## Μη παραμετρική στατιστική: Εισαγωγή

### Τι είναι

* **η παραμετρική στατιστική** **ξεκινάει πάντοτε από την** **υπόθεση ότι μπορούμε να περιγράψουμε το μηχανισμό που γεννά τα δεδομένα με λίγες παραμέτρους**

π.χ.

*  οπού αρκούν (δυο παράμετροι) για να περιγράψουν την (κατανομή των)
* , όπου α, β,  αρκούν (3 παράμετροι)…
*  η f ανήκει σε παραμετρική οικογένεια, που ορίζεται από (pπαράμετροι) από εκεί : , συμπερασματολογία , κλπ
* Έτσι κάθε φορά **γίνονται ισχυρές υποθέσεις** για τις ιδιότητες του μοντέλου που γεννά τα δεδομένα: κανονικότητα, γραμμικότητα στη σχέση, μορφή της (πυκνότητα της κατανομής των). **Σε αυτές στηρίζεται η θεωρία για την κατανομή των , η συμπερασματολογία** κ.λπ.
* ***Αν αυτές πληρούνται η συμπερασματολογία είναι σωστή****.* ***Αν όχι, η συμπερασματολογία*** *(διαστήματα εμπιστοσύνης, αποφάσεις ελέγχων)* ***μπορεί να είναι εσφαλμένη ή παραπλανητική****.*
* **Συνέπεια:** *Χρειαζόμαστε* ***μεθόδους που απαιτούν ελάχιστες υποθέσεις*** *και να οδηγούν παρ’ όλα αυτά* ***σε σωστή συμπερασματολογία****. Αυτές πραγματεύεται η μη-παραμετρική στατιστική.*
* Αποφεύγεται έτσι η απλουστευτική απομείωση της αβεβαιότητας που αφορά στην κατανομή του  σε 2 (ή p) αριθμούς.
* Συνήθως αυτό σημαίνει ότι τα μοντέλα που θα υποθέσουμε χρειάζονται άπειρα πολλές παραμέτρους (όχι πεπερασμένα πολλές - λίγες») για να περιγραφούν.
* Γι αυτό το λόγο η **μη-παραμετρική στατιστική** θα μπορούσε να αποδοθεί και ως «**συμπερασματολογία με άπειρα πολλές παραμέτρους**»

Έτσι η **«κλασσική υπόθεση»** στη μη-παραμετρική στατιστική είναι:

* **iid, όπου συνεχής αθροιστική συνάρτηση κατανομής** (, και επίσης ίσωςσυμμετρική:
* και ενδιαφερόμαστε για εκτίμηση  [ή γενικότερα ]

ή στην παλινδρόμηση να υποθέσουμε

* , όπου **συνεχής, ή καλύτερα λεία/ομαλή)**, αντί για γραμμική και
* εκτιμούμε την (άπειρα πολλές παραμέτρους) αντί για συντελεστές , ή γενικότερα,.

Κάποιες φορές ενδιαφερόμαστε και για **ήμι-παραμετρικά μοντέλα** που συνδυάζουν την παρουσία παραμέτρων σε κάποιες πτυχές του μοντέλου, ενώ άλλες περιγράφονται μη-παραμετρικά]

### Ιδιότητες που ενδιαφέρουν τη μη παραμετρική στατιστική

που μας ενδιαφέρουν και που συχνά είναι προβληματικές στη παραμετρική προσέγγιση/μέθοδο:

* **Ευστάθεια (robustness)**:

***έλλειψη ευαισθησίας*** *(insensibility) ως προς αποκλίσεις από τις αρχικές μου υποθέσεις* και ιδιαίτερα ως προς τη παρουσία «αποκλινουσών» /έκτοπων παρατηρήσεων(outliers).

π.χ. μέσος έναντι διάμεσου:

ο **μέσος επηρεάζεται πολύ όταν μία παρατήρηση είναι «πολύ» έκτοπη**. Η διάμεσος δεν επηρεάζεται ακόμα και αν «πειραχτούν» λίγο λιγότερες από τις μισές παρατηρήσεις.

* **Ελεύθερη κατανομής (distribution free):**

Οι μη παραμετρικές στατιστικές μέθοδοι **συχνά** στηρίζονται σε *στατιστικά η κατανομή των οποίων* ***δεν εξαρτάται από την αρχική κατανομή*** * των παρατηρήσεων*

Αυτό επιτρέπει να φτιαχτούν έλεγχοι ή **διαστήματα εμπιστοσύνης που είναι ορθά όποια και αν ήταν η αρχική κατανομή των δεδομένων μου**.

Π.χ.

:

Απορρίπτω όταν  η δεν εξαρτάται από  στην

### Άλλα πλεονεκτήματα μη παραμετρικών μεθόδων

* **Συχνά η *κατανομή στατιστικών της μη-παραμετρικής προσέγγισης*, *είναι ακριβώς γνωστή* (για πεπερασμένο μέγεθος δείγματος),** πράγμα που επιτρέπει τον υπολογισμό των p-values με ακρίβεια.

π.χ. στατιστικών που βασίζονται μόνο στις τάξεις των παρατηρήσεων στο δείγμα (Rank- statistics).

Η «απώλεια πληροφορίας» που οφείλεται στο πέρασμα σε τάξεις τελικά είναι μικρή: οι μέθοδοι αυτές αποδεικνύεται ότι είναι «λίγο χειρότερες» εκεί όπου οι παραμετρικές αντίστοιχες είναι βέλτιστες, αλλά πολύ καλύτερες όταν δεν τηρούνται οι παραμετρικές υποθέσεις.

* Μέθοδοι όπως το **jackknife και το bootstrap** *δίνουν λύσεις σε σύνθετες καταστάσεις* (με αρκετά ευρεία εφαρμοσιμότητα), όπου η παραμετρική θεωρία θα ήταν πολύ δύσκολη.
* Οι μη παραμετρική θεώρηση επιτρέπει τη ***προσέγγιση προβλημάτων όπου η διατύπωση σε όρους παραμέτρων θα ήταν προβληματική ή ανέφικτη***.

Π.χ:(“goodness of fit test”)

ή έλεγχος κανονικότητας :

* Οι μη-παραμετρικές μέθοδοι **«επιτρέπουν» το *πραγματικό μοντέλο να μην ανήκει στο σύνολο των μοντέλων στα οποία θα ανήκει το εκτιμώμενο μοντέλο.*** Αυτό είναι πιο ρεαλιστικό καθώς συχνά θεωρούμε τα μοντέλα που εκτιμάμε σαν «προσέγγιση» στην πραγματικότητα. Η παραπάνω θεωρία επιτρέπει να υπάρχει μεροληψία (bias) στις εκτιμήτριες, μια μεροληψία που προσπαθούμε να εξισορροπήσουμε με τη διακύμανση, *ώστε να ελαχιστοποιήσουμε το μέσο τετραγωνικό σφάλμα*.
* Επίσης συχνά επιτρέπουν να θεωρούμε προβλήματα που η πολυπλοκότητα τους αυξάνει με το δείγμα (γέφυρα μεταξύ παραμετρικής και μη παραμετρικής στατιστικής.)

Π.χ.

,,

ανεξ. με 

### Επιμέρους ζητήματα/στόχοι της μη παραμετρικής στατιστικής.

* **Εκτίμηση (με διαστήματα κλπ.) της αθροιστικής συνάρτησης κατανομής** (cumulative distribution function, cdf) από δείγμα ανεξάρτητων iid
* **Εκτίμηση συναρτησιακών**  της  π.χ.





ποσοστιαία σημεία

συμπεριλαμβανομένων και **των bootstrap/ jackknife** ως μεθόδων κατασκευής διαστημάτων εμπιστοσύνης για τις παραπάνω εκτιμήσεις.

* **Εκτίμηση συνάρτησης παλινδρόμησης** από της (τεχνικές εξομάλυνσης, smoothing)
* **Εκτίμηση πυκνότητας**  από 
* **Έλεγχοι στηριγμένοι στις τάξεις (ranks)** των  «rank tests»

### Αναλυτικό πρόγραμμα μαθήματος:

1. Μη παραμετρική στατιστική: Εισαγωγή (0.5 βδομ.)
2. Υπενθύμιση κάποιων ορισμών και προτάσεων (0.5 βδομ.)
3. Εκτίμηση της αθροιστικής συνάρτησης κατανομής (cdf) (0.5 βδομ.)
4. Εκτίμηση συναρτησιακών της cdf (αθρ. συναρτ. καταν.) (2.5 βδομ.)
5. Το Jackknife και το Βootstrap (2.5 βδομ.)
6. Μη παραμετρική εκτίμηση πυκνότητας (2 βδομ.)
7. Μη παραμετρική παλινδρόμηση Ι: τοπική εξομάλυνση (local smoothing) (1.5 βδομ.)
8. Μη παραμετρική παλινδρόμηση ΙΙ: splines και penalized regression (1 βδομ.)
9. Μη παραμετρική παλινδρόμηση ΙΙΙ: τοπική πιθανοφάνεια (local likelihood) (1 βδομ.)
10. Πολλαπλή μη παραμετρική παλινδρόμηση ΙV: GAM και Regression Trees (1 βδομ.)

### Σημειώσεις

Κορμός + εργαστήριο, αλλά καλύτερα files διαδικτυακού μαθήματος (αλλαγές, συμπληρωματικά στοιχεία)

### Εργαστήριο

* Rank tests (Θεωρία επιγραμματικά + εφαρμογή στην R)
* Eφαρμογή κορμού στην R

### Εργασίες

* +0.5 έως +2 αν θεωρητική εξέταση (γραπτή / προφορική) >=4

### Εγχειρίδια:

* Larry Wasserman (2006). All of Nonparametric Statistics, Springer

Που είναι διαθέσιμο και σαν electronic resource στην

<http://link.springer.com/book/10.1007/0-387-30623-4>.

* Breheny, P. (2012) , Nonparametric Statistics (<https://web.as.uky.edu/statistics/users/pbreheny/621/F12/notes.html>)
* Α. Μπατσίδης, Π. Παπασταμούλης, Κ. Πετρόπουλος, Α. Ρακιτζής (2022). Μη Παραμετρική Στατιστική, Θεωρία και εφαρμογές με χρήση R και S.P.S.S., Κάλλιπος ανοιχτές ακαδημαικές εκδόσεις.
* Κούτρας, Μ. και Τριανταφύλλου, Ι. (2022). Μη Παραμετρική Στατιστική, Θεωρία και Εφαρμογές, Εκδόσεις Τσότρας
* Zou, C. Lecture Notes on Nonparametric Statistics. (https://www.dropbox.com/sh/ff6xkunvb9emlc1/AACVVO60q5YQBpMpV97JP2y8a/Changliang%20Zou%20-%20Lecture%20Notes%20on%20Nonparametric%20Statistics.pdf?dl=0*)*
* Hollander, M. and Wolfe, D. A., (1999). Nonparametric Statistical Methods.

### Προτεινόμενη συμπληρωματική βιβλιογραφία:

Ως συνοδευτική βιβλιογραφία συνιστώνται και τα ακόλουθα βιβλία:

#### Bootstrap

* Efron ,B. and Tibshirani, R,(1998). An introduction to the Bootstrap, Chapman and Hall
* Shao, J. and Tu, D, (1996). The Jackknife and the Bootstrap, Springer

#### Simple non-parametric Regression

* Todd, P. (2014), Lecture Notes on Nonparametric Density and Regression Estimation (<http://athena.sas.upenn.edu/petra/class721/nonpar3new.pdf>)
* Fan, J. and Gijbels, I. (1996). Local polynomial modeling and its applications, Chapman & Hall.
* Fox, J. (2000). Nonparametric Simple Regression: Smoothing Scatterplots. Sage Publications.
* Loader, C (1999). Local Regression and Likelihood, Springer
* Silverman, B.W.(1986). Density Estimation for Statistics and Data Analysis. Chapman and Hall.
* Wand, M. P. and Jones, M. C. (1994). Kernel Smoothing. Chapman and Hall.

#### Multiple non-parametric Regression/smoothing

* Hastie, T. J. and Tibshirani, R. J. (1990). Generalized Additive Models. Chapman and Hall.
* Wood, S (2006). Generalized Additive Models, An introduction with R, Chapman & Hall.

#### Rank tests

* Gibbons and Chakraborti, (2003). Nonparametric Statistical Inference, Marcel Dekker
* Kloke J. and McKean, J (2015), Nonparametric Statistical Methods Using R, CRC press.
* Sidak, Z., Sen, P. K. and Hajek, J. (1999). Theory of Rank Tests. Academic Press.
* Ξεκαλάκη , Ε. (2001). Μη παραμετρική στατιστική.

#### Other classics

* Hajek, J. (1969). A Course in Nonparametric Statistics. Holden Day.
* Higgins, J. (2004). Introduction to Modern Nonparametric Statistics. Thomson/Brooks/Cole, New York.
* Hettmansperger, T. and McKean, J. (2011). Robust Νonparametric Statistical Methods. Boca Ration : CRC/Taylor & Francis.