Hands-On 1: Παράρτημα

Κωνσταντίνος Χατζηλυγερούδης

29 Οκτωβρίου 2026

Διαβάστε τα παραρτήματα μετά τη συνεδρία. Δίνουν τις πλήρεις αποδείξεις και περαιτέρω ανάλυση για τις έξι ενότητες του φυλλαδίου. Το προαιρετικό notebook περιέχει αριθμητικές ασκήσεις στα ίδια θέματα.

Τα παραρτήματα E–G χρησιμοποιούν τις πραγματικές ετικέτες βλάβης του πειράματος λειτουργίας: ο αισθητήρας 7 είναι ελαττωματικός από τη λήψη 200 και μετά. Τις χρησιμοποιούμε εκεί μόνο για να εξηγήσουμε τι συνέβη σε αυτό το λυμένο παράδειγμα. Το κύριο notebook δεν χρησιμοποιεί ποτέ αυτές τις ετικέτες για τη διάγνωση· τις αποκαλύπτει μόνο στην τελική επικύρωση (ενότητα 6 του φυλλαδίου).

Περιεχόμενα

A Η Βαθμονόμηση ως Πρόβλημα Ελαχίστων Τετραγώνων (Διαλέξεις 1 και 5) 3
A.1 Ένας αισθητήρας . . . 3
A.2 Πώς διαβάζουμε τα υπόλοιπα . . . 4
A.3 Όλοι οι αισθητήρες μαζί . . . 5
A.4 Από counts σε θερμοκρασία . . . 6

B Κακή Αριθμητική Κατάσταση και Ridge 6
B.1 Δύο σημεία αναφοράς που είναι σχεδόν ίδια . . . 6
B.2 Απλή Ridge: ευσταθής, αλλά πού; . . . 8
B.3 Τα ελάχιστα τετράγωνα είναι ML· η Ridge είναι MAP . . . 8
B.4 Μια εκ των προτέρων κατανομή από τους υπόλοιπους αισθητήρες . . . 9
B.5 Μεροληψία και διακύμανση σε πολλές σύντομες επαναβαθμονομήσεις . . . 10

C Η Συνδιακύμανση ανάμεσα στους Αισθητήρες (Διαλέξεις 4 και 5) 10
C.1 Εκτίμηση από τα υπόλοιπα . . . 11
C.2 Ιδιοανάλυση και κύριες συντεταγμένες . . . 11
C.3 Οι εκτιμημένες συνδιακυμάνσεις υπερεκτιμούν τις αποστάσεις . . . 12

D Η Κατανομή χ^2 13
D.1 Ορισμός και πρώτες ιδιότητες . . . 13
D.2 Γιατί το κατώφλι πρέπει να μεγαλώνει με τη διάσταση . . . 14

E Ελλείψεις, Λευκοποίηση και Ανίχνευση (Διάλεξη 4) 15
E.1 Ελλείψεις συνδιακύμανσης . . . 15
E.2 Λευκοποίηση, και πώς τη ελέγχουμε με εντιμότητα . . . 15
E.3 Η απόσταση Mahalanobis και η κατανομή της . . . 16
E.4 Γιατί η απόσταση Mahalanobis βλέπει ό,τι χάνει ένα z-score . . . 16
E.5 Ποιος αισθητήρας; Αφαιρούμε έναν κάθε φορά . . . 18

F Μία Θερμοκρασία με Αξιόπιστο Περιθώριο Σφάλματος **19**
F.1 Ο βέλτιστος γραμμικός συνδυασμός 19
F.2 Γιατί ο μέσος όρος δεν νικά ένα κοινό σφάλμα 20
F.3 Αξιόπιστα και αναξιόπιστα περιθώρια σφάλματος 20

G Χωρίς Θερμόμετρο Αναφοράς **20**

Σημειογραφία.

$k = 1, \dots, N$; $j = 1, \dots, C$	δείκτης λήψης· δείκτης αισθητήρα ($N = 400$, $C = 10$)
T_k	θερμοκρασία αναφοράς της λήψης k [°C]
y_{kj}; $\boldsymbol{Y} \in \mathbb{R}^{N \times C}$	η ένδειξη (σε counts) του αισθητήρα j στη λήψη k· όλες οι ενδείξεις, μία στήλη ανά αισθητήρα
a_j, b_j; $\boldsymbol{\theta}_j = [a_j,\, b_j]^\top$	μετατόπιση [counts] και κέρδος [counts/°C] του αισθητήρα j
$\boldsymbol{X} \in \mathbb{R}^{N \times 2}$, $p = 2$	πίνακας σχεδιασμού με γραμμές $[1,\ T_k]$· παράμετροι ανά αισθητήρα
$\boldsymbol{\Theta} = [\boldsymbol{\theta}_1\ \cdots\ \boldsymbol{\theta}_C]$	όλοι οι συντελεστές, $2 \times C$
$\boldsymbol{n}_k \in \mathbb{R}^C$; $\boldsymbol{\Sigma}$	ο θόρυβος της λήψης k σε όλους τους αισθητήρες· η συνδιακύμανσή του σε $^\circ\mathrm{C}^2$
$\boldsymbol{E} \in \mathbb{R}^{N \times C}$	τα υπόλοιπα της βαθμονόμησης, μετατρεμμένα σε °C
$\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$	η προβολή στον χώρο στηλών του $\boldsymbol{X}$ (Διάλεξη 1)
χ^2_ν; $\chi^2_\nu(q)$	κατανομή χ-τετράγωνο (chi-squared) με ν βαθμούς ελευθερίας· το q-ποσοστημόριό της

Το μοντέλο θορύβου. Το σφάλμα κάθε αισθητήρα, μετρημένο σε °C, είναι το άθροισμα τριών μερών: ένα που μοιράζονται όλοι οι αισθητήρες (η τάση αναφοράς του ADC και η τροφοδοσία), ένα που μεταβάλλεται κατά μήκος της συστοιχίας (ο ανεμιστήρας) και ο δικός θόρυβος του αισθητήρα. Τα σφάλματα είναι ανεξάρτητα από λήψη σε λήψη, αλλά *όχι* από αισθητήρα σε αισθητήρα μέσα στην ίδια λήψη:

$$\boldsymbol{n}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}) \ \text{ ανεξάρτητα ως προς } k, \qquad \text{ο } \boldsymbol{\Sigma} \text{ δεν είναι διαγώνιος.} \tag{0.1}$$

Σε counts, ο θόρυβος του αισθητήρα j είναι b_j φορές ο θόρυβός του σε °C.

A Η Βαθμονόμηση ως Πρόβλημα Ελαχίστων Τετραγώνων (Διαλέξεις 1 και 5)

Βαθμονόμηση σημαίνει να βρούμε, για κάθε αισθητήρα, την ευθεία που μετατρέπει τα counts του σε βαθμούς. Το πείραμα βαθμονόμησης δίνει 400 λήψεις σε γνωστές θερμοκρασίες, πολύ περισσότερα δεδομένα από όσα χρειάζονται δύο άγνωστοι ανά αισθητήρα, οπότε προσαρμόζουμε την ευθεία με ελάχιστα τετράγωνα.

A.1 Ένας αισθητήρας

Το μοντέλο τύπου LM35 της Διάλεξης 1, γραμμένο με έξοδο τα counts, είναι

$$y_{kj} = a_j + b_j\, T_k + n_{kj} \quad \Longleftrightarrow \quad \boldsymbol{y}_j = \boldsymbol{X}\boldsymbol{\theta}_j + \boldsymbol{n}_j, \qquad \boldsymbol{X} = \begin{bmatrix} 1 & T_1 \\ \vdots & \vdots \\ 1 & T_N \end{bmatrix}. \tag{A.1}$$

Εδώ το a_j είναι η ένδειξη που θα έδινε ο αισθητήρας στους 0 °C, και το b_j είναι ο αριθμός των counts ανά βαθμό.

Παρατήρηση A.1 (Ποια μεταβλητή μπαίνει σε ποια πλευρά)**.** Τα ελάχιστα τετράγωνα υποθέτουν ότι ο θόρυβος βρίσκεται στο $\boldsymbol{y}$ και ότι το $\boldsymbol{X}$ είναι ακριβώς γνωστό. Το θερμόμετρο αναφοράς είναι πολύ ακριβέστερο από τους αισθητήρες, οπότε οι θερμοκρασίες αναφοράς μπαίνουν στο $\boldsymbol{X}$ και τα counts στο $\boldsymbol{y}$. Προσαρμόζουμε τα counts ως συνάρτηση της θερμοκρασίας και αντιστρέφουμε την ευθεία στη συνέχεια (Ενότητα A.4). Αν προσαρμόζαμε τη θερμοκρασία ως συνάρτηση των counts,

η θορυβώδης ποσότητα θα βρισκόταν μέσα στο $\boldsymbol{X}$, και η κλίση θα είχε μεροληψία προς το μηδέν. Ο όρος-κλειδί για αυτό το πρόβλημα είναι *errors-in-variables*.

Τα ελάχιστα τετράγωνα δίνουν $\hat{\boldsymbol{\theta}}_j = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y}_j$ (οι κανονικές εξισώσεις, Διάλεξη 1). Με $S_{TT} = \sum_k (T_k - \bar{T})^2$, μπορούμε να γράψουμε αναλυτικά τον αντίστροφο του 2×2 πίνακα:

$$(\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \frac{1}{S_{TT}} \begin{bmatrix} S_{TT}/N + \bar{T}^2 & -\bar{T} \\ -\bar{T} & 1 \end{bmatrix}. \tag{A.2}$$

Για να το επαληθεύσουμε, πολλαπλασιάζουμε με $\boldsymbol{X}^\top \boldsymbol{X} = \begin{bmatrix} N & N\bar{T} \\ N\bar{T} & \sum_k T_k^2 \end{bmatrix}$ και χρησιμοποιούμε ότι $\sum_k T_k^2 = S_{TT} + N\bar{T}^2$.

Πρόταση A.2 (Περιθώρια σφάλματος μιας αφινικής βαθμονόμησης)**.** *Αν ο θόρυβος του αισθητήρα j είναι λευκός με διακύμανση σ_j^2 (σε counts), τότε $\mathrm{Cov}(\hat{\boldsymbol{\theta}}_j) = \sigma_j^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}$ (Διάλεξη 5). Ειδικότερα,*

$$\mathrm{Var}(\hat{b}_j) = \frac{\sigma_j^2}{S_{TT}}, \qquad \mathrm{Var}(\hat{a}_j) = \sigma_j^2 \Big(\frac{1}{N} + \frac{\bar{T}^2}{S_{TT}} \Big), \tag{A.3}$$

και η προσαρμοσμένη ευθεία στη θερμοκρασία T, $\boldsymbol{x}_T^\top \hat{\boldsymbol{\theta}}_j$ με $\boldsymbol{x}_T = [1, T]^\top$, έχει

$$\mathrm{Var}\left(\boldsymbol{x}_T^\top \hat{\boldsymbol{\theta}}_j\right) = \sigma_j^2 \, h(T), \qquad h(T) = \boldsymbol{x}_T^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_T = \frac{1}{N} + \frac{(T - \bar{T})^2}{S_{TT}}. \tag{A.4}$$

Για την απόδειξη διαβάζουμε την (A.2) στοιχείο προς στοιχείο και χρησιμοποιούμε τη σχέση

$$\mathrm{Var}(\boldsymbol{x}_T^\top \hat{\boldsymbol{\theta}}) = \boldsymbol{x}_T^\top \, \mathrm{Cov}(\hat{\boldsymbol{\theta}}) \boldsymbol{x}_T.$$

Τον αριθμό $h(T)$ τον λέμε *μόχλευση* (leverage) του T. Είναι ελάχιστος στο κέντρο του εύρους βαθμονόμησης και αυξάνεται τετραγωνικά καθώς απομακρυνόμαστε από αυτό.

Πώς διαβάζουμε τους τύπους. Το κέρδος καθορίζεται από το πόσο *απλωμένες* είναι οι θερμοκρασίες αναφοράς. Η μετατόπιση είναι η τιμή της ευθείας στο $T = 0$, που βρίσκεται έξω από το εύρος βαθμονόμησης, οπότε κληρονομεί τη μόχλευση του $T = 0$. Για τη συστοιχία ($N = 400$, δεξαμενές στους 10, 20, 30 και 40 °C, $\bar{T} = 25\,^\circ\mathrm{C}$, $S_{TT} = 400 \cdot 125 = 50\,000\,^\circ\mathrm{C}^2$) και για τον θόρυβο του αισθητήρα 3, $s_3 = 8.2$ counts, παίρνουμε ± 0.037 counts/°C και ± 1.0 count.

A.2 Πώς διαβάζουμε τα υπόλοιπα

Πρόταση A.3 (Ιδιότητες των υπολοίπων)**.** *Έστω $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\theta}}$ με $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$. Τότε $\boldsymbol{r} = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{n}$ και*

$$\mathbb{E}[\boldsymbol{r}] = \boldsymbol{0}, \qquad \mathrm{Cov}(\boldsymbol{r}) = \sigma^2 (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}}), \qquad \mathbb{E}\|\boldsymbol{r}\|^2 = (N - p)\, \sigma^2.$$

Έτσι, το $s^2 = \|\boldsymbol{r}\|^2/(N-p)$ είναι αμερόληπτος εκτιμητής του σ^2 (Διάλεξη 5), και το υπόλοιπο k έχει διακύμανση $\sigma^2(1 - h(T_k))$, ελαφρώς μικρότερη από σ^2.

Προαιρετικό: απόδειξη

Ισχύει $\boldsymbol{X}\hat{\boldsymbol{\theta}} = \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y}$ και $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X} = \boldsymbol{X}$, άρα $\boldsymbol{r} = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})(\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{n}) = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{n}$. Η μέση τιμή του είναι μηδέν. Η συνδιακύμανσή του είναι $(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})\, \sigma^2 \boldsymbol{I}\, (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})^\top = \sigma^2 (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})$, επειδή ο $\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}}$ είναι συμμετρική προβολή, οπότε $(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})^2 = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}}$. Τα διαγώνια στοιχεία του $\boldsymbol{P}_{\boldsymbol{X}}$ είναι οι μοχλεύσεις $h(T_k)$. Τέλος, $\mathbb{E}\|\boldsymbol{r}\|^2 = \mathrm{tr}\,\mathrm{Cov}(\boldsymbol{r}) = \sigma^2\,\mathrm{tr}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})$, και $\mathrm{tr}\,\boldsymbol{P}_{\boldsymbol{X}} = \mathrm{tr}\left((\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{X}\right) = \mathrm{tr}\,\boldsymbol{I}_p = p$.

Η Πρόταση A.3 λέει πώς μοιάζουν τα υπόλοιπα όταν όλα πάνε καλά: κανένα συστηματικό μοτίβο και διασπορά περίπου σ. Το notebook τα διαβάζει με τρεις τρόπους.

1. **Ως προς το T.** Μια καμπυλότητα θα φαινόταν ως μέσο υπόλοιπο κάθε δεξαμενής που απομακρύνεται από το μηδέν. Οι τέσσερις μέσες τιμές των δεξαμενών κάθονται στο μηδέν, οπότε το αφινικό μοντέλο (affine model) είναι επαρκές.

2. **Η κατανομή τους.** Αν ο θόρυβος είναι Gaussian, τα τυποποιημένα υπόλοιπα r_{kj}/s_j ακολουθούν κατά προσέγγιση την $\mathcal{N}(0,1)$, και το 4.6% από αυτά θα πρέπει να ξεπερνά το 2 κατά απόλυτη τιμή. Στη βαθμονόμηση της συστοιχίας το ποσοστό είναι 4.4%. Αυτό έχει σημασία, επειδή το «τα ελάχιστα τετράγωνα είναι μέγιστη πιθανοφάνεια» (Ενότητα B) στηρίζεται σε Gaussian θόρυβο.

3. **Ανάμεσα στους αισθητήρες.** Τα υπόλοιπα των αισθητήρων 1 και 2 έχουν συσχέτιση 0.95. Αυτή είναι η δομή ανάμεσα στους αισθητήρες της (0.1), την οποία μελετάμε στην Ενότητα C.

A.3 Όλοι οι αισθητήρες μαζί

Όλοι οι αισθητήρες έχουν τον ίδιο πίνακα $\boldsymbol{X}$, επειδή βλέπουν τις ίδιες δεξαμενές. Άρα τα δέκα προβλήματα είναι οι στήλες μίας εξίσωσης πινάκων:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\Theta} + \boldsymbol{\mathcal{N}}, \qquad \hat{\boldsymbol{\Theta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{Y}. \tag{A.5}$$

Αν στοιβάξουμε τις στήλες του $\boldsymbol{Y}$ σε ένα μακρύ διάνυσμα $\operatorname{vec}\boldsymbol{Y} \in \mathbb{R}^{CN}$, παίρνουμε ένα μεγάλο πρόβλημα ελαχίστων τετραγώνων. Ο πίνακας σχεδιασμού του είναι μπλοκ διαγώνιος,

$$\operatorname{vec}\boldsymbol{Y} = (\boldsymbol{I}_C \otimes \boldsymbol{X})\operatorname{vec}\boldsymbol{\Theta} + \operatorname{vec}\boldsymbol{\mathcal{N}}, \qquad \boldsymbol{I}_C \otimes \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X} & & \\ & \ddots & \\ & & \boldsymbol{X} \end{bmatrix}, \tag{A.6}$$

όπου $\otimes$ είναι το γινόμενο Kronecker: το $\boldsymbol{A}\otimes\boldsymbol{B}$ αντικαθιστά κάθε στοιχείο A_{ij} με το μπλοκ $A_{ij}\boldsymbol{B}$. Ο πίνακας των κανονικών εξισώσεων είναι επίσης μπλοκ διαγώνιος, οπότε το μεγάλο πρόβλημα χωρίζεται στα C μικρά.

Ο θόρυβος διαφορετικών αισθητήρων είναι συσχετισμένος. Σε counts,

$$\operatorname{Cov}(\boldsymbol{n}_k^{\text{counts}}) = \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B} =: \boldsymbol{\Sigma}_y, \qquad \boldsymbol{B} = \operatorname{diag}(b_j).$$

Θα έπρεπε λοιπόν να χρησιμοποιήσουμε σταθμισμένα ελάχιστα τετράγωνα (WLS, Διάλεξη 5) με την πλήρη συνδιακύμανση $\boldsymbol{\Sigma}_y \otimes \boldsymbol{I}_N$ του $\operatorname{vec}\boldsymbol{\mathcal{N}}$; Αποδεικνύεται ότι δεν αλλάζει τίποτα.

Πρόταση A.4 (Κοινός πίνακας σχεδιασμού: τα WLS ταυτίζονται με τα LS ανά αισθητήρα)**.** *Αν κάθε αισθητήρας έχει τον ίδιο πίνακα σχεδιασμού $\boldsymbol{X}$ και $\operatorname{Cov}(\operatorname{vec}\boldsymbol{\mathcal{N}}) = \boldsymbol{\Sigma}_y \otimes \boldsymbol{I}_N$ με $\boldsymbol{\Sigma}_y \succ 0$, τότε η εκτίμηση σταθμισμένων ελαχίστων τετραγώνων του $\operatorname{vec}\boldsymbol{\Theta}$ ταυτίζεται με την εκτίμηση ελαχίστων τετραγώνων στήλη προς στήλη* (A.5)*.*

Προαιρετικό: απόδειξη της Πρότασης A.4

Χρησιμοποιούμε τους κανόνες του γινομένου Kronecker $(\boldsymbol{A}\otimes\boldsymbol{B})^\top = \boldsymbol{A}^\top\otimes\boldsymbol{B}^\top$, $(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{C}\otimes\boldsymbol{D}) = \boldsymbol{AC}\otimes\boldsymbol{BD}$ και $(\boldsymbol{A}\otimes\boldsymbol{B})^{-1} = \boldsymbol{A}^{-1}\otimes\boldsymbol{B}^{-1}$. Με $\boldsymbol{\mathcal{X}} = \boldsymbol{I}_C\otimes\boldsymbol{X}$ και $\boldsymbol{W} = (\boldsymbol{\Sigma}_y\otimes\boldsymbol{I}_N)^{-1} = \boldsymbol{\Sigma}_y^{-1}\otimes\boldsymbol{I}_N$,

$$\boldsymbol{\mathcal{X}}^\top\boldsymbol{W}\boldsymbol{\mathcal{X}} = (\boldsymbol{I}_C\otimes\boldsymbol{X}^\top)(\boldsymbol{\Sigma}_y^{-1}\otimes\boldsymbol{I}_N)(\boldsymbol{I}_C\otimes\boldsymbol{X}) = \boldsymbol{\Sigma}_y^{-1}\otimes\boldsymbol{X}^\top\boldsymbol{X}, \qquad \boldsymbol{\mathcal{X}}^\top\boldsymbol{W} = \boldsymbol{\Sigma}_y^{-1}\otimes\boldsymbol{X}^\top.$$

Επομένως η εκτίμηση WLS είναι

$$(\boldsymbol{\Sigma}_y\otimes(\boldsymbol{X}^\top\boldsymbol{X})^{-1})(\boldsymbol{\Sigma}_y^{-1}\otimes\boldsymbol{X}^\top)\operatorname{vec}\boldsymbol{Y} = (\boldsymbol{I}_C\otimes(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)\operatorname{vec}\boldsymbol{Y},$$

που εφαρμόζει τον $(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$ σε κάθε στήλη του $\boldsymbol{Y}$ χωριστά. Ο $\boldsymbol{\Sigma}_y$ έχει απαλειφθεί.

Η ίδια άλγεβρα δίνει τη συνδιακύμανση όλων των συντελεστών:

$$\operatorname{Cov}(\operatorname{vec}\hat{\boldsymbol{\Theta}}) = \boldsymbol{\Sigma}_y \otimes (\boldsymbol{X}^\top \boldsymbol{X})^{-1}. \tag{A.7}$$

Οι συντελεστές *διαφορετικών* αισθητήρων είναι συσχετισμένοι, ακριβώς όπως και οι θόρυβοί τους. Επανερχόμαστε σε αυτό στην Παρατήρηση F.6.

Κεντρική ιδέα

Η συσχέτιση ανάμεσα στους αισθητήρες δεν αλλάζει τους συντελεστές βαθμονόμησης. Αλλάζει όμως ό,τι κάνουμε *με* τις βαθμονομημένες ενδείξεις: πώς τις συνδυάζουμε, πόσο αβέβαιες είναι και τι θεωρούμε ακραία τιμή.

A.4 Από counts σε θερμοκρασία

Αντιστρέφοντας την ευθεία βαθμονόμησης παίρνουμε την ένδειξη θερμοκρασίας κάθε αισθητήρα,

$$\hat{T}_j = \frac{y_j - \hat{a}_j}{\hat{b}_j}. \tag{A.8}$$

Αν οι συντελεστές ήταν ακριβείς, θα είχαμε $\hat{T}_j = T + n_j$, με n_j τον θόρυβο σε °C. Τα υπόλοιπα της βαθμονόμησης σε °C είναι επομένως $e_{kj} = r_{kj}/\hat{b}_j$, ο πίνακας $\boldsymbol{E}$ της Ενότητας C.

Για να δούμε τι κερδίζουμε με τη βαθμονόμηση, τη συγκρίνουμε με την ευθεία του φύλλου προδιαγραφών (μετατόπιση 200 counts και κέρδος 40 counts/°C για κάθε αισθητήρα). Στα δεδομένα βαθμονόμησης αυτή η ευθεία αφήνει σε κάθε αισθητήρα μέσο σφάλμα από -1.4 ως $+1.4\,^\circ\mathrm{C}$. Μετά τη βαθμονόμηση τα μέσα σφάλματα είναι μηδέν, και αυτό που μένει είναι ο θόρυβος, περίπου $0.2\,^\circ\mathrm{C}$ ανά αισθητήρα.

B Κακή Αριθμητική Κατάσταση, Ridge και Εκ των Προτέρων Κατανομές (Διαλέξεις 2 και 6)

Μια πλήρης βαθμονόμηση χρειάζεται αρκετές θερμοκρασίες δεξαμενής, και καθεμία θέλει χρόνο για να σταθεροποιηθεί. Είναι λοιπόν δελεαστικό να ελέγξουμε έναν αισθητήρα μετά τη συντήρηση σε μία βολική θερμοκρασία. Σε αυτή την ενότητα ρωτάμε τι μπορεί να μας πει ένας τέτοιος έλεγχος και πώς η προηγούμενη γνώση από τους υπόλοιπους αισθητήρες μπορεί να συμπληρώσει ό,τι δεν μπορεί.

B.1 Δύο σημεία αναφοράς που είναι σχεδόν ίδια

Για τον αισθητήρα 3 έχουμε μόνο μια σύντομη επαναβαθμονόμηση: $n = 10$ μετρήσεις στους $25.0\,^\circ\mathrm{C}$ και $n = 10$ στους $25.2\,^\circ\mathrm{C}$. *Με μία μόνο δεξαμενή*, και οι δύο στήλες του $\boldsymbol{X}$ είναι πολλαπλάσια του $\mathbf{1}$, οπότε $\operatorname{rank}\boldsymbol{X} = 1$ και ο $\boldsymbol{X}^\top\boldsymbol{X}$ έχει μηδενική ιδιοτιμή. Μόνο ο συνδυασμός $a + 25\,b$ είναι αναγνωρίσιμος.

Με τις δύο δεξαμενές, $\operatorname{rank}\boldsymbol{X} = 2$, αλλά $S_{TT} = 20 \cdot 0.1^2 = 0.2\,^\circ\mathrm{C}^2$ αντί για 50 000. Από την (A.3),

$$\operatorname{sd}(\hat{b}) = \frac{8.2}{\sqrt{0.2}} = 18.3\ \tfrac{\text{counts}}{^\circ\mathrm{C}}, \qquad \operatorname{sd}(\hat{a}) = 8.2\sqrt{\tfrac{1}{20} + \tfrac{25.1^2}{0.2}} = 460\ \text{counts}. \tag{B.1}$$

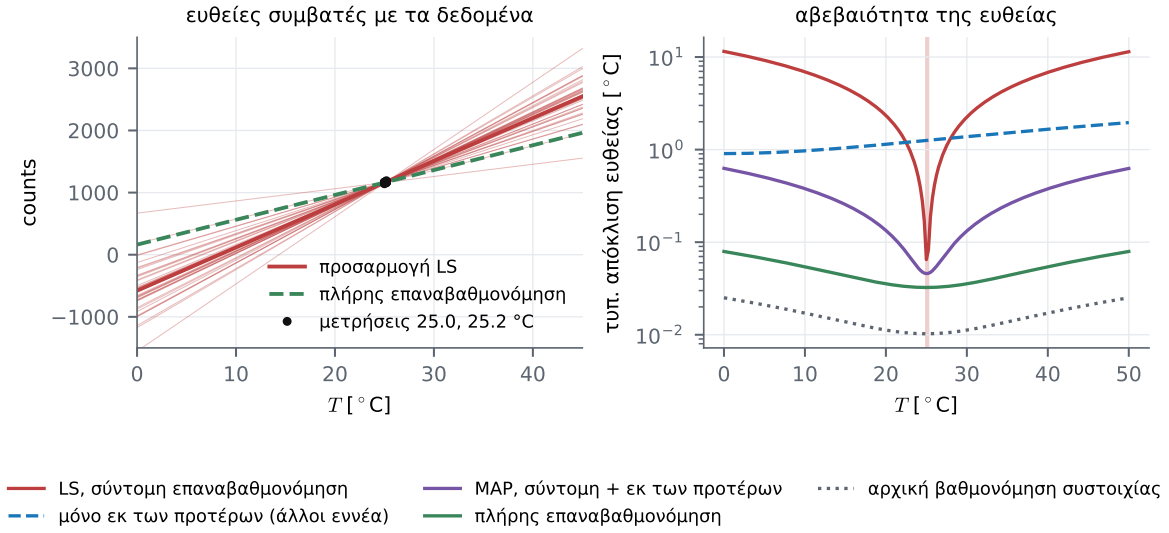
Αυτές οι τιμές είναι 500 και 460 φορές μεγαλύτερες από αυτές της βαθμονόμησης της συστοιχίας. Οι συντελεστές εκρήγνυνται: $\hat{a} = -578$ και $\hat{b} = 69.5$, ενώ η πλήρης επαναβαθμονόμηση δίνει 162 και 40.0. Κι όμως, η προσαρμοσμένη ευθεία εξακολουθεί να περνά μέσα από τα δεδομένα.

Ο μοχλοβραχίονας. Η εξίσωση (A.4), διαιρεμένη με το κέρδος, δίνει την αβεβαιότητα της ευθείας βαθμονόμησης σε °C:

$$\mathrm{sd}\left(\text{ευθεία στο } T\right) = \frac{\sigma}{b}\sqrt{\frac{1}{n} + \frac{(T-\bar{T})^2}{S_{TT}}}. \tag{B.2}$$

Στο $\bar{T} = 25.1\,^{\circ}\mathrm{C}$ μετρά μόνο ο πρώτος όρος, που δίνει $\pm 0.06\,^{\circ}\mathrm{C}$: η σύντομη επαναβαθμονόμηση καθορίζει καλά την ευθεία *εκεί όπου μετρήθηκε*. Στους 10 ή στους 40 °C κυριαρχεί ο δεύτερος όρος, $15/\sqrt{0.2} \approx 34$, και η ευθεία είναι αβέβαιη κατά περίπου $\pm 7\,^{\circ}\mathrm{C}$. Η κλίση λειτουργεί σαν μοχλός, και ένας κοντός μοχλός μεγεθύνει κάθε σφάλμα στην άκρη του (Σχήμα 1).

Η δειγματοληπτική κατανομή. Η Πρόταση A.2 είναι μια πρόταση για επαναλαμβανόμενα πειράματα. Το notebook προσομοιώνει 5000 σύντομες επαναβαθμονομήσεις, με αληθινές τιμές τους συντελεστές του αισθητήρα 3 από τη βαθμονόμηση της συστοιχίας. Τα προσαρμοσμένα κέρδη είναι σωστά κεντραρισμένα, επειδή τα ελάχιστα τετράγωνα είναι αμερόληπτα, αλλά απλώνονται με τυπική απόκλιση περίπου 18 counts/°C. Στο 1.7% των ελέγχων το προσαρμοσμένο κέρδος βγαίνει αρνητικό.



Σχήμα 1: Η σύντομη επαναβαθμονόμηση του αισθητήρα 3. Αριστερά: 40 ευθείες που προέρχονται από την $\mathcal{N}(\hat{\boldsymbol{\theta}}, \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1})$. Όλες προσαρμόζονται στα δεδομένα στους 25 °C, αλλά ανοίγουν σαν βεντάλια σε όλο το εύρος όπου θα χρησιμοποιηθεί ο αισθητήρας. Δεξιά: η τυπική απόκλιση (B.2) της ευθείας βαθμονόμησης για κάθε εκτιμητή αυτής της ενότητας (λογαριθμική κλίμακα). Η κόκκινη λωρίδα σημειώνει τις δύο δεξαμενές της σύντομης επαναβαθμονόμησης.

Η ματιά μέσω SVD. Οι ιδιόμορφες τιμές του $\boldsymbol{X}$ είναι $s_1 = 112.3$ και $s_2 = 0.018$, οπότε $\kappa(\boldsymbol{X}) = s_1/s_2 \approx 6300$ και $\kappa(\boldsymbol{X}^\top\boldsymbol{X}) = \kappa(\boldsymbol{X})^2 \approx 4 \times 10^7$ (Διάλεξη 2). Το δεξί ιδιάζον διάνυσμα της s_2 είναι

$$\boldsymbol{v}_2 \approx \pm\frac{1}{\sqrt{626}}\begin{bmatrix}25\\-1\end{bmatrix}.$$

Αν μετακινήσουμε το $\boldsymbol{\theta}$ κατά μήκος του $\boldsymbol{v}_2$, το κέρδος αυξάνεται και η μετατόπιση μειώνεται έτσι ώστε οι προβλέψεις κοντά στους 25 °C να μην αλλάζουν. Τα δεδομένα δεν βλέπουν αυτή τη διεύθυνση, και ο θόρυβος ενισχύεται κατά μήκος της με συντελεστή $1/s_2$.

B.2 Απλή Ridge: ευσταθής, αλλά πού;

Η κανονικοποίηση Ridge (regularization) λύνει την $(\boldsymbol{X}^\top \boldsymbol{X} + \lambda \boldsymbol{I})\hat{\boldsymbol{\theta}} = \boldsymbol{X}^\top \boldsymbol{y}$ (Διάλεξη 2).

Πρόταση B.1 (Η Ridge μέσω της SVD)**.** *Με τη λεπτή (thin) SVD* $\boldsymbol{X} = \sum_i s_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$*,*

$$\hat{\boldsymbol{\theta}}_\lambda = \sum_i f_i \, \frac{\boldsymbol{u}_i^\top \boldsymbol{y}}{s_i} \, \boldsymbol{v}_i, \qquad f_i = \frac{s_i^2}{s_i^2 + \lambda}. \tag{B.3}$$

Προαιρετικό: απόδειξη

Στη βάση $\boldsymbol{v}_1, \boldsymbol{v}_2$ έχουμε $\boldsymbol{X}^\top \boldsymbol{X} = \sum_i s_i^2 \boldsymbol{v}_i \boldsymbol{v}_i^\top$ και $\boldsymbol{X}^\top \boldsymbol{y} = \sum_i s_i (\boldsymbol{u}_i^\top \boldsymbol{y}) \boldsymbol{v}_i$. Ο πίνακας της Ridge γίνεται $\boldsymbol{X}^\top \boldsymbol{X} + \lambda \boldsymbol{I} = \sum_i (s_i^2 + \lambda) \boldsymbol{v}_i \boldsymbol{v}_i^\top$, με αντίστροφο $\sum_i (s_i^2 + \lambda)^{-1} \boldsymbol{v}_i \boldsymbol{v}_i^\top$. Πολλαπλασιάζουμε με $\boldsymbol{X}^\top \boldsymbol{y}$.

Στα ελάχιστα τετράγωνα κάθε συντελεστής φίλτρου είναι $f_i = 1$. Η Ridge αποσβένει κάθε διεύθυνση κατά f_i. Για $s_2^2 \ll \lambda \ll s_1^2$, δηλαδή περίπου $10^{-2} \lesssim \lambda \lesssim 10^2$ εδώ, η ασθενής διεύθυνση σβήνει ($f_2 \approx 0$) και η ισχυρή μένει ανέγγιχτη ($f_1 \approx 1$). Τότε $\hat{\boldsymbol{\theta}}_\lambda \approx (\boldsymbol{u}_1^\top \boldsymbol{y}/s_1)\boldsymbol{v}_1$, που είναι ακριβώς η λύση ελάχιστης νόρμας του προβλήματος με μία δεξαμενή. Αυτό είναι το πλατό της διαδρομής της Ridge στο notebook, στα $a \approx 2$ και $b \approx 46$. Η Ridge διάλεξε το μικρότερο $\boldsymbol{\theta}$ που ικανοποιεί το $a + 25b \approx 1160$. Είναι ευσταθής, αλλά λάθος κατά περίπου 2.4 °C στους 10 και στους 40 °C.

B.3 Τα ελάχιστα τετράγωνα είναι ML· η Ridge είναι MAP

Πρόταση B.2 (Διαλέξεις 5 και 6)**.** *Έστω* $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{n}$ *με* $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$*.*

1. *Η εκτίμηση μέγιστης πιθανοφάνειας (ML) είναι η εκτίμηση ελαχίστων τετραγώνων.*
2. *Με εκ των προτέρων κατανομή (prior)* $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0^{-1})$*, η εκ των υστέρων κατανομή (posterior) είναι Gaussian με πίνακα ακρίβειας (precision) και μέση τιμή*

$$\boldsymbol{P} = \frac{\boldsymbol{X}^\top \boldsymbol{X}}{\sigma^2} + \boldsymbol{\Lambda}_0, \qquad \hat{\boldsymbol{\theta}}_{\mathrm{MAP}} = \boldsymbol{P}^{-1}\Big(\frac{\boldsymbol{X}^\top \boldsymbol{y}}{\sigma^2} + \boldsymbol{\Lambda}_0 \boldsymbol{\mu}_0\Big), \tag{B.4}$$

   και $\mathrm{Cov}(\boldsymbol{\theta} \mid \boldsymbol{y}) = \boldsymbol{P}^{-1}$*.*

3. *Η απλή Ridge με παράμετρο* λ *είναι η περίπτωση* $\boldsymbol{\mu}_0 = \boldsymbol{0}$*,* $\boldsymbol{\Lambda}_0 = \boldsymbol{I}/\sigma_0^2$*, με* $\lambda = \sigma^2/\sigma_0^2$*.*

Προαιρετικό: απόδειξη

Ισχύει $-2 \log p(\boldsymbol{y} \mid \boldsymbol{\theta}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|^2/\sigma^2 + \text{σταθ.}$, οπότε η μεγιστοποίηση της πιθανοφάνειας ελαχιστοποιεί το τετράγωνο του υπολοίπου: αυτό είναι το μέρος 1. Για το μέρος 2 προσθέτουμε $-2 \log p(\boldsymbol{\theta}) = (\boldsymbol{\theta} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Lambda}_0 (\boldsymbol{\theta} - \boldsymbol{\mu}_0) + \text{σταθ}$. Το άθροισμα είναι τετραγωνική μορφή ως προς $\boldsymbol{\theta}$ με Εσσιανό πίνακα $2\boldsymbol{P}$. Μηδενίζοντας την παράγωγο παίρνουμε την (B.4), και η συμπλήρωση τετραγώνου δείχνει ότι η εκ των υστέρων κατανομή είναι Gaussian με πίνακα ακρίβειας $\boldsymbol{P}$ (Διάλεξη 6). Για το μέρος 3, πολλαπλασιάζουμε την (B.4) με σ^2: με $\boldsymbol{\mu}_0 = \boldsymbol{0}$ και $\boldsymbol{\Lambda}_0 = \boldsymbol{I}/\sigma_0^2$ γίνεται $(\boldsymbol{X}^\top \boldsymbol{X} + \lambda \boldsymbol{I})\hat{\boldsymbol{\theta}} = \boldsymbol{X}^\top \boldsymbol{y}$.

Το μέρος 3 λέει τι *ισχυρίστηκε* η απλή Ridge: πριν δει οποιαδήποτε δεδομένα, ότι η μετατόπιση και το κέρδος είναι και τα δύο κοντά στο μηδέν, με την ίδια διασπορά, παρόλο που μετριούνται σε διαφορετικές μονάδες. Για έναν αισθητήρα του οποίου τα counts είναι εκατοντάδες, αυτός ο ισχυρισμός είναι λάθος, και η απάντηση τον κληρονομεί.

B.4 Μια εκ των προτέρων κατανομή από τους υπόλοιπους αισθητήρες

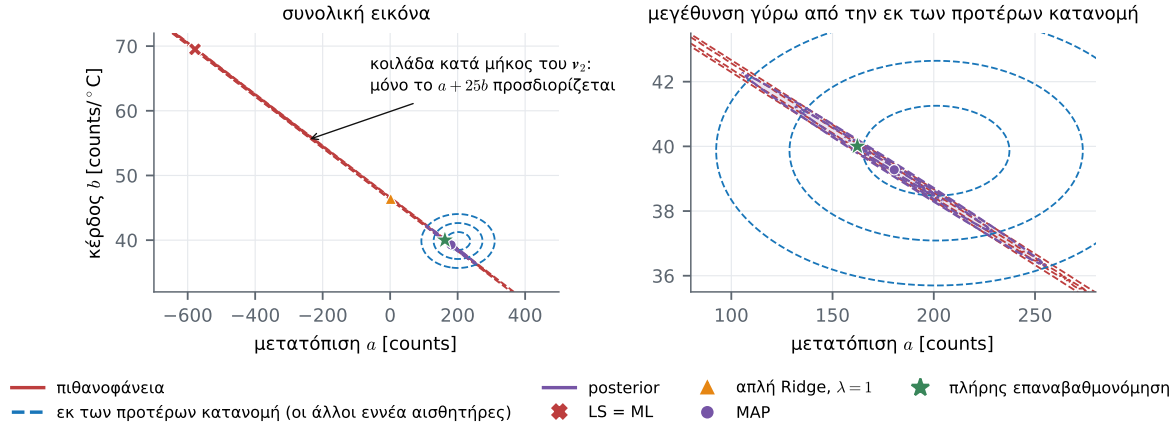
Οι άλλοι εννέα αισθητήρες είναι ίδιου τύπου και προέρχονται από την ίδια παρτίδα παραγωγής. Οι προσαρμοσμένοι συντελεστές τους μάς δίνουν λοιπόν εννέα παραδείγματα για το πώς είναι συνήθως τα (a, b) ενός τέτοιου αισθητήρα. Η δειγματική μέση τιμή και οι τυπικές αποκλίσεις τους δίνουν

$$\boldsymbol{\mu}_0 = \begin{bmatrix} 201 \\ 39.87 \end{bmatrix}, \qquad \boldsymbol{\Lambda}_0 = \operatorname{diag}\left(\frac{1}{36^2}, \frac{1}{1.39^2}\right).$$

Αυτή η εκ των προτέρων κατανομή εκφράζει γνώση για τους ίδιους τους αισθητήρες. Στη γλώσσα του μέρους 3 έχει ένα λ ανά παράμετρο, $\lambda_a = \sigma^2/\sigma_{0,a}^2 = 0.051$ και $\lambda_b = \sigma^2/\sigma_{0,b}^2 = 34.8$. Διαφέρουν κατά έναν παράγοντα σχεδόν 700, οπότε κανένα μεμονωμένο λ δεν θα μπορούσε να είναι σωστό. Αυτή είναι η προειδοποίηση της Διάλεξης 6 ότι το $\lambda\|\boldsymbol{\theta}\|^2$ προσθέτει ποσότητες με διαφορετικές μονάδες. Με $\sigma = s_3 = 8.2$ counts, η (B.4) δίνει

$$\hat{a}_{\text{MAP}} = 180 \pm 25 \text{ counts}, \qquad \hat{b}_{\text{MAP}} = 39.3 \pm 1.0 \text{ counts}/^\circ\text{C},$$

μέσα σε μία εκ των υστέρων τυπική απόκλιση από την πλήρη επαναβαθμονόμηση (162, 40.0). Η ευθεία βαθμονόμησης απέχει τώρα λιγότερο από 0.3 °C από την πλήρη επαναβαθμονόμηση σε όλο το εύρος από 10 ως 40 °C. Το δικό της περιθώριο σφάλματος, από τον $\boldsymbol{P}^{-1}$ στην (A.4), είναι ± 0.38 °C στα άκρα του εύρους (Σχήμα 1, δεξιά).



Σχήμα 2: Η σύντομη επαναβαθμονόμηση του αισθητήρα 3 στο επίπεδο (a, b). Η πιθανοφάνεια (κόκκινο) είναι μια μακριά κοιλάδα κατά μήκος της ασθενούς διεύθυνσης $\boldsymbol{v}_2$. Η εκ των προτέρων κατανομή από τους άλλους εννέα αισθητήρες (μπλε) είναι πλατιά σε άλλη διεύθυνση. Η εκ των υστέρων κατανομή (μωβ, σκιασμένη) είναι η μικρή περιοχή όπου συμφωνούν και οι δύο. Ισοϋψείς στα $\Delta \log = -0.5, -2, -4.5$, δηλαδή 1, 2 και 3 τυπικές αποκλίσεις.

Παρατήρηση B.3 (Γιατί η εκ των υστέρων κατανομή είναι τόσο μικρότερη από κάθε παράγοντα)**.** Οι πίνακες ακρίβειας προστίθενται, όπως στην (B.4). Η ακρίβεια της πιθανοφάνειας είναι τεράστια κατά μήκος του $\boldsymbol{v}_1$ και σχεδόν μηδενική κατά μήκος του $\boldsymbol{v}_2$. Η ακρίβεια της εκ των προτέρων κατανομής είναι μέτρια σε κάθε διεύθυνση. Το άθροισμά τους είναι μεγάλο σε κάθε διεύθυνση. Αυτή είναι η μορφή πινάκων του «η εκ των προτέρων κατανομή μπορεί να καλύψει τα τυφλά σημεία των δεδομένων» της Διάλεξης 6. Εδώ η εκ των υστέρων τυπική απόκλιση του b (1.0) καθορίζεται κυρίως από την εκ των προτέρων κατανομή (1.39), ενώ το $a + 25b$ το καθορίζουν τα δεδομένα.

B.5 Μεροληψία και διακύμανση σε πολλές σύντομες επαναβαθμονομήσεις

Πρόταση B.4 (Μέση τιμή και συνδιακύμανση της εκτίμησης MAP)**.** *Για σταθερό αληθινό* $\boldsymbol{\theta}$ *και θόρυβο* $\boldsymbol{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})$*,*

$$\mathbb{E}[\hat{\boldsymbol{\theta}}_{\mathrm{MAP}}] = \boldsymbol{\theta} + \boldsymbol{P}^{-1}\boldsymbol{\Lambda}_0(\boldsymbol{\mu}_0 - \boldsymbol{\theta}), \qquad \mathrm{Cov}(\hat{\boldsymbol{\theta}}_{\mathrm{MAP}}) = \boldsymbol{P}^{-1}\frac{\boldsymbol{X}^\top \boldsymbol{X}}{\sigma^2}\boldsymbol{P}^{-1}.$$

Προαιρετικό: απόδειξη

Αντικαθιστούμε το $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{n}$ στην (B.4): $\hat{\boldsymbol{\theta}}_{\mathrm{MAP}} = \boldsymbol{P}^{-1}\big(\frac{\boldsymbol{X}^\top \boldsymbol{X}}{\sigma^2}\boldsymbol{\theta} + \boldsymbol{\Lambda}_0\boldsymbol{\mu}_0\big) + \boldsymbol{P}^{-1}\frac{\boldsymbol{X}^\top \boldsymbol{n}}{\sigma^2}$. Γράφουμε $\frac{\boldsymbol{X}^\top \boldsymbol{X}}{\sigma^2} = \boldsymbol{P} - \boldsymbol{\Lambda}_0$ στον πρώτο όρο, ο οποίος γίνεται $\boldsymbol{\theta} + \boldsymbol{P}^{-1}\boldsymbol{\Lambda}_0(\boldsymbol{\mu}_0 - \boldsymbol{\theta})$, δηλαδή η μέση τιμή. Ο δεύτερος όρος έχει συνδιακύμανση $\boldsymbol{P}^{-1}\frac{\boldsymbol{X}^\top(\sigma^2\boldsymbol{I})\boldsymbol{X}}{\sigma^4}\boldsymbol{P}^{-1}$.

Η μεροληψία είναι η έλξη της εκ των προτέρων κατανομής, $\boldsymbol{P}^{-1}\boldsymbol{\Lambda}_0$, εφαρμοσμένη στο πόσο λάθος είναι η εκ των προτέρων μέση τιμή, $\boldsymbol{\mu}_0 - \boldsymbol{\theta}$. Μια εκ των προτέρων κατανομή κεντραρισμένη στην αλήθεια δεν δίνει μεροληψία. Μια κεντραρισμένη μακριά δίνει μεγάλη. Η προσομοίωση του notebook μετρά το σφάλμα της ευθείας βαθμονόμησης στους 40 °C:

εκτιμητής	μεροληψία [°C]	τυπ. απόκλιση [°C]	rms σφάλμα [°C]
ελάχιστα τετράγωνα	+0.12	6.90	6.90
απλή Ridge, $\lambda = 1$	+2.40	0.07	2.40
MAP με εκ των προτέρων	−0.28	0.06	0.29

Σημαντική διάκριση

Η εκ των υστέρων τυπική απόκλιση στους 40 °C είναι 0.38 °C, αλλά η εκτίμηση MAP μεταβάλλεται από έλεγχο σε έλεγχο μόνο κατά 0.06 °C. Οι δύο αριθμοί απαντούν σε διαφορετικά ερωτήματα. Η δειγματοληπτική διασπορά κρατά την εκ των προτέρων κατανομή σταθερή και ρωτά πόσο μετακινείται η εκτίμηση με τον θόρυβο. Η εκ των υστέρων κατανομή παραδέχεται επιπλέον ότι δεν ξέρουμε πόσο απέχει ο αισθητήρας 3 από τον μέσο αισθητήρα. Από εκεί ακριβώς προέρχεται η μεροληψία των 0.28 °C, και την καλύπτει το 0.38 της εκ των υστέρων κατανομής.

Παρατήρηση B.5 (Όροι-κλειδιά)**.** Όταν κατασκευάζουμε την εκ των προτέρων κατανομή από τα άλλα μέλη ενός πληθυσμού, η μέθοδος λέγεται *empirical Bayes*. Η πλήρως Μπεϋζιανή εκδοχή της, που βάζει εκ των προτέρων κατανομή και στα $\boldsymbol{\mu}_0$ και $\boldsymbol{\Lambda}_0$, είναι ένα *ιεραρχικό μοντέλο* (hierarchical model).

C Η Συνδιακύμανση ανάμεσα στους Αισθητήρες (Διαλέξεις 4 και 5)

Η βαθμονόμηση δίνει την ευθεία κάθε αισθητήρα, και η Ενότητα Α έδειξε ότι η συσχέτιση ανάμεσα στους αισθητήρες δεν την αλλάζει. Όλα όσα ακολουθούν όμως εξαρτώνται από αυτή τη συσχέτιση: ο συνδυασμός των δέκα ενδείξεων σε μία θερμοκρασία, το περιθώριο σφάλματός της και η διάκριση ενός χαλασμένου αισθητήρα από τον συνηθισμένο κοινό θόρυβο. Χρειαζόμαστε λοιπόν τη συνδιακύμανση $\boldsymbol{\Sigma}$ των σφαλμάτων των αισθητήρων. Τα υπόλοιπα της βαθμονόμησης μάς δίνουν 400 δείγματά της.

C.1 Εκτίμηση από τα υπόλοιπα

Η Διάλεξη 4 εκτίμησε μια συνδιακύμανση από κεντραρισμένες επαναλήψεις με διαιρέτη $N-1$. Τα υπόλοιπα της βαθμονόμησης είναι κεντραρισμένα από μια προσαρμογή με p παραμέτρους ανά αισθητήρα, και ο σωστός διαιρέτης γίνεται $N-p$.

Πρόταση C.1 (Αμερόληπτη συνδιακύμανση από υπόλοιπα παλινδρόμησης)**.** *Έστω $\boldsymbol{R} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\Theta}}$, και ας υποθέσουμε ότι οι γραμμές του πίνακα θορύβου είναι ανεξάρτητες με συνδιακύμανση $\boldsymbol{\Sigma}_y$. Τότε $\mathbb{E}[\boldsymbol{R}^\top \boldsymbol{R}] = (N-p)\,\boldsymbol{\Sigma}_y$, οπότε ο*

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N-p}\boldsymbol{E}^\top \boldsymbol{E} \tag{C.1}$$

εκτιμά τον $\boldsymbol{\Sigma}$ χωρίς μεροληψία, με εξαίρεση το πολύ μικρό σφάλμα από τη χρήση του $\hat{b}_j$ αντί του b_j.

Προαιρετικό: απόδειξη

Όπως στην Πρόταση Α.3, $\boldsymbol{R} = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{\mathcal{N}}$, οπότε $\boldsymbol{R}^\top\boldsymbol{R} = \boldsymbol{\mathcal{N}}^\top(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{\mathcal{N}}$. Το στοιχείο (i,j) είναι $\sum_{k,l}(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{X}})_{kl}\,\mathcal{N}_{ki}\mathcal{N}_{lj}$. Οι γραμμές είναι ανεξάρτητες με μηδενική μέση τιμή, άρα $\mathbb{E}[\mathcal{N}_{ki}\mathcal{N}_{lj}] = \delta_{kl}(\boldsymbol{\Sigma}_y)_{ij}$, και η αναμενόμενη τιμή είναι $\operatorname{tr}(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{X}})\,(\boldsymbol{\Sigma}_y)_{ij} = (N-p)(\boldsymbol{\Sigma}_y)_{ij}$. Η διαίρεση της στήλης j με το b_j μετατρέπει τα counts σε °C.

Για τη συστοιχία, κάθε συσχέτιση στον $\hat{\boldsymbol{\Sigma}}$ βρίσκεται ανάμεσα στο 0.56 και στο 0.95. Ο πίνακας απέχει πολύ από το να είναι διαγώνιος.

C.2 Ιδιοανάλυση και κύριες συντεταγμένες

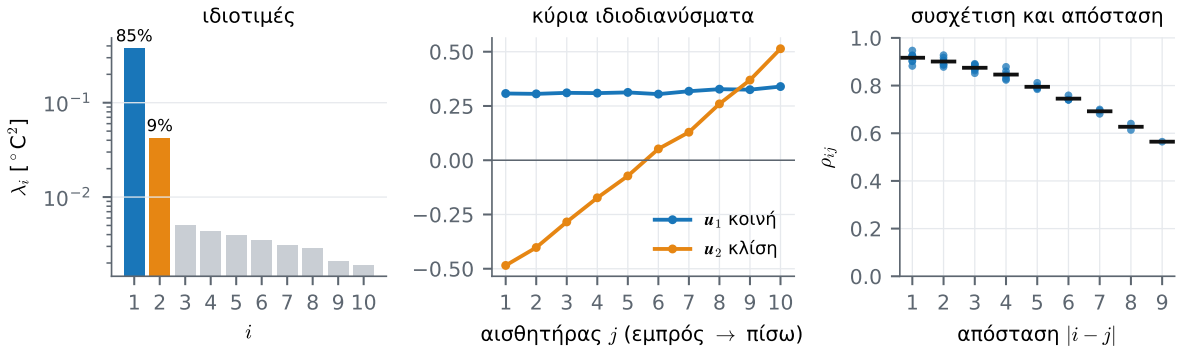
Γράφουμε $\hat{\boldsymbol{\Sigma}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$, με ιδιοτιμές $\lambda_1 \geq \cdots \geq \lambda_C$ και ορθοκανονικά ιδιοδιανύσματα $\boldsymbol{u}_i$ (ιδιοανάλυση (eigendecomposition), Διάλεξη 4).

Πρόταση C.2 (Κύριες συντεταγμένες)**.** *Αν $\operatorname{Cov}(\boldsymbol{e}) = \boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$, οι συντεταγμένες $\boldsymbol{p} = \boldsymbol{U}^\top\boldsymbol{e}$ (δηλαδή $p_i = \boldsymbol{u}_i^\top\boldsymbol{e}$) ικανοποιούν $\operatorname{Cov}(\boldsymbol{p}) = \boldsymbol{\Lambda}$: είναι ασυσχέτιστες και $\operatorname{Var}(p_i) = \lambda_i$. Επιπλέον $\sum_i \lambda_i = \operatorname{tr}\boldsymbol{\Sigma} = \sum_j \operatorname{Var}(e_j)$.*

Η απόδειξη είναι η $\boldsymbol{U}^\top\boldsymbol{\Sigma}\boldsymbol{U} = \boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{\Lambda}$, και το ίχνος δεν αλλάζει με την περιστροφή. Άρα το $\lambda_i / \sum_l \lambda_l$ είναι το *μερίδιο της συνολικής διακύμανσης* κατά μήκος του μοτίβου $\boldsymbol{u}_i$. Για τη συστοιχία,

$$\lambda = (0.373,\ 0.042,\ 0.0050,\ \ldots,\ 0.0019)\ ^\circ\mathrm{C}^2,$$

και τα δύο πρώτα μοτίβα φέρουν το 85% και το 9% της διακύμανσης (Σχήμα 3).



Σχήμα 3: Αριστερά: οι ιδιοτιμές του $\hat{\boldsymbol{\Sigma}}$ (λογαριθμική κλίμακα), με το μερίδιο της συνολικής διακύμανσης για τις δύο μεγαλύτερες. Κέντρο: τα δύο κύρια ιδιοδιανύσματα, μια κοινή συνιστώσα $\boldsymbol{u}_1$ και μια κλίση από εμπρός προς τα πίσω $\boldsymbol{u}_2$. Δεξιά: η συσχέτιση κάθε ζεύγους αισθητήρων ως συνάρτηση της απόστασής τους κατά μήκος της συστοιχίας· οι οριζόντιες γραμμές είναι μέσες τιμές.

Ένα μοντέλο που εξηγεί την εικόνα. Το μοντέλο της Διάλεξης 4 για μια κοινή διαταραχή είναι $\boldsymbol{n} = \boldsymbol{L}\boldsymbol{w}$ με $\mathrm{Cov}(\boldsymbol{w}) = \boldsymbol{I}$, οπότε $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top$. Παίρνουμε

$$\boldsymbol{n} = \sigma_c\,\mathbf{1}\,w_c + \sigma_g\,\boldsymbol{g}\,w_g + \boldsymbol{\epsilon}, \qquad \boldsymbol{\Sigma} = \sigma_c^2\mathbf{1}\mathbf{1}^\top + \sigma_g^2\boldsymbol{g}\boldsymbol{g}^\top + \boldsymbol{D}, \tag{C.2}$$

με μια κοινή συνιστώσα (common mode), ένα μοτίβο κλίσης (gradient) $\boldsymbol{g}$ που πηγαίνει γραμμικά από -1 μπροστά ως $+1$ πίσω, και ανεξάρτητο θόρυβο με διαγώνιο $\boldsymbol{D}$. Αφού $\mathbf{1}^\top\boldsymbol{g} = 0$, όταν ο $\boldsymbol{D}$ είναι μικρός τα κύρια ιδιοδιανύσματα είναι κοντά στα $\mathbf{1}/\sqrt{C}$ και $\boldsymbol{g}/\|\boldsymbol{g}\|$, με

$$\lambda_1 \approx C\sigma_c^2, \qquad \lambda_2 \approx \sigma_g^2\|\boldsymbol{g}\|^2.$$

Από τα δεδομένα, $\sigma_c \approx 0.19\,^\circ\mathrm{C}$ και $\sigma_g \approx 0.10\,^\circ\mathrm{C}$. Οι υπόλοιπες ιδιοτιμές βρίσκονται ανάμεσα στο 0.002 και στο $0.005\,^\circ\mathrm{C}^2$. Αυτές είναι ο ανεξάρτητος θόρυβος, περίπου $0.06\,^\circ\mathrm{C}$ ανά αισθητήρα. Το μοντέλο προβλέπει επίσης τις συσχετίσεις:

$$\rho_{ij} = \frac{\sigma_c^2 + \sigma_g^2 g_i g_j}{\sqrt{(\sigma_c^2 + \sigma_g^2 g_i^2 + D_{ii})(\sigma_c^2 + \sigma_g^2 g_j^2 + D_{jj})}}.$$

Οι γειτονικοί αισθητήρες έχουν $g_i g_j > 0$, οπότε η κλίση προσθέτει στη συσχέτισή τους. Τα δύο άκρα έχουν $g_1 g_{10} = -1$, οπότε η κλίση *αφαιρεί*: $\rho_{1,10} \approx (0.037 - 0.010)/(0.037 + 0.010 + 0.003) \approx 0.53$, έναντι 0.56 στη μέτρηση. Γι' αυτό η συσχέτιση φθίνει με την απόσταση, αλλά ποτέ δεν πέφτει κάτω από περίπου 0.5 (Σχήμα 3, δεξιά).

C.3 Οι εκτιμημένες συνδιακυμάνσεις υπερεκτιμούν τις αποστάσεις

Παρατήρηση C.3. Αν ο $\hat{\boldsymbol{\Sigma}}$ εκτιμάται από ν βαθμούς ελευθερίας και το $\boldsymbol{e}$ είναι ένα νέο, ανεξάρτητο δείγμα από την $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, τότε

$$\mathbb{E}\big[\boldsymbol{e}^\top\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{e}\big] = C\,\frac{\nu}{\nu - C - 1}.$$

Αναφέρουμε αυτό το αποτέλεσμα χωρίς απόδειξη (όροι-κλειδιά: *κατανομή Wishart*, *αντίστροφη Wishart*). Ο λόγος είναι ότι οι μικρότερες ιδιοτιμές μιας δειγματικής συνδιακύμανσης τείνουν να βγαίνουν υπερβολικά μικρές, και ο $\hat{\boldsymbol{\Sigma}}^{-1}$ μεγεθύνει ακριβώς αυτές τις διευθύνσεις. Εδώ $\nu = N - p = 398$ και $C = 10$, οπότε η υπερεκτίμηση είναι κατά μέσο όρο μόνο περίπου 3%. Ένας συγκεκριμένος $\hat{\boldsymbol{\Sigma}}$ μπορεί να απέχει περισσότερο, και η μέση διακύμανση 1.10 στα δεδομένα ελέγχου (held-out) του notebook συνδυάζει αυτό το φαινόμενο με τη δειγματοληπτική διασπορά. Με μόνο 100 λήψεις βαθμονόμησης ($\nu = 98$) η μέση υπερεκτίμηση θα ήταν ήδη 13%.

D Η Κατανομή χ^2

Στο υπόλοιπο μέρος της συνεδρίας ανάγουμε ένα διάνυσμα σφαλμάτων σε έναν αριθμό, «πόσο απέχει αυτή η λήψη από το κανονικό;», και στη συνέχεια πρέπει να αποφασίσουμε *πόσο μεγάλο είναι υπερβολικά μεγάλο*. Για μία Gaussian μεταβλητή γνωρίζουμε ήδη την απάντηση.

Το z-score. Αν μια ένδειξη x έχει μέση τιμή μ και τυπική απόκλιση σ, το *z-score* (τυποποιημένη τιμή) της είναι

$$z = \frac{x - \mu}{\sigma},$$

δηλαδή ο αριθμός των τυπικών αποκλίσεων κατά τις οποίες το x διαφέρει από τη μέση τιμή του. Αν το x είναι Gaussian, τότε $z \sim \mathcal{N}(0,1)$, όποια κι αν είναι τα μ και σ, οπότε ένας πίνακας αρκεί για κάθε μεταβλητή: το $|z| > 1.96$ συμβαίνει στο 5% των περιπτώσεων και το $|z| > 2.576$ στο 1%. Τα τυποποιημένα υπόλοιπα r_{kj}/s_j της Ενότητας A είναι z-scores. Για το σφάλμα ενός αισθητήρα, η μέση τιμή είναι μηδέν όταν ο αισθητήρας είναι υγιής, οπότε το z-score του είναι e_j/σ_j. Ένας έλεγχος 1% σημαίνει συναγερμό όταν $|z| > 2.576$, και ένας υγιής αισθητήρας σημαίνεται στο 1% των περιπτώσεων. Αυτή η ενότητα επεκτείνει την απάντηση σε πολλές μεταβλητές μαζί.

D.1 Ορισμός και πρώτες ιδιότητες

Ορισμός D.1 (Κατανομή χ-τετράγωνο)**.** Αν οι $z_1, \dots, z_\nu$ είναι ανεξάρτητες Gaussian μεταβλητές με μηδενική μέση τιμή και μοναδιαία διακύμανση, $z_i \sim \mathcal{N}(0,1)$, η κατανομή του

$$d^2 = z_1^2 + \cdots + z_\nu^2 = \|\boldsymbol{z}\|^2$$

είναι η *κατανομή χ-τετράγωνο* (chi-squared) *με ν βαθμούς ελευθερίας*, που τη συμβολίζουμε χ^2_ν. Το q-ποσοστημόριό της, δηλαδή η τιμή που ξεπερνιέται με πιθανότητα $1-q$, το γράφουμε $\chi^2_\nu(q)$.

Άρα το d^2 είναι το τετράγωνο της απόστασης από την αρχή ενός σημείου που προέρχεται από την $\mathcal{N}(\mathbf{0}, \boldsymbol{I}_\nu)$, ένα τέλεια στρογγυλό Gaussian νέφος. Για $\nu = 1$ είναι απλώς το z^2, και ο έλεγχος 1% $|z| > 2.576$ είναι ο έλεγχος $z^2 > 2.576^2 = 6.63 = \chi^2_1(0.99)$.

Πρόταση D.2 (Μέση τιμή και διακύμανση)**.** *Αν $d^2 \sim \chi^2_\nu$, τότε $\mathbb{E}[d^2] = \nu$ και $\mathrm{Var}(d^2) = 2\nu$.*

Προαιρετικό: απόδειξη

Για μια μεταβλητή $z \sim \mathcal{N}(0,1)$ ισχύει $\mathbb{E}[z^2] = 1$. Με ολοκλήρωση κατά παράγοντες και $\varphi'(z) = -z\varphi(z)$, όπου φ η πυκνότητα της $\mathcal{N}(0,1)$, παίρνουμε $\mathbb{E}[z^4] = \int z^3 \cdot z\varphi(z)\,dz = \int 3z^2\varphi(z)\,dz = 3$. Άρα $\mathrm{Var}(z^2) = 3 - 1 = 2$. Οι ν όροι είναι ανεξάρτητοι, οπότε οι μέσες τιμές τους προστίθενται, και το ίδιο οι διακυμάνσεις τους.

Μια τιμή της χ^2_ν είναι επομένως συνήθως $\nu \pm \sqrt{2\nu}$· για $\nu = 10$ αυτό είναι 10 ± 4.5. Κάθε συντεταγμένη συνεισφέρει κατά μέσο όρο περίπου 1, οπότε το σύνολο μεγαλώνει με τον αριθμό των συντεταγμένων.

Πρόταση D.3 (Δύο βαθμοί ελευθερίας)**.** *Αν $d^2 \sim \chi^2_2$, τότε $P(d^2 > x) = e^{-x/2}$ για $x \geq 0$. Επομένως $\chi^2_2(q) = -2\ln(1-q)$· για παράδειγμα, $\chi^2_2(0.95) = 5.99$.*

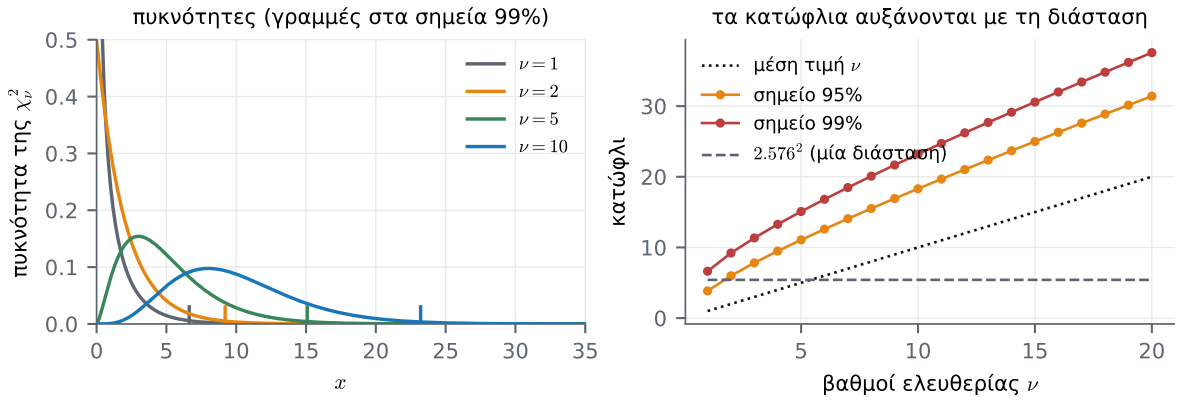
Προαιρετικό: απόδειξη

Η πυκνότητα του (z_1, z_2) είναι $\frac{1}{2\pi}e^{-(z_1^2+z_2^2)/2}$. Σε πολικές συντεταγμένες, με $\rho^2 = z_1^2 + z_2^2$

και $dz_1\,dz_2 = \rho\,d\rho\,d\phi$,

$$P(\rho^2 > x) = \int_0^{2\pi}\int_{\sqrt{x}}^{\infty} \frac{1}{2\pi} e^{-\rho^2/2}\rho\,d\rho\,d\phi = \Big[-e^{-\rho^2/2}\Big]_{\sqrt{x}}^{\infty} = e^{-x/2}.$$

Για άλλα ν δεν υπάρχει τόσο απλός τύπος, και χρησιμοποιούμε πίνακες ή λογισμικό. Στην Python, η `scipy.stats.chi2.ppf(q, nu)` δίνει το $\chi^2_\nu(q)$ και η `scipy.stats.chi2.sf(x, nu)` δίνει το $P(d^2 > x)$.



Σχήμα 4: Αριστερά: πυκνότητες της χ^2_ν· οι μικρές κατακόρυφες γραμμές σημειώνουν τα σημεία 99%. Καθώς μεγαλώνει το ν, η κατανομή μετακινείται προς τα δεξιά και απλώνεται. Δεξιά: τα σημεία 95% και 99% ως συνάρτηση του ν, σε σύγκριση με τη μέση τιμή ν και με το μονοδιάστατο σημείο 99%, 2.576^2.

D.2 Γιατί το κατώφλι πρέπει να μεγαλώνει με τη διάσταση

Το Σχήμα 4 (δεξιά) δείχνει το σημείο 99% να μεγαλώνει από 6.63 για $\nu = 1$ σε 23.21 για $\nu = 10$. Αν χρησιμοποιούσαμε το μονοδιάστατο κατώφλι σε δέκα διαστάσεις, θα είχαμε πάρα πολλούς συναγερμούς. Το ίδιο φαινόμενο εμφανίζεται αν ελέγξουμε κάθε συντεταγμένη χωριστά. Ακόμη και για δέκα *ανεξάρτητους* αισθητήρες, ο καθένας ελεγμένος στο 1%, η πιθανότητα να σημάνει συναγερμό τουλάχιστον ένας όταν όλα είναι εντάξει είναι

$$1 - 0.99^{10} \approx 9.6\%.$$

Ένας ενιαίος έλεγχος χ^2_ν εξετάζει όλες τις συντεταγμένες μαζί και κρατά τον συνολικό ρυθμό ψευδών συναγερμών (false alarms) ακριβώς στο επίπεδο που επιλέγουμε.

Κεντρική ιδέα

Η συνταγή της χ^2_ν εφαρμόζεται σε ένα άθροισμα ν τετραγώνων μεταβλητών που είναι *ανεξάρτητες* και ακολουθούν *την* $\mathcal{N}(0,1)$. Τα ακατέργαστα σφάλματα των αισθητήρων δεν είναι τίποτα από τα δύο: είναι συσχετισμένα και έχουν διαφορετικές διακυμάνσεις. Η λευκοποίηση (whitening, Ενότητα E) είναι ακριβώς το βήμα που τα μετατρέπει σε τέτοιες μεταβλητές.

Τα ποσοστημόρια που χρησιμοποιούμε στη συνεδρία.

	$\chi^2_1(0.99)$	$\chi^2_2(0.95)$	$\chi^2_8(0.99)$	$\chi^2_9(0.99)$	$\chi^2_{10}(0.99)$
τιμή	6.63	5.99	20.09	21.67	23.21
χρήση	μία μεταβλητή	ελλείψεις	χωρίς αναφορά, LOO	χωρίς αναφορά	με αναφορά

LOO σημαίνει leave one out, αφαιρούμε έναν αισθητήρα κάθε φορά (Ενότητα E.5)· «χωρίς αναφορά» είναι η Ενότητα G.

E Ελλείψεις, Λευκοποίηση και Ανίχνευση (Διάλεξη 4)

Περνάμε τώρα στο πείραμα επικύρωσης και ρωτάμε, για κάθε λήψη, αν οι δέκα ενδείξεις είναι συμβατές με μια υγιή συστοιχία. Η απάντηση πρέπει να λαμβάνει υπόψη τον κοινό θόρυβο: όταν όλοι οι αισθητήρες αποκλίνουν μαζί προς την ίδια κατεύθυνση, αυτό είναι φυσιολογικό, αλλά όταν ένας αισθητήρας αποκλίνει αντίθετα από τους άλλους, κάτι δεν πάει καλά.

E.1 Ελλείψεις συνδιακύμανσης

Για δύο αισθητήρες, το μπλοκ $\boldsymbol{\Sigma}_2$ του $\boldsymbol{\Sigma}$ ορίζει τις ελλείψεις $\boldsymbol{e}^\top \boldsymbol{\Sigma}_2^{-1} \boldsymbol{e} = c^2$. Οι άξονές τους δείχνουν κατά μήκος των ιδιοδιανυσμάτων του $\boldsymbol{\Sigma}_2$, με ημιάξονες $c\sqrt{\lambda_i}$ (Διάλεξη 4). Από την Πρόταση E.2 παρακάτω, εφαρμοσμένη σε δύο διαστάσεις, $\boldsymbol{e}^\top \boldsymbol{\Sigma}_2^{-1} \boldsymbol{e} \sim \chi^2_2$. Από την Πρόταση D.3, η έλλειψη περιέχει τότε ένα ποσοστό $1 - e^{-c^2/2}$ των δειγμάτων:

c	1	2	3	$\sqrt{5.99} = 2.45$
ποσοστό στο εσωτερικό	39%	86%	99%	95%

Η έλλειψη «μίας τυπικής απόκλισης» περιέχει μόνο το 39%, όχι το 68% ενός διαστήματος μίας τυπικής απόκλισης σε μία διάσταση. Το notebook μετρά τα υπόλοιπα της βαθμονόμησης μέσα σε κάθε έλλειψη και βρίσκει 40%, 88% και 98% για τους αισθητήρες 1 και 2.

E.2 Λευκοποίηση, και πώς τη ελέγχουμε με εντιμότητα

Με $\hat{\boldsymbol{\Sigma}} = \boldsymbol{L}\boldsymbol{L}^\top$ (παραγοντοποίηση Cholesky), το λευκοποιημένο διάνυσμα $\boldsymbol{q}$ λύνει την $\boldsymbol{L}\boldsymbol{q} = \boldsymbol{e}$, οπότε $\mathrm{Cov}(\boldsymbol{q}) = \boldsymbol{L}^{-1}\boldsymbol{\Sigma}\boldsymbol{L}^{-\top} = \boldsymbol{I}$ (Διάλεξη 4).

Πρόταση E.1 (Δύο πίνακες λευκοποίησης, ένα μήκος)**.** *Έστω $\boldsymbol{W}$ οποιοσδήποτε πίνακας με $\boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{W}^\top = \boldsymbol{I}$, για παράδειγμα ο $\boldsymbol{L}^{-1}$ ή ο $\boldsymbol{W}_{\mathrm{sym}} = \boldsymbol{U}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}^\top$. Τότε $\|\boldsymbol{W}\boldsymbol{e}\|^2 = \boldsymbol{e}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{e}$ για κάθε $\boldsymbol{e}$. Δύο τέτοιοι πίνακες λευκοποίησης διαφέρουν κατά έναν ορθογώνιο πίνακα.*

Προαιρετικό: απόδειξη

Από την $\boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{W}^\top = \boldsymbol{I}$, ο $\boldsymbol{W}$ είναι αντιστρέψιμος και $\boldsymbol{\Sigma} = \boldsymbol{W}^{-1}\boldsymbol{W}^{-\top}$, άρα $\boldsymbol{\Sigma}^{-1} = \boldsymbol{W}^\top\boldsymbol{W}$ και $\|\boldsymbol{W}\boldsymbol{e}\|^2 = \boldsymbol{e}^\top\boldsymbol{W}^\top\boldsymbol{W}\boldsymbol{e} = \boldsymbol{e}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{e}$. Αν οι $\boldsymbol{W}_1$ και $\boldsymbol{W}_2$ λευκοποιούν και οι δύο, ο $\boldsymbol{O} = \boldsymbol{W}_2\boldsymbol{W}_1^{-1}$ ικανοποιεί την $\boldsymbol{O}\boldsymbol{O}^\top = \boldsymbol{W}_2\boldsymbol{\Sigma}\boldsymbol{W}_2^\top = \boldsymbol{I}$.

Οι συντεταγμένες του $\boldsymbol{q}$ εξαρτώνται από τον πίνακα λευκοποίησης, το μήκος του όμως όχι. Παρακάτω χρησιμοποιούμε μόνο το μήκος.

Σημαντική διάκριση

Αν λευκοποιήσουμε *τα ίδια* υπόλοιπα που έδωσαν τον $\hat{\boldsymbol{\Sigma}}$, παίρνουμε ακριβώς $\boldsymbol{I}$, για οποιαδήποτε δεδομένα. Με $\boldsymbol{Q} = \boldsymbol{E}\boldsymbol{L}^{-\top}$,

$$\frac{\boldsymbol{Q}^\top\boldsymbol{Q}}{N-p} = \boldsymbol{L}^{-1}\frac{\boldsymbol{E}^\top\boldsymbol{E}}{N-p}\boldsymbol{L}^{-\top} = \boldsymbol{L}^{-1}\hat{\boldsymbol{\Sigma}}\boldsymbol{L}^{-\top} = \boldsymbol{I}.$$

Αυτό ελέγχει τον κώδικα, όχι το μοντέλο. Το μοντέλο το ελέγχουμε σε δεδομένα που ο $\hat{\boldsymbol{\Sigma}}$ δεν έχει δει ποτέ.

Σε δεδομένα ελέγχου (held-out), ο $\mathrm{Cov}(\boldsymbol{q})$ δεν θα είναι ακριβώς $\boldsymbol{I}$, για δύο λόγους. Ο πρώτος είναι η συνηθισμένη δειγματοληπτική διασπορά: με M δείγματα, περίπου $\pm 1/\sqrt{M}$ για τα εκτός διαγωνίου στοιχεία και $\pm\sqrt{2/M}$ για τα διαγώνια. Ο δεύτερος είναι ότι ο $\hat{\boldsymbol{\Sigma}}$ είναι εκτίμηση (Παρατήρηση C.3). Στις πρώτες 200 λήψεις του πειράματος επικύρωσης, η διαγώνιος του $\mathrm{Cov}(\boldsymbol{q})$ έχει μέση τιμή 1.10.

E.3 Η απόσταση Mahalanobis και η κατανομή της

Στο πείραμα επικύρωσης, το θερμόμετρο αναφοράς δίνει το T, οπότε κάθε λήψη μάς δίνει το διάνυσμα σφάλματος $\boldsymbol{e} = \hat{\boldsymbol{t}} - T\mathbf{1} \in \mathbb{R}^C$. Η απόσταση Mahalanobis του είναι

$$d^2 = \boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e} = \|\boldsymbol{q}\|^2, \qquad \boldsymbol{L}\boldsymbol{q} = \boldsymbol{e}. \tag{E.1}$$

Πρόταση E.2. *Αν η συστοιχία είναι υγιής, $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, τότε $\boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_C)$ και $d^2 \sim \chi^2_C$.*

Απόδειξη. Το $\boldsymbol{q} = \boldsymbol{L}^{-1}\boldsymbol{e}$ είναι γραμμική απεικόνιση ενός Gaussian διανύσματος, άρα είναι Gaussian, με συνδιακύμανση $\boldsymbol{L}^{-1}\boldsymbol{\Sigma}\boldsymbol{L}^{-\top} = \boldsymbol{I}$. Από κοινού Gaussian μεταβλητές που είναι ασυσχέτιστες είναι ανεξάρτητες, οπότε το $d^2 = \sum_i q_i^2$ είναι άθροισμα των τετραγώνων C ανεξάρτητων μεταβλητών $\mathcal{N}(0,1)$. □

Η Πρόταση E.2 δίνει έναν έλεγχο με γνωστό ρυθμό ψευδών συναγερμών: σημαίνουμε μια λήψη όταν $d^2 > \chi^2_C(0.99) = 23.21$, και για μια υγιή συστοιχία έχουμε συναγερμό στο 1% των λήψεων. Στις πρώτες 200 λήψεις το notebook μετρά μέσο d^2 ίσο με 11.0, έναντι $\nu = 10$, και ρυθμό συναγερμών 1.5%. Και τα δύο είναι ελαφρώς υψηλά, κάτι συμβατό με την Παρατήρηση C.3.

Παρατήρηση E.3 (Πολλές λήψεις, πολλοί έλεγχοι)**.** Ένας έλεγχος 1% σε 300 υγιείς λήψεις δίνει περίπου τρεις συναγερμούς. Ένας μεμονωμένος συναγερμός είναι επομένως ασθενής ένδειξη. Βλάβη σημαίνει ένας *ρυθμός* πολύ πάνω από 1%, ή συναγερμοί που επιμένουν. Έτσι διαβάζει τα αποτελέσματα το notebook.

E.4 Γιατί η απόσταση Mahalanobis βλέπει ό,τι χάνει ένα z-score

Ο προφανής ανιχνευτής ελέγχει κάθε αισθητήρα χωριστά: υπολογίζει το z-score του, $z_j = e_j/\sqrt{\Sigma_{jj}}$, και σημαίνει συναγερμό όταν $|z_j| > 2.576$. Ας υποθέσουμε ότι ο αισθητήρας j αποκτά μια μετατόπιση δ, οπότε $\boldsymbol{e} \sim \mathcal{N}(\delta\,\boldsymbol{\epsilon}_j, \boldsymbol{\Sigma})$, όπου $\boldsymbol{\epsilon}_j$ είναι το j-οστό μοναδιαίο διάνυσμα.

Πρόταση E.4 (Τι κάνει μια μετατόπιση σε έναν αισθητήρα)**.** *1. Η μετατόπιση μετακινεί κατά $\delta/\sqrt{\Sigma_{jj}}$ το z-score του ίδιου του αισθητήρα, $z_j = e_j/\sqrt{\Sigma_{jj}}$, όπου Σ_{jj} είναι η συνολική διακύμανση του αισθητήρα j.*

2. Η απόσταση Mahalanobis έχει μέση τιμή $\mathbb{E}[d^2] = C + \delta^2(\boldsymbol{\Sigma}^{-1})_{jj}$, και

$$\frac{1}{(\boldsymbol{\Sigma}^{-1})_{jj}} = \mathrm{Var}\left(e_j \mid e_i,\ i \neq j\right), \tag{E.2}$$

η διακύμανση του αισθητήρα j που απομένει όταν οι άλλοι αισθητήρες είναι γνωστοί.

Προαιρετικό: απόδειξη του μέρους 2

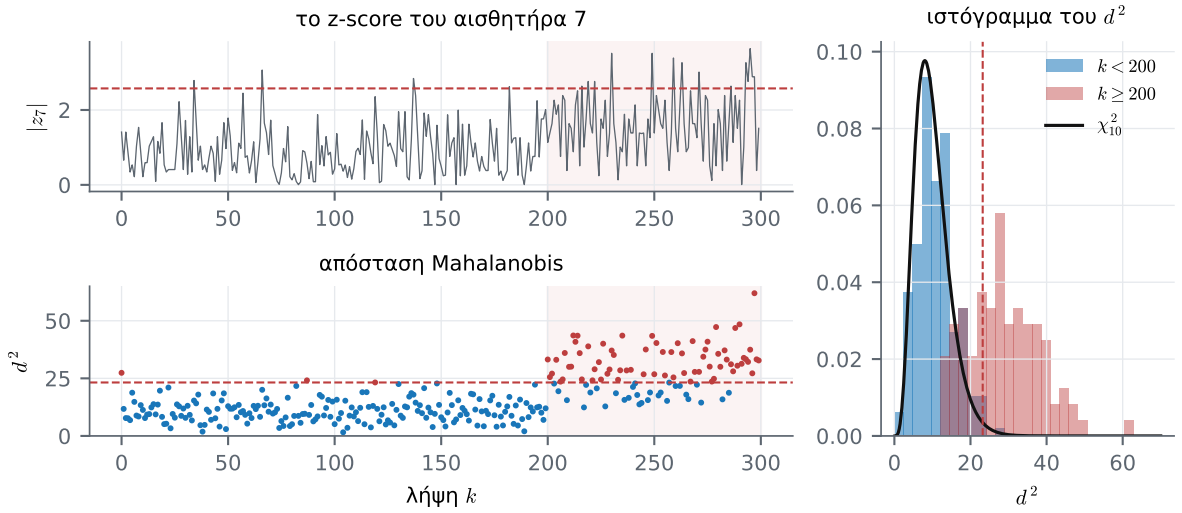
Γράφουμε $\boldsymbol{e} = \delta\boldsymbol{\epsilon}_j + \boldsymbol{e}_0$ με $\boldsymbol{e}_0 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Τότε $\mathbb{E}[d^2] = \mathbb{E}[\boldsymbol{e}_0^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_0] + \delta^2\boldsymbol{\epsilon}_j^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}_j = C + \delta^2(\boldsymbol{\Sigma}^{-1})_{jj}$, αφού ο μικτός όρος έχει μηδενική μέση τιμή. Για την (E.2), διατάσσουμε

τους αισθητήρες ώστε ο j να είναι τελευταίος και διαμερίζουμε $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{oo} & \boldsymbol{s} \\ \boldsymbol{s}^\top & \Sigma_{jj} \end{bmatrix}$. Ο τύπος αντιστροφής κατά μπλοκ δίνει το τελευταίο διαγώνιο στοιχείο του $\boldsymbol{\Sigma}^{-1}$ ίσο με $1/(\Sigma_{jj} - \boldsymbol{s}^\top \boldsymbol{\Sigma}_{oo}^{-1} \boldsymbol{s})$. Ο παρονομαστής είναι η διακύμανση του σφάλματος της βέλτιστης γραμμικής πρόβλεψης του e_j από τους άλλους, που για μια Gaussian κατανομή είναι η δεσμευμένη διακύμανση.

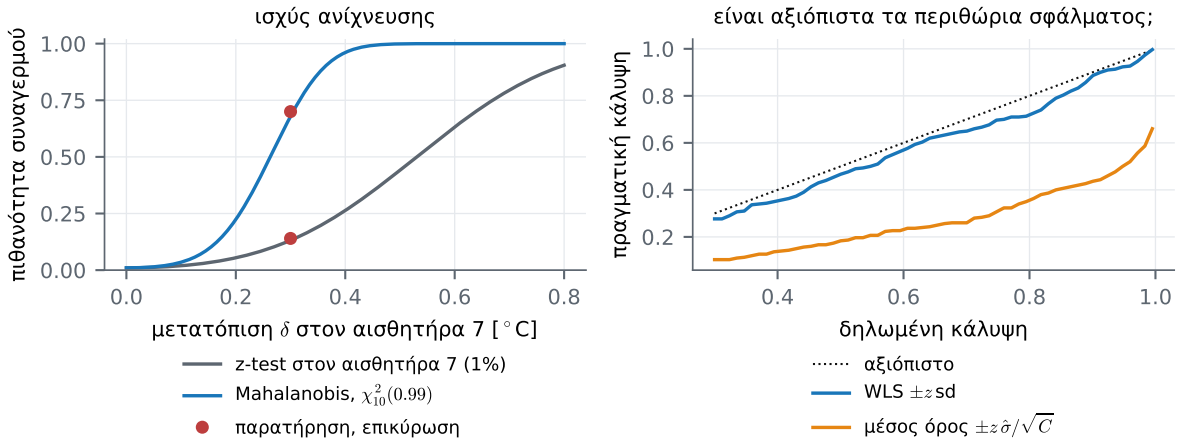
Ορισμός Ε.5 (Μη κεντρική χ^2)**.** Αν $\boldsymbol{q} \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{I}_\nu)$, η κατανομή του $\|\boldsymbol{q}\|^2$ είναι η *μη κεντρική κατανομή χ-τετράγωνο* (noncentral chi-squared) με ν βαθμούς ελευθερίας και παράμετρο μη κεντρικότητας $\|\boldsymbol{m}\|^2$. Στην Python είναι η `scipy.stats.ncx2`.

Μετά τη λευκοποίηση, μια μετατόπιση δ στον αισθητήρα j μετακινεί τη μέση τιμή του $\boldsymbol{q}$ κατά $\boldsymbol{L}^{-1}\delta\boldsymbol{\epsilon}_j$, του οποίου το τετράγωνο του μήκους είναι $\delta^2(\boldsymbol{\Sigma}^{-1})_{jj}$. Άρα το d^2 ακολουθεί μη κεντρική χ^2_C με αυτή την παράμετρο μη κεντρικότητας, και μπορούμε να υπολογίσουμε την πιθανότητα συναγερμού για οποιοδήποτε δ.

Για τον αισθητήρα 7, η συνολική τυπική απόκλιση είναι 0.21 °C, αλλά η δεσμευμένη είναι μόνο 0.07 °C: οι άλλοι εννέα αισθητήρες προβλέπουν το μεγαλύτερο μέρος του θορύβου του. Μια μετατόπιση $\delta = 0.3\,^\circ$C είναι 1.5 συνολικές τυπικές αποκλίσεις, αλλά 4.3 δεσμευμένες. Ο z-έλεγχος στο 1% σημαίνει επομένως συναγερμό με πιθανότητα 13%, ενώ το d^2 έχει παράμετρο μη κεντρικότητας $4.3^2 \approx 18.4$ και ξεπερνά το 23.21 με πιθανότητα 68%. Το notebook παρατηρεί 14% και 70%. Για να πιάσει τις μισές βλάβες, ο z-έλεγχος χρειάζεται $\delta \approx 0.53\,^\circ$C, ενώ ο έλεγχος Mahalanobis χρειάζεται $\delta \approx 0.26\,^\circ$C (Σχήμα 6, αριστερά).



Σχήμα 5: Πείραμα επικύρωσης. Αριστερά: το z-score του αισθητήρα 7 μόλις που αντιδρά στη βλάβη των 0.3 °C από τη λήψη 200 (σκιασμένη περιοχή). Η απόσταση Mahalanobis ξεπερνά το $\chi^2_{10}(0.99)$ (διακεκομμένη) στο 1.5% των υγιών λήψεων και στο 70% των λήψεων με βλάβη. Δεξιά: πριν από τη βλάβη, το ιστόγραμμα του d^2 ακολουθεί την πυκνότητα της χ^2_{10}· μετά από αυτήν, το ιστόγραμμα μετακινείται προς τα δεξιά.



Σχήμα 6: Αριστερά: πιθανότητα συναγερμού ως συνάρτηση του μεγέθους μιας μετατόπισης στον αισθητήρα 7, για τον z-έλεγχο ανά αισθητήρα και για τον έλεγχο Mahalanobis, από την Πρόταση E.4· οι τελείες είναι οι ρυθμοί που παρατηρούνται μετά τη λήψη 200. Δεξιά: δηλωμένη έναντι πραγματικής κάλυψης των διαστημάτων $\pm z \cdot \mathrm{sd}$ για τη θερμοκρασία που αναφέρουμε (Ενότητα F).

Κεντρική ιδέα

Ένα z-score συγκρίνει έναν αισθητήρα με τη δική του διασπορά. Η απόσταση Mahalanobis τον συγκρίνει με *αυτό που οι άλλοι αισθητήρες λένε ότι θα έπρεπε να δείχνει*. Όταν ο θόρυβος είναι κοινός, η δεύτερη διασπορά είναι πολύ μικρότερη, οπότε οι βλάβες ξεχωρίζουν.

E.5 Ποιος αισθητήρας; Αφαιρούμε έναν κάθε φορά

Πρόταση E.6. *Αν* $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ *και το* $\boldsymbol{e}_{-j}$ *προκύπτει αφαιρώντας τη συνιστώσα* j*, τότε* $\boldsymbol{e}_{-j} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{-j})$*, όπου ο* $\boldsymbol{\Sigma}_{-j}$ *προκύπτει αφαιρώντας τη γραμμή και τη στήλη* j*. Επομένως* $\boldsymbol{e}_{-j}^\top \boldsymbol{\Sigma}_{-j}^{-1} \boldsymbol{e}_{-j} \sim \chi^2_{C-1}$.

Ο πρώτος ισχυρισμός ισχύει επειδή $\boldsymbol{e}_{-j} = \boldsymbol{S}\boldsymbol{e}$ για έναν πίνακα επιλογής $\boldsymbol{S}$, και ο $\mathrm{Cov}(\boldsymbol{S}\boldsymbol{e}) = \boldsymbol{S}\boldsymbol{\Sigma}\boldsymbol{S}^\top$ απομονώνει το αντίστοιχο μπλοκ. Ο δεύτερος ισχυρισμός είναι η Πρόταση E.2 σε $C-1$ διαστάσεις. Αν ο αισθητήρας j είναι ο χαλασμένος, η αφαίρεσή του αφήνει εννέα υγιείς αισθητήρες, και οι συναγερμοί πέφτουν στο επίπεδο του 1%. Αν αφαιρέσουμε οποιονδήποτε άλλο αισθητήρα, η βλάβη μένει μέσα, και οι συναγερμοί παραμένουν. Στο notebook, η αφαίρεση του αισθητήρα 7 εξαφανίζει το 97% των συναγερμών μετά τη λήψη 200, ενώ η αφαίρεση οποιουδήποτε άλλου εξαφανίζει το πολύ το 7%.

Η βλάβη, εντοπισμένη

Ο **αισθητήρας 7** αποκτά μια μετατόπιση περίπου $+0.3\,^\circ$C στη **λήψη 200**: το μέσο σφάλμα του είναι $-0.04\,^\circ$C πριν και $+0.32\,^\circ$C μετά. Αυτό είναι 1.5 φορές η δική του τυπική απόκλιση, οπότε ο δικός του z-έλεγχος πιάνει μόνο το 14% των λήψεων με βλάβη. Ο έλεγχος Mahalanobis πιάνει το 70%, επειδή, δεδομένων των άλλων εννέα αισθητήρων, η βλάβη είναι 4.3 τυπικές αποκλίσεις. Στη συνέχεια η αφαίρεση ενός αισθητήρα κάθε φορά δείχνει τον ένοχο: η αφαίρεση του αισθητήρα 7 εξαφανίζει το 97% των συναγερμών, ενώ η αφαίρεση οποιουδήποτε άλλου το πολύ το 7%. Η Ενότητα G βρίσκει τον ίδιο αισθητήρα χωρίς το θερμόμετρο αναφοράς.

F Μία Θερμοκρασία με Αξιόπιστο Περιθώριο Σφάλματος (Διάλεξη 5)

Δέκα ενδείξεις της ίδιας θερμοκρασίας είναι πλεονασμός. Ο χρήστης της συστοιχίας θέλει έναν αριθμό ανά λήψη, με περιθώριο σφάλματος που μπορεί να εμπιστευτεί.

F.1 Ο βέλτιστος γραμμικός συνδυασμός

Κάθε υγιής αισθητήρας δείχνει $\hat{T}_j = T + e_j$ με $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Αυτό είναι ένα γραμμικό μοντέλο με μία μόνο παράμετρο, το T, και πίνακα σχεδιασμού $\mathbf{1}$, οπότε τα σταθμισμένα ελάχιστα τετράγωνα της Διάλεξης 5 εφαρμόζονται απευθείας.

Πρόταση F.1 (Συνδυασμός συσχετισμένων αισθητήρων)**.** *Ανάμεσα σε όλες τις εκτιμήσεις $\hat{T} = \boldsymbol{w}^\top \hat{\boldsymbol{t}}$ που είναι αμερόληπτες ($\boldsymbol{w}^\top \mathbf{1} = 1$), η διακύμανση $\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}$ γίνεται ελάχιστη για*

$$\boldsymbol{w} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}, \qquad \operatorname{Var}(\hat{T}) = \frac{1}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}. \tag{F.1}$$

Αυτή είναι η εκτίμηση WLS $(\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{t}}$.

Προαιρετικό: απόδειξη

Έστω $\boldsymbol{w}_\star$ τα βάρη της (F.1) και $\boldsymbol{w}$ οποιαδήποτε άλλα αμερόληπτα βάρη. Γράφουμε $\boldsymbol{w} = \boldsymbol{w}_\star + \boldsymbol{d}$ με $\mathbf{1}^\top \boldsymbol{d} = 0$. Τότε $\boldsymbol{d}^\top \boldsymbol{\Sigma} \boldsymbol{w}_\star = \boldsymbol{d}^\top \mathbf{1}/(\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}) = 0$, οπότε $\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} = \boldsymbol{w}_\star^\top \boldsymbol{\Sigma} \boldsymbol{w}_\star + \boldsymbol{d}^\top \boldsymbol{\Sigma} \boldsymbol{d} \geq \boldsymbol{w}_\star^\top \boldsymbol{\Sigma} \boldsymbol{w}_\star$. Τέλος, $\boldsymbol{w}_\star^\top \boldsymbol{\Sigma} \boldsymbol{w}_\star = \mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}/(\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1})^2 = 1/(\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1})$.

Παράδειγμα F.2 (Δύο αισθητήρες και ένα αρνητικό βάρος)**.** Παίρνουμε δύο αισθητήρες με τυπικές αποκλίσεις $\sigma_1 = 0.1\,^\circ\mathrm{C}$ και $\sigma_2 = 0.2\,^\circ\mathrm{C}$ και συσχέτιση $\rho = 0.8$. Για δύο αισθητήρες, η (F.1) γίνεται

$$w_1 = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}, \qquad w_2 = 1 - w_1, \qquad \operatorname{Var}(\hat{T}) = \frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}.$$

Εδώ ο παρονομαστής είναι $0.01 + 0.04 - 0.032 = 0.018$, οπότε $w_1 = 0.024/0.018 = 4/3$ και $w_2 = -1/3$. Ο συνδυασμός έχει τυπική απόκλιση $\sqrt{0.0004 \cdot 0.36/0.018} = 0.089\,^\circ\mathrm{C}$, καλύτερη από τον καλύτερο αισθητήρα μόνο του. Ο θορυβώδης αισθητήρας παίρνει αρνητικό βάρος: το μεγαλύτερο μέρος του σφάλματός του είναι κοινό με τον αισθητήρα 1, οπότε αν αφαιρέσουμε το ένα τρίτο του, αφαιρούμε μέρος του σφάλματος του αισθητήρα 1. Γενικά $w_2 < 0$ ακριβώς όταν $\rho > \sigma_1/\sigma_2$.

Η συστοιχία συμπεριφέρεται με τον ίδιο τρόπο. Χωρίς τον αισθητήρα 7, τρεις αισθητήρες παίρνουν αρνητικά βάρη (οι 1, 8 και 10), και ο αισθητήρας 6 παίρνει σχεδόν το μισό βάρος. Οι αισθητήρες 1 και 10 βρίσκονται στα δύο άκρα της κλίσης του ανεμιστήρα, και η καλύτερη χρήση τους είναι ως αναφορές για την αφαίρεσή της.

Πρόταση F.3 (Πότε ο απλός μέσος όρος είναι βέλτιστος)**.** *Αν το $\mathbf{1}$ είναι ιδιοδιάνυσμα του $\boldsymbol{\Sigma}$, τότε $\boldsymbol{w} = \mathbf{1}/C$: ο απλός μέσος όρος είναι ο καλύτερος συνδυασμός.*

Απόδειξη. Αν $\boldsymbol{\Sigma}\mathbf{1} = \mu\mathbf{1}$, τότε $\boldsymbol{\Sigma}^{-1}\mathbf{1} = \mathbf{1}/\mu$, και η (F.1) δίνει $\boldsymbol{w} = \mathbf{1}/C$. □

Στο μοντέλο (C.2), το $\mathbf{1}$ είναι ιδιοδιάνυσμα όταν κάθε αισθητήρας έχει τον ίδιο ανεξάρτητο θόρυβο και η κλίση είναι ισορροπημένη ($\mathbf{1}^\top \boldsymbol{g} = 0$). Η πραγματική συστοιχία παραβιάζει και τις δύο συνθήκες: τα επίπεδα του δικού θορύβου των αισθητήρων διαφέρουν, και η αφαίρεση του αισθητήρα 7 χαλά την ισορροπία της κλίσης. Γι' αυτό τα βάρη είναι άνισα.

F.2 Γιατί ο μέσος όρος δεν νικά ένα κοινό σφάλμα

Πρόταση F.4 (Το κατώτατο όριο της κοινής συνιστώσας)**.** *Αν* $\boldsymbol{\Sigma} = \sigma_c^2 \mathbf{1}\mathbf{1}^\top + \boldsymbol{\Sigma}_r$ *με* $\boldsymbol{\Sigma}_r \succeq 0$*, τότε κάθε αμερόληπτος συνδυασμός έχει* $\mathrm{Var}(\hat{T}) \geq \sigma_c^2$*, όσους αισθητήρες κι αν χρησιμοποιήσουμε.*

Απόδειξη. Για $\boldsymbol{w}^\top \mathbf{1} = 1$, $\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} = \sigma_c^2 (\boldsymbol{w}^\top \mathbf{1})^2 + \boldsymbol{w}^\top \boldsymbol{\Sigma}_r \boldsymbol{w} = \sigma_c^2 + \boldsymbol{w}^\top \boldsymbol{\Sigma}_r \boldsymbol{w} \geq \sigma_c^2$. □

Εδώ $\sigma_c \approx 0.19\,^\circ\mathrm{C}$, και η (F.1) δίνει $\mathrm{sd}(\hat{T}) = 0.190\,^\circ\mathrm{C}$. Η εκτίμηση WLS βρίσκεται ακριβώς στο κατώτατο όριο. Ο απλός μέσος όρος είναι ελάχιστα χειρότερος, με τυπική απόκλιση $\sqrt{\mathbf{1}^\top \boldsymbol{\Sigma} \mathbf{1}}/C = 0.193\,^\circ\mathrm{C}$.

F.3 Αξιόπιστα και αναξιόπιστα περιθώρια σφάλματος

Σημαντική διάκριση

Το αφελές περιθώριο σφάλματος $\hat{\sigma}/\sqrt{C} = 0.070\,^\circ\mathrm{C}$ αντιμετωπίζει τους αισθητήρες σαν ανεξάρτητες γνώμες. Η πραγματική διακύμανση του απλού μέσου όρου είναι $\mathbf{1}^\top \boldsymbol{\Sigma} \mathbf{1}/C^2$, που περιέχει όλες τις συνδιακυμάνσεις.

Ένα περιθώριο σφάλματος μπορούμε να το ελέγξουμε όποτε η αλήθεια είναι γνωστή. Αν είναι αξιόπιστο, το διάστημα $\pm z \cdot \mathrm{sd}$ με $z = \Phi^{-1}((1+q)/2)$ καλύπτει την αλήθεια σε ποσοστό q των περιπτώσεων, για κάθε q. Απέναντι στο θερμόμετρο αναφοράς (Σχήμα 6, δεξιά), το αφελές περιθώριο δηλώνει 95% και καλύπτει 50%, και δηλώνει 68% και καλύπτει 26%. Το περιθώριο WLS καλύπτει 92% και 64%. Βρίσκεται λίγο κάτω από τη διαγώνιο επειδή ο $\hat{\boldsymbol{\Sigma}}$ είναι λίγο μικρότερος από όσο θα έπρεπε (Παρατήρηση C.3).

Παρατήρηση F.5 (Τι κάνει η βλάβη στη θερμοκρασία που αναφέρουμε)**.** Αν ο αισθητήρας 7 *δεν* αφαιρεθεί, η μετατόπισή του μετακινεί την εκτίμηση WLS των δέκα αισθητήρων κατά $w_7 \delta = 0.169 \times 0.3 \approx 0.05\,^\circ\mathrm{C}$. Αυτό είναι το ένα τέταρτο του περιθωρίου σφάλματος, πολύ λίγο για να το προσέξουμε στις ενδείξεις. Άρα η βλάβη είναι πολύ πιο ορατή στον έλεγχο ακραίων τιμών παρά στη θερμοκρασία που αναφέρουμε. Μια βλάβη που μεγαλώνει σιγά σιγά θα μεροληπτούσε τελικά κάθε ένδειξη, και γι' αυτό συμφέρει να την πιάσουμε νωρίς.

Παρατήρηση F.6 (Αβεβαιότητα της ίδιας της βαθμονόμησης)**.** Η τυπική απόκλιση που αναφέρουμε θεωρεί τα $\hat{a}_j, \hat{b}_j$ ακριβή. Σε πρώτη τάξη, ένα σφάλμα στους συντελεστές προσθέτει $-[(\hat{a}_j - a_j) + (\hat{b}_j - b_j)T]/b_j = -\boldsymbol{x}_T^\top (\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j)/b_j$ στο $\hat{T}_j$. Από την (A.7), η συνδιακύμανση αυτών των προσθέσεων ανάμεσα στους αισθητήρες είναι $h(T)\,\boldsymbol{\Sigma}$, όπου $h(T)$ η μόχλευση (A.4). Η πλήρης συνδιακύμανση είναι επομένως $(1 + h(T))\boldsymbol{\Sigma}$. Για τη συστοιχία, $h = 0.0025$ στους $25\,^\circ\mathrm{C}$ και 0.007 στους 10 ή $40\,^\circ\mathrm{C}$, οπότε αν τη συμπεριλαμβάναμε, το περιθώριο σφάλματος θα μεγάλωνε λιγότερο από 0.4%. Αυτό είναι αμελητέο εδώ, αλλά δεν θα ήταν μετά από μια σύντομη επαναβαθμονόμηση σαν αυτή της Ενότητας B. Αυτό το σφάλμα είναι επίσης *σταθερό* μόλις γίνει η βαθμονόμηση: δεν εξαλείφεται με τον μέσο όρο πολλών λήψεων.

G Χωρίς Θερμόμετρο Αναφοράς

Πεδίο εφαρμογής. Οι υπολογισμοί παρακάτω αφορούν συναγερμούς λήψεων. Το κύριο hands-on ελέγχει επιπλέον το συνολικό πλήθος συναγερμών και τη μοναδικότητα υποψήφιου πριν αποκλείσει αισθητήρα. Μεμονωμένοι συναγερμοί δεν αρκούν.

Ο ανιχνευτής βλαβών της Ενότητας E συγκρίνει τις ενδείξεις με το θερμόμετρο αναφοράς. Στην κανονική λειτουργία δεν υπάρχει θερμόμετρο αναφοράς: το T είναι άγνωστο, οπότε το $\boldsymbol{e}$ δεν είναι διαθέσιμο. Το φυσικό υποκατάστατο είναι το υπόλοιπο γύρω από την εκτίμηση WLS,

$$\boldsymbol{r} = \hat{\boldsymbol{t}} - \hat{T}\,\mathbf{1} = (\boldsymbol{I} - \mathbf{1}\boldsymbol{w}^\top)\hat{\boldsymbol{t}} = (\boldsymbol{I} - \mathbf{1}\boldsymbol{w}^\top)\boldsymbol{e}, \tag{G.1}$$

όπου το τελευταίο βήμα χρησιμοποιεί το $(\boldsymbol{I} - \mathbf{1}\boldsymbol{w}^\top)\mathbf{1} = \mathbf{0}$: το άγνωστο T απαλείφεται.

Πρόταση G.1 (Ένας βαθμός ελευθερίας για την εκτιμημένη θερμοκρασία)**.** *Αν* $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ *και το* $\boldsymbol{w}$ *δίνεται από την* (F.1)*, τότε* $\boldsymbol{r}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{r} \sim \chi^2_{C-1}$.

Προαιρετικό: απόδειξη

Λευκοποιούμε, θέτοντας $\boldsymbol{q} = \boldsymbol{L}^{-1}\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ και $\boldsymbol{g} = \boldsymbol{L}^{-1}\mathbf{1}$. Αφού

$$\boldsymbol{w}^\top \boldsymbol{e} = \frac{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}} = \frac{\boldsymbol{g}^\top \boldsymbol{q}}{\boldsymbol{g}^\top \boldsymbol{g}},$$

έχουμε

$$\boldsymbol{L}^{-1}\boldsymbol{r} = \boldsymbol{q} - \boldsymbol{g}\,\frac{\boldsymbol{g}^\top \boldsymbol{q}}{\boldsymbol{g}^\top \boldsymbol{g}} = (\boldsymbol{I} - \boldsymbol{\Pi}_{\boldsymbol{g}})\boldsymbol{q},$$

όπου $\boldsymbol{\Pi}_{\boldsymbol{g}}$ η ορθογώνια προβολή στο $\boldsymbol{g}$ (Διάλεξη 1). Άρα $\boldsymbol{r}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{r} = \|(\boldsymbol{I} - \boldsymbol{\Pi}_{\boldsymbol{g}})\boldsymbol{q}\|^2$. Επιλέγουμε μια ορθοκανονική βάση της οποίας το πρώτο διάνυσμα είναι το $\boldsymbol{g}/\|\boldsymbol{g}\|$. Σε αυτήν, το $\boldsymbol{q}$ εξακολουθεί να έχει ανεξάρτητες συντεταγμένες $\mathcal{N}(0,1)$, επειδή μια περιστροφή της $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$ είναι πάλι $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Η προβολή αφαιρεί ακριβώς την πρώτη συντεταγμένη. Αυτό που μένει είναι άθροισμα των τετραγώνων $C-1$ ανεξάρτητων μεταβλητών $\mathcal{N}(0,1)$.

Αυτή είναι η λευκοποιημένη εκδοχή του «κάθε προσαρμοσμένη παράμετρος απορροφά περίπου ένα σ^2» της Διάλεξης 5: η προσαρμογή του T καταναλώνει μία από τις C διαστάσεις. Ο έλεγχος στην κανονική λειτουργία σημαίνει μια λήψη όταν $\boldsymbol{r}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{r} > \chi^2_{C-1}(0.99) = 21.67$, και στη συνέχεια η αφαίρεση ενός αισθητήρα κάθε φορά συγκρίνει με το $\chi^2_{C-2}(0.99) = 20.09$. Στο πείραμα επικύρωσης αυτός ο έλεγχος, που δεν βλέπει ποτέ το T, σημαίνει το 1.5% των υγιών λήψεων και το 70% των λήψεων με βλάβη, τους ίδιους ρυθμούς με τον έλεγχο που χρησιμοποιεί την αναφορά.