Θεωρία Αποφάσεων

Σ. Λυκοθανάσης, Καθηγητής

Δ. Κοσμόπουλος, Αναπληρωτής Καθηγητής

Τμήμα Μηχανικών Η/Υ & Πληροφορικής -
Εργαστήριο Αναγνώρισης Προτύπων

Διευθυντής: Σ. Λυκοθανάσης, Καθηγητής

Μη παραμετρικές μέθοδοι

- Ούτε η σ.π.π είναι γνωστή, ούτε η διακρίνουσα συνάρτηση

Συμβαίνει συχνά!

Το μόνο που έχουμε είναι τα δεδομένα (με ετικέτες)

Σολομός Πέρκα Σολομός Σολομός

Θέλουμε να εκτιμήσουμε την σ.π.π. από τα δεδομένα

Μη παραμετρικές μέθοδοι

- Σε προηγούμενα μαθήματα υποθέσαμε ότι

1. είτε είναι γνωστή η p(x)

Δεν συμβαίνει σχεδόν ποτέ στην πράξη

2. είτε είναι γνωστή η p(x|c)

Συμβαίνει κάποιες φορές, **αλλά:**

• Στην πράξη οι περισσότερες κατανομές είναι multimodal περισσότερα από ένα μέγιστα), άρα πιο σύνθετα μοντέλα για παραμετροποίηση.

• Η προσέγγιση των πολυδιάστατων κατανομών σαν γινόμενο μονοδιάστατων δεν δουλεύει τόσο καλά στην πράξη.

Μη παραμετρικές τεχνικές

- Μη παραμετρικές τεχνικές μπορούν να χρησιμοποιηθούν με αυθαίρετες κατανομές χωρίς καθόλου υποθέσεις για την μορφή τους

Θα ασχοληθούμε με δυο μεθόδους μη παραμετρικής εκτίμησης:

Παράθυρα Parzen

Εκτιμούν πιθανοφάνεια $p(x|c_j)$

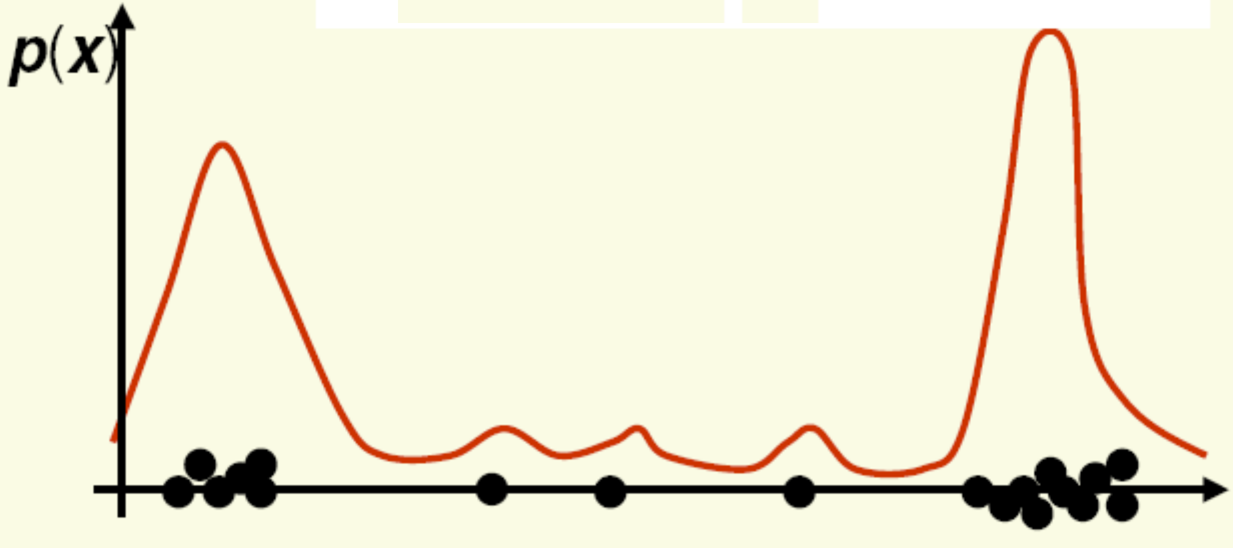
Μέθοδος των πλησιέστερων Γειτόνων (Nearest Neighbors)

Μπορούν απ' ευθείας να υπολογίσουν τις εκ των υστέρων πιθανότητες $P(c_j|x)$

Μη παραμετρικές τεχνικές

Ιδέα: Όσα περισσότερα τα δεδομένα σε μια περιοχή, τόσο μεγαλυτέρη η τιμή της σ.π.π

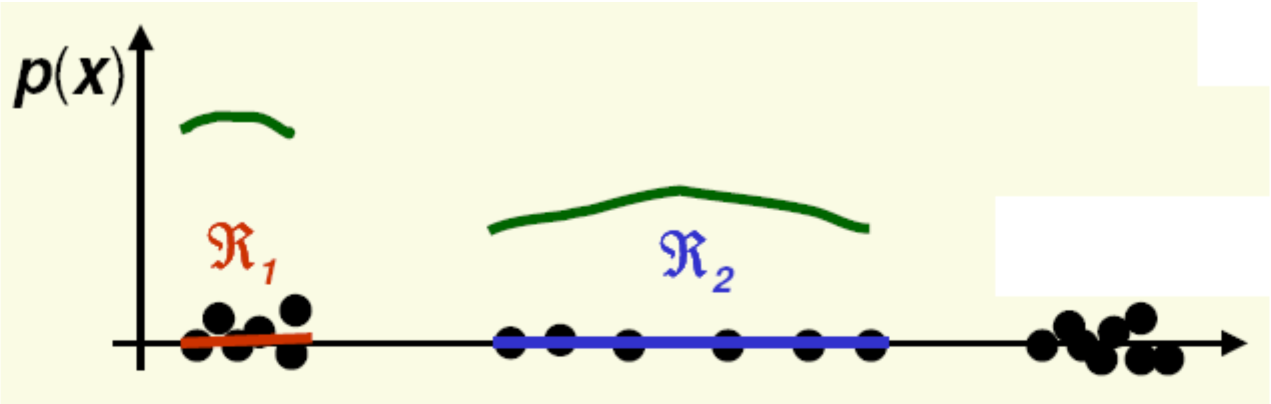
$$Pr[X \in \Re] = \int_{\Re} f(x)dx$$



Μη παραμετρικές τεχνικές

$Pr[X \in \Re_1]$ = ? $Pr[X \in \Re_2]$ = ?

$$Pr[X \in \Re] = \int_{\Re} f(x)dx$$

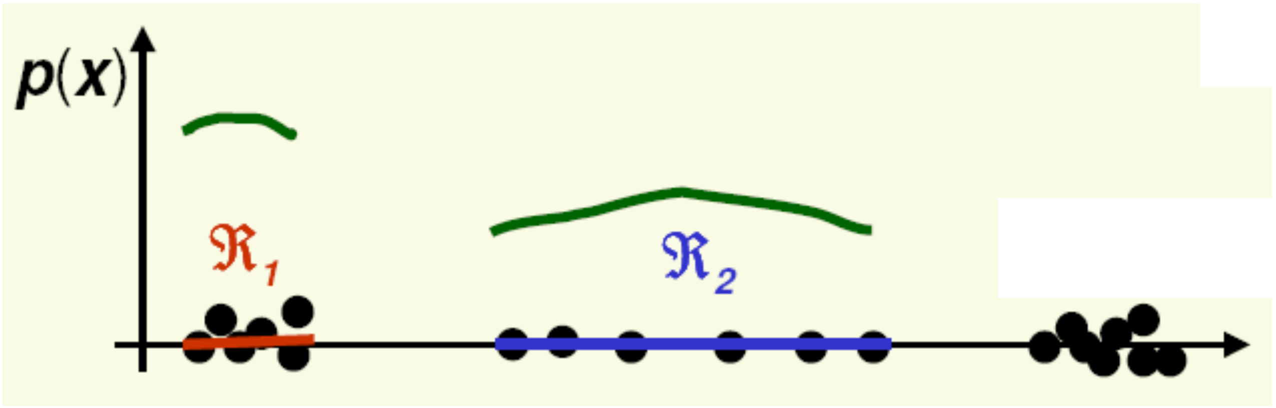


Μη παραμετρικές τεχνικές

$$Pr[X \in \Re] = \int_{\Re} f(x)dx$$

$$Pr[X \in \Re_1] \approx \frac{6}{20} \qquad Pr[X \in \Re_2] \approx \frac{6}{20}$$

Η τιμή της πυκνότητας είναι ίση για τις δύο περιοχές?



Μη παραμετρικές τεχνικές

$$Pr[X \in \Re] = \int_{\Re} f(x)dx$$

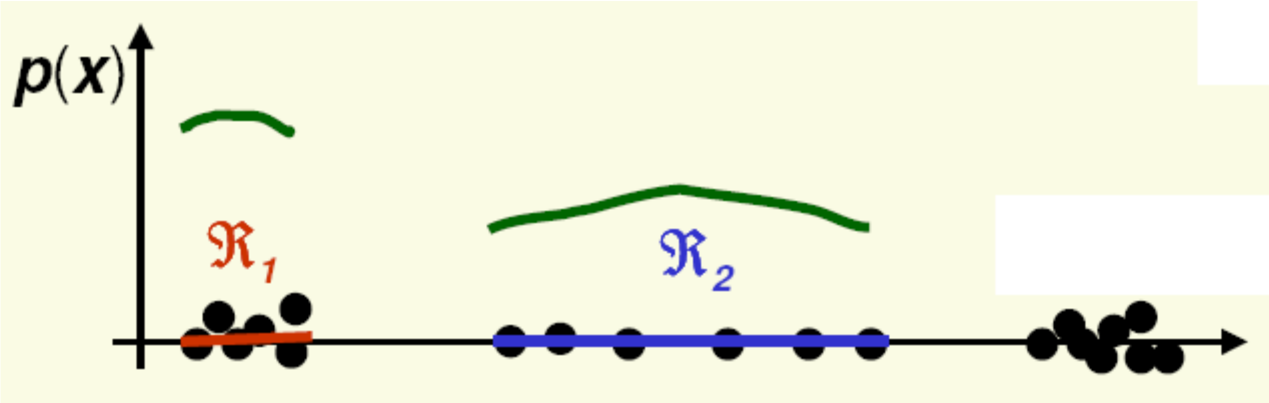
$$Pr[X \in \Re_1] \approx \frac{6}{20} \qquad Pr[X \in \Re_2] \approx \frac{6}{20}$$

Η τιμή της πυκνότητας είναι ίση για τις δύο περιοχές?

Όχι διότι η περιοχή R_1 είναι μικρότερη από την R_2

$$Pr[X \in \Re_1] = \int_{\Re_1} f(x)dx \approx \int_{\Re_2} f(x)dx = Pr[X \in \Re_2]$$

Τι πρέπει να κάνω για να βρω την πυκνότητα?



Μη παραμετρικές τεχνικές

$$Pr[X \in \Re] = \int_{\Re} f(x)dx$$

$$Pr[X \in \Re_1] \approx \frac{6}{20} \qquad Pr[X \in \Re_2] \approx \frac{6}{20}$$

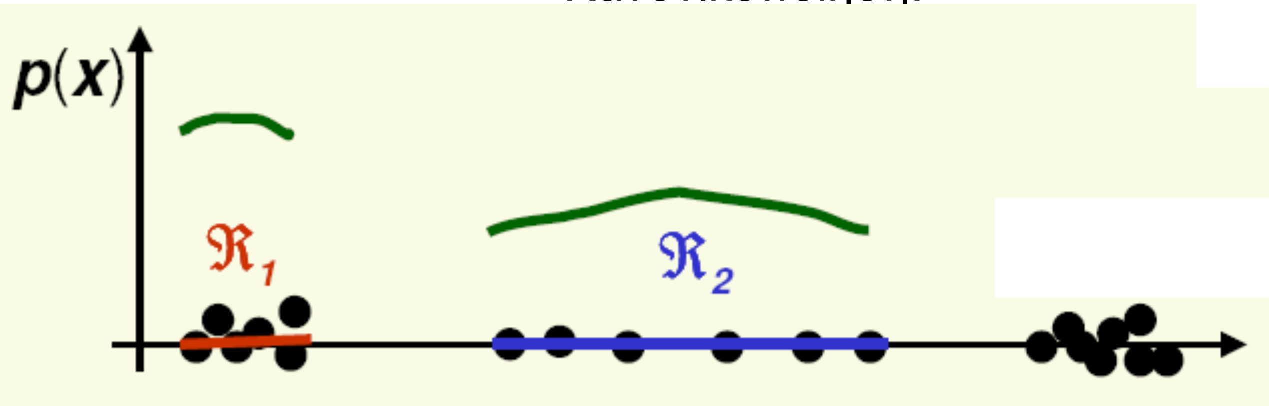
Η τιμή της πυκνότητας είναι ίση για τις δύο περιοχές?

Όχι διότι η περιοχή R_1 είναι μικρότερη από την R_2

$$Pr[X \in \Re_1] = \int_{\Re_1} f(x)dx \approx \int_{\Re_2} f(x)dx = Pr[X \in \Re_2]$$

Τι πρέπει να κάνω για να βρω την πυκνότητα?

Κανονικοποίηση!



Μη παραμετρικές τεχνικές

- Θεωρώντας ότι η f(x) είναι σχετικά επίπεδη μέσα στην περιοχή R

$$\frac{\#\text{δειγμάτων στη περιοχή R}}{\text{Συνολικός \# των δειγμάτων}} \approx Pr[X \in \Re] = \int_{\Re} f(y)dy \approx f(x) * Volume(\Re)$$

Άρα η τιμή της πυκνότητας σε ένα σημείο **x** μέσα στην περιοχή R μπορεί να εκτιμηθεί ως:

$$f(x) \approx \frac{\#\text{δειγμάτων στη περιοχή R}}{\text{Συνολικός \# των δειγμάτων}} \frac{1}{Volume(\Re)}$$

Και τώρα τα ίδια με πιο «μαθηματικό» τρόπο...

Μη παραμετρικές τεχνικές

Πιθανότητα P ότι ένα διάνυσμα $\mathbf{x}$ θα «πέσει» στην περιοχή R $$P = \int_R p(\mathbf{x}')d\mathbf{x}'$$

Υποθέτουμε ότι έχουμε δείγματα $\mathbf{x}_1\ \mathbf{x}_2 \ldots \mathbf{x}_n$ από την σ.π.π. $p(\mathbf{x})$. Η πιθανότητα ότι k δείγματα από τα n θα «πέσουν» στην περιοχή R δίνονται από την διωνυμική κατανομή:

$$P(k) = P_k = \binom{n}{k} P^k (1-P)^{n-k}$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Άρα k είναι τυχαία μεταβλητή που ακολουθεί τη διωνυμική κατανομή. Από τις ιδιότητες της διωνυμικής κατανομής:

$$E[k] = nP \quad E[k/n] = P$$

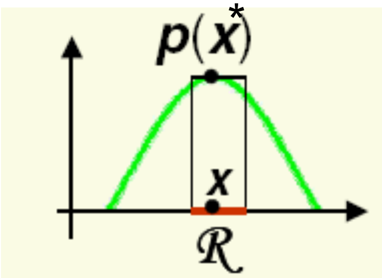
$$Var[k/n] = E[(k/n - P)^2] = \frac{P(1-P)}{n}$$

Όταν $n \to \infty$ Ο δεξιός όρος μηδενίζεται, άρα και ο αριστερός, αρα: Εκτίμηση της P: $P \approx k/n$

Μη παραμετρικές τεχνικές

Υποθέτουμε ότι η *p*(x) είναι συνεχής και ότι η περιοχή R είναι τόσο μικρή ώστε η *p*(x) να μπορεί να θεωρηθεί σταθερή στην περιοχή:

$$P = P(\mathbf{x} \in R) = \int_{\mathbf{x}' \in R} p(\mathbf{x}')d\mathbf{x}'$$



$$P = \int_{\mathbf{x}' \in R} p(\mathbf{x}')d\mathbf{x}' \approx p(\mathbf{x}^*) \cdot V \approx k / n$$

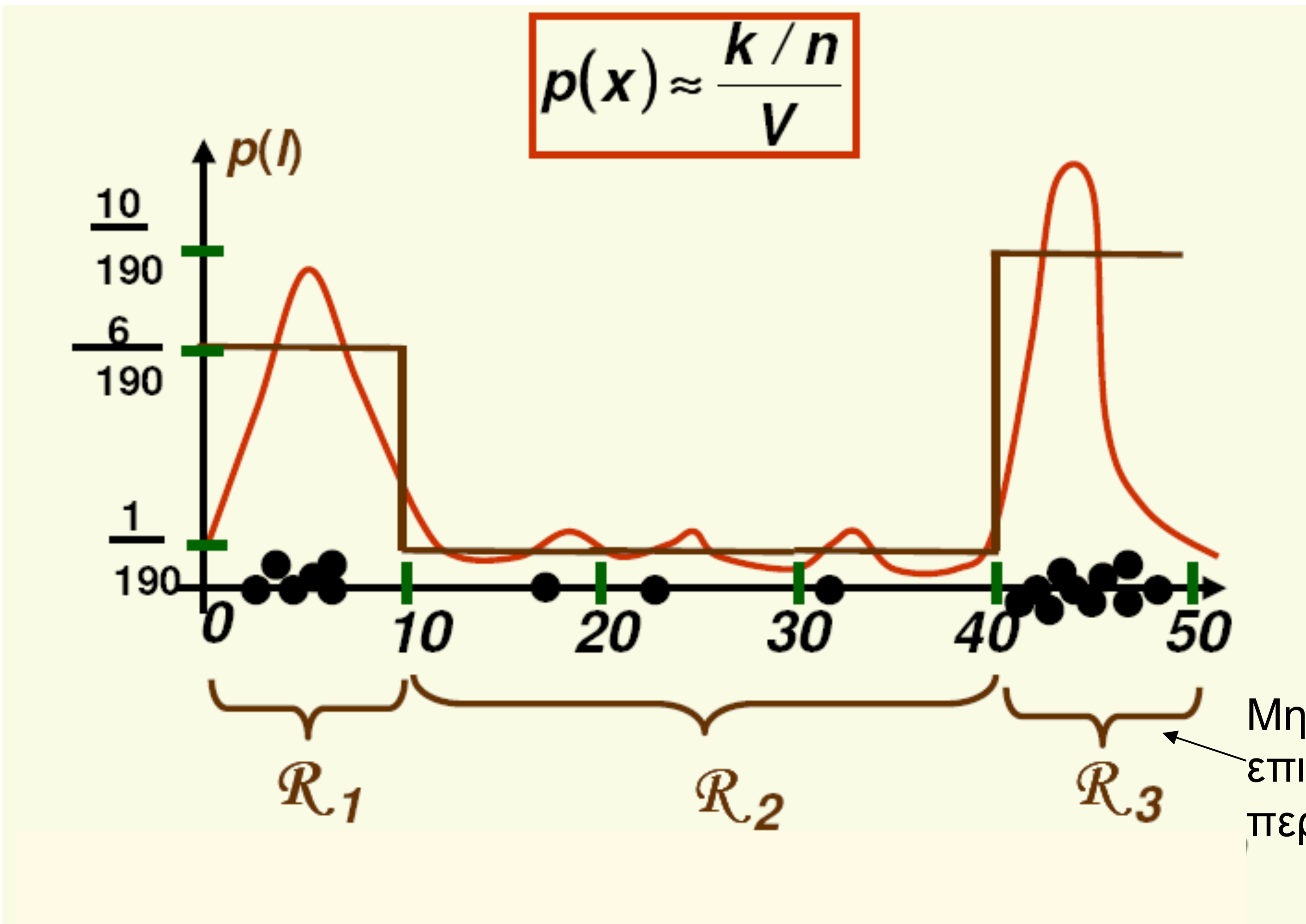
Από την προηγούμενη διαφάνεια

V: Εμβαδό της περιοχής R. Στην μονοδιάστατη περίπτωση, V= μήκο
k: Πλήθος δειγμάτων που βρίσκονται εντός της περιοχής R
n: Συνολικό πλήθος δειγμάτων

Οπότε

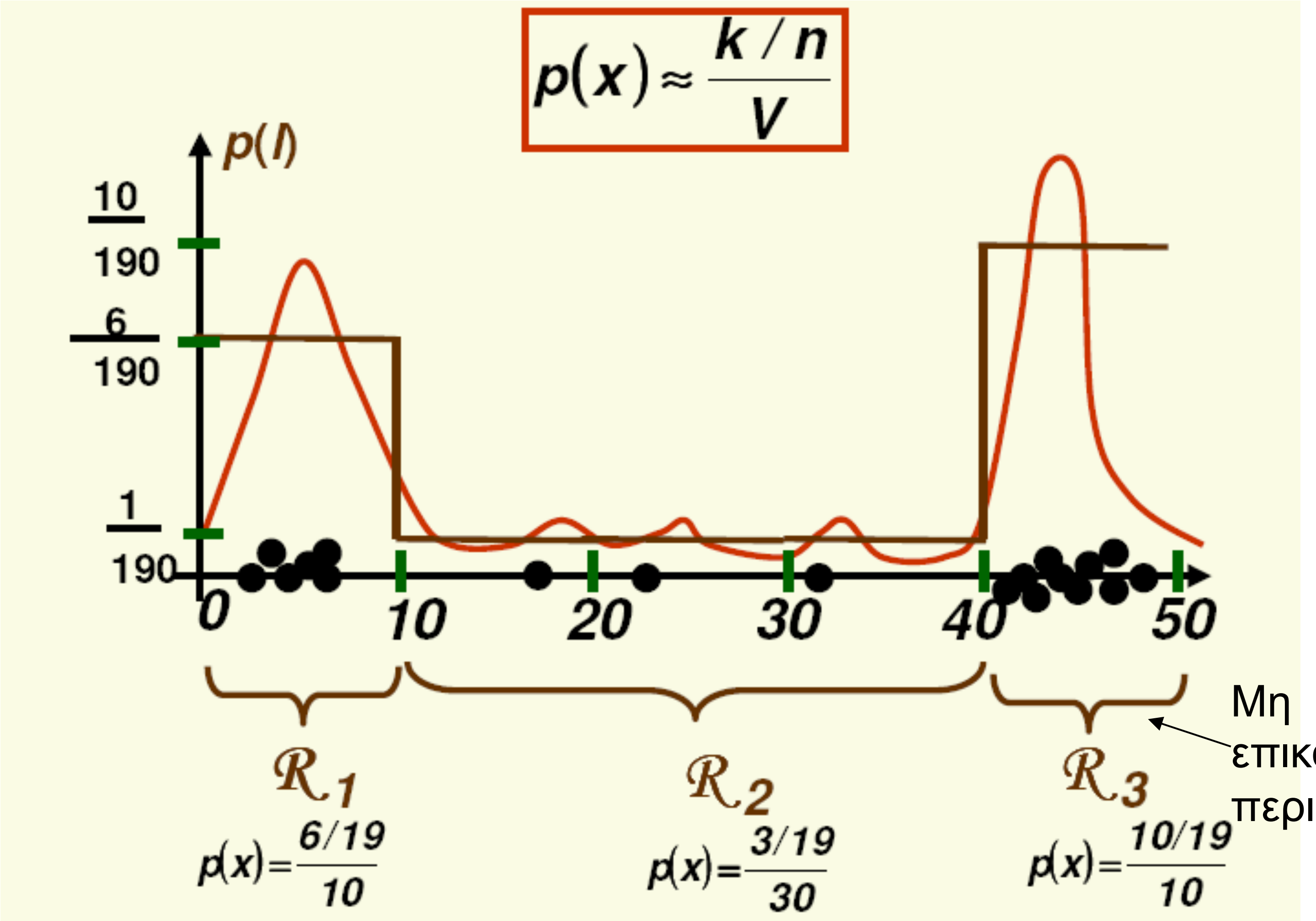
$$\boxed{p(x) \approx \frac{k/n}{V}}$$

Ιστόγραμμα

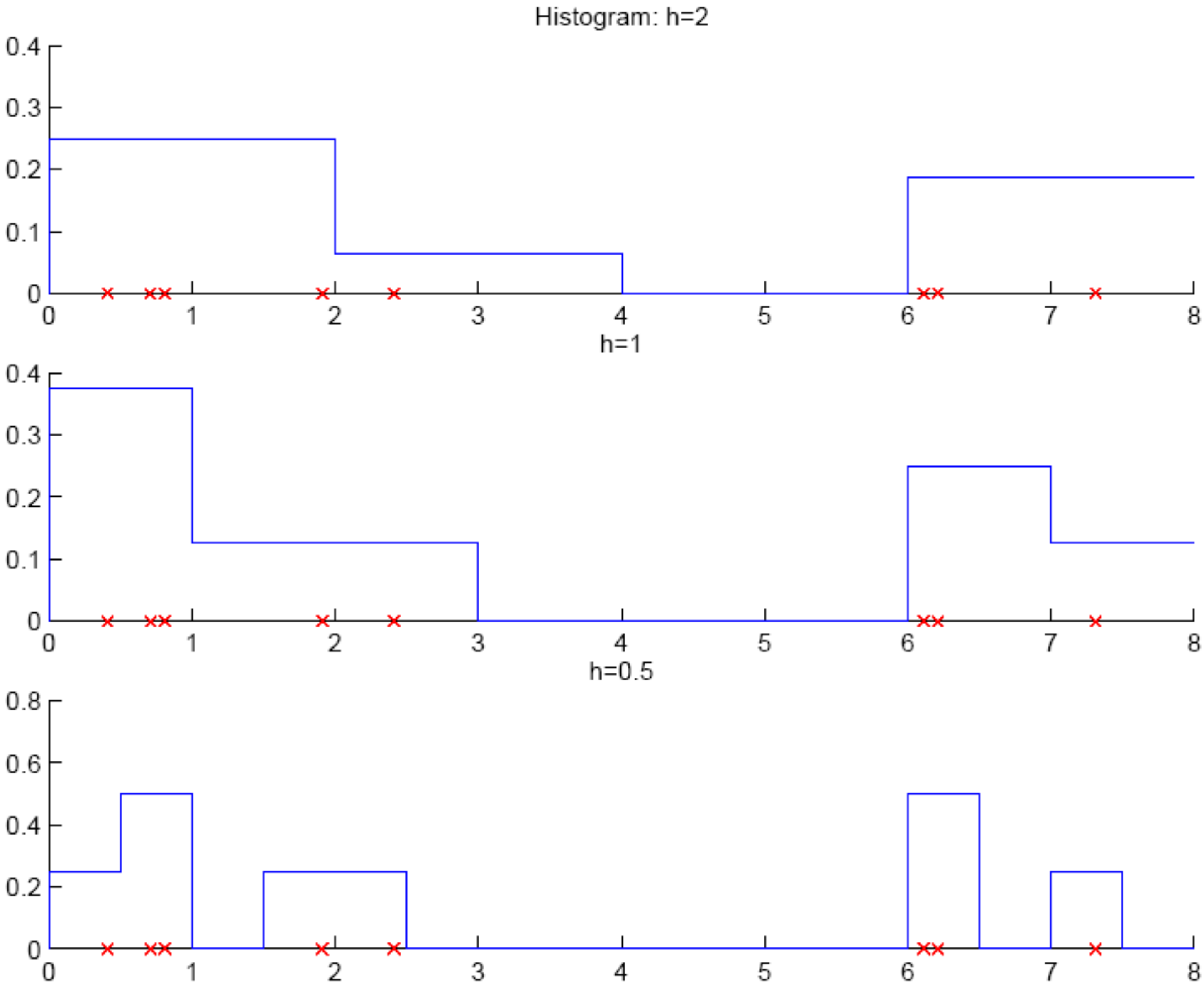


Μη επικαλυπτόμενες περιοχές

Ιστόγραμμα



Μη επικαλυπτόμενες περιοχές



h: το μέγεθος του bin. Εκτός από την δυσκολία επιλογής του κατάλληλου bin τι άλλα προβλήματα παρουσιάζουν τα ιστογράμματα?

Τα ιστογράμματα είναι ένας πολύ απλός τρόπος εκτίμησης της πυκνότητας, αλλά έχουν πολλά μειονεκτήματα:

- Η εκτίμηση της πυκνότητας εξαρτάται από την αρχική θέση των στηλών (bins).
- Οι ασυνέχειες των εκτιμήσεων δεν οφείλονται στην υποκείμενη πυκνότητα, αλλά συχνά είναι αποτέλεσμα της επιλογής των στηλών.
- Οι ασυνέχειες κάνουν δύσκολη την κατανόηση της δομής των δεδομένων.
- Ένα πιο σοβαρό πρόβλημα είναι η κατάρα της διαστασιμότητας, γιατί ο αριθμός των στηλών αυξάνει εκθετικά με τον αριθμό των διαστάσεων.
- Στις μεγάλες διαστάσεις θα απαιτούσαμε ένα πολύ μεγάλο αριθμό παραδειγμάτων, διαφορετικά οι περισσότερες στήλες θα είναι άδειες.

- Όλα αυτά τα μειονεκτήματα κάνουν τα ιστογράμματα ακατάλληλα για τις περισσότερες πρακτικές εφαρμογές, εκτός από τη γρήγορη οπτικοποίηση αποτελεσμάτων σε μία ή δύο διαστάσεις.

Παράδειγμα

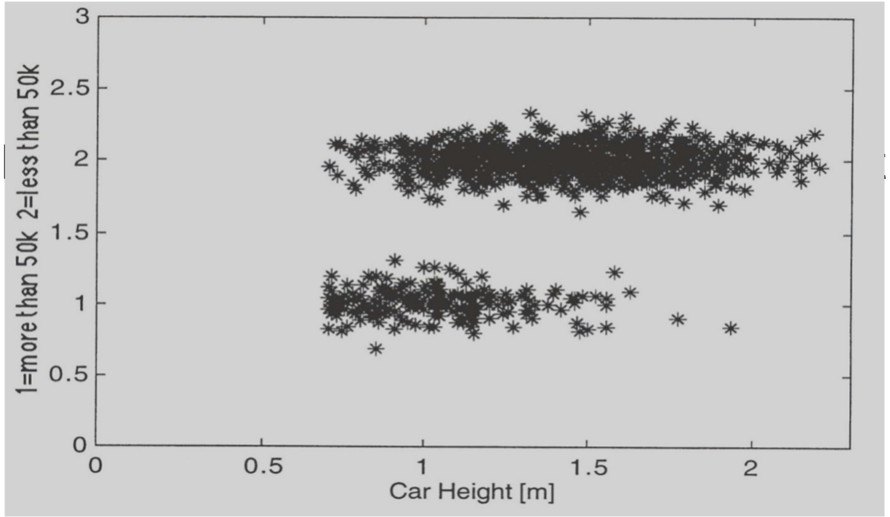
- Ταξινόμησε αυτοκίνητα στην πανεπιστημιούπολη ανάλογα με το αν κοστίζουν περισσότερο ή λιγότερο από €50K:
  - C_1: τιμή > € 50K
  - C_2: τιμή < € 50K
  - Χαρακτηριστικό x: ύψος αυτοκινήτου
- Από τον κανόνα του Bayes γνωρίζουμε ότι:

$$P(C_i / x) = \frac{p(x / C_i) P(C_i)}{p(x)}$$

- Πρέπει να υπολογίσουμε *p(x/C_1), p(x/C_2), P(C_1), P(C_2)*

Παράδειγμα

- Καθόρισε τις εκ των προτέρων πιθανότητες
  - Συλλογή δεδομένων
  - πχ., 1209 δείγματα: $\#C_1=221$ $\#C_2=988$



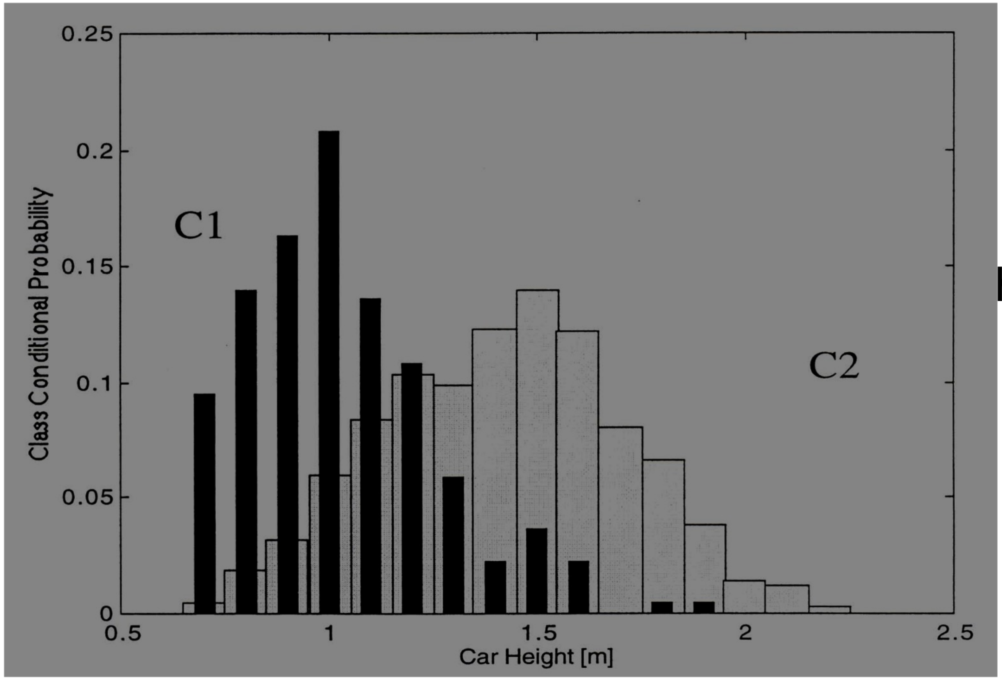
$$P(C_1) = \frac{221}{1209} = 0.183$$

$$P(C_2) = \frac{988}{1209} = 0.817$$

Παράδειγμα

- Καθόρισε πιθανοφάνειες
  - Διακριτοποίησε τις τιμές του ύψους και δημιούργησε κανονικοποιημένο ιστόγραμμα.

$P(x|C_i)$



ht

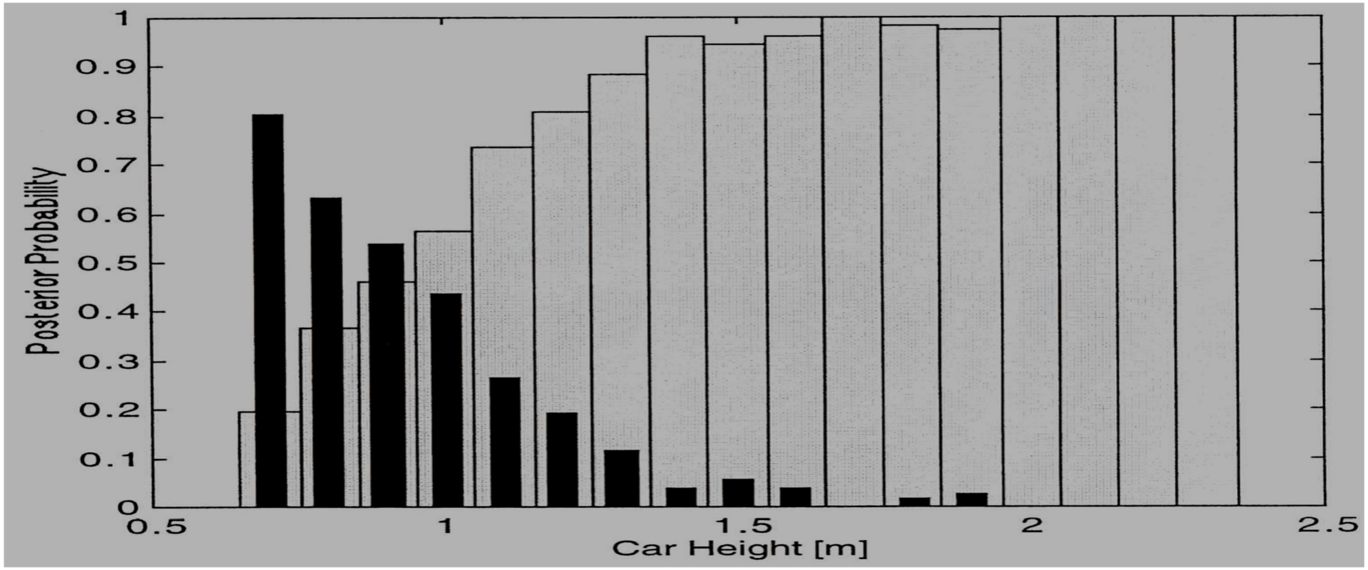
Παράδειγμα

- Υπολόγισε την εκ των υστέρων πιθανότητα (για κάθε bin):

$$P(C_1 / x = 1.0) = \frac{p(x = 1.0 / C_1) P(C_1)}{p(x = 1.0 / C_1) P(C_1) + p(x = 1.0 / C_2) P(C_2)} =$$

$$= \frac{0.2081 * 0.183}{0.2081 * 0.183 + 0.0597 * 0.817} = 0.438$$

- Για ταξινόμηση αρκεί ο αριθμητής. Γιατί;



Ακρίβεια

Δεύτερη
(σταθερή σππ)

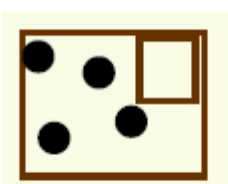
Πόσο ακριβής είναι η εκτίμηση $\boxed{p(x) \approx \dfrac{k/n}{V}}$

Πρώτη: (δείγματα διαθέσιμα στην περιοχή δειγματοληψίας)

Κάναμε δυο υποθέσεις: $$P = \int_{\mathbf{x}' \in R} p(\mathbf{x}')d\mathbf{x}' \approx p(\mathbf{x}^*)\cdot V \approx k/n$$

V: Εμβαδό της περιοχής R. Στην μονοδιάστατη περίπτωση, V= μήκο
k: Πλήθος δειγμάτων που βρίσκονται εντός της περιοχής R
n: Συνολικό πλήθος δειγμάτων

Καθώς μικραίνει η περιοχή R η εκτίμηση γίνεται πιο ακριβής



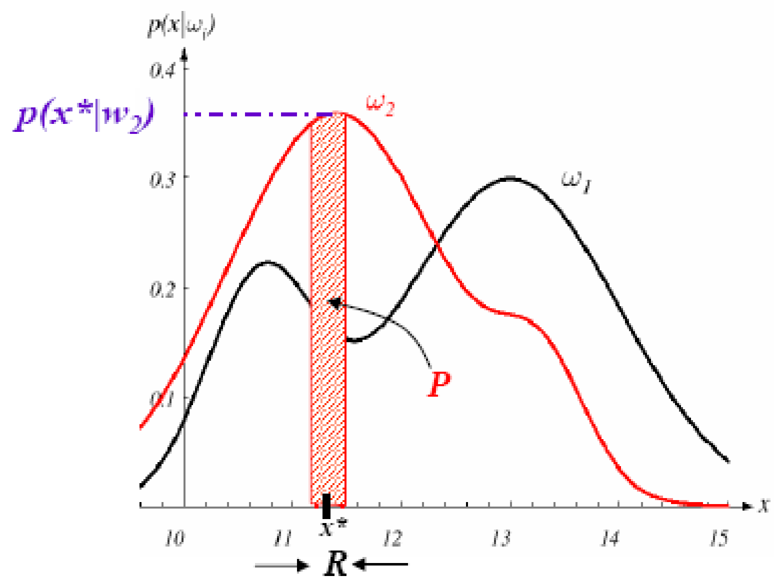
Στην πράξη όμως πρέπει να προσέχουμε ώστε να υπάρχουν κάποια δείγματα μέσα στην περιοχή R διαφορετικά p(x)=0 για όλα τα x στην περιοχή R

Άρα σε περίπτωση που είχαμε άπειρο αριθμό δειγμάτων η εκτίμηση θα γινόταν πιο ακριβής, όσο θα αυξάναμε τον αριθμό των δειγμάτων, ενώ ταυτόχρονα θα μειώναμε την περιοχή, όχι όμως πάρα πολύ, ώστε και πάλι να περιέχει έναν αρκετά μεγάλο αριθμό δειγμάτων

Εκτίμηση Κατανομών

Για την εκτίμηση της κατανομής στο *x, επιλέγουμε μια σειρά περιοχών R_1, R_2,…,R_n,* που περιέχουν το *x. Έστω V_n ο όγκος της R_n, k_n το πλήθος των δειγμάτων που βρίσκονται εντός της n-στής περιοχής, και $p_n(x)$* η *n-στή εκτίμηση της p(x). Τότε,*

$$\boxed{p_n(\mathbf{x}) = \frac{k_n/n}{V_n} \xrightarrow[n\to\infty]{} p(\mathbf{x})}$$



Για να είναι το *k/n μια καλή εκτίμηση του P,* και άρα το *$p_n(x)$ μια καλή εκτίμηση του p(x),* θα πρέπει να ικανοποιούνται τα εξής:

$$\lim_{n\to\infty} V_n = 0$$

$$\lim_{n\to\infty} k_n = \infty$$

$$\lim_{n\to\infty} k_n/n = 0$$

Ακρίβεια

- Στην πράξη ο αριθμός των δειγμάτων είναι περιορισμένος

Άρα η μόνη επιλογή για να αυξηθεί η ακρίβεια είναι να μειωθεί το μέγεθος της R (να μειωθεί ο όγκος V)

Αν V είναι πολύ μικρό p(x)=0 για τα περισσότερα x, διότι οι περιοχές δεν θα περιέχουν δείγματα

Άρα πρέπει να βρούμε συμβιβαστική λύση για τον όγκο V:

✓Όχι τόσο μικρός ώστε να περιέχει αρκετά δείγματα

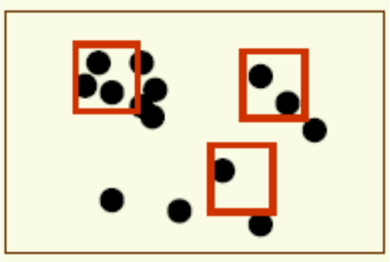
✓Όχι τόσο μεγάλος ώστε η p(x) να μπορεί να θεωρηθεί σταθερή μέσα στον όγκο

Δυο προσεγγίσεις

$$p(x) \approx \frac{k/n}{V}$$

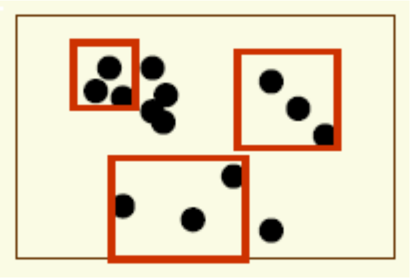
Παράθυρα Parzen

Επίλεξε σταθερό *V* και υπολόγισε *k* από τα δεδομένα



Εκτίμηση πλησιέστερων γειτόνων

Επίλεξε σταθερό *k* και υπολόγισε *V* από τα δεδομένα



Υπό κατάλληλες συνθήκες και καθώς ο αριθμός των δειγμάτων τείνει προς το άπειρο και οι δυο εκτιμήσεις μπορεί να αποδειχθεί ότι συγκλίνουν στην πραγματική p(x)

Εκτίμηση Κατανομών

- Υπάρχουν δύο τρόποι για τη δημιουργία ακολουθιών περιοχών R_i ώστε να συγκλίνει η $\boldsymbol{p_n(x)}$ *στην* $\boldsymbol{p(x)}$:

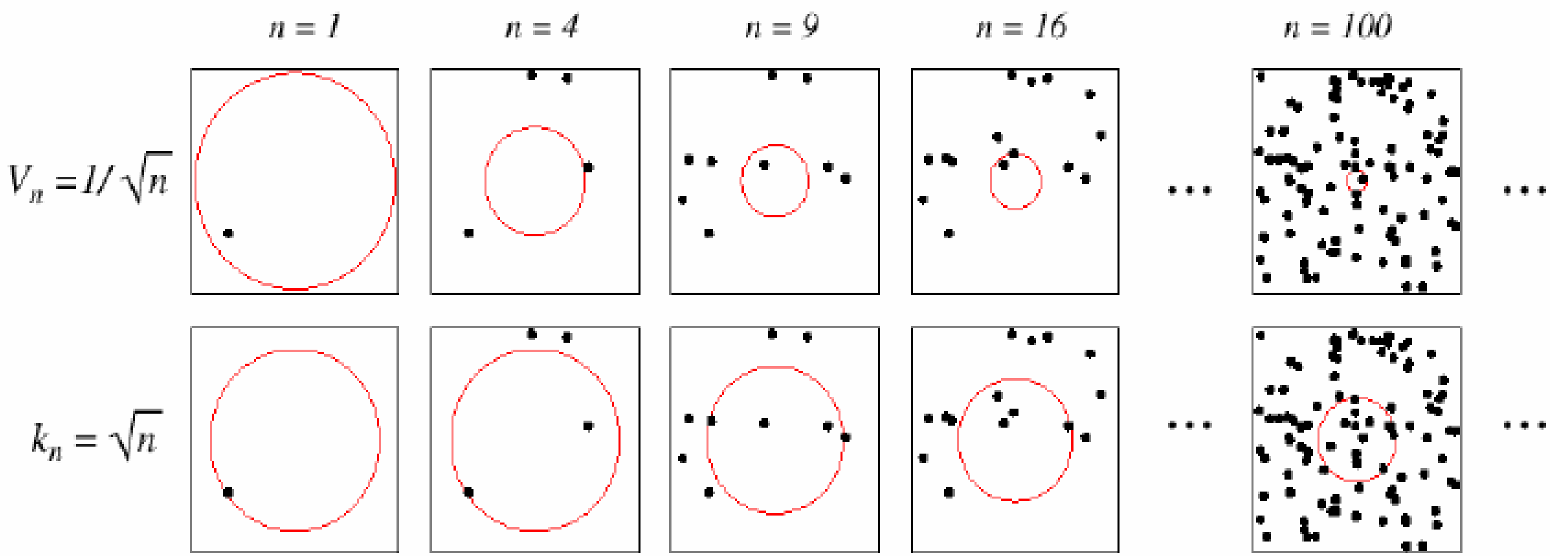
Μειώνουμε τον όγκο μιας αρχικής περιοχής ορίζοντας μια ακολουθία όγκων V_n ως συναρτήσεων του n, π.χ. $V_n = V_1 / \sqrt{n}$

→ Εκτίμηση πυκνότητας με τη μέθοδο των Παραθύρων Parzen (Parzen Windows)

Ορίζουμε το k_n σαν συνάρτηση του n, $k_n = \sqrt{n}$ οπότε το V_n αυξάνει έως ότου περιλάβει k_n δείγματα.

→ Εκτίμηση πυκνότητας με τη μέθοδο των k_n Πλεισιέστερων Γειτόνων (k_n-Nearest Neighbor)

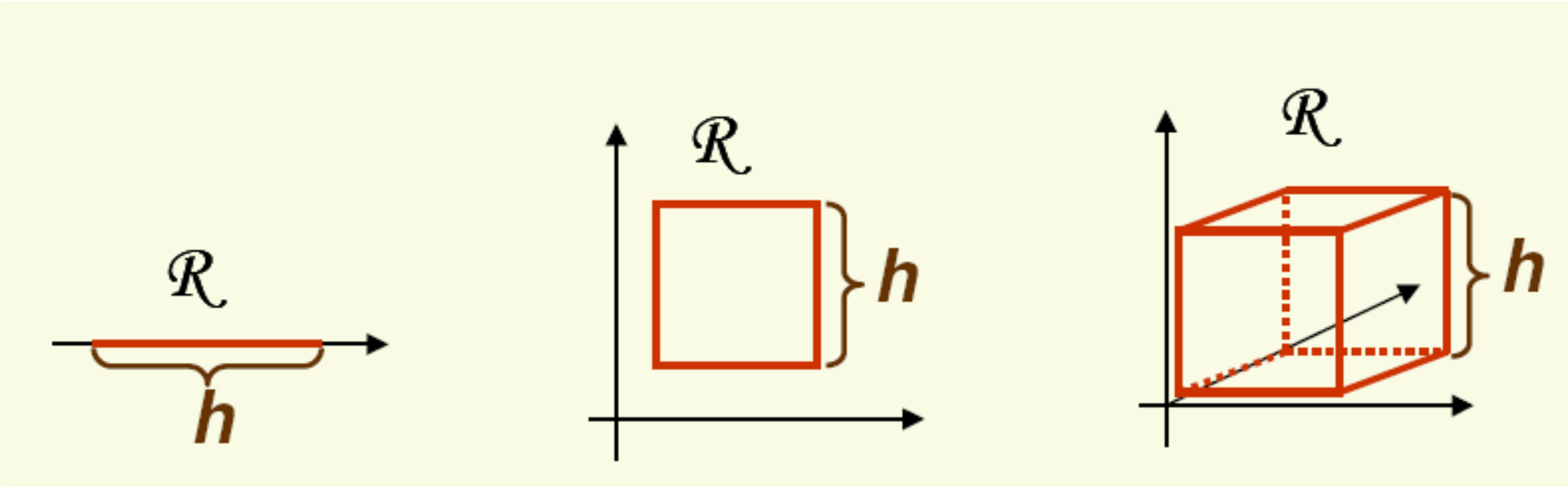
Δύο Προσεγγίσεις



Παράθυρα Parzen

• Στην προσέγγιση των παραθύρων Parzen προεπιλέγουμε το μέγεθος και την μορφή της περιοχής R.

• Υποθέτουμε ότι η περιοχή R είναι ένα *d*-διάστατος υπερκύβος με πλευρά μήκους **h** και άρα όγκο $\mathbf{h}^{d}$

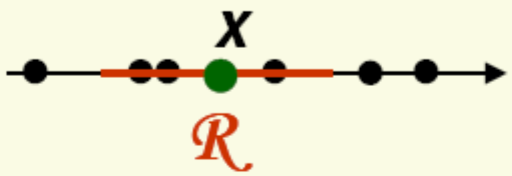


1-D 2-D 3-D

Παράθυρα Parzen

Για να εκτιμήσουμε την σ.π.π. στο **x** , «κεντράρουμε» την περιοχή R στο **x**, μετράμε τα δείγματα που βρίσκονται στην περιοχή και αντικαθιστούμε τις τιμές στον τύπο:

$$p(x) \approx \frac{k/n}{V}$$



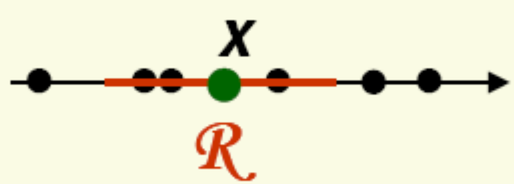
V=10

$$p(x) \approx \ ?$$

Παράθυρα Parzen

Για να εκτιμήσουμε την σ.π.π. στο **x** , «κεντράρουμε» την περιοχή R στο **x**, μετράμε τα δείγματα που βρίσκονται στην περιοχή και αντικαθιστούμε τις τιμές στον τύπο:

$$p(x) \approx \frac{k/n}{V}$$



$$p(x) \approx \frac{3/6}{10}$$

Παράθυρα Parzen

Θέλουμε αναλυτική έκφραση για την σ.π.π.

Ορίζουμε μια συνάρτηση παραθύρου

$$\varphi(u) = \begin{cases} 1 & |u_j| \le \frac{1}{2} \quad j = 1, \ldots, d \\ 0 & \text{διαφορετικά} \end{cases}$$

?

1-D

?

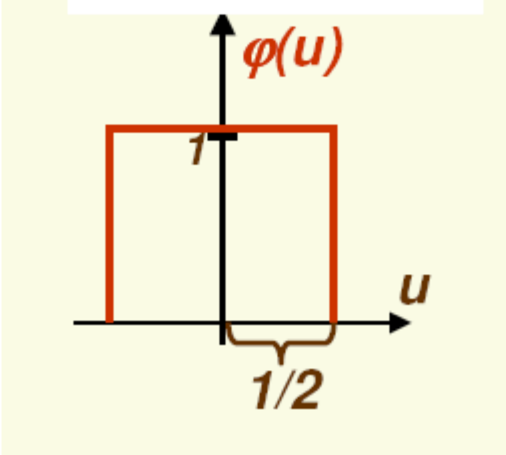
2-D

Παράθυρα Parzen

Θέλουμε αναλυτική έκφραση για την σ.π.π.

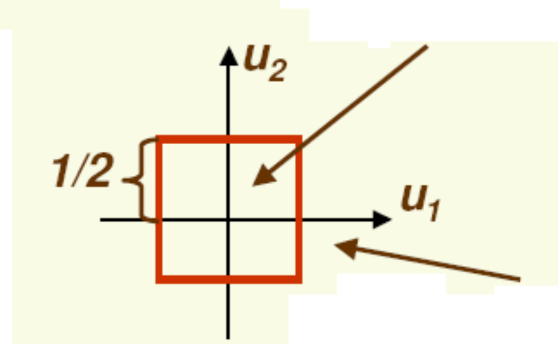
Ορίζουμε μια συνάρτηση παραθύρου

$$\varphi(u) = \begin{cases} 1 & |u_j| \leq \frac{1}{2} \quad j = 1, \dots, d \\ 0 & \text{διαφορετικά} \end{cases}$$



1-D

φ=1 μέσα



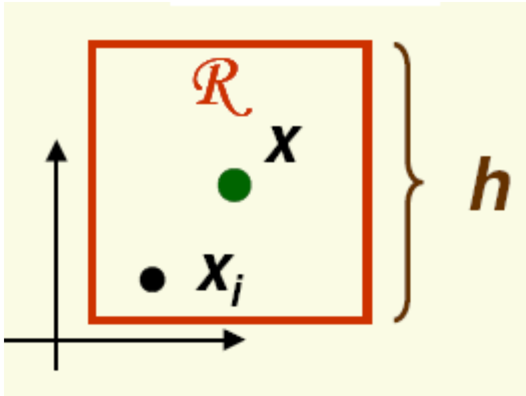
φ=0 έξω

2-D

Παράθυρα Parzen

Για τα δείγματα μας $x_1, x_2, \ldots, x_n$

$$\varphi\left(\frac{x - x_i}{h}\right) = \begin{cases} 1 & \left|x - x_i\right| \le \dfrac{h}{2} \quad j = 1, \ldots, d \\ 0 & \text{διαφορετικά} \end{cases}$$

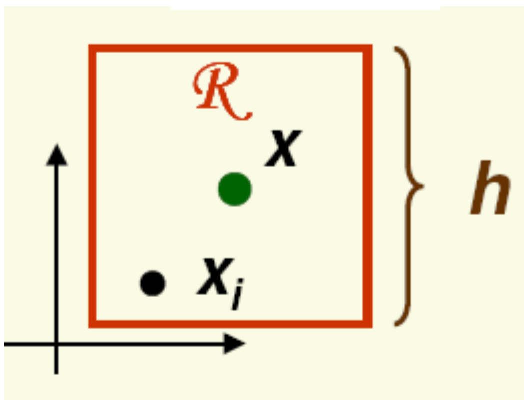


$$\varphi\left(\frac{x - x_i}{h}\right) = \begin{cases} 1 & \text{Όταν ... ?} \\ 0 & \text{Όταν ... ?} \end{cases}$$

Παράθυρα Parzen

Για τα δείγματά μας $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$

$$\varphi\left(\frac{x - x_i}{h}\right) = \begin{cases} 1 & \left|x - x_i\right| \le \dfrac{h}{2} \quad j = 1, \ldots, d \\ 0 & \text{διαφορετικά} \end{cases}$$



$$\varphi\left(\frac{x - x_i}{h}\right) = \begin{cases} 1 & \text{Όταν το } x_i \text{ βρίσκεται μέσα σε υπερκύβο «κεντραρισμένο» στο x και με πλευρά h} \\ 0 & \text{διαφορετικά} \end{cases}$$

Παράθυρα Parzen

$$\varphi\left(\frac{x - x_i}{h}\right) = \begin{cases} 1 \\ 0 \end{cases}$$

Πώς να μετρήσουμε τον συνολικό αριθμό των δειγμάτων $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ που βρίσκονται στον υπερκύβο με κέντρο το **x** και με πλευρά h

Παράθυρα Parzen

Πώς να μετρήσουμε τον συνολικό αριθμό των δειγμάτων $x_1, x_2, \ldots, x_n$ που βρίσκονται στον υπερκύβο με κέντρο το **x** και με πλευρά h :

$$k = \sum_{i=1}^{i=n} \varphi\left(\frac{x - x_i}{h}\right)$$

Οπότε βρίσκουμε την αναλυτική έκφραση της $p_\varphi(x)$ ως ?

Παράθυρα Parzen

Πώς να μετρήσουμε τον συνολικό αριθμό των δειγμάτων $x_1, x_2, \ldots, x_n$ που βρίσκονται στον υπερκύβο με κέντρο το **x** και με πλευρά h

$$k = \sum_{i=1}^{i=n} \varphi\left(\frac{x - x_i}{h}\right)$$

Θυμηθείτε: $p(x) \approx \frac{k/n}{V}$

Οπότε βρίσκουμε την αναλυτική έκφραση της $p_\varphi(x)$ ως

$$p_\varphi(x) = \frac{\overbrace{\sum_{i=1}^{i=n} \varphi\left(\frac{x - x_i}{h}\right)}^{=k} / n}{\underbrace{h^d}_{=V}} = \frac{1}{n} \sum_{i=1}^{i=n} \frac{1}{h^d} \varphi\left(\frac{x - x_i}{h}\right)$$

Παράθυρα Parzen

$$p_\varphi(x) = \frac{1}{n}\sum_{i=1}^{i=n} \frac{1}{h^d}\,\varphi\!\left(\frac{x - x_i}{h}\right)$$

Ας σιγουρευτούμε ότι η $p_\varphi(x)$ είναι πράγματι μια σ.π.π.

Τι πρέπει να ισχύει?

Παράθυρα Parzen

$$p_\varphi(x) = \frac{1}{n}\sum_{i=1}^{i=n}\frac{1}{h^d}\varphi\left(\frac{x - x_i}{h}\right)$$

Ας σιγουρευτούμε ότι η $p_\varphi(x)$ είναι πράγματι μια σ.π.π.

$$p_\varphi(x) \geq 0 \qquad \forall x$$

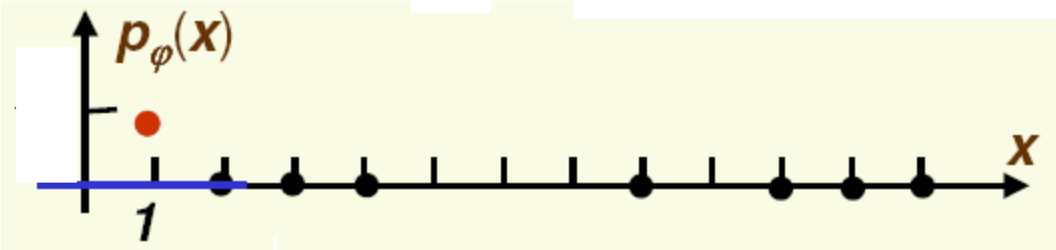
$$\int p_\varphi(x)dx = \int \frac{1}{n}\sum_{i=1}^{i=n}\frac{1}{h^d}\varphi\left(\frac{x - x_i}{h}\right)dx = \frac{1}{h^d n}\sum_{i=1}^{i=n}\overbrace{\int \varphi\left(\frac{x - x_i}{h}\right)dx}^{\text{Όγκος υπερκύβου}}$$

$$= \frac{1}{n}\frac{1}{h^d}\sum_{i=1}^{i=n} h^d = 1$$

Παράδειγμα 1-D

$$p_{\varphi}(x) = \frac{1}{n}\sum_{i=1}^{i=n}\frac{1}{h^d}\varphi\left(\frac{x-x_i}{h}\right)$$

Υποθέτουμε ότι έχουμε 7 δείγματα $D = \{2,3,4,8,10,11,12\}$

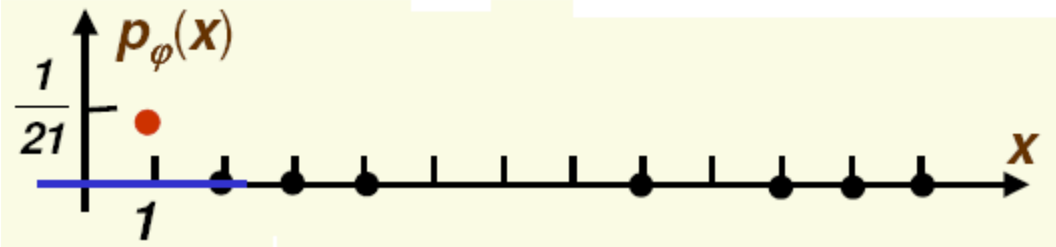


Θέλουμε να βρούμε την πυκνότητα στο σημείο $x=1$ για $h=3$

Παράδειγμα 1-D

$$p_\varphi(x) = \frac{1}{n}\sum_{i=1}^{i=n}\frac{1}{h^d}\varphi\left(\frac{x-x_i}{h}\right)$$

Υποθέτουμε ότι έχουμε 7 δείγματα $D=\{2,3,4,8,10,11,12\}$



Θέλουμε να βρούμε την πυκνότητα στο σημείο $x=1$ για $h=3$

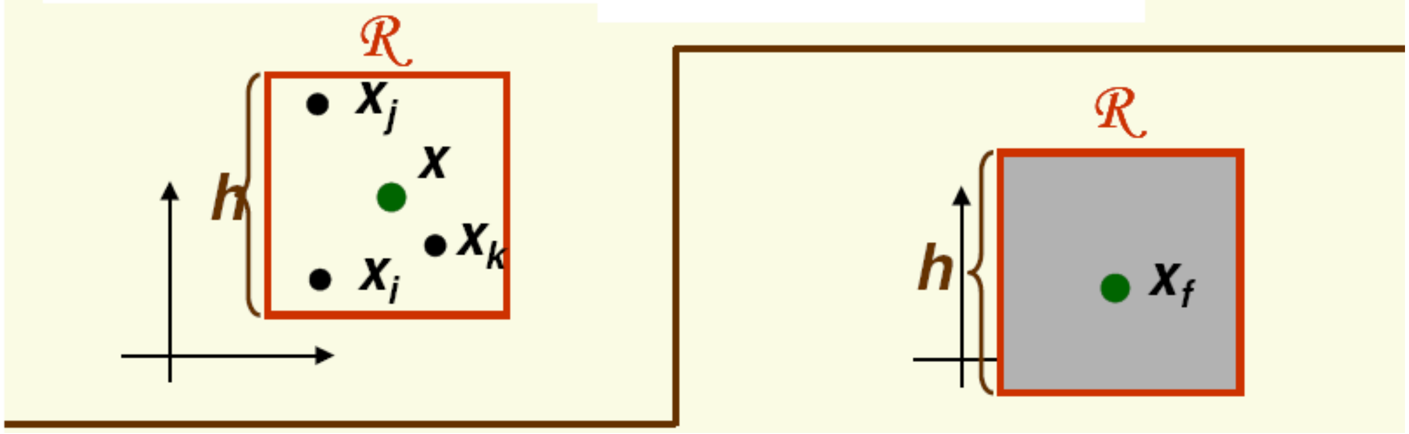
$$p_\varphi(1) = \frac{1}{7}\sum_{i=1}^{i=7}\frac{1}{3}\varphi\left(\frac{1-x_i}{3}\right) = \frac{1}{21}\left[\varphi\left(\frac{1-2}{3}\right)+\varphi\left(\frac{1-3}{3}\right)+\varphi\left(\frac{1-4}{3}\right)+\ldots+\varphi\left(\frac{1-12}{3}\right)\right]$$

$$\left|-\frac{1}{3}\right| \le 1/2 \qquad \left|-\frac{2}{3}\right| > 1/2 \qquad \left|-1\right| > 1/2 \qquad \left|-\frac{11}{3}\right| > 1/2$$

$$p_\varphi(1) = \frac{1}{7}\sum_{i=1}^{i=7}\frac{1}{3}\varphi\left(\frac{1-x_i}{3}\right) = \frac{1}{21}[1+0+0+\ldots+0] = \frac{1}{21}$$

Παράθυρα Parzen

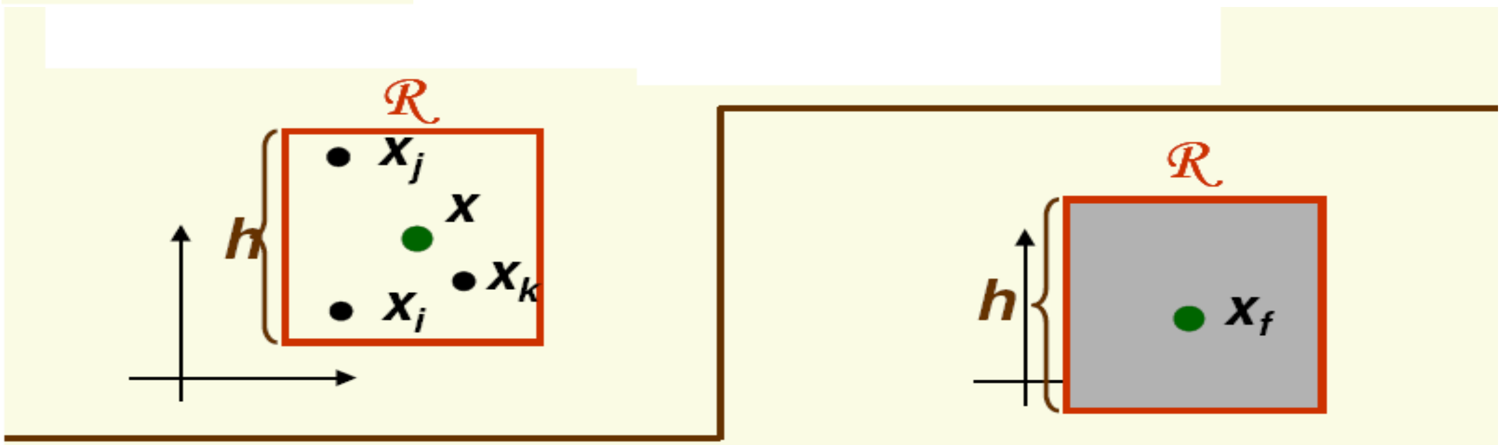
$$\varphi\left(\frac{x - x_i}{h}\right) = 1\ ?$$



$$\varphi\left(\frac{x - x_f}{h}\right) = 1\ ?$$

Παράθυρα Parzen

$$\varphi\left(\frac{x-x_i}{h}\right)=1\ ?$$



$$\varphi\left(\frac{x-x_f}{h}\right)=1\ ?$$

Οπότε η $\varphi\left(\frac{x-x_f}{h}\right)=1$: Συνάρτηση η οποία έχει τιμή 1 για τιμές του x μέσα στο τετράγωνο με κέντρο το x_f και πλευρά h και τιμή 0 για άλλες τιμές του x

Παράθυρα Parzen Άθροισμα Συναρτήσεων

- Ας ξαναπαρατηρήσουμε την έκφραση:

$$p_\varphi(x) = \frac{1}{n}\sum_{i=1}^{i=n}\frac{1}{h^d}\varphi\left(\frac{x-x_i}{h}\right) = \sum_{i=1}^{i=n}\frac{1}{nh^d}\underbrace{\varphi\left(\frac{x-x_i}{h}\right)}$$

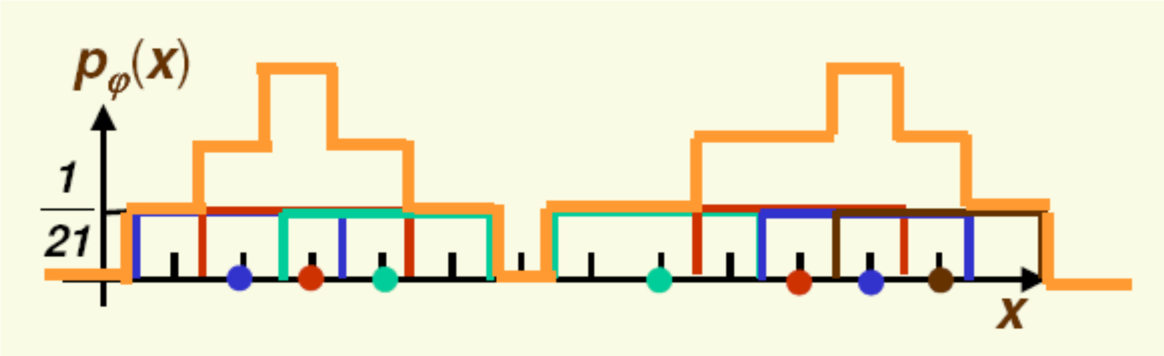
1 μέσα στο τετράγωνο με κέντρο το x
0 διαφορετικά

- Οπότε η $p_\varphi(x)$ είναι το άθροισμα **n** συναρτήσεων «κύβου» (box-like functions)

  η κάθε μία με ύψος $\dfrac{1}{nh^d}$

Παράδειγμα

$\boldsymbol{D}$={2,3,4,8,10,11,12}, $\boldsymbol{h}$=3

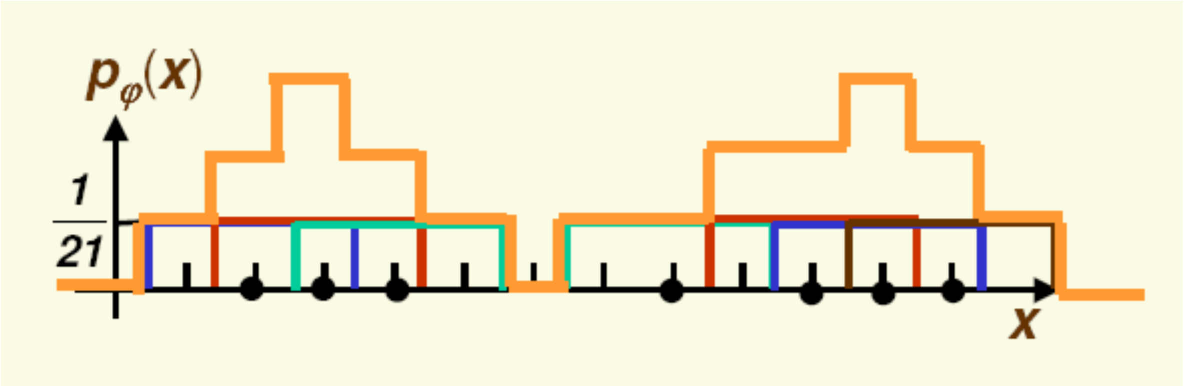


7 box-like συναρτήσεις, οι οποίες αθροίζονται

Ύψος συναρτήσεων: $\frac{1}{nh^d} = \frac{1}{21}$

Παρεμβολή

Ουσιαστικά η συνάρτηση παραθύρου φ χρησιμοποιείται για παρεμβολή:
Κάθε δείγμα x_i συνεισφέρει στην πυκνότητα στο **x,** αν το **x** είναι αρκετά κοντά στο x_i

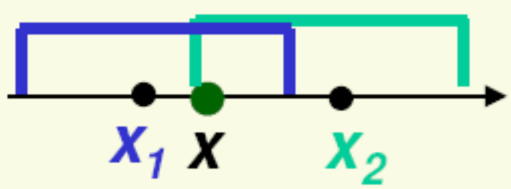


Μειονεκτήματα υπερκύβου φ

- Όσο το δείγμα $\mathbf{x_i}$ είναι στον ίδιο υπερκύβο με το $\mathbf{x}$ η συνεισφορά του $\mathbf{x_i}$ για τον υπολογισμό της πυκνότητας στο $\mathbf{x}$ είναι…?

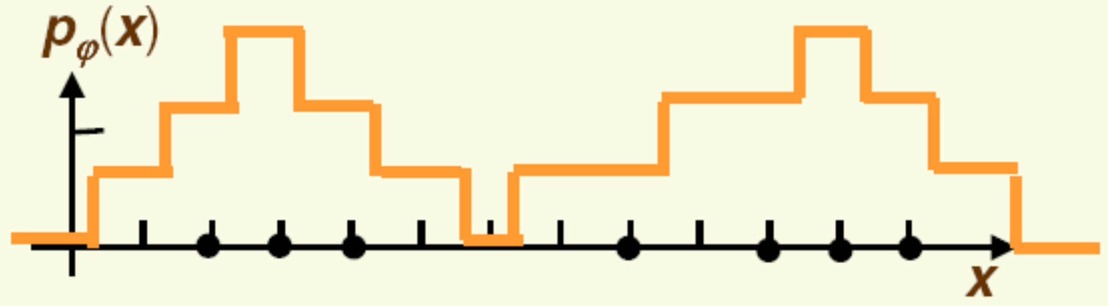
Μειονεκτήματα υπερκύβου φ

- Όσο το δείγμα $\mathbf{x_i}$ είναι στον ίδιο υπερκύβο με το $\mathbf{x}$ η συνεισφορά του $\mathbf{x_i}$ για τον υπολογισμό της πυκνότητας στο $\mathbf{x}$ είναι σταθερή, ανεξάρτητα από το πόσο κοντά είναι το $\mathbf{x_i}$ στο $\mathbf{x}$



$$\varphi\left(\frac{x - x_1}{h}\right) = \varphi\left(\frac{x - x_2}{h}\right) = 1$$

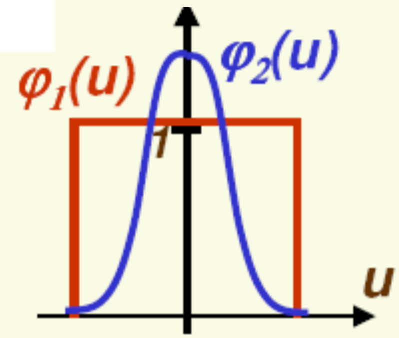
Η σ.π.π που προκύπτει δεν είναι «ομαλή» (smooth) έχει ασυνέχειες



Γενική συνάρτηση παραθύρου φ

$$p_{\varphi}(x) = \frac{1}{n}\sum_{i=1}^{i=n}\frac{1}{h^d}\varphi\left(\frac{x - x_i}{h}\right)$$

Μπορούμε να χρησιμοποιήσουμε γενικό παράθυρο φ αρκεί η $p_{\varphi}(x)$ να είναι σ.π.π., να ισχύει δηλαδή:



1. $p_{\varphi}(u) \geq 0$
   - $\varphi(u) \geq 0$
2. $\int p_{\varphi}(x)dx = 1$
   - $\int \varphi(u)du = 1$

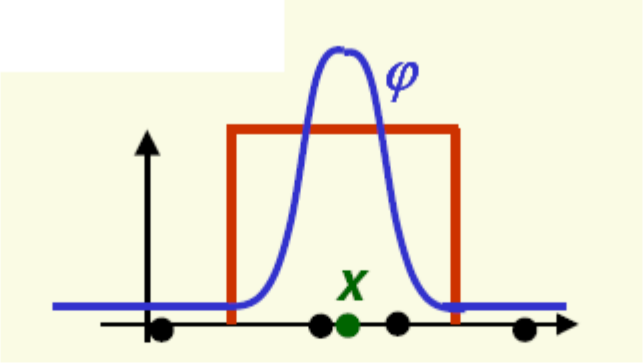
$$\int p_{\varphi}(x)dx = \frac{1}{nh^d}\sum_{i=1}^{i=n}\int \varphi\left(\frac{x - x_i}{h}\right)dx = 1$$

Γενική συνάρτηση παραθύρου φ

$$p_{\varphi}(x) = \frac{1}{n}\sum_{i=1}^{i=n}\frac{1}{h^d}\varphi\left(\frac{x-x_i}{h}\right)$$

•Χρησιμοποιώντας ένα γενικό παράθυρο φ δεν υπολογίζουμε πλέον τον αριθμό των δειγμάτων στην περιοχή R

•Υπολογίζουμε τον ?

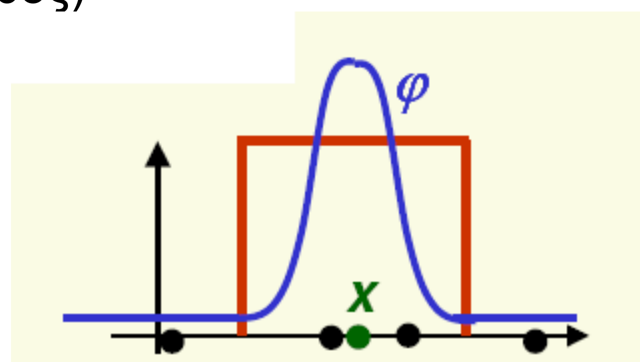


Γενική συνάρτηση παραθύρου φ

$$p_\varphi(x) = \frac{1}{n}\sum_{i=1}^{i=n}\frac{1}{h^d}\varphi\left(\frac{x - x_i}{h}\right)$$

- Χρησιμοποιώντας ένα γενικό παράθυρο φ δεν υπολογίζουμε πλέον τον αριθμό των δειγμάτων στην περιοχή R

- Υπολογίζουμε τον σταθμισμένο μέσο όρο (κάθε δείγμα έχει διαφορετικό βάρος) πιθανώς όλων των δειγμάτων (αν και μόνο τα δείγματα σε απόσταση h έχουν σημαντικό βάρος)



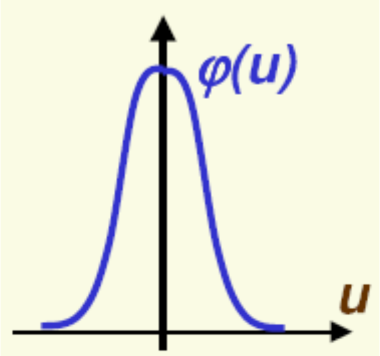
Για άπειρο αριθμό δειγμάτων και κατάλληλες συνθήκες μπορεί να αποδειχθεί και πάλι ότι: $p_\varphi^n(x) \to p(x)$

Gaussian φ

$$p_\varphi(x) = \frac{1}{n}\sum_{i=1}^{i=n}\frac{1}{h^d}\varphi\left(\frac{x-x_i}{h}\right)$$

- Μια συνήθης επιλογή για την φ είναι η N(0,1) (h=1)

$$\varphi(u) = \frac{1}{\sqrt{2\pi}}e^{-u^2/2}$$



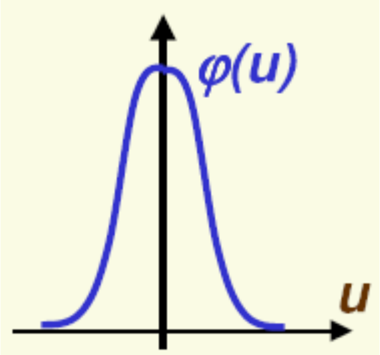
- Επιλύει και τα δυο προβλήματα της συνάρτησης υπερκύβου:

Gaussian φ

$$p_\varphi(x) = \frac{1}{n}\sum_{i=1}^{i=n}\frac{1}{h^d}\varphi\left(\frac{x - x_i}{h}\right)$$

- Μια συνήθης επιλογή για την φ είναι η N(0,1)

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$



- Επιλύει και τα δυο προβλήματα της συνάρτησης υπερκύβου:

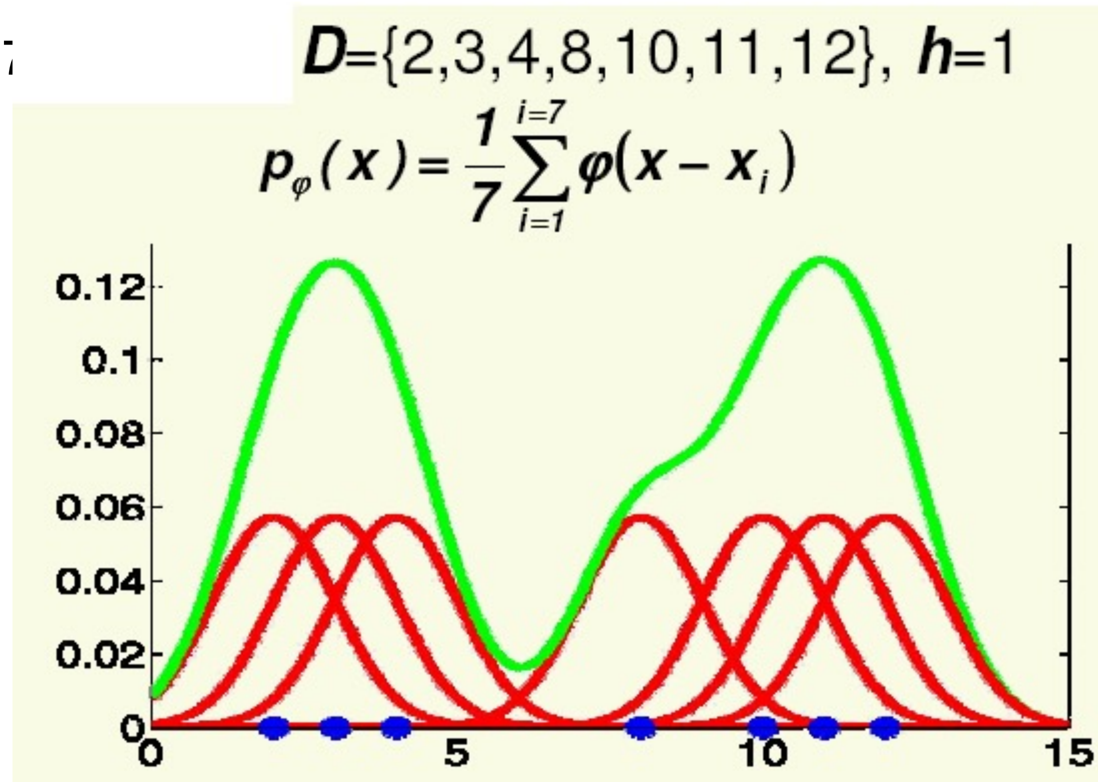
• Σημεία x που βρίσκονται κοντά στο δειγματικό σημείο x_i έχουν μεγαλύτερο βάρος

• Η συνάρτηση π.π που προκύπτει είναι ομαλή (smooth)

Παράδειγμα

7 δείγματα $D=\{2,3,4,8,10,11,12\}$, $h=1$

Παράδειγμα



$p_\varphi(x)$ Είναι το άθροισμα 7 συναρτήσεων Gauss κάθε μία μέ κέντρο ένα από τα δειγματικά σημεία και πολλαπλασιασμένη με 1/7

$$p_{\varphi}(x) = \frac{1}{n}\sum_{i=1}^{i=n}\frac{1}{h^d}\varphi\left(\frac{x - x_i}{h}\right)$$

Αριθμητικό Παράδειγμα

5 δείγματα: D={2, 2.5, 3, 1, 6}

Υπολογίστε την σ.π.π. για x=3 χρησιμοποιώντας ως συνάρτηση παραθύρου συνάρτηση Gauss με σ=1

$$p_\varphi(x) = \frac{1}{n}\sum_{i=1}^{i=n}\frac{1}{h^d}\varphi\left(\frac{x-x_i}{h}\right)$$

Αριθμητικό Παράδειγμα

5 δείγματα: D={2, 2.5, 3, 1, 6}

Υπολογίστε την σ.π.π. για x=3 χρησιμοποιώντας ως συνάρτηση παραθύρου συνάρτηση Gauss με σ=1 (h=1)

Λύση:

$x_1=2$

$$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x_1-x)^2}{2}\right)$$

$$= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(2-3)^2}{2}\right) = 0.2420$$

$x_2=2.5$

$$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x_2-x)^2}{2}\right)$$

$$= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(2.5-3)^2}{2}\right) = 0.3521$$

$$p_\varphi(x) = \frac{1}{n}\sum_{i=1}^{i=n}\frac{1}{h^d}\varphi\left(\frac{x - x_i}{h}\right)$$

Αριθμητικό Παράδειγμα

$x_3=3$ $$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x_3 - x)^2}{2}\right) = 0.3989$$

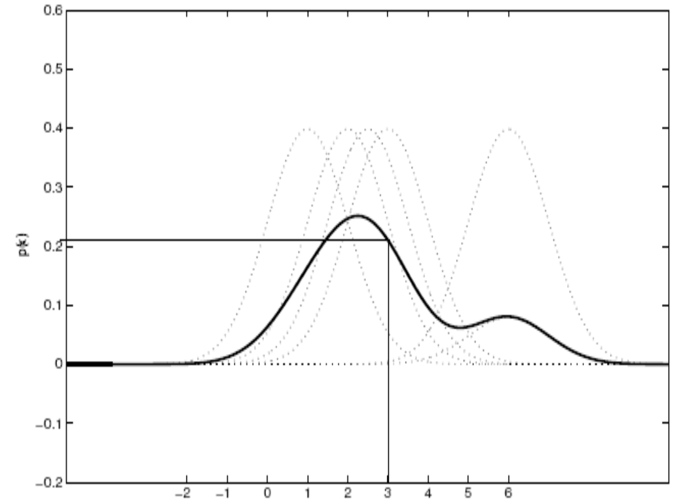
$x_4=1$ $$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x_4 - x)^2}{2}\right) = 0.0540$$

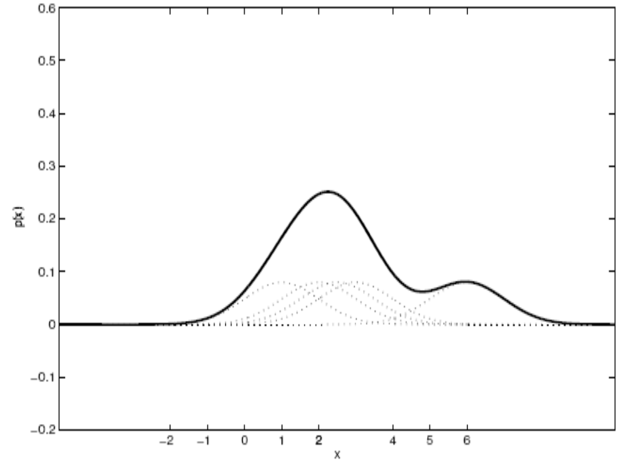
$x_5=6$ $$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x_5 - x)^2}{2}\right) = 0.0044$$

Οπότε:

$$p(x=3) = (0.2420 + 0.3521 + 0.3989$$

$$+0.0540 + 0.0044)/5 \quad = 0.2103$$



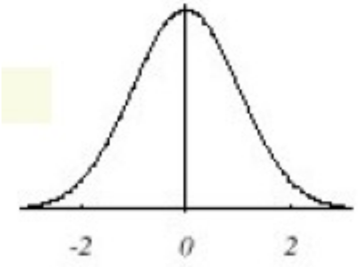


Παραδείγματα

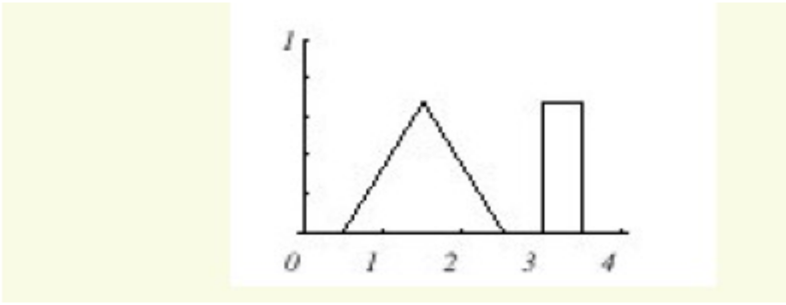
1. Παίρνουμε δείγματα από γνωστή σ. π. π.

2. Χρησιμοποιούμε την μέθοδο μας για να εκτιμήσουμε την σ.π.π. και συγκρίνουμε με την πραγματική σ.π.π.

Μεταβάλλουμε τον άριθμο των δειγμάτων **n** και το πλάτος του παραθύρου **h**

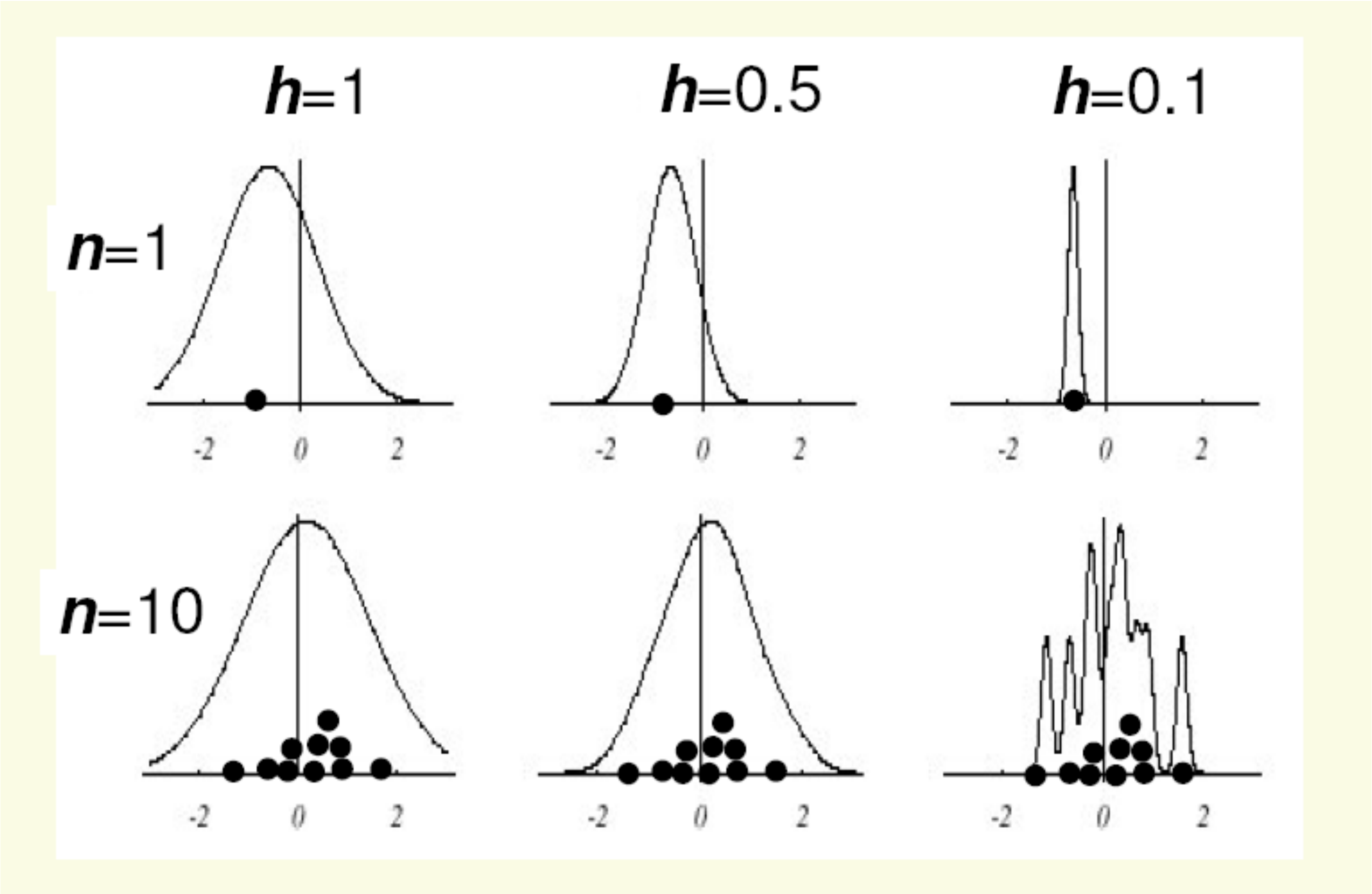


$N(0,1)$

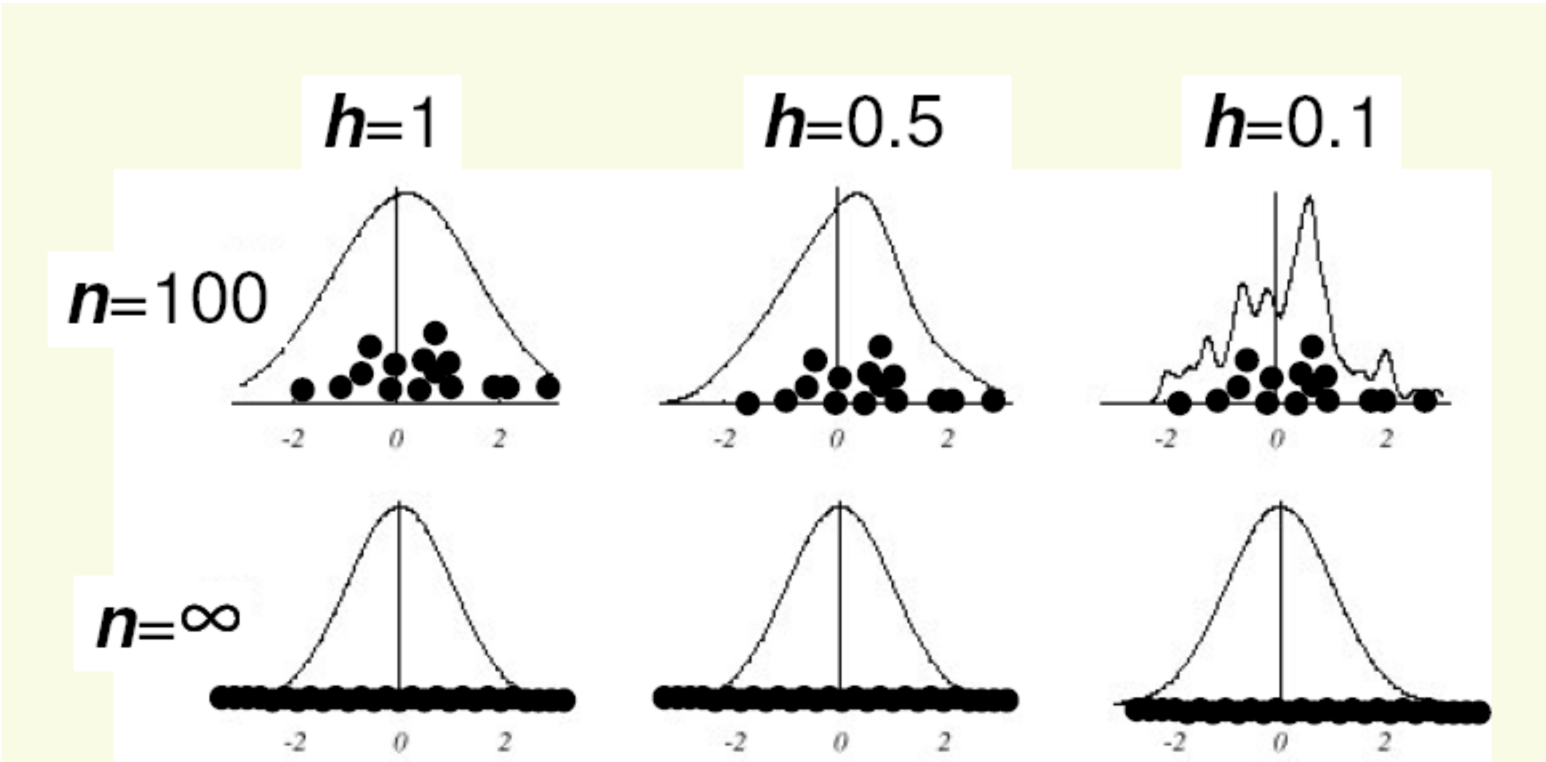


Μίξη τριγωνικής και ομοιόμορφης

N(0,1)

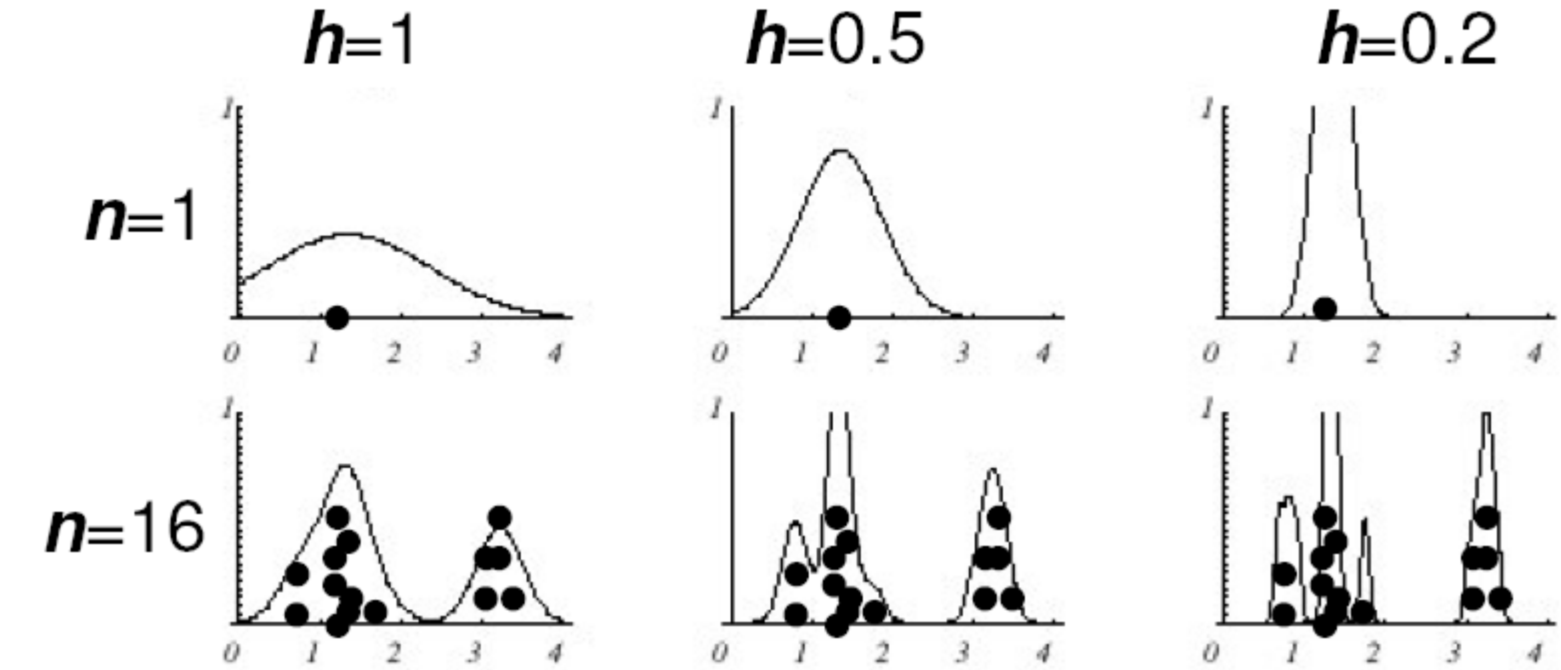


N(0,1)

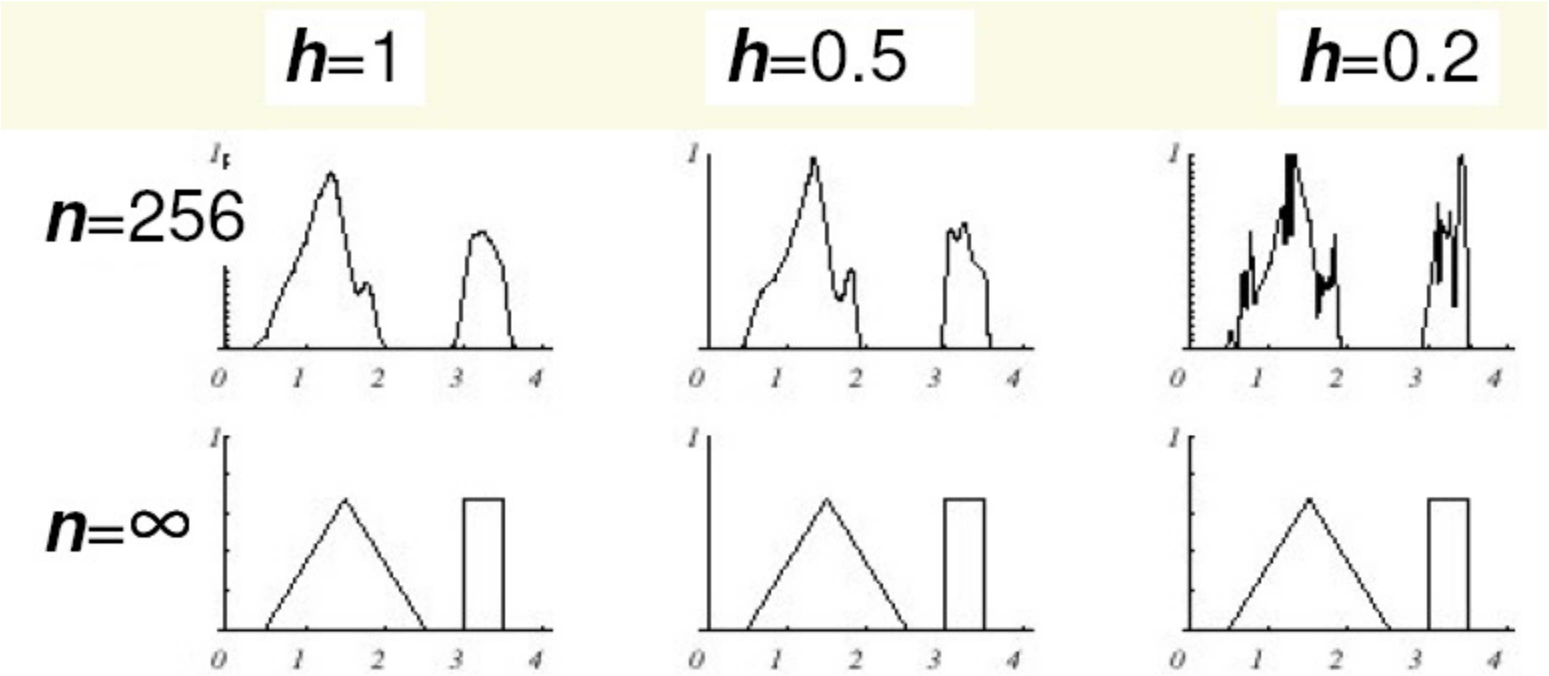


Καθώς το n τείνει στο άπειρο, η εκτίμηση γίνεται ακριβής, ανεξάρτητα από το μήκος του παραθύρου!

Μίξη τριγωνικής και ομοιόμορφης



Μίξη τριγωνικής και ομοιόμορφης



Καθώς το n τείνει στο άπειρο, η εκτίμηση γίνεται ακριβής, ανεξάρτητα από το μήκος του παραθύρου!

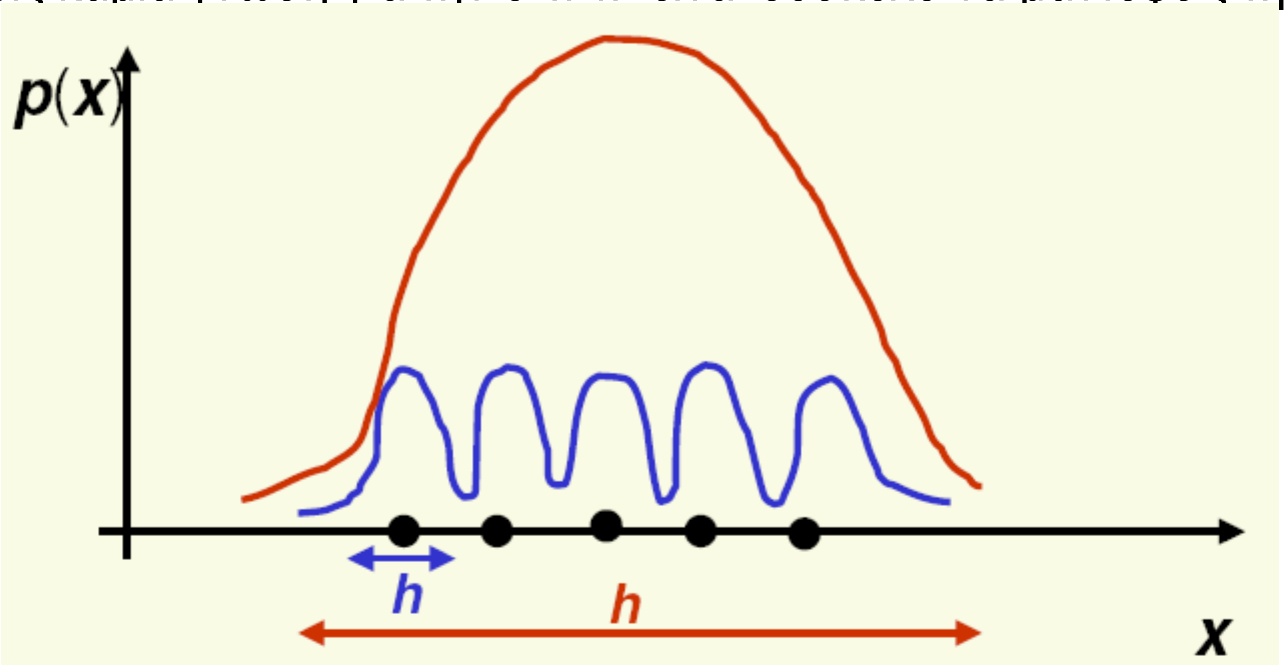
Το πλάτος h

Επιλέγοντας το h «μαντεύουμε» την περιοχή όπου ?

Το πλάτος h

Επιλέγοντας το h «μαντεύουμε» την περιοχή όπου η σ.π.π είναι σταθερή

Χωρίς καμία γνώση για την σ.π.π. είναι δύσκολο να μαντέψεις την περιοχή....

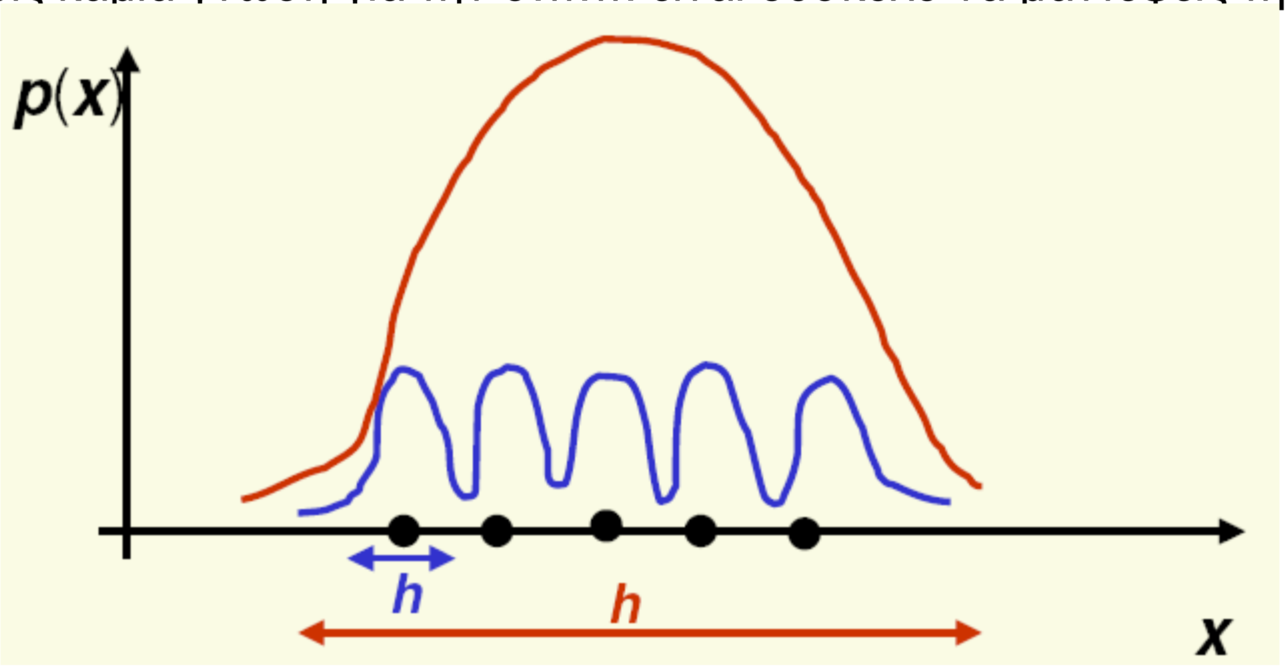


Μπορούμε να χρησιμοποιήσουμε τεχνικές ...?

Το πλάτος h

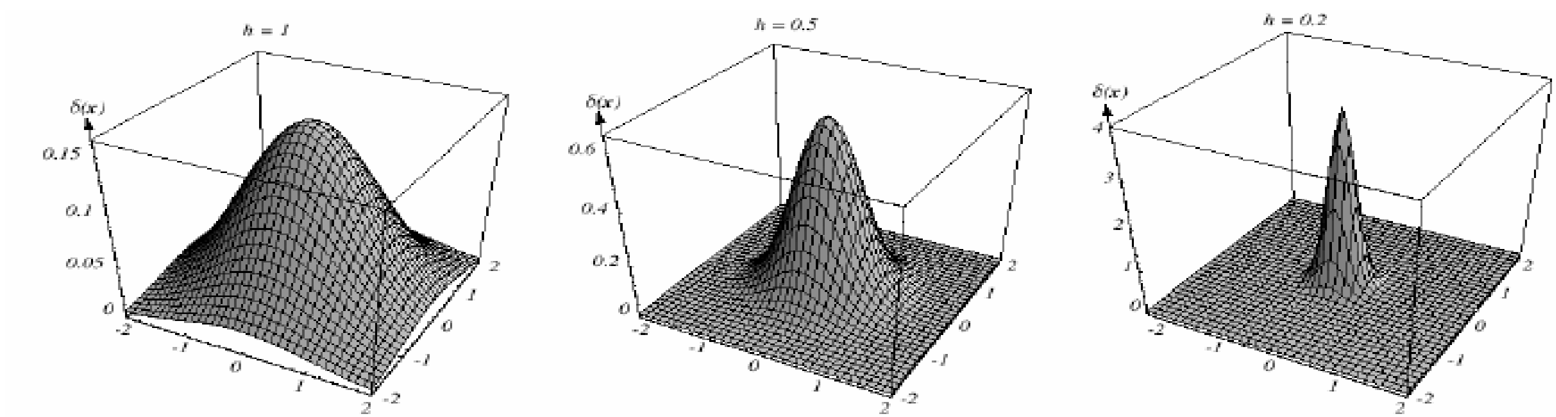
Επιλέγοντας το h «μαντεύουμε» την περιοχή όπου η σ.π.π είναι σταθερή

Χωρίς καμία γνώση για την σ.π.π. είναι δύσκολο να μαντέψεις την περιοχή....



Μπορούμε να χρησιμοποιήσουμε τεχνικές cross-validation για την εκτίμηση του h

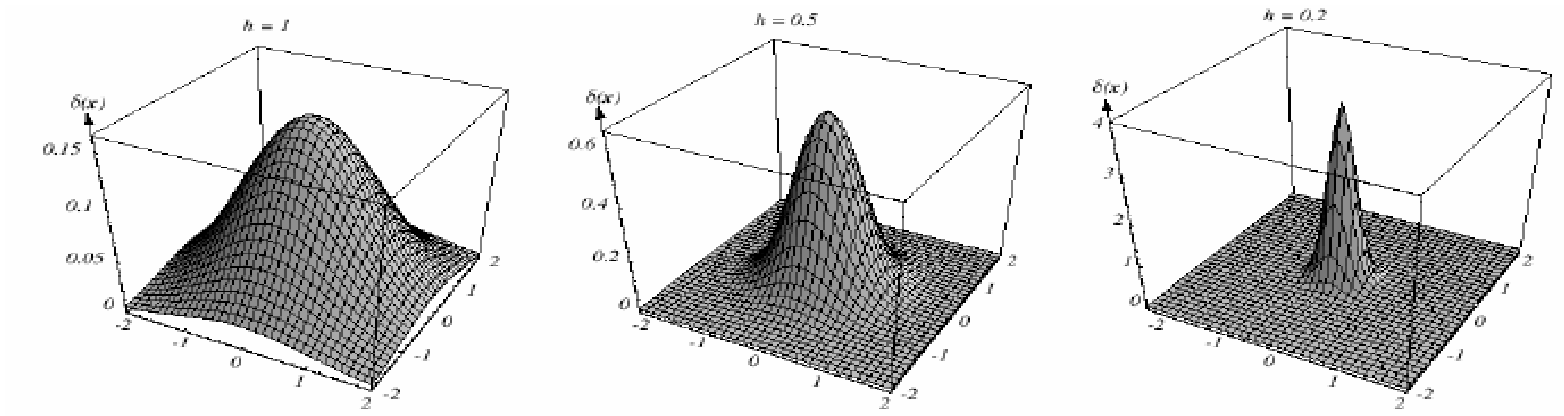
Επίδραση του Πλάτους Παραθύρου, h_n



Η παράμετρος h_n επιδρά και στο πλάτος του παραθύρου αλλά και στο μέτρο του:

- **Όταν το h_n είναι μεγάλο (μικρό), το παράθυρ ? (στενό), το μέτρο του παραθύ ? ό (μεγάλο) και το x πρέπε ? κρυά (κοντά) από το x_i πρίν η τιμή της συνάρτησης αλλάξει αρκετά από την μέγιστη τιμή της**

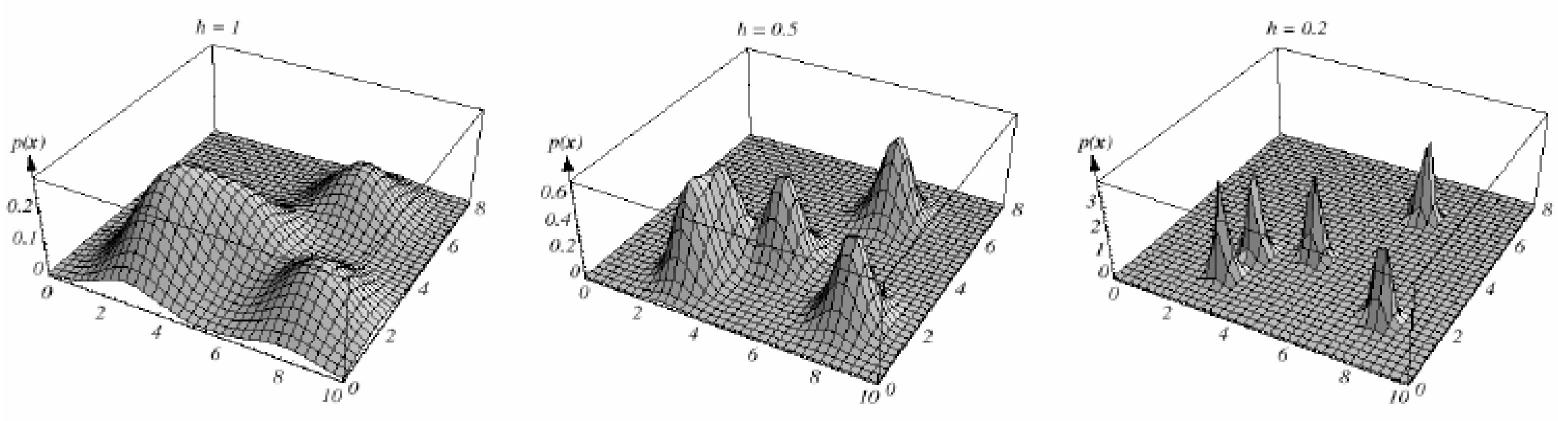
Επίδραση του Πλάτους Παραθύρου, h_n



Η παράμετρος h_n επιδρά και στο πλάτος του παραθύρου αλλά και στο μέτρο του:

- **Όταν το h_n είναι μεγάλο (μικρό), το παράθυρο είναι πλατύ (στενό), το μέτρο του παραθύρου είναι μικρό (μεγάλο) και το x πρέπει να είναι μακρυά (κοντά) από το x_i πρίν η τιμή της συνάρτησης αλλάξει αρκετά από την μέγιστη τιμή της**

Επίδραση του Πλάτους Παραθύρου, h_n



Πώς το πλάτος του παραθύρου επιδρά στην εκτίμηση της σ.π.π. p(x) :

- ✓ **Όταν το h_n είναι <u>μεγάλο</u>, η εκτιμήτρια $p_n(x)$ είναι η υπέρθεση n πλατιών συναρτήσεων επικεντρωμένων στα δείγματα εκπαίδευσης και αποτελεί μία <u>ομαλή, "out-of-focus"</u> εκτίμηση του p(x), <u>χωρίς μεγάλη ανάλυση.</u>**
- ✓ **Όταν το h_n είναι <u>μικρό</u>, η $p_n(x)$ είναι η υπέρθεση n στενών συναρτήσεων, και είναι ίσως μία <u>θορυβώδης, "erratic or noisy"</u> εκτίμηση του p(x).**

Ταξινόμηση με Χρήση Παραθύρων Parzen

- Εκτίμηση της πιθανοφάνειας $p(x|\omega_i)$ *από τα δεδομένα, χρησιμοποιώντας τη* μέθοδο παραθύρων Parzen, και χρήση του κανόνα του Bayes για την ταξινόμηση, δηλ. υπολογισμός των εκ των υστέρων πιθανοτήτων, και επιλογή της κλάσης με την μεγαλύτερη πιθανότητα.

- Οι περιοχές απόφασεις για έναν ταξινομητή χρήση παραθύρων Parzen εξαρτάται από ...?

Ταξινόμηση με Χρήση Παραθύρων Parzen

- Εκτίμηση της πιθανοφάνειας $p(x|\omega_i)$ *από τα δεδομένα, χρησιμοποιώντας τη* μέθοδο παραθύρων Parzen, και χρήση του κανόνα του Bayes για την ταξινόμηση, δηλ. υπολογισμός των εκ των υστέρων πιθανοτήτων, και επιλογή της κλάσης με την μεγαλύτερη πιθανότητα.

- Οι περιοχές απόφασης για έναν ταξινομητή με χρήση παραθύρων Parzen εξαρτάται από την επιλογή του παραθύρου

- Το λάθος εκπαίδευσης μπορεί να γίνει αρκούντως μικρό (ακόμα και μηδέν), επιλέγοντας αρκετά μικρά παράθυρα! Παρ' όλα αυτά, δεν είναι επιθυμητό...?

Ταξινόμηση με Χρήση Παραθύρων Parzen

- Εκτίμηση της πιθανοφάνειας $p(x|\omega_i)$ *από τα δεδομένα, χρησιμοποιώντας τη* μέθοδο παραθύρων Parzen, και χρήση του κανόνα του Bayes για την ταξινόμηση, δηλ. υπολογισμός των εκ των υστέρων πιθανοτήτων, και επιλογή της κλάσης με την μεγαλύτερη πιθανότητα.

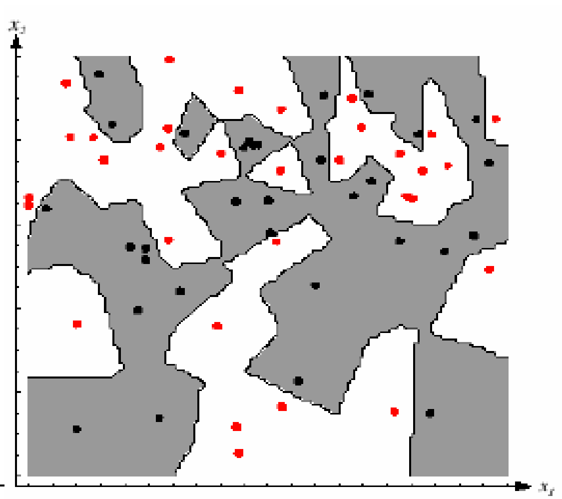
- Οι περιοχές απόφασης για έναν ταξινομητή με χρήση παραθύρων Parzen εξαρτάται από την επιλογή του παραθύρου

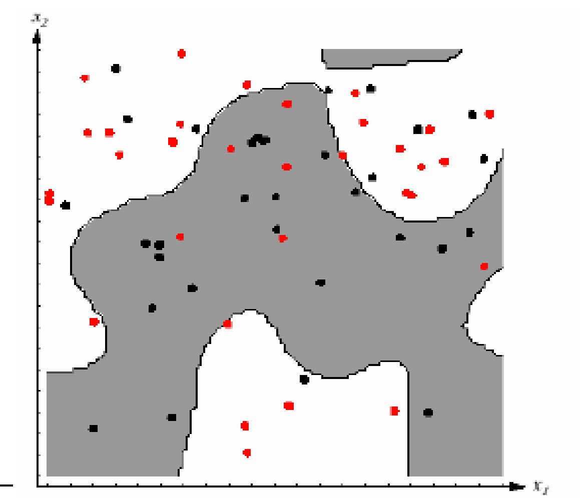
- Το λάθος εκπαίδευσης μπορεί να γίνει αρκούντως μικρό (ακόμα και μηδέν), επιλέγοντας αρκετά μικρά παράθυρα! Παρ' όλα αυτά, δεν είναι επιθυμητό, επειδή σίγουρα θα προκαλέσει overfitting (υπερταίριασμα) και θα μειώσει την απόδοση στα νέα αταξινόμητα δεδομένα ελέγχου (test data).

Ταξινόμηση με Χρήση Παραθύρων Parzen

Πολύ μικρό παράθυρο → Πολύ μικρή διαμέριση του χώρου χαρακτηριστικών, πράγμα μη επιθυμητό!

Μεγαλύτερο παράθυρο →Υψηλότερο λάθος κατά την εκπαίδευση, αλλά καλύτερη απόδοση γενίκευσης! Καλύτερη απόδοση γενίκευσης: Επιθυμητή ιδιότητα.





Στην πράξη, αυτό που θα θέλαμε είναι παράθυρα μικρού πλάτους στις περιοχές με υψηλή πυκνότητα δεδομένων, και παράθυρα μεγάλου πλάτους στις περιοχές όπου τα δεδομένα είναι αραιά!

Συμπεράσματα

Πλεονεκτήματα:

• Δεν προϋποθέτει καμία γνώση για το πρόβλημα, εκτός από την ύπαρξη του συνόλου δειγμάτων εκπαίδευσης!

• Θεωρητικά συγκλίνει στην πραγματική σ.π.π. καθώς ο αριθμός των δειγμάτων τείνει στο άπειρο

Μειονεκτήματα:

• Απαιτεί (πολλά)^d δεδομένα για να εξασφαλίσει ότι η εκτίμηση συγκλίνει στην πραγματική κατανομή.

- Επιπλέον, καθώς η διάσταση αυξάνει, η απαίτηση για (πολλά)^d δεδομένα γίνεται ((πολλά)$^{\text{πολλά (n)}}$)^n !!!! ➔ Πρόβλημα διάστασης (Curse of dimensionality) !
- Ο μόνος τρόπος για την αντιμετώπιση του είναι η ύπαρξη εκ των προτέρων, σωστής πληροφορίας για τα δεδομένα!

• Υπολογιστικός φόρτος: για την ταξινόμηση ενός σημείου πρέπει να υπολογιστεί η τιμή μιας συνάρτησης η οποία πιθανώς να εξαρτάται από όλα τα δείγματα

• Δεν είναι εύκολη η επιλογή του πλάτους του παραθύρου h