

ΚΕΦΑΛΑΙΟ 3: ΠΡΟΒΛΕΨΗ

Ως τώρα είχαμε θεωρήσει γνωστή τη ζήτηση ή τις πιθανότητες ζήτησης. Πώς προκύπτουν αυτές οι πληροφορίες; Προκύπτουν από

Εκτιμήσεις: πληροφορίες για ένα **μη μετρήσιμο** μέγεθος (π.χ. η αυριανή ζήτηση) που αντλούμε από **μετρήσεις από ένα άλλο συναφές μέγεθος**.

"Δείξε μου τους φίλους σου να σου πω ποιος είσαι".

Ο όρος **εκτίμηση** (estimation) αναφέρεται σε όλα τα ακόλουθα:

- εξομάλυνση (smoothing): τι έγινε στο παρελθόν
- εκτίμηση τρέχουσας κατάστασης (state estimation, filtering)
- πρόβλεψη (prediction, forecasting): τι θα συμβεί στο μέλλον.

Όταν το μετρήσιμο μέγεθος είναι στοχαστική διαδικασία, X_t , οι **μετρήσεις** σχηματίζουν **χρονοσειρά** (time series) ήτοι μία ακολουθία τιμών $x_1, x_2, \dots, x_n$ που είχε η X_t στους χρόνους $t=1, 2, \dots, n$ ή στους χρόνους $t_1 < t_2 < \dots < t_n$.

Συμβολισμοί: **Κεφαλαία** για τη στοχαστική διαδικασία (π.χ. ημερήσια ζήτηση X_t) και **πεζά** για τις πιθανές τιμές της, π.χ. $P(X_t > x) = 0.85$.

$E(X_t)$, $\text{Var}(X_t)$, $\text{Cov}(X_t, X_{t+k})$: μέση τιμή, διασπορά, συμμεταβλητότητα της X_t .

Εξαίρεση: πεζά a_t ή $\varepsilon(t)$ ο λευκός θόρυβος, $E(a_t) = \text{Cov}(a_t, a_{t+k}) = 0$ για $k \neq 0$.

$\bar{x} = (x_1 + x_2 + \dots + x_n)/n$: μέσος όρος χρονοσειράς, t_n : πιο πρόσφατος χρόνος

$\hat{X}_{n+k}$ ή $\hat{x}_{n+k}$: πρόβλεψη της ζήτησης για τον χρόνο t_{n+k} στο μέλλον ($k \geq 1$).

Το πλέον γενικό μοντέλο: $X_t = \alpha(t) + \varepsilon(t)$, όπου $\alpha(t)$ = άγνωστη, μη γραμμική συνάρτηση

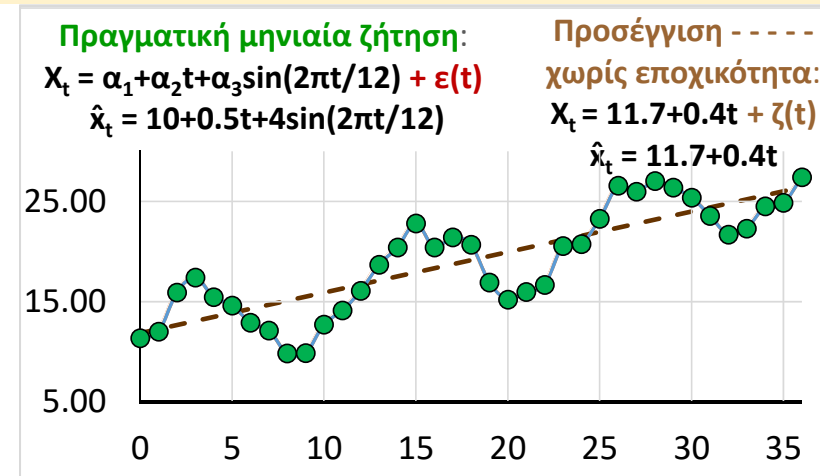
Γενικό μοντέλο γραμμικής παλινδρόμησης

Δεδομένα

- Μία χρονοσειρά τιμών $x_1, x_2, \dots, x_n$ της X_t στους χρόνους $t_1 < t_2 < \dots < t_n$.
- Ένα γραμμικό μοντέλο εκτίμησης της X_t (πραγματικό ή **προσεγγιστικό**):

$$X_t = \alpha_1 f_1(t) + \dots + \alpha_m f_m(t) + \varepsilon(t) \quad \text{ή} \quad \text{εκτίμηση: } \hat{x}_t = \alpha_1 f_1(t) + \dots + \alpha_m f_m(t)$$
- όπου $f_1(t), \dots, f_m(t)$ **γνωστές** συναρτήσεις (π.χ. σταθερή $t^0=1$, t^5 , e^{-2t} , ...)
- $\alpha_1, \dots, \alpha_m$ **άγνωστοι συντελεστές** που χρειαζόμαστε για να έχουμε την $\hat{x}_t$
- $\varepsilon(t) = X_t - \hat{x}_t$ το **σφάλμα του μοντέλου**, μία μη προβλέψιμη ποσότητα.

t	$\varepsilon(t)$	Ζήτηση τον μήνα t: $X_t = 10 + 0.5t + 4\sin(2\pi t/12) + \varepsilon(t)$
0	1.35	11.35
1	-0.49	12.01
2	1.46	15.92
3	1.92	17.42
4	-0.02	15.44
5	0.11	14.61
...	...	...



Το $\varepsilon(t)$ λαμβάνεται ως **λευκός θόρυβος**. Γιατί; Αναζητούμε την **καλύτερη συνάρτηση** $\hat{x}_t$. Αφού η $\hat{x}_t$ δεν είναι τυχαία, $E(\hat{x}_t) = \hat{x}_t$ και $E[\varepsilon(t)] = (X_t - \hat{x}_t) = E(X_t) - \hat{x}_t$. Αν καταφέρουμε $\hat{x}_t = E(X_t)$ τότε $E[\varepsilon(t)] = 0$ και, όπως θα δούμε, τα $\varepsilon(t)$ έχουν μικρά μεγέθη. Επίσης, οι $\varepsilon(t)$ είναι μεταξύ τους ασυσχέτιστες γιατί αλλιώς η $\hat{x}_t$ θα επιδεχόταν βελτίωσης. Άρα **Cov=0**. Άρα **λευκός θόρυβος**.

Γραμμική παλινδρόμηση $x_t = \alpha_1 f_1(t) + \dots + \alpha_m f_m(t) + \varepsilon(t)$

Ζητούμε $\alpha_1, \dots, \alpha_m$: $\min q_1 \varepsilon^2(t_1) + \dots + q_n \varepsilon^2(t_n)$. $q_i \geq 0$ είναι βάρη σημαντικότητας των $\varepsilon(t_i)$. Ίδια βάρη: $q_i = 1$.

$$\varepsilon^T = [\varepsilon(t_1) \ \varepsilon(t_2) \ \dots \ \varepsilon(t_n)] \quad Q = Q^T = \begin{bmatrix} q_1 & 0 & \dots & 0 \\ 0 & q_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q_n \end{bmatrix} \Rightarrow \varepsilon^T Q \varepsilon = [\varepsilon(t_1)q_1 \ \varepsilon(t_2)q_2 \ \dots \ \varepsilon(t_n)q_n]$$

$$\varepsilon = \begin{bmatrix} \varepsilon(t_1) \\ \varepsilon(t_2) \\ \vdots \\ \varepsilon(t_n) \end{bmatrix} \Rightarrow \varepsilon^T Q \varepsilon = q_1 \varepsilon^2(t_1) + \dots + q_n \varepsilon^2(t_n)$$

$$\text{Επίσης: } X = \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{\text{χρονοσειρά}} \quad \alpha = \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix}}_{\text{συντελεστές}} \quad A = \underbrace{\begin{bmatrix} f_1(t_1) & f_1(t_2) & \dots & f_1(t_n) \\ f_2(t_1) & f_2(t_2) & \dots & f_2(t_n) \\ \vdots & \vdots & \ddots & \vdots \\ f_m(t_1) & f_m(t_2) & \dots & f_m(t_n) \end{bmatrix}}_{\text{Ο Α σχηματίζεται αντικαθιστώντας } t=t_i \text{ στις } f_j(t)}$$

$$\Rightarrow \varepsilon = X - A^T \alpha$$

$$\varepsilon^T = X^T - \alpha^T A$$

$$\Rightarrow \min \varepsilon^T Q \varepsilon = (X^T - \alpha^T A) Q (X - A^T \alpha) = X^T Q X - X^T Q A^T \alpha - \alpha^T A Q X + \alpha^T A Q A^T \alpha. \text{ Ακρότατο (*):}$$

$$\frac{\partial \varepsilon^T Q \varepsilon}{\partial \alpha} = 0 = \underbrace{\frac{\partial (X^T Q X)}{\partial \alpha}}_0 - \underbrace{\frac{\partial (X^T Q A^T) \alpha}{\partial \alpha}}_{(X^T Q A^T)} - \underbrace{\frac{\partial \alpha^T (A Q X)}{\partial \alpha}}_{(X^T Q A^T)} + \underbrace{\frac{\partial \alpha^T (A Q A^T) \alpha}{\partial \alpha}}_{2 \alpha^T (A Q A^T)}$$

$$\frac{\partial^2 \varepsilon^T Q \varepsilon}{\partial \alpha^2} = 2 A Q A^T > 0$$

Θετικά ορισμένος $\Rightarrow \varepsilon^T Q \varepsilon$ κυρτή
ολικό ελάχιστο

$$2 \alpha^T (A Q A^T) - 2 X^T Q A^T = 0 \Rightarrow \alpha^T = X^T Q A^T (A Q A^T)^{-1} \text{ ή } \alpha = (A Q A^T)^{-1} A Q X$$

(*) Ορίζεται στο Παράρτημα $\text{grad}(f) \triangleq \partial f(\alpha) / \partial \alpha$ συνάρτησης f ως διάνυσμα γραμμής. Για πίνακα B , προκύπτει $\partial(B\alpha) / \partial \alpha = B$. Αν ορίζαμε $\partial f(\alpha) / \partial \alpha$ ως στήλη, τότε $\dots = B^T$. Έχουμε ίδιο ακρότατο. 4

Παράδειγμα

Μοντέλο μηνιαίας ζήτησης: $\hat{z}_t = at^2 + bt + \varepsilon(t)$. Δεδομένα: $z_1=5, z_2=20, z_3=43, z_4=71$. Ζητούνται τα a, b που ελαχιστοποιούν το άθροισμα τετραγώνων των σφαλμάτων και η πρόβλεψη $\hat{z}_7$.

Το βέλτιστο $\alpha^T = [a \ b]$ είναι $\alpha = (AQA^T)^{-1}AQX$, όπου

$$A = \begin{bmatrix} 1^2 & 2^2 & 3^2 & 4^2 \\ 1 & 2 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 9 & 16 \\ 1 & 2 & 3 & 4 \end{bmatrix} \quad \text{ίσα βάρη: } Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad X = \begin{bmatrix} 5 \\ 20 \\ 43 \\ 71 \end{bmatrix}$$

Άρα $\alpha = (AA^T)^{-1}AX$. Πράξεις με πίνακες:

$$AA^T = \begin{bmatrix} 1 & 4 & 9 & 16 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 4 & 2 \\ 9 & 3 \\ 16 & 4 \end{bmatrix} = \begin{bmatrix} 354 & 100 \\ 100 & 30 \end{bmatrix} \quad AX = \begin{bmatrix} 1 & 4 & 9 & 16 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 5 \\ 20 \\ 43 \\ 71 \end{bmatrix} = \begin{bmatrix} 1608 \\ 458 \end{bmatrix}$$

$$(AA^T)^{-1} = \frac{1}{\underbrace{354 \cdot 30 - 100 \cdot 100}_{620}} \begin{bmatrix} 30 & -100 \\ -100 & 354 \end{bmatrix} \quad \begin{bmatrix} a \\ b \end{bmatrix} = (AA^T)^{-1}AX = \frac{1}{620} \begin{bmatrix} 2440 \\ 1332 \end{bmatrix} \approx \begin{bmatrix} 3.94 \\ 2.15 \end{bmatrix}$$

$$\hat{z}_7 = 3.94(7^2) + 2.15(7) = 208.11$$

Στο διαγώνισμα, να ξέρετε να αντιστρέψετε πίνακες $n \times n$, και ειδικά 2×2 και 3×3 .

Κινούμενος μέσος όρος

Θεωρούμε το μοντέλο $\hat{x}_t = \alpha$ για $x_1 \dots x_n$. Το άθροισμα τετραγώνων των σφαλμάτων είναι : $\epsilon^2 = \sum (x_i - \alpha)^2$. Τοπικό ακρότατο έχουμε όταν:

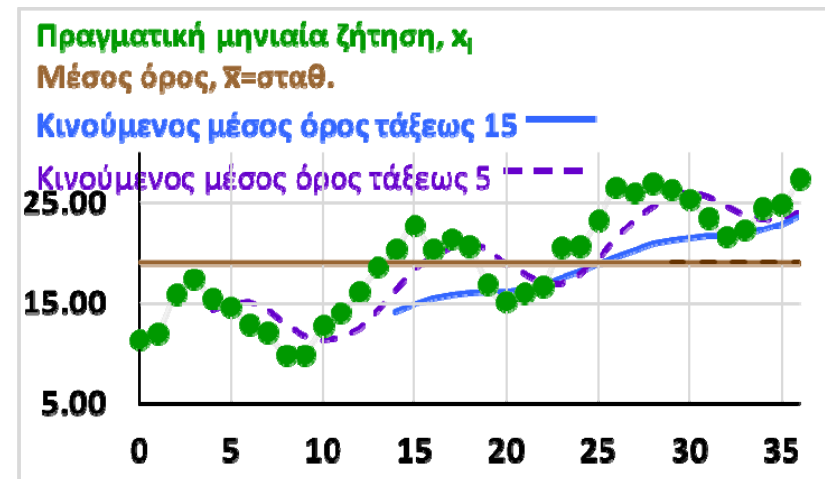
$$0 = \partial \epsilon^2 / \partial \alpha = -2(x_1 - \alpha) \dots -2(x_n - \alpha) \Rightarrow \alpha = \bar{x} = (\sum x_i) / n \quad \partial^2 \epsilon^2 / \partial \alpha^2 = 2n > 0 \Rightarrow \epsilon^2 \text{ κυρτή} \\ \Rightarrow \bar{x} \text{ είναι ολικό ελάχιστο}$$

Από όλες τις σταθερές εκτιμήσεις, ο μέσος όρος $\bar{x}$ (δειγματική μέση τιμή) ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα ϵ^2/n .

Όμως αυτή η εκτίμηση είναι απλοϊκή, βασίζεται το ίδιο σε παλιές και πρόσφατες πληροφορίες και δεν παρακολουθεί την πρόσφατη τάση.

Γι' αυτό συχνά εφαρμόζεται μία παραλλαγή: ο **κινούμενος μέσος όρος** τάξεως k , αντί για όλες τις μετρήσεις χρησιμοποιεί τις k πιο πρόσφατες:

$$\alpha = (x_n + \dots + x_{n-k+1}) / k$$



Ο κινούμενος μέσος όρος εφαρμόζεται στην πρόβλεψη ζήτησης όταν δεν έχουμε άλλη πληροφορία. Για παράδειγμα, όταν δεν γνωρίζουμε αν η ζήτηση εμφανίζει εποχικές κινήσεις ή τάση. Οι προβλέψεις φαίνονται να έχουν μία χρονική **καθυστέρηση σε σχέση με τις πραγματικές τιμές x_t** .

Εφαρμόζεται και στην εκτίμηση της αξίας μετοχών στο χρηματιστήριο. Αποτελεί βάση της μεθοδολογίας της Τεχνικής Ανάλυσης.

Απλή εκθετική εξομάλυνση (simple exponential smoothing)

Ξεκινάμε πάλι με την εκτίμηση $\hat{x}_t = \alpha$ για $x_1 \dots x_n$. Επιλέγουμε έναν αριθμό $0 \leq \beta \leq 1$ και κριτήριο $\varepsilon^2 = \beta^0(x_n - \alpha)^2 + \dots + \beta^{n-1}(x_1 - \alpha)^2$. Ολικό ελάχιστο όταν:

$$0 = \partial \varepsilon^2 / \partial \alpha = -2(x_n - \alpha)\beta^0 \dots - 2(x_1 - \alpha)\beta^{n-1} \Rightarrow \alpha = \frac{1x_n + \beta x_{n-1} + \dots + \beta^{n-1}x_1}{1 + \beta + \dots + \beta^{n-1}} = \frac{A_n}{B_n} = Y_n$$

Αντί για αποκοπή των παλαιών δεδομένων με το σταθερό παράθυρο k του κινούμενου μέσου όρου, εδώ εφαρμόζεται μία πιο ομαλή **εκθετική απόσβεση** (μικραίνουν τα βάρη β^{n-i} όταν το i έχει μικρές-παλαιές τιμές).

Είδαμε ότι $\alpha = Y_n = A_n / B_n$ όπου $A_n = x_n + \beta(x_{n-1} + \beta x_{n-2} + \dots) = x_n + \beta A_{n-1}$
και $B_n = 1 + \beta(1 + \beta + \dots) = 1 + \beta B_{n-1}$ (αρχικές τιμές: $A_0 = B_0 = 0$)

Αν $\beta = 0 \Rightarrow Y_n = x_n$ όχι παρελθόν. Αν $\beta = 1 \Rightarrow Y_n = \bar{x}_n$. Αν $0 < \beta < 1 \Rightarrow B_n = (1 - \beta^n) / (1 - \beta)$.
Όταν $\beta < 1$ και $n \rightarrow \infty$, τότε έχουμε $\beta^n \rightarrow 0$ και τα ακόλουθα:

$$Y_n = (1 - \beta)A_n = (1 - \beta)x_n + \beta A_{n-1}(1 - \beta)$$

Όμως, $Y_{n-1} = A_{n-1}(1 - \beta)$. Με αντικατάσταση $\Rightarrow$ **απλή εκθετική εξομάλυνση:**
 $Y_n = (1 - \beta)x_n + \beta Y_{n-1}$ (αρχικές τιμές: $Y_1 = x_1$ και $n = 2, 3, \dots$ ή $Y_0 = \bar{x}$ και $n = 1, 2, \dots$)

Η Y_n που βρήκαμε είναι

- 1) εκτίμηση της στιγμιαίας μέσης τιμής $E(X_n)$, για $t = n$
- 2) εκτίμηση (φίλτρο) κατάστασης του συστήματος τη στιγμή $t = n$.
- 3) εκτίμηση της μέσης τιμής για κάθε t , αν η X_t είναι αμετάβλητη
- 4) πρόβλεψη για τις $X_{n+1}, X_{n+2}, \dots$ με βάση πληροφορίες **ως τη** στιγμή $t = n$

Παράδειγμα (διαφορετικό από των σημειώσεων)

Δεδομένα:

Εβδομάδα	1	2	3	4	5	6	7	8	9	10
Ζήτηση	15	18	10	12	20	17	22	16	14	20

Για το $Y_n = (1-\beta)x_n + \beta Y_{n-1}$ ζητούμε το β που $\min \varepsilon^2$ χωρίς βάρη. Όμως $\min \varepsilon^2$ επιτυγχάνουμε με το $Y_n = \bar{X} = (x_1 + \dots + x_n)/n$ και τότε $\beta = 1$. **Άρα τελειώσαμε;**

Όχι: Το βέλτιστο $\bar{x}$ προκύπτει με **η δεδομένα**. Όμως στα πρώτα στάδια της πρόβλεψης **έχουμε λιγότερα**. Αν είχαμε x_1, x_2, x_3 , ο μέσος $(x_1 + x_2 + x_3)/3$ **διαφέρει από τον $\bar{x}$** . Οι πρώτοι $n-1$ μέσοι όροι **δεν είναι βέλτιστοι!**

Ζητούμε το β που δίνει $\min$ σφάλμα και με λίγα και με πολλά δεδομένα.

Αλγόριθμος αντιπαραβολής (cross validation)--εκτός ύλης

Παραλείπουμε τους δείκτες στα Y . Ξεκινάμε με $\varepsilon_{opt} > (\infty)$.

Για $\beta = 0.00, 0.05, 0.10, \dots, 0.90, 0.95, 1.00$:

Ξεκινάμε με $Y = x_1$ και $\varepsilon^2 = 0$

Για $t=2, \dots, n$: $\varepsilon^2 = \varepsilon^2 + (x_t - Y)^2$ και νέο $Y = (1-\beta)x_t + \beta Y$ (επανάληψη με νέο t)

Αν $\varepsilon^2 < \varepsilon_{opt}$, τότε νέο $\varepsilon_{opt} = \varepsilon^2$ και $\beta_{opt} = \beta$.

Όταν έχουν δοκιμασθεί όλα τα β , επιλέγεται το β_{opt} .

Π.χ. για $\beta=0.8$

Εβδομάδα	1	2	3	4	5	6	7	8	9	10
Ζήτηση	15	18	10	12	20	17	22	16	14	20
$Y_t = (1-0.8)x_{t-1} + 0.8Y_{t-1}$	15	15.60	14.48	13.98	15.19	15.55	16.84	16.67	16.14	16.91
$\varepsilon^2 = 150.36$	σφάλμα: 3.00	-5.60	-2.48	6.02	1.81	6.45	-0.84	-2.67	3.86	

Με το πρόσθετο Excel Solver, βρήκα $\beta_{opt} = 0.95$ και $\varepsilon_{opt} = 146.49$.