##  Μη παραμετρική παλινδρόμηση Ι: τοπική εξομάλυνση (local smoothing)

### Εκτίμηση διακύμανσης

* Έστω  ένας γραμμικός εξομαλυντής : . Θέτουμε ,  και εκτιμάμε τη διακύμανση των καταλοίπων με:



**Πρόταση**. Αν  «αρκούντως ομαλή» και  ,  τότε η  είναι συνεπής για το 

Σκίτσο απόδειξης



Άρα 





 ,

Τώρα: , και τελικά



αλλά αν  αρκούντως ομαλή  και άρα 

 Ετεροσκεδαστικότητα

Αν **η  εξαρτάται από το ,** δηλαδή  τότε μπορεί και αυτή να εκτιμηθεί

**παραδειγμα:** figure 5.4 από wassermann



**A) τρόπος:** **απλά με την εκτίμηση  που παίρνουμε από το τοπικό γραμμικό / πολυωνυμικό μοντέλο** με το οποίο εκτιμούμε την 

**B) τρόπος:** **Εναλλακτικά** προτείνεται π.χ. η εξής μέθοδος:

αν , τότε θέτοντας

και 

παίρνουμε 

**Η ιδέα είναι να εκτιμηθεί η  με παλινδρόμηση  των πάνω στα **

Δηλαδή :

1. Εκτίμησε την  με κάποια μέθοδο και πάρε 

2. Σχημάτισε τα 

3. Με τοπική παλινδρόμηση (όπως στον υπολογισμό των) των  έναντι των  πάρε εκτιμήτρια  του 

4. Θέσε 

**Παράδειγμα:** figure 5.9 Wassermann



### Διαστήματα εμπιστοσύνης για  και για

(παρόμοια όπως στην εκτίμηση και τα δ.ε. για την πυκνότητα)

Έστω  ή 

και ,

δηλαδή  ή 

#### Σταθερό , δ.ε. για

Αν  , , θα έχουμε



 όπου 

άρα για  θα παίρναμε ασυμπτωτικά δ.ε. για τον 



όπου  μια εκτίμηση της διακύμανσης , η οποία μπορεί να εξαρτάται από το 

#### Σταθερό  , δ. ε. για

Αν θέλαμε δ. ε. για θα έπρεπε να ισχύει 

Αυτό θα συμβαίνει για  μόνο αν



Όμως  (συνήθως) . Άρα για να κατασκευάσουμε δ.ε. για  θα χρειαζόμασταν

 ή ισοδύναμα 

**Αν ίσχυε ** **το bias**  **θα ήταν «αμελητέο»** συγκρινόμενο με τη διακύμανση **και τότε τα δ.ε.**

 **θα ήταν και δ.ε. για το **

Όμως αυτό **δεν ισχύει για την «βέλτιστη» bandwidth ** , η οποία ακριβώς επιλέγεται ώστε bias και διακύμανση να είναι ίδιας τάξης μεγέθους, και είναι τάξης . Άρα για την  θα έχουμε . Άρα η συνθήκη  ανταποκρίνεται σε καταστάσεις όπου κάνουμε **under smoothing**.

Υπάρχουν προσεγγίσεις για να υπερβληθεί η δυσκολία αυτή μέσω εκτίμησης του bias. Αυτό προϋποθέτει εκτίμηση της  και ενέχει αρκετές δυσκολίες. Εδώ απλά θα αποδεχτούμε ότι κάνουμε δ.ε. για την 

**Παράδειγμα** figure 5.10 Wassermann



#### Ζώνη εμπιστοσύνης για

Θέλουμε να κατασκευάσουμε **«ζώνη» της μορφής**



**που θα εμπεριέχει την για όλα τα  με πιθανότητα τουλάχιστον .**

Πως θα βρούμε κατάλληλο  ?

##### Αν  σταθερό και γνωστό

 με  Gaussian process

Sun & Loader:  (\*) , με  όπου .

Θα αρκούσε λοιπόν **να βρεθεί , τέτοιο ώστε η (\*) να είναι .**

Ως προς τον υπολογισμό του : έχει δειχτεί ότι αν τα  έχουν ίση απόσταση μεταξύ τους (ομοιόμορφο design ), μπορούμε να προσεγγίσουμε  , με  , για NW και  για loess.

##### Αν σ άγνωστο

το εκτιμούμε με  με  β.ε. (), και η (\*) γίνεται .

##### Αν σ εξαρτάται από x

τότε  και τα δ.ε. θα είχαν τη μορφή

.

αν η είναι ομαλή στο τότε αυτό ισούται περίπου με



παράδειγμα : figure 5.11 Wassermann + 5.4 Breheny

**figure 5.11 Wassermann**



figure 5.4 Breheny







#### Bootstrap στη μη παραμετρική παλινδρόμηση

##### Α. Bootstrap στα κατάλοιπα για σταθερό  και  ίδιο για όλα τα .

(δ.ε. για  (ή και για , αν )

* Έστω .
* Θέτουμε  χωρίς τα  πρώτα και τελευταία, και  και επίσης  και 
* Τραβάμε  bootstrap δείγματα  , από τα  και θέτουμε .
* Υπολογίζουμε την  και  από .

Τότε  Hall (AS,1992)

Συμβολίζοντας με  την cdf της ποσότητας  παίρνουμε 

δ.ε. για  (ή για αν ) τα:



##### Β. Σημειακές ζώνες εμπιστοσύνης με χρήση Bootstrap

Hall & Horowitz (AS, 2013)

(δ.ε. για  εξαιρώντας ποσοστό των )

* Κλασσικά δ.ε. για την  για τιμές της  που είναι κοντά στο βέλτιστο τείνουν να υπο-καλύπτουν την  λόγω του bias  που έχει την ίδια τάξη μεγέθους με τη διακύμανση.
* Στη πραγματικότητα ασυμπτωτικά η πιθανότητα κάλυψης της , αντί για , θα είναι , που ισούται με  μόνο αν .
* Η παρακάτω μέθοδος προσπαθεί με τη χρήση bootstrap να διορθώσει το ονομαστικό  που χρησιμοποιείται στη κατασκευη των δ.ε. ώστε αυτά να έχουν πραγματική πιθανότητα κάλυψης της  κοντά στην ονομαστική (π.χ 95%) για όλα τα  (πλην  των ), αλλά όχι ομοιόμορφα στα .

Δύο παράμετροι

* : λάθος τύπου Ι και
*  ποσοστό των  που «εξαιρούνται», δηλαδή ζητάμε «μόνο» . Αυτή η «εξαίρεση» αφορά κυρίως τα  για τα οποία η  έχει «κορυφές» και «τρύπες» δηλαδή υψηλό bias.

Έστω  ,   (υπάρχει και γενίκευση για  να εξαρτάται από ).

* **Βήμα 1** Υπολογισμός  και , π.χ. με local regression , NW , …. Έπεται .
* **Βήμα 2** Υπολογισμός καταλοίπων  ,  και .
* **Βήμα 3** Δημιουργία bootstrap δειγμάτων  όπου  bootstrap δείγμα από .
* **Βήμα 4** Υπολογισμός bootstrap εκτιμητριών  **με ίδιο ** όπως ,  και ζωνών  *«ονομαστικού» α* με .
* **Βήμα 5** Bootstrap εκτίμηση της «πραγματικής» πιθανότητας κάλυψης της , δηλαδή της . Έστω



το ποσοστό των bootstrap δειγμάτων που η ζώνη τους  καλύπτει την  στο .

* **Βήμα 6** Έστω
	+  η λύση στο  της εξίσωσης . Άρα .

*Δηλαδή* * είναι εκείνο το «ονομαστικό » που θα έδινε στο  σωστή (δηλαδή μεγαλύτερη ίση με ) «πραγματική» (εκτιμώμενη) πιθανότητα κάλυψης*, και

* +  το -ποσοστιαίο σημείο του συνόλου των  (, προσεγγιζόμενο με ένα λεπτό πλέγμα πάνω στο ).

*Δηλαδή*  *είναι εκείνο το «ονομαστικό » που θα έδινε για  των  σωστή (δηλαδή μεγαλύτερη ίση με ) «πραγματική» (εκτιμώμενη) πιθανότητα κάλυψης*.

* **Βήμα 7** Τέλος θέτουμε για το δ.ε. πραγματικής πιθανότητας κάλυψης * :*

,

 όπου 

Σε αυτή τη προσέγγιση το Bootstrap χρησιμοποιείται μόνο **για να «διορθώσει»** **τις κλασσικές ζώνες που κατασκευάζονται για μεμονωμένο **, **προσαρμόζοντας το «ονομαστικό »**, έτσι ώστε να προκύπτει ζώνη που να έχει τελικά πραγματικό ποσοστό κάλυψης  σε «σχεδόν όλα» τα , δηλαδή σε  των . Η εκτίμηση αυτού του «πραγματικού ποσοστού» είναι αυτή που γίνεται με το bootstrap.