## Το Jackknife και το Βootstrap

* Είναι **μέθοδοι κατασκευής διαστημάτων εμπιστοσύνης για παράμετρο** (χωρίς παραμετρικές υποθέσεις μοντέλου),
  + όπως και η ασυμπτωτική θεωρία των προηγούμενων κεφαλαίων που στηρίζεται στην influence function.
  + Είναι πιο «απλές» στην εφαρμογή, διότι απαιτούν λιγότερη θεωρητική δουλειά, αλλά είναι πιο απαιτητικές υπολογιστικά.



### Το Jackknife



**Στόχος: *προσεγγιστικός υπολογισμός της μεροληψίας και της διακύμανσης μιας εκτιμήτριας*,** ώστε αφενός να διορθωθεί η μεροληψία και αφετέρου να κατασκευαστούν διαστήματα εμπιστοσύνης



* Έστω μια εκτιμήτρια μίας ποσότητας  : .



* Θέτουμε , η μεροληψία της εκτιμήτριας και



*  την εκτιμήτρια που προκύπτει αν παραλείψουμε την i-παρατήρηση.



* Επίσης θέτουμε , δηλαδή παραλείπω πρώτα την πρώτη, μετά την δεύτερη, κλπ παρατήρηση και παίρνω το μέσο όρο.



* Αν η  είναι αμερόληπτη, το ίδιο θα συμβαίνει και με τις (για κάθε ) και άρα και με την 



* Αν όμως η  έχει μια μεροληψία, τότε για πολλά στατιστικά μπορεί να δειχτεί ότι έχουμε ένα ανάπτυγμα της μεροληψίας σε όρους του  της μορφής

,



όπου σημαίνει κάτιγια κάποια σταθερά 

* Π.χ. αν και . Τότε και άρα 



Άρα στο ανάπτυγμα του bias.



Τότε:



 και άρα







Επίσης:











Όπου  είναι  και άρα:



θέτοντας  θα έχουμε 



δηλαδή η  συμφωνεί με τα  με λάθος , δηλαδή σωστό για τάξη μεγέθους .



**Ιδέα :*Διόρθωσε την μεροληψία της , χρησιμοποιώντας την ***

**Ορίζω διορθωμένη (bias-corrected) εκτιμήτρια**



****



τότε:











Δηλαδή **το bias της είναι μικρότερης τάξης κατά παράγοντα **

Ενώ bias του  θα έχουμε 



Για παράδειγμα αποδεικνύεται ότι για την  η , δηλαδή παίρνουμε την αμερόληπτη εκτιμήτρια



Ας ορίσουμε τώρα τις **«ψευδο-τιμές»**





παρατηρείστε ότι αν π.χ.  τότε  και θα έχουμε







Απόδειξη











Με βάση τις ψευδο-τιμές , και χρησιμοποιώντας τις σαν να ήταν ανεξάρτητες τυχαίες μεταβλητές, μπορούμε **να ορίσουμε μια εκτιμήτρια της διακύμανσης της αρχικής μας εκτιμήτριας , δηλαδή της **:



και 



**η  είναι η δειγματική διακύμανση των ψευδο-τιμών γύρω από τον μέσο τους**



##### Σημείωση



,

διότι











,







όπου

**παράδειγμα**: Για είδαμε ότι  και άρα η  είναι η δειγματική διακύμανση.



Επομένως **η  είναι η συνεπής εκτιμήτρια της **. Παρόμοια ισχύουν για γραμμικά στατιστικά ή και για «ομαλές» συναρτήσεις των δεδομένων



Πρόταση: αν και  και έστω  με συνεχώς παραγωγίσιμη και . Τότε:



 , με και







Όμως η δεν είναι συνεπής π.χ. για στατιστικά «μη-ομαλά» όπως τα ποσοστιαία σημεία 

Έτσι για τη διάμεσο, αποδεικνύεται ότι , όπου  άρα δεν συγκλίνει στο 1.



Σημειώστε ότι αν  το δείγμα, με  μονό τότε η, ενώ οι  παίρνουν όλες τρεις μόνο τιμές: τις τρείς κεντρικές παρατηρήσεις. Δεν είναι δυνατό να στηρίξουμε σε τέτοιου είδους «μεταβλητότητα» μια καλή εκτιμήτρια της .



### Σχέση μεταξύ Jackknife και συνάρτησης επιρροής

Υπενθυμίζουμε .



**Προσέγγιση της  με  και :**







* Η  θέτει «μάζα»  στο 



* Η θέτει «μάζα»  στο 



* Άρα η  είναι η 



 και τέλος



, όπου 



Υπενθύμιση 



***Περίληψη ορισμών-σχέσεων jackknife***

, 



, 



, ****

bias του  

, 

, 

