} \Sigma_{N-1} + \frac{N-1}{N^2} (x_N - \tilde{x}_N)' \\ &= \frac{N-1}{N} \Sigma_{N-1} + \frac{N-1}{N^2} \Phi_N \Phi_N' \end{aligned} \quad (3.7)$$

يمكننا أن نظهر بالمثل ، عبر معادلة ريكاتي ، أن معكوس Σ_N له شكل متكرر:

$$\Sigma_N^{-1} = \frac{N}{N-1} \Sigma_{N-1}^{-1} - \frac{N}{(N-1)(N+p_N)} \Sigma_{N-1}^{-1} \Phi_N \Phi_N' \Sigma_{N-1}^{-1} \quad (3.8)$$

$$p_N = \Phi_N' \Sigma_{N-1}^{-1} \Phi_N \text{ حيث}$$

الشكل التكراري Γ_N

في الحالة العامة التي يكون فيها العدد H من الشرائح عشوائية $H=2$ ، من أجل إعطاء الشكل التكراري لمقدار Γ_N الخاص بالمصفوفة Γ ، نفترض أن الملاحظة الجديدة (x_N, y_N) هي أن y_N تنتمي إلى شريحة s_{h^*}

لاحظ أولاً أن شكل متكرر للمقدار $p_{h.N}$ من p_h

$$p_{h.N} = \begin{cases} \frac{N-1}{N} p_{h^*.N-1} + \frac{1}{N} & \text{si } h = h^* \\ \frac{N-1}{N} p_{h^*.N-1} & \text{sinon} \end{cases} \quad (3.9)$$

الشكل التكراري ل $m_{h.N}$ حيث $m_{h^*.N-1} = x_N - \Phi_{h^*.N}$ كالتالي :

$$m_{h.N} = \begin{cases} m_{h^*.N-1} + \frac{1}{N_{h^*.N-1}+1} \Phi_{h^*.N} & \text{si } h = h^* \\ m_{h.N-1} & \text{sinon} \end{cases} \quad (3.10)$$

من العلاقات (3.6) و (3.10) نحصل على شكل التكراري ل $z_{h.N} = m_{h.N} - \bar{x}_N$

$$z_{h.N} = \begin{cases} z_{h^*.N-1} - \frac{1}{N}\Phi_N + \frac{1}{N_{h^*.N-1}+1}\Phi_{h^*.N} & \text{si } h = h^* \\ z_{h^*.N-1} - \frac{1}{N}\Phi_N & \text{sinon} \end{cases} \quad (3.11)$$

بكتابة Γ_N في النموذج :

$$\Gamma_N = \sum_{h \neq h^*} p_{h.N} z_{h.N} z'_{h.N} + p_{h^*.N} z_{h^*.N} z'_{h^*.N}$$

يمكننا استنتاج من العلاقات (3.9) و (3.11) النموذج التكراري التالي ل Γ_N :

$$\begin{aligned} \Gamma_N &= \frac{N-1}{N}\Gamma_{N-1} - \frac{N-1}{N^2} \sum_{h=1}^H p_{N,N-1} (z_{h.N-1}\Phi'_N + \Phi_N z'_{h.N}) \\ &+ \frac{N-1}{N^3}\Phi_N\Phi'_N + \frac{N-1}{N}p_{h^*.N-1}A_{h^*.N} + \frac{1}{N}B_{h^*.N}B'_{h^*.N} \end{aligned} \quad (3.12)$$

حيث

$$A_{h^*.N} = \frac{1}{N_{h^*.N-1}}(z_{h.N-1}\Phi'_N + \Phi_N z'_{h.N}) - \frac{1}{N(N_{h^*.N-1}+1)}(\Phi_N\Phi'_{h^*.N} + \Phi_{h^*.N}\Phi'_N) + \frac{1}{N(N_{h^*.N-1}+1)^2}\Phi_{h^*.N}\Phi'_{h^*.N}$$

و

$$B_{h^*.N} = z_{h^*.N-1} - \frac{1}{N}\Phi_N + \frac{1}{N_{h^*.N-1}}\Phi_{h^*.N}$$

الشكل التكراري لمصفوفة الفائدة $\Sigma_N^{-1}\Gamma_N$

من العلاقات (3.8) و (3.12) . يمكننا الحصول على شكل التكراري ل $\Sigma_N^{-1}\Gamma_N$

(3.13)

$$\Sigma_N^{-1}\Gamma_N = \Sigma_{N-1}^{-1}\Gamma_{N-1} + \Sigma_{N-1}^{-1}C_{h^*.N} - \frac{1}{N+p_N}\Sigma_{N-1}^{-1}\Phi_N\Phi'_N\Sigma_{N-1}^{-1}[\Gamma_{N-1} + C_{h^*.N}]$$

و نضع

$$C_{h^*.N} = -\frac{1}{N} \sum_{h=1}^H p_{h.N-1} (z_{h.N-1}\Phi'_N + \Phi_N z'_{h.N-1}) + \frac{1}{N^2}\Phi_N\Phi'_N + p_{h^*.N-1}A_{h^*.N} + \frac{1}{N-1}B_{h^*.N}B'_{h^*.N}$$

التقدير التكراري المباشر ل θ

ننتقل الآن إلى الحالة حيث $H = 2$. في هذه الحالة ، قدمنا في (3,5) التعبير التحليلي للأشكال غير التكرارية للمقدرات. λ_N و $\tilde{\theta}_N$ سوف نستنتج منهم الآن أشكالهم التكرارية، ونحن نفترض هنا مرة أخرى أن الملاحظة الجديدة (x_n, y_n) حيث $y_n \in s_{h^*}$ بالاضافة الى h^* يساوي 1 او 2

 الشكل التكراري ل $\tilde{\theta}_N$

الشكل التكراري لتقدير $\tilde{\theta}_N$ ل $\tilde{\theta}$ من العلاقات (3.8) و (3.11) هو:

$$\begin{aligned} \tilde{\theta}_N = & \frac{N}{N-1} \tilde{\theta}_{N-1} - \frac{N}{(N-1)(N+p_N)} \Sigma_{N-1}^{-1} \Phi_N \Phi_N' \tilde{\theta}_{N-1} \\ & - \frac{(-1)^{h^*} N}{(N_{h^*.N-1} + 1)(N-1)} \left(\Sigma_{N-1}^{-1} - \frac{1}{N+p_N} \Sigma_{N-1}^{-1} \Phi_N \Phi_N' \Sigma_{N-1}^{-1} \right) \Phi_{h^*.N} \end{aligned} \quad (3.14)$$

 الشكل التكراري ل λ_N

الشكل التكراري لتقدير λ_N ل λ :

$$\lambda_N = \frac{p_{1.N-1} + \frac{1}{N-1} \mathbb{I}[h^* = 1]}{p_{2.N-1} + \frac{1}{N-1} \mathbb{I}[h^* = 2]} V_N' \left(\frac{N}{N-1} \Sigma_{N-1}^{-1} - \frac{N}{(N-1)(N+p_N)} \Sigma_{N-1}^{-1} \Phi_N \Phi_N' \Sigma_{N-1}^{-1} \right) V_N$$

حيث

$$V_N = z_{1.N-1} - \frac{1}{N} \Phi_N + \frac{1}{V_{1.N-1}} \Phi_{1.N} \mathbb{I}[h^* = 1]$$

5

2.3 نتائج مقارب

نحن نركز في هذا القسم على نتائج التقارب للمقدر $\tilde{\theta}_n$ من اتجاه EDR في حالة $H = 1, 2$. الافتراضات التالية و ضرورتها لتحديد شبه التقارب للمقدر المقترح.

(A1) الملاحظات $(x_i, y_i), i = 1, \dots, n$, يتم أخذ عينات منها بشكل مستقل عن النموذج

$$y = f(\theta'x, \varepsilon)$$

⁵ Thi Mong Ngoc Nguyen et Jérôme Saracco, Estimation récursive en régression inverse par tranches, لطريقة المتكرر النهج

(A2) نقوم ينقسم y إلى شريحتين ثابتتين s_1 و s_2 حيث $p_h \neq 0$ ل $h = 1, 2$.

نقدم هنا اثنين من نتائج التقارب , أول نقوم يحدد سرعة تقارب شبه مؤكد من $\tilde{\theta}_N$ والثاني الحالة الطبيعية لمقاربة هذا المقدر. سيكون من الممكن الحصول على نتائج تقارب مماثلة من λ_N إلى λ

باستخدام اللوغاريتم المتكرر ل Hartman-Wintner وتطبيق معادلة Riccati للمصفوفة العكسية , نظهر المبرهنة التالية.

نظريه 7 تحت الافتراضات (H) , (A1) و (A2) لدينا

$$\|\tilde{\theta}_n - \tilde{\theta}\| = O\left(\sqrt{\frac{\log(\log n)}{n}}\right) \quad p, s$$

نظريه 8 تحت الافتراضات (H) , (A1) و (A2) لدينا

$$\sqrt{N}(\tilde{\theta}_n - \tilde{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma^{-1} \Delta_3 \Sigma^{-1})$$

Δ_3 و البرهان على النظرية موجود في الملحق

3.3 بعض نتائج المحاكاة

المقارنة بين أوقات الحساب ل SIR الكلاسيكي , SIR المتكرر , و SIR غير التكراري $H = 1, 2$ شريحة

لقد تم قياس أوقات حساب المقدرات $\tilde{\theta}_n$ من اتجاه ل n انطلاقاً من N_0 الى غاية N لكل من SIR المتكرر (الذي يُسمى SIR التكراري) و SIR غير التكراري (يدعى classic SIR (على أساس $H = 2$ شرائح)). تحصلو على أن نهج SIR المتكرر أسرع بكثير في متوسط $(N - N_0)$ للمقدار $\tilde{\theta}_n$ مثل نهج SIR غير تكراري. أعطيت (3.5) تعبيراً تحليلياً عن الشكل غير التكراري للمقدّر $\tilde{\theta}_n$ (حيث نعتبر أن $H = 2$ شرائح). و تم قياس أوقات حساب المقدرات ل n من $\tilde{\theta}_n$ تتراوح من N_0 إلى N لهذا المقدر (تسمى SIR 2 شرائح). لمقارنة أوقات الحساب بين SIR الكلاسيكي , SIR التكرارية , ونهج SIR غير التكرارية (على أساس $H=1,2$ شرائح) ,

$P = 40$	$P = 20$	$p = 5$		
(0.014) 4.10	(0.014) 1.97	(0.008) 0.99	SIR كلاسيكي	N=300
(0.009) 1.93	(0.010) 1.05	(0.004) 0.42	SIR 2 شرائح	
(0.004) 0.17	(0.001) 0.13	(0.001) 0.11	SIR تكراري	
(0.020) 9.81	(0.013) 4.57	(0.012) 2.11	SIR كلاسيكي	N=600
(0.009) 5.10	(0.012) 2.64	(0.009) 0.95	SIR 2 شرائح	
(0.006) 0.35	(0.001) 0.26	(0.001) 0.22	SIR تكراري	
(0.051) 16.76	(0.019) 7.66	(0.019) 3.35	SIR كلاسيكي	N=900
(0.043) 9.39	(0.005) 4.71	(0.032) 1.63	IR 2 شرائح	
(0.009) 0.54	(0.007) 0.39	(0.002) 0.33	SIR تكراري	

الجدول 2.1: حساب الوقت بالتواني من المقدرات $\tilde{\theta}_n$ لاتجاه θ (من أجل θ من N_0 إلى N) باستخدام طريقة SIR التقليدية وبنهج SIR التكراري أو غير التكراري: الوسائل والانحرافات - بين الأشكال بين القوسين المحسوبة على $B = 500$ نسخة طبق الأصل من العينات مع قيم مختلفة من p

و منه نجد أن نهج الشريحة 2 SIR غير المتكرر هو أسرع مرتين من نهج SIR التقليدي. يكون نهج SIR التكراري أكثر من 4 مرات أسرع من 2 شرائح SIR ، عندما $p = 5$. كلما ازداد حجم العينة ، يبدو أن زمن الحساب يزيد بشكل متناسب لطريقة التكرارية SIR وأكثر نسبياً لنهج SIR غير التكراري (SIR) كلاسيكي و 2 شرائح SIR).

وبالتالي ، تم استنتاج ميزة كبيرة من هذا النهج المتكررة مقارنة سرعتها مع SIR غير تكراري وأساليب SIR التقليدية. هذه الميزة أكثر أهمية لأن البعد p ل x كبير. لاحظ أيضاً أنه مع زيادة حجم العينة ، يبدو أن الوقت الحسابي يزداد بشكل متناسب لطريقة التكرارية SIR وأكثر من ذلك بشكل نسبي لمقاربة SIR ونهج SIR التقليدي غير التكراري ⁶.

⁶ نتائج المحاكات ، من اطروحة Thi Mong Ngoc Nguyen صفحة 30-31

4.3 نتائج المقاربة

1.4.3 نموذج ومقدرات

لدينا عينة (X_n) تتكون من متجهات عشوائية من \mathbb{R}^p المستقلة و قانون كثافة الاحتمال f_X . نعتبر نموذج الانحدار التالي:

$$Y_{n+1} = f(\Phi_n) + \varepsilon_{n+1} \quad (3.16)$$

حيث

• $\Phi_n = \theta' X_n$ و المعلمة θ هو ناقل معرف على \mathbb{R}^p و احداثياتها ليست معدومة .

• (ε_n) هو عبارة عن سلسلة من مصطلحات الخطأ العشوائي ، مستقلة عن (X_n) . و لدينا الوسيط الشرطي التالي $\mathbb{E}[\varepsilon_{n+1}|F_n] = 0$ و $\mathbb{E}[\varepsilon_{n+1}^2|F_n] = \sigma^2$.

• الدالة ربط غير معروفة و f هي دالة ذات قيم في \mathbb{R} .

في هذا النموذج ، نريد تقدير المتجه θ والمعلمة الوظيفية f . تذكر $\hat{\theta}_n$ هي تقدير θ بالنموذج SIR التكراري الذي تم دراسته في الجزء الاول من الفصل , يمكننا بعد ذلك التنبؤ ب Φ_n بواسطة $\hat{\Phi}_n = \hat{\theta}'_{n-1} X_n$ و يمكننا الجمع بين تكراري Nadaraya-Watson المقدر f (الذي قدم في الفصل الثاني) إلى المقيم التكراري لاتجاه EDR ل θ .

لدينا

$$\hat{f}_n(y) = \frac{1}{\sum_{i=2}^n \frac{1}{h_{i-1}} K\left(\frac{y - \hat{\Phi}_{i-1}}{h_{i-1}}\right)} \sum_{i=2}^n \frac{1}{h_{i-1}} K\left(\frac{y - \hat{\Phi}_{i-1}}{h_{i-1}}\right) Y_i \quad (3.17)$$

سوف ندرس الخصائص المقاربة لهذا المقدر $\hat{f}_n(y)$ لاحظ أنه يمكننا أيضاً الكتابة $\hat{f}_n(y)$ في النموذج التكراري التالي:

$$(3.18)$$

$$\hat{f}_n(y) = \hat{f}_{n-1}(y) + \frac{1}{\sum_{i=2}^n \frac{1}{h_{i-1}} K\left(\frac{y - \hat{\Phi}_{i-1}}{h_{i-1}}\right)} \frac{1}{h_{n-1}} K\left(\frac{y - \hat{\Phi}_{n-1}}{h_{n-1}}\right) (Y_n - \hat{f}_{n-1}(y))$$

من أجل تبسيط الكتابات في بقية هذا الجزء الثالث ، نكتب مقدر $f(y)$ الوارد في (3,16) في الشكل التالي :

$$\hat{f}_n(y) = \frac{1}{N_{n-1}(y)} \sum_{i=2}^n \frac{1}{h_{i-1}} K\left(\frac{y - \hat{\theta}'_{i-2} X_{i-1}}{h_{i-1}}\right) Y_i \quad (3.19)$$

حيث

$$N_{n-1}(y) = \sum_{i=2}^n W_{i-1}(y)$$

و

$$W_{i-1}(y) = \frac{1}{h_{i-1}} K \left(\frac{y - \hat{\theta}'_{i-2} X_{i-1}}{h_{i-1}} \right)$$

لدينا التحليل التالي :

$$\begin{aligned} \hat{f}_n(y) - f(y) &= \frac{1}{N_{n-1}(y)} \sum_{i=2}^n W_{i-1}(y) Y_i - f(y) \\ &= \frac{1}{N_{n-1}(y)} \sum_{i=2}^n W_{i-1}(y) \{f(\Phi_{i-1}) + \varepsilon_i - f(y)\} \quad (3.20) \\ &= \frac{1}{N_{n-1}(y)} \{P_n(y) + Q_{n-1}(y)\} \end{aligned}$$

حيث

$$P_n(y) = \sum_{i=2}^n W_{i-1}(y) \varepsilon_i$$

و

$$Q_{n-1}(y) = \sum_{i=2}^n W_{i-1}(y) \{f(\Phi_{i-1}) + \varepsilon_i - f(y)\}$$

 المتغير العشوائي $P_n(y)$ و $Q_{n-1}(y)$ سيلعب دوراً مهماً في دراسة الخصائص التقاربية لمقدّرنا.

2.4.3 الفرضيات

 قبل إعطاء النتائج المقاربة ل $\hat{f}_n(y)$ ، نفترض أن الفرضيات الخمس التالية مستوفاة:

- (H1) النافذة (h_n) تدل على تسلسل تنازلي للأرقام الحقيقية الموجبة بدقة $0 \rightarrow h_n$ و $nh_n \rightarrow \infty$ عندما $n \rightarrow \infty$.

• (H2) K هي النواة المدجة , و هي دالة قابلة للقياس محدودة و موجبة حيث :

$$\int_{\mathbb{R}} K(x)dx = 1, \int_{\mathbb{R}} K^2(x)dx = \nu^2, \int_{\mathbb{R}} K^4(x)dx = \gamma^2, \int_{\mathbb{R}} xK(x)dx = 0, \int_{\mathbb{R}} xK^2(x)dx = 0,$$

$$\int_{\mathbb{R}} xK^4(x)dx = 0, \int_{\mathbb{R}} x^2K(x)dx = \tau^2, \int_{\mathbb{R}} x^2K^2(x)dx = \mu^2, \int_{\mathbb{R}} x^2K^4(x)dx = \eta^4$$

• (H3) f هي دالة Lipschitzian

$$f_X \in C^2(\mathbb{R}^p) \text{ معرفة على كل } 1 \leq i \leq p \text{ (H4) .}$$

$$\sup_{x \in \mathbb{R}} \left| \frac{\delta^2}{\delta x^2} f_X(x) \right| < \infty$$

• (H5) السلسلة (X_n) محدودة بشكل شبه مؤكد .

5.3 النظريات

1.5.3 قانون الأعداد الكبرى

نظريه 9 نفرض (ε_n) لديه لحظة شرطية من الرتبة $a > 2$. في ظل الفرضيات الخمس (H1) ، ... ، (H5) ، لدينا كل $y \in \mathbb{R}$

$$\hat{f}_n(y) \xrightarrow{P.S.} f(y)$$

• و تكون النتيجة ادق اذا كان $0 < \alpha < 1/3$

$$\hat{f}_n(y) - f(y) = O\left(\sqrt{n^{-\alpha}}\right) \quad p.s.$$

بينما اذا كان $1/3 \leq \alpha < 1$

$$\hat{f}_n(y) - f(y) = O\left(\sqrt{n^{\alpha-1} \log(n)}\right) \quad p.s.$$

2.5.3 نظرية الحدود المركزية

نظريه 10 نفرض (ε_n) لديه لحظة شرطية من الرتبة $a > 2$. و تحت الفرضيات الخمس (H1) ، ... ، (H5) ، لكل $\alpha \in]1/3, 1[$ لدينا :

$$\sqrt{nh_n} \left(\hat{f}_n(x) - f(x) \right) \xrightarrow{\mathcal{L}} N \left(0, \frac{\sigma^2 \nu^2}{(1 + \alpha)h(\theta, y)} \right)$$

6.3 بعض نتائج المحاكاة

في هذه المحاكاة ، قدمنا نماذج الانحدار التالية:

$$(M1) : y = (\theta'x)^3 + \varepsilon$$

$$(M2) : y = (\theta'x) \exp(-\theta'x) + \varepsilon$$

و x يتبع القانون العادي المتعدد $N_p(0; 1)$ ، $\theta = (1; -1; 0; \dots; 0)' / \sqrt{2} \in \mathbb{R}^p$ ، ومصطلح الخطأ ε مستقل عن x ، يتبع القانون الطبيعي $N(0; \sigma^2)$. وحسب البعد p بالنسبة ل x التي هي $p = 2$ فضلا عن القيمة σ^2 لتباين الضوضاء ε .

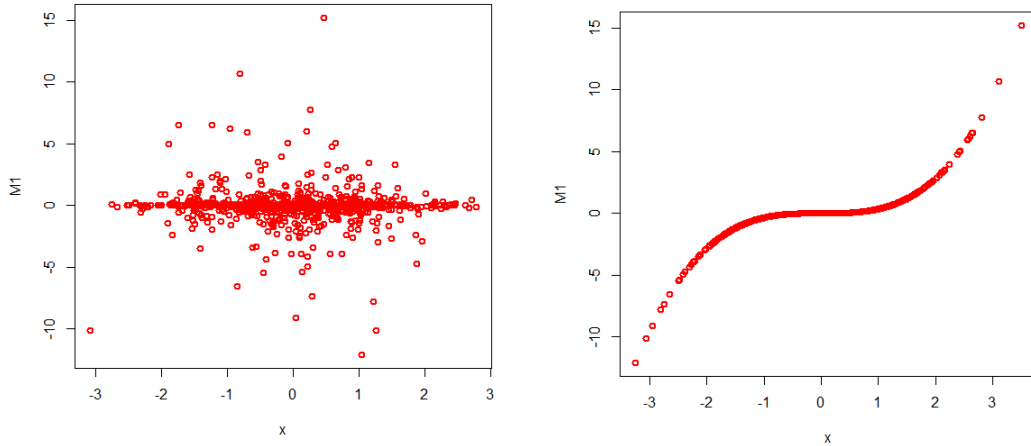
استخدمنا النواة الغوسية

$$(\mathbf{K}_{Gauss}(x) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{x^2}{2} \right))$$

و النافذة $h_n = n^{-\alpha}$ حيث $0 < \alpha < 1$

نقدم في الشكل 3.1 اثنين من البيانات ل $n = 1000$ نقطة $\{(x'_i \theta, y_i); i = 1, \dots, 1000\}$ محاكاة من نماذج (M1) ل البعد $p = 2$ و $\sigma = 1$.

باستعمال برنامج R تحصلنا على المحاكات التالية :

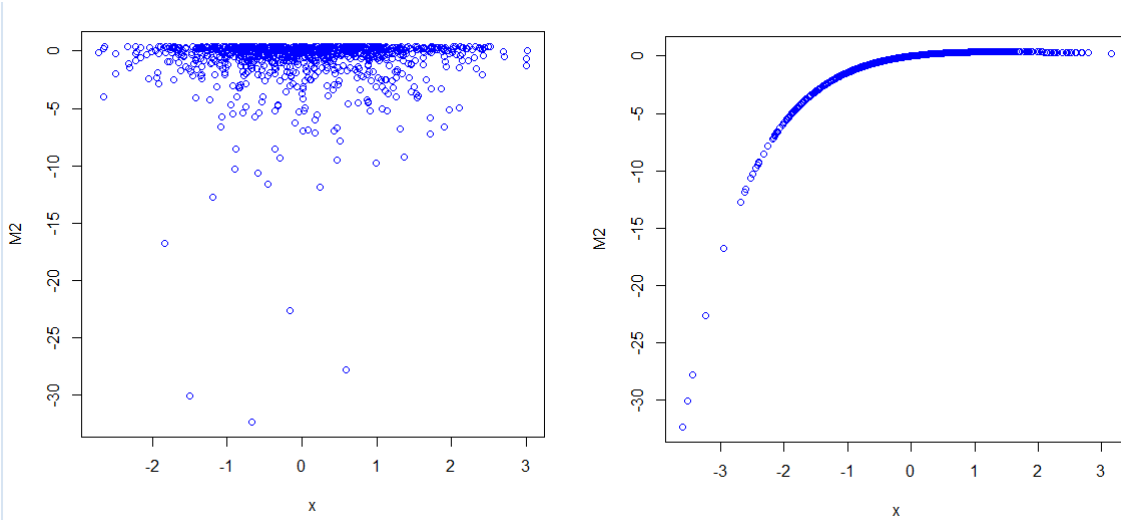


المنحنى العادي للنموذج M1 انتشار البيانات في شكل سحابة

الشكل 3.1 : تبين من النقاط $\{(x'_i, y_i); i = 1, \dots, 1000\}$ لمحاكاة من النموذج حيث (M1) البعد

$$\sigma = 1 \text{ و } p = 2$$

نلاحظ في الشكل 3.1 : تبين للنموذج (M1) في البعد $p = 2$ حيث يتم تقديرها من خلال التباين نجد تركز الكلي للبيانات على طول محور الفواصل وانتشار بعضها في الجزء العلوي والسفلي بالنسبة لل صفر



المنحنى العادي للنموذج M2 انتشار البيانات في شكل سحابة

الشكل 3.2 : تبين من النقاط $\{(x'_i, y_i); i = 1, \dots, 1000\}$ لمحاكاة من النموذج M2 حيث البعد

$$\sigma = 1 \text{ و } p = 2$$

نلاحظ في الشكل 3.2 : تبين للنموذج (M2) في البعد $p = 2$ حيث يتم تقديرها من خلال التباين نجد تركز الكلي للبيانات على طول محور الفواصل وانتشار بعضها في الجزء السفلي بالنسبة لل صفر

ومنه نستنتج ان تقدير النموذج M1 و M2 من خلال التباين الذي هو موضح في الشكل 3.1 والشكل 3.2 تركز البيانات على طول محور الفواصل لكلي النموذجين ونستنتج ان تقدير النموذجين يعتمد على التباين

الملاحق

تطبيق طريقة دلتا

نفرق عمود المتجهة يكون في الشكل التالي :

$$\bar{U}_1 = (p_{1,N}, P_{2,N}, m'_{1,N}, m'_{2,N}, \bar{x}', \frac{1}{N} \sum_{i=1}^N \text{vec}(x_i, x'_i)')$$

حيث : $m_{h,N} = \frac{1}{N p_{h,N}} \sum_{i=1}^N x_i \mathbb{I}_{h(i)} = \frac{1}{p_{h,N}} \tilde{m}_{h,N}$ \bar{U}_1 asymptotic

لتكن الدالة h_n المعرفة كالتالي :

$$\begin{cases} h_1 : \mathbb{R}^{2+3P+P^2} & \longmapsto \mathbb{R}^{2+3P+P^2} \\ u = (a_1, a_2, b'_1, b'_2, c', d') & \longmapsto (a_1, a_2, b'_1/a_1, b'_2/a_2, c', d')' \end{cases}$$

نلاحظ ان $\bar{U}_1 = h_1(\bar{U})$ ونفرض ان $\mu_{U_1} := h_1(\mu_U) = (p_1, P_2, m'_1, m'_2, \mu', \text{vac}(\Sigma + \mu\mu'))'$ من خلال طريقة ديلتا نحصل على:

$$\sqrt{N}(h_1(\bar{U}) - h_1(\mu_U)) \xrightarrow{\mathcal{L}} \mathcal{N}_{2+3P+P^2}(0, J_1, \Sigma_U, J'_1)$$

حيث : $J_1 = \partial h_1 / \partial u |_{\mathbb{E}}$ و $g|_{\mathbb{E}}$ تمثل قيمة الدالة g بالنسبة للوسيط

و I_s هي مصفوفة الوحدة في البعد $s \times s$ و $0_{s_1, s_2}$ مصفوفة صفرية في البعد $s \times s$ بعدما قننا بالحسابات نحصل على :

$$J'_1 = \partial h'_1 / \partial u |_{\mathbb{E}} = \begin{pmatrix} I_2 & A_1 & 0_{2,p} & 0_{2,p^2} \\ 0_{2p,2} & A_2 & 0_{2p,p} & 0_{2p,p^2} \\ 0_{p,2} & 0_{p,2p} & I_p & 0_{p,p^2} \\ 0_{p^2,2} & 0_{p^2,2p} & 0_{p^2,p} & I_{p^2} \end{pmatrix}$$

$$A_2 = \begin{pmatrix} \frac{1}{p_1} I_p & 0_{p,p} \\ 0_{p,p} & \frac{1}{p_2} I_p \end{pmatrix} \text{ و } A_1 = \begin{pmatrix} -\tilde{m}'_1/p_1^2 & 0_{1,p} \\ 0_{1,p} & -\tilde{m}'_2/p_2^2 \end{pmatrix} : \text{حيث}$$

نستنتج المصفوفة covariance للتوزيع asymptotic ل \bar{U}_1 التي تكتب على الشكل التالي :

$$\Delta_1 = J_1 \Sigma_U J'_1 = \begin{pmatrix} B_{11} & 0_{2,2p} & B_{13} & B_{14} \\ 0_{2p,2} & B^*_{22} & B^*_{23} & B^*_{24} \\ B'_{13} & B'^*_{23} & B_{33} & B_{34} \\ B'_{14} & B'^*_{24} & B'_{34} & B_{44} \end{pmatrix}$$

حيث : $B_{23}^* = A_1' B_{13} + A_2 B_{23}$ و $B_{22}^* = A_1' B_{11} A_1 + A_1' B_{12} A_2 + A_2 B_{12}' A_1 + A_2 B_{22} A_2$
و $B_{24}^* = A_1' B_{14} + A_2 B_{24}$
نعطي الان الدالة h_2 :

$$\begin{cases} h_2 : \mathbb{R}^{2+3P+P^2} & \longmapsto \mathbb{R}^{p+p+p^2} \\ u = (a_1, a_2, b_1', b_2', c', \text{vec}(d)') & \longmapsto (b_1', b_2', (\text{vec}(d) - \text{vec}(cc'))')' \end{cases}$$

$$h_2(\mu_{U_1}) = \begin{pmatrix} m_1 \\ m_2 \\ \text{vec}(\Sigma_N) \end{pmatrix} \text{ و } h_2(\bar{U}_1) = \begin{pmatrix} m_{1,N} \\ m_{2,N} \\ \text{vec}(\Sigma_N) \end{pmatrix} \text{ نلاحظ ان}$$

نستنتج من طريقة دلتا :

$$\sqrt{N} \left(\begin{pmatrix} m_{1,N} \\ m_{2,N} \\ \text{vec}(\Sigma_N) \end{pmatrix} - \begin{pmatrix} m_1 \\ m_2 \\ \text{vec}(\Sigma_N) \end{pmatrix} \right) \xrightarrow{\mathcal{L}} \mathcal{N}_{2P+P^2}(0, \Delta_2)$$

حيث :

$$\Delta_2 = J_2 \Delta_1 J_2'$$

$$J_2' = \partial h_2' / \partial u \mid \mathbb{E} \text{ و}$$

بعد حساب مشتقات المتجهة تحصل على :

$$J_2' = \begin{pmatrix} 0_{2,p} & 0_{2,p} & 0_{2,p^2} \\ I_p & 0_{p,p} & 0_{p,p^2} \\ 0_{p,p} & I_p & 0_{33} \\ 0_{p,p} & 0_{p,p} & -(I_p \otimes \mu' + \mu' \otimes I_p) \\ 0_{p^2,p} & 0_{p^2,p} & I_{p^2} \end{pmatrix}$$

لدينا : $Z = \sqrt{N} \{ (m_{1,N} - m_1) - (m_{2,N} - m_2) - D^* \text{vec}(\Sigma_N - \Sigma) \}$ و منه نستنتج المصفوفة

ذات البعد $p \times (2p + p^2)$ في الشكل التالي : $E = (I_p \quad -I_p \quad D^*)$

$$Z = E \sqrt{N} \left(\begin{pmatrix} m_{1,N} \\ m_{2,N} \\ \text{vec}(\Sigma_N) \end{pmatrix} - \begin{pmatrix} m_1 \\ m_2 \\ \text{vec}(\Sigma_N) \end{pmatrix} \right)$$

و منه القانون التقاربي asymptotique ل Z يكون كالتالي :

$$Z \xrightarrow{\mathcal{L}} \mathcal{N}_{2P+P^2}(0, \Delta_3)$$

و

$$\Delta_3 = E\Delta_2E'$$

$\sqrt{N}\tilde{\theta}_N - \tilde{\theta}$ de asymptotique loi

لدينا : $Z : \sqrt{N}\tilde{\theta}_N - \tilde{\theta} = \Sigma_N^{-1}$ و Σ_N^{-1} تتقارب باحتمال الى Σ^{-1} و Z هي الشبه الطبيعي للوسط الصفري و للمصفوفة التباين والتغاير Δ_3

و منه نستنتج

$$\sqrt{N}(\tilde{\theta}_n - \tilde{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma^{-1}\Delta_3\Sigma^{-1})$$

خوارزمية النموذجين M1, M2 :

1- المنحنى العادي للنموذج

```
n=1000
x<-rnorm(n)
z<-c(-1/sqrt(2),1/sqrt(2))
m<-length(z)
M2<-0
for(i in (1:m))
{M2<-(z[i]*x)*exp(-z[i]*x)}
plot(x,M2,col="blue", lwd="2")
```

خوارزمية النموذج M2

```
n=1000
x<-rnorm(n)
z<-c(-1/sqrt(2),1/sqrt(2))
m<-length(z)
M1<-0
for(i in (1:m))
{M1<-(z[i]*x)^3}
plot(x, M1, col = "red", lwd = "2")
```

خوارزمية النموذج M1

2- انتشار البيانات في شكل سحابة

```

n=1000
x<-rnorm(n)
z<-c(-1/sqrt(2),1/sqrt(2))
m<-length(z)
k<-length(x)
for (j in (j=m))
{
for (i in (i=k))
{M2[j]<-(z[j]*x[i])*exp(-z[j]*x[i])}
}
plot(x,M2,col="blue",lod="2")

```

خوارزمية النموذج M2

```

n=1000
x<-rnorm(n)
z<-c(-1/sqrt(2),1/sqrt(2))
m<-length(z)
k<-length(x)
for (j in (j=m))
{
for (i in (i=k))
{M1[j]<-(z[j]*x[i])3}
}
plot(x, M1,col = "red",lod = "2")

```

خوارزمية النموذج M1

الخلاصة

لقد تم تطوير في هذه المذكرة نهجاً يستند إلى طريقة قدمها Li (1991) ، تسمى معكوس الانحدار (SIR) تقوم هذه الطريقة بتقييم الجزء المعلمي θ من نموذج شبه المعلمي دون الحاجة إلى تقدير الدالة f ، و تم مقارنة السلوك العددي للمقدرين الثلاثة Nadaraya-Watson ، Nadaraya-Watson التكراري و Révész و تم اقتراح نهجاً تكرارياً لطريقة SIR تم اقتراح طريقة للجمع بين التقدير المتكرر لدالة الربط f مع تقدير Nadaraya-Watson التكراري وتقدير معلمات θ عبر مقيم SIR التكراري، ولقد تم وضع قانون الأعداد الكبيرة بالإضافة إلى نظرية الحد المركزي للنقطة، و تم وضع بعض المحاكات لتقدير بعض نماذج الانحدار .

المراجع العلمية

- [1] Li, K. C. (1991). Sliced inverse regression for dimension reduction, with discussion. Journal of the American Statistical Association, 86, 316-342.
- [2] Paul Deheuvels : Sur l'estimation séquentielle de la densité. C. R. Acad. Sci. Paris Sér. A-B, 276:A1119-A1121, 1973.
- [3] E. A. Nadaraja : On a regression estimate. Teor. Verojatnost. i Primenen., 9:157-159, 1964.
- [4] Geoffrey S. Watson : Smooth regression analysis. Sankhya Ser. A, 26:359-372, 1964.
- [5] Gérard Collomb : Quelques propriétés de la méthode du noyau pour l'estimation non paramétrique de la régression en un point fixé. C. R. Acad. Sci. Paris Sér. A-B, 285(4):A289-A292, 1977.
- [6] Eugene F. Schuster : Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. Ann. Math. Statist., 43:84-88, 1972.

- [7] P. Révész : How to apply the method of stochastic approximation in the nonparametric estimation of a regression function. *Math. Operationsforsch. Statist. Ser. Statist.*, 8(1):119-126, 1977.
- [8] Abdelkader Mokkadem, Mariane Pelletier et Yousri Slaoui : Revisiting Révész's stochastic approximation method for the estimation of a regression function. *ALEA Lat. Am. J. Probab. Math. Stat.*, 6:63-114, 2009.
- [9] Duan, N. and Li, K. C. (1991). Slicing regression : a link-free regression method. *The Annals of Statistics*, 19, 505-530.

Résumé

Dans ce mémoire nous avons proposer l'estimateur récursive de la fonction de lien dans un modèle semi-paramétrique et nous donnons quelques propriétés asymptotiques pour un estimateur récursif de noyau, et nous proposons une méthode permettant de combiner l'estimation récursive de la fonction de lien f par l'estimateur de Nadaraya-Watson récursif et l'estimation du paramètre θ via l'estimateur SIR récursif (Sliced Inverse Régression récursif)

.

Mots Clés : *estimation récursive, modèle semi-paramétriques de régression, méthode de*