## Εκτίμηση συναρτησιακών της cdf (αθρ. συναρτ. καταν.)

Ενδιαφερόμαστε για **την εκτίμηση παραμέτρων που γράφονται ως συναρτήσεις της cdf των παρατηρήσεων μας**,   
π.χ.

* .Έχουμε 
* 
* το α-ποσοστιαίο σημείο της 







Σε όλες αυτές (και άλλες) περιπτώσεις

* έχουμε ένα **συναρτησιακό **, που έχει πεδίο ορισμού ένα σύνολο συναρτήσεων  που περιλαμβάνει τις cdf και
* πεδίο τιμών το  και
* **ενδιαφερόμαστε να εκτιμήσουμε την , όπου είναι η «πραγματική» cdf των παρατηρήσεων μας**.

Ορίζουμε την **plug-in εκτιμήτρια** (εκτιμήτρια αντικατάστασης) **αντικαθιστώντας την με την εκτιμήτρια της :**

****

Σε πολλές περιπτώσεις **παίρνουμε ως εκτιμήτρια το δειγματικό «εμπειρικό» ανάλογο** της ποσότητας που μας ενδιαφέρει.

* 
* 
* . Έτσι π.χ. αν  τότε .
* **Γραμμικά συναρτησιακά:** αυτά που γράφονται ως για κατάλληλη συνάρτηση (βλ. π.χ. ) . Τότε



### Διαστήματα Εμπιστοσύνης

* Αν είχαμε μια εκτίμηση της τυπικής απόκλισης και
* αν επιπλέον 

θα μπορούμε να φτιάξουμε **-διαστήματα εμπιστοσύνης ως**:

,

όπου το -ποσοστιαία σημασία της .

Π.χ. για  έχουμε:  και , με  μία εκτιμήτρια του 

Άρα-δ.ε. για :.

#### Δύο ερωτήματα:

Α) Γνωρίζουμε  «κοντά» στην . Αυτό είναι **επαρκές** για να έχουμε κοντά στην **;**

B) Όταν η απάντηση στο Α είναι θετική **πώς φτιάχνουμε δ.ε.;**

**«Προσέγγιση» σε απαντήσεις**

**Στο Α): Όχι πάντα**: Αντιπαράδειγμα π.χ. έστω παραγωγίσιμη με πυκνότητακαι έστω .

Όμως η  δεν είναι παραγωγίσιμη στα , ενώ για , το οποίο δεν είναι κοντά στο .

* Παραμετρικό ανάλογο: δ-method.

&.

* Αυτό μετά δίνει και **απάντηση στο Β**):



**Άρα χρειάζομαι κάτι σα παραγωγισιμότητα της  δηλ. της .**

**Υπενθυμ.: Παράγωγος κατά κατεύθυνση.**

Αν  η παραγωγος της  στο σημείο  στη κατεύθυνση του  είναι η κλιση της  στο  όταν το πλησιάzω ερχόμενος σε ευθεία από το :



Το σύνολο των  για είναι το «ευθύγραμμο τμήμα» που συνδέει τα .

### Συναρτήσεις επιρροής (Influence Functions)

**Τι σημαίνει όμως παράγωγος της ,** δηλαδή μιας συνάρτησης  με πεδίο ορισμού το σύνολο  των cdf συναρτήσεων;



* Αν  δύο cdf τότε  μία ενδιάμεση cdf: για παίρνω την ενώ για την .
* Το σύνολο των  για είναι το «ευθύγραμμο τμήμα» που συνδέει τις  αν τις φανταστούμε ως στοιχεία του (σχήμα…)

Η **στατιστική ερμηνεία** του  για  είανι ότι πρόκειται για μια **μίξη κατανομών**: Έστω έχω δύο τυχαίες μεταβλητές  Που έχουν cdf τις  και μία τρίτη ανεξάρτητη τυχαία μεταβλήτη  που είναι Bernoulli(). Δημιουργώ τη μίξη των  θέτοντας  όταν  και  όταν . Τότε η cdf της  θα είναι



=

=

=

* Μάλιστα, γράφοντας παίρνουμε 

**Ορισμός Gateaux παράγωγος** της  στην  με κατεύθυνση :



Ειδικά αν ,η cdf μιας τυχαίας μεταβλητής παίρνει τη τιμή  με πιθανότητα 1, τότε η



ονομάζεται **συνάρτηση επιρροής της (στο )**

* Η **συνάρτηση επιρροής ενός στατιστικού μας δίνει** **πόσο θα επηρεαστεί το στατιστικό μας αν τα δεδομένα μας έχουν προσμίξεις** (είναι ακάθαρτα/μολυσμένα/νοθευμένα, ***contaminated***) **κατά ένα μικρό ποσοστό από «ξένα» δεδομένα**. (π.χ. έκτοπες παρατηρήσεις, outliers).
* Ένας τρόπος να οριστούν «**Ανθεκτικά**» (***robust***) στατιστικά είναι ως *«εκείνα που έχουν φραγμένη συνάρτηση επιρροής».*

Η συνάρτηση επιρροής **μπορεί να εκτιμηθεί με την δειγματική συνάρτηση επιρροής** , που ορίζεται όπως η συνάρτηση επιρροής με αντικατάσταση της  με ):



#### Παραδείγματα συναρτήσεων επιρροής/παραγώγισης

##### Εκτίμηση πυκνότητας







**Γραμμικά συναρτησιακά**

Δηλαδή: 





Και 

π.χ. ο μέσος  έχει  και άρα μη-φραγμένη  μη ανθεκτικό.

### Break down point

Άλλο μέτρο ανθεκτικότητας (robustness) ενός στατιστικού , π.χ. , (πλην της Influence function):το break down point.

**Ορίζεται ως το ελάχιστο ποσοστό του δείγματος μου το οποίο αν αντικατασταθεί από ακραίους outliers θα δώσει άπειρη διαφορά στο .**

Ακριβέστερα, αν το αρχικό μας δείγμα με cdf  και  το δείγμα που προκύπτει αν αντικαταστήσω με  των  με άλλες «μολυσμένες» παρατηρήσεις και έστω  η cdf των , τότε το

**Break down point της  ορίζεται ως το **

Παρατήρηση

* Ο δειγματικός μέσος  έχει break down point, διότι ακόμα και μια παρατήρηση αν στείλουμε στο όλος ο .
* Η διάμεσος έχει break down point
* Η -trimmed mean (διαγράφει τις μεγαλύτερες και  μικρότερες παρατηρήσεις και παίρνει μέσο των υπολοίπων) έχει break down point.

Παρατήρηση:

* Η ***influence function***εξετάζει το ***ρυθμό μεταβολής* του στατιστικού σε *μικρή contamination*** των δεδομένων 
* Το ***break down point***είναι ένα ***άνω γενικό φράγμα στο ποσοστό της contamination που γίνεται «ανεκτό»*** από το στατιστικό μου πριν πάρει αλλοπρόσαλλες τιμές.

### Παραδείγματα συναρτήσεων επιρροής

#### Ο δειγματικός μέσος

* (σχήμα...)
* 
* 

#### Το Ποσοστιαίο σημείο

 τότε 

Π.χ.

διάμεσος 



**Απόδειξη (εκτός ύλης):** Έστω 

Τότε για  θα έχουμε:



Παραγωγίζοντας κατά ε παίρνουμε:





Αποτιμώντας στο παίρνουμε:



με 

,