ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ
ΣΗΜΕΙΩΣΕΙΣ

Καθηγ. Ι. Σίσκος

ΜΑΘΗΜΑ 1ο

ΑΝΑΛΥΣΗ ΧΡΟΝΟΣΕΙΡΩΝ & ΠΡΟΒΛΕΨΗ

TIME SERIES ANALYSIS & FORECASTING

SERIES CHRONOLOGIQUES & PREVISION A COURT TERME

- Πρόβλημα πρόβλεψης
- Παράδειγμα
- Συνιστώσες μιας χρονολογικής σειράς
- Το γραμμικό μοντέλο
- Μεθοδολογία
- Μέθοδος των ελαχίστων τετραγώνων (εκτίμηση της τάσης)

Το Πρόβλημα της πρόβλεψης

- Πρόβλεψη ενός αριθμητικού μεγέθους στο χρόνο (ανεργία, πληθυσμιακή αύξηση / μείωση, ζήτηση προϊόντος, κλπ.) από την εξέλιξή του στο παρελθόν, σε διάφορες χρονικές περιόδους (χρονοσειρά).
- X_t : Τυχαία μεταβλητή που μοντελοποιεί την τιμή της σειράς κατά την περίοδο t (ζήτηση προϊόντος κατά τη χρονική στιγμή t)
- Διαθέτουμε ένα ιστορικό – σειρά του μεγέθους που μελετάται σε t συνεχόμενες περιόδους : $x_1, x_2, \ldots, x_{t-1}, x_t$.
- Πρόβλημα : Να προβλεφθεί η τιμή της σειράς θ περιόδους μετά την t-οστή περίοδο, δηλαδή να βρεθεί η τιμή $x_{t+\theta}$, $\theta = 1, 2, \ldots$ θ περιόδους μετά την t-οστή.

- $\hat{x}_{t+\theta}$: τιμή προβλεφθείσα κατά την περίοδο $t+\theta$.

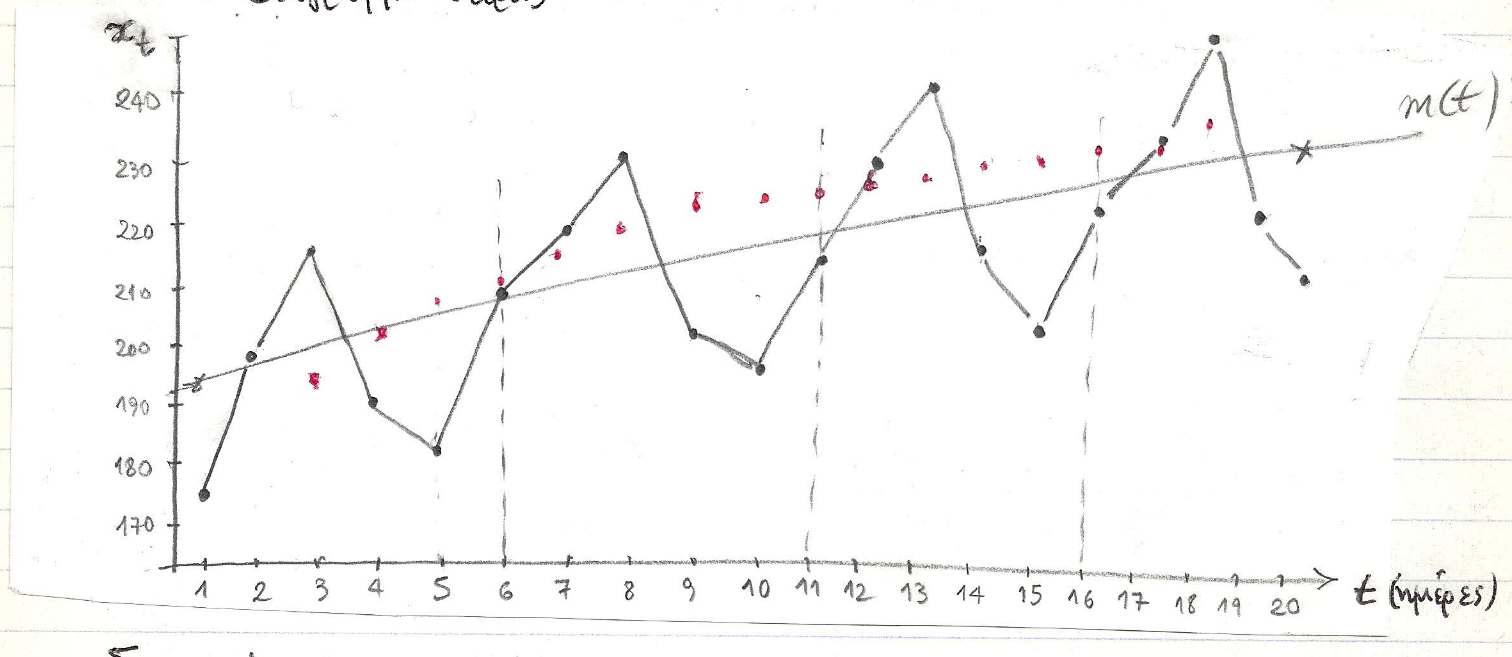
- $(\hat{x}_{t+\theta} - \varepsilon_1 , \hat{x}_{t+\theta} + \varepsilon_2)$: διάστημα εμπιστοσύνης μέσα στο οποία η εκτίμηση $\hat{x}_{t+\theta}$ έχει μια συγκεκριμένη πιθανότητα (π.χ. 0,95) να βρεθεί.

Παράδειγμα : Χρονοσειρά της εβδομαδιαίας ζήτησης προϊόντος

t	X_t	tX_t	t	X_t	tX_t	t	X_t	tX_t	t	X_t	tX_t
1	176	176	6	210	1260	11	216	2376	16	223	3568
2	197	394	7	220	1540	12	229	2748	17	234	3978
3	215	645	8	231	1848	13	240	3120	18	248	4464
4	190	760	9	202	1818	14	210	2940	19	220	4180
5	182	910	10	196	1960	15	200	3000	20	211	4220

$\sum_t X_t = 4.250$ $\sum_t tX_t = 45.905$

- Διάγραμμα σειράς



Συνιστώσες της Χρονοσειράς

Αποσύνθεση της χρονοσειράς σε ένα άθροισμα ή γινόμενο συνιστωσών που ερμηνεύουν τα διάφορα μέρη της σειράς :

$$X_t = m(t) + s(t) + E_t \quad , t=1,2,\ldots$$

1. <u>Η τάση</u>: $m(t)$

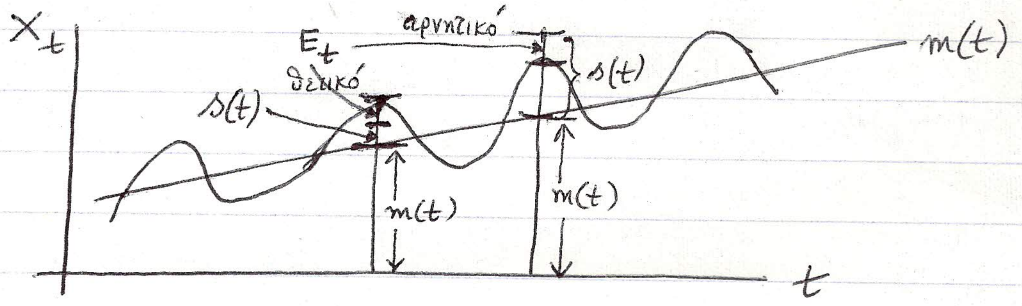
Παριστάνει την εξέλιξη (αύξουσα ή φθίνουσα) της σειράς μεσοπρόθεσμα.
Πρδμα: γραμμική εξέλιξη ; $m(t) = at + b$.

2. <u>Οι εποχικές μεταβολές</u>: $s(t)$

Πρόκειται για διακυμάνσεις περιοδικού χαρακτήρα, οι οποίες επηρρεάζουν την τάση κάθε T περιόδους η καθεμία, δηλαδή:
$s(t) = s(t+T)$, π.χ. $T = 12$ μήνες.

3. <u>Τα τυχαία σφάλματα</u>: E_t, $t=1,2,\ldots$

Είναι τα τυχαία μέρη της σειράς X_t που δεν μπορούν να εξηγηθούν ούτε από την τάση, ούτε από τις εποχικότητες.

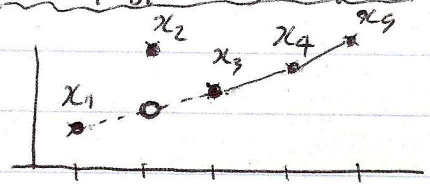


<u>Διόρθωση μιας σειράς</u>

- <u>Αριθμός εργάσιμων ημερών</u>

Όταν ο αριθμός των εργάσιμων ημερών του μήνα είναι 22, θέτουμε: $\frac{20}{22} x_t$ στη θέση του x_t για τη ζήτηση του μήνα.

- <u>Εξωπραγματικές τιμές</u>



x_2: εξωπραγμ. τιμή
Θέτουμε: $x_2 = \frac{x_1 + x_3}{2}$

<u>Το γραμμικό μοντέλο</u>

Μοντέλα
- → μη αιτιοκρατικά (χωρίς επεξηγηματικές μεταβλητές)
  - → στάσιμα σφάλματα E_t
  - → μη στάσιμα σφάλματα E_t
- → αιτιοκρατικά (με επεξηγηματικές μεταβλητές): $X_t = f(\Psi_1, \Psi_2, \ldots, \Psi_k)$

- E_t στάσιμα: Ο νόμος που τα διέπει (μαθημ. ελπίδα, διασπορά) παραμένει αναλλοίωτος μέσα στο χρόνο.

- Γραμμικό μοντέλο: $X_t = m(t) + s(t) + E_t$
- Πολλαπλασιαστικό μοντέλο: $X_t = m(t) \cdot s(t) \cdot E_t$ το οποίο ανάγεται σε γραμμικό μετά από λογαρίθμηση: $\mathrm{Log}\, X_t = \mathrm{Log}\, m(t) + \mathrm{Log}\, s(t) + \mathrm{Log}\, E_t$.

- $m(t) = at + b$
- $s(t)$: περιοδική εποχικότητα περιόδου T: $s(t) = s(t+T) \quad \forall t$.
- E_t: στάσιμη τυχαία μεταβλητή, που επαληθεύει τις σχέσεις:

$$E(E_t) = E(E_{t+h}) = 0 \quad , \quad \forall t, h \neq 0$$
$$V(E_t) = V(E_{t+h}) = \sigma^2 \quad , \quad \forall t, h \neq 0$$
$$Cov(E_t, E_{t+h}) = 0$$ (ιδιότητα ανεξαρτησίας των σφαλμάτων)

Μεθοδολογία της πρόβλεψης

- Εκτίμηση της τάσης $m(t) = at + b$ (δηλαδή της a και b).
- Εκτίμηση των εποχικών συντελεστών $s(t)$. Ο αριθμός τους ισούται με T.
- Επέκταση του μοντέλου στις επόμενες περιόδους.

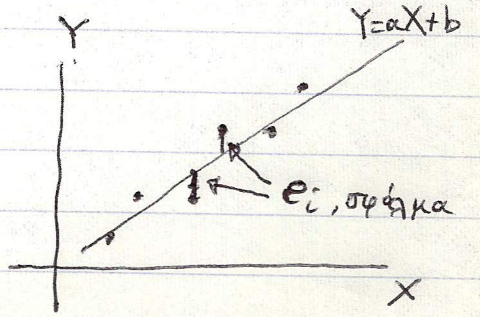
Μέθοδος των ελαχίστων τετραγώνων

$X_t = at + b + [s(t) + E_t]$, οι εποχικότητες συμπεριλαμβάνονται στα σφάλματα. Θα εκτιμηθούν τα a και b από το σύνολο των παρατηρήσεων (t, X_t) της σειράς.

Γενική μέθοδος

Έστω το γενικό μοντέλο:

$Y = aX + b + E$



Οι συντελεστές a και b θα εκτιμηθούν από το σύνολο των n παρατηρήσεων (x_i, y_i), έτσι ώστε το άθροισμα των τετραγώνων των σφαλμάτων να είναι ελάχιστο.

$$\min_{\alpha,b} F(\alpha,b) = \sum_i e_i^2 = \sum_i (y_i - \alpha x_i - b)^2$$

$$\frac{\partial F}{\partial \alpha} = 0 \quad \text{κ} \quad \frac{\partial F}{\partial b} = 0 \quad , \text{απ' όπου :}$$

$$-2 \sum_i (y_i - \alpha x_i - b) x_i = 0 \qquad (1)$$
$$-2 \sum_i (y_i - \alpha x_i - b) = 0 \qquad (2)$$

Η (2) γράφεται : $\bar{y} = \alpha \bar{x} + b$, $\bar{y} = \frac{1}{n} \sum_i y_i$, $\bar{x} = \frac{1}{n} \sum_i x_i$

$$(1) \Rightarrow \sum_i x_i y_i - \alpha \sum_i x_i^2 - (\bar{y} - \alpha \bar{x}) \sum_i x_i = 0 \Rightarrow$$

$$\boxed{\alpha = \frac{\frac{1}{n}\sum_i x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_i x_i^2 - \bar{x}^2} = \frac{Cov(X,Y)}{Var(X)} \quad , \quad b = \bar{y} - \alpha \bar{x}}$$

<u>Εφαρμογή στις χρονοσειρές</u> $(x_1, x_2, \ldots, x_h)$ με h παρατηρήσεις.

Θέτουμε : t αντί x_i

x_t αντί y_i

Τότε ισχύουν :

$$\sum_{t=1}^{h} t = \frac{h(h+1)}{2}$$

$$\sum_{t=1}^{h} t^2 = \frac{h(h+1)(2h-1)}{6}$$

Απ' όπου προκύπτουν οι νέοι τύποι :

$$\boxed{\begin{aligned} \alpha &= \frac{6}{h(h-1)} \left[\frac{2}{h+1} \sum_{t=1}^{h} t x_t - \sum_{t=1}^{h} x_t \right) \\ b &= \bar{x} - \alpha \frac{h+1}{2} \end{aligned}}$$

Εφαρμογή στο Παράδειγμα της Ζήτησης $(h=20)$

$$a = \frac{6}{20 \times 19}\left(\frac{2}{21} \times 45.905 - 4.250\right)$$

$$= 1{,}925$$

$$b = \frac{4.250}{20} - 1{,}925 \times \frac{21}{2}$$

$$= 192{,}288$$

Τελικά:

$$\boxed{m(t) = 1{,}925\,t + 192{,}288}$$

Στο γράφημα της σειράς, έχουν υπολογιστεί:

$m(1) = 194{,}213$

$m(20) = 230{,}788$

ΑΝΑΛΥΣΗ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ & ΠΡΟΒΛΕΨΗ

$\boxed{\text{ΜΑΘΗΜΑ } 2^{o}}$

- Κριτική της μεθόδου των ελαχίστων τετραγώνων
- Μέθοδος των κινητών μέσων
- Εκτίμηση της τάσης
- Παράδειγμα

Μέθοδος των ελαχίστων τετραγώνων (εκτίμηση της τάσης)

Γενικός όρος μιας χρονοσειράς:

$$X_t = at + b + s(t) + E_t$$

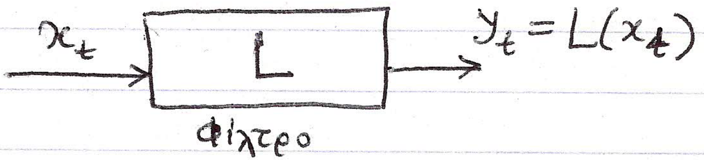
$$\boxed{\begin{aligned} \hat{a} &= \frac{6}{n(n-1)}\left(\frac{2}{n+1}\sum_{t=1}^{n} t x_t - \sum_{t=1}^{n} x_t\right) \\ \hat{b} &= \bar{x} - \hat{a}\,\frac{n+1}{2} \end{aligned}}$$

Κριτική

- Αναγκαιότητα επικαιροποίησης της τάσης της σειράς, π.χ. επανεκτίμηση του $\hat{a}$ κάθε T περιόδους (T: περιοδικότητα σειράς)
- Η μέθοδος δεν είναι παραδεκτή όταν η τάση αλλάζει απότομα κλίση (π.χ. περίπτωση νέων αγορών). Εφαρμόζεται κυρίως σε ομαλές σειρές.

Φιλτράρισμα της σειράς μέσω της μεθόδου των κινητών μέσων

Φίλτρο "κινητός μέσος" (μετασχηματισμός της σειράς)



- 1η Περίπτωση : Εύρος ομαδοποίησης περιττό $p = 2m+1$

$$y_t = \frac{1}{2m+1} \sum_{i=-m}^{+m} x_{t+i}$$

Παράδειγμα : $p = 5 (= 2m+1) \Rightarrow m = 2$

Ο μετασχηματισμός αρχίζει από τον 3° όρο :

$$y_3 = \frac{1}{5}(x_1 + x_2 + x_3 + x_4 + x_5)$$

$$= \frac{1}{5}(176 + 197 + 215 + 190 + 182)$$

$$= 192,0$$

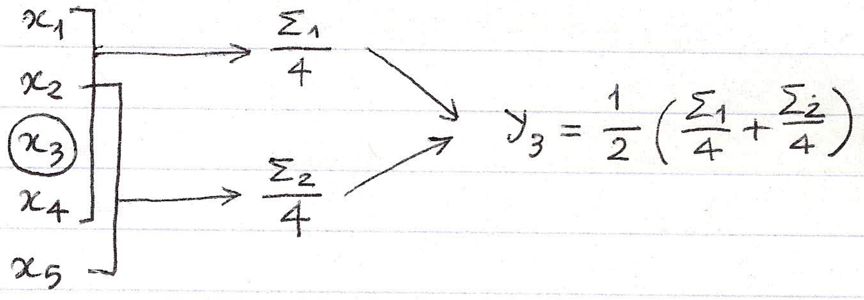
t	X_t	Y_t	tY_t	t^2	$m(t) = (1,97t + 193)$	$X_t - m(t)$
1	176	—	—	—	194,97	
2	197	—	—	—	196,94	
3	215	192,0	576,0	9	198,91	
4	190	198,8	795,2	16	200,88	
5	182	203,4	1.017,0	25	202,85	
6	210	206,6	1.239,6	36	204,82	
7	220	209,0	1.463,0	49	206,79	
8	231	211,8	1.694,4	64	208,76	
9	202	213,0	1.917,0	81	210,73	
10	196	214,8	2.148,0	100	212,70	
11	216	216,6	2.382,6	121	214,67	
12	229	218,2	2.618,4	144	216,64	
13	240	219,0	2.847,0	169	218,61	
14	210	220,4	3.085,6	196	220,58	
15	200	221,4	3.321,0	225	222,55	
16	223	223,0	3.568,0	256	224,52	
17	234	225,0	3.825,0	289	226,49	
18	248	227,2	4.089,6	324	228,46	
19	220	—	—	—	230,43	
20	211	—	—	—	232,40	
Σ 168		3.420,2	36.587,4	2.104		

– 2^{η} Περίπτωση: Εύρος ομαλοποίησης άρτιο $p = 2m$

$$y_t = \frac{1}{2}\left(\frac{1}{2m}\sum_{i=-m+1}^{m} x_{t+i} + \frac{1}{2m}\sum_{i=-m}^{m-1} x_{t+i}\right)$$

Παράδειγμα: $p = 4\ (= 2m) \Rightarrow m = 2$

$$y_3 = \frac{1}{2}\left[\frac{1}{4}(x_2+x_3+x_4+x_5) + \frac{1}{4}(x_1+x_2+x_3+x_4)\right]$$



Ιδιότητες

1. Ο κινητός μέσος L είναι ένας γραμμικός μετασχηματισμός. Άρα ισχύουν οι σχέσεις:

$$L(x_t + x'_t) = L(x_t) + L(x'_t)$$
$$L(\alpha x_t) = \alpha L(x_t)$$

2. Η γραμμική σειρά παραμένει αμετάβλητη:

$$\begin{aligned} y_t &= L\left[at + b + s(t) + E_t\right] \\ &= \underbrace{at+b}_{\text{όμοια ίσον}} + L[s(t)] + L(E_t) \end{aligned}$$

3. Όταν το εύρος ομαλοποίησης ισούται με την περιοδικότητα της σειράς T, οι εποχικοί συντελεστές εξαφανίζονται:

$$L\left[s(t)\right] = 0$$

Τούτο προκύπτει από τον ορισμό των εποχικοτήτων.

4. Το φιλτραρισμένο σφάλμα γίνεται:

$$A_t = L(E_t) = \frac{1}{2m+1} \sum_{i=-m}^{+m} E_{t+i}$$, απ' όπου:

- $E(A_t) = \frac{1}{2m+1} \sum_i E(E_{t+i}) = 0$

- $V(A_t) = V\left[\frac{1}{2m+1} \sum_i E_{t+i}\right] = \frac{1}{(2m+1)^2} \sum_i V(E_{t+i})$

$$= \frac{1}{2m+1} \sigma^2 < \sigma^2 = V(E_t)$$

- $Cov(A_t, A_{t+h}) \neq 0$, που μπορεί να προκαλέσει "παράσιτα" μέσα στη σειρά (εφφέ Yule-Slutzki)

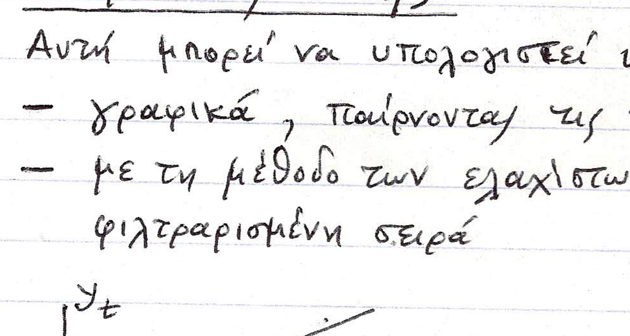
<u>Συμπερασματικά</u>

$$Y_t = at + b + L(E_t)$$

Ο μετασχηματισμός "κινητών μέσων":
- Εξαφανίζει (αποεποχοποιεί) τις εποχικότητες της σειράς,
- μειώνει το εύρος των σφαλμάτων,
- αφήνει αναλλοίωτη την τάση.

<u>Εκτίμηση της τάσης</u>

Αυτή μπορεί να υπολογιστεί ως εξής:
- γραφικά, παίρνοντας τις τιμές $m(t) = L(x_t) := at + b$,
- με τη μέθοδο των ελαχίστων τετραγώνων πάνω στην φιλτραρισμένη σειρά



Παράδειγμα

$$\tilde{a} = \frac{\frac{1}{n}\sum_i t y_i - \bar{t}\bar{y}}{\frac{1}{n}\sum t^2 - \bar{t}^2} = \frac{\frac{1}{16} \times 36.587,4 - 10,5 \times 213,8}{\frac{1}{16} \times 2.104 - (10,5)^2}$$

$$= \frac{2.286,7 - 2.244,9}{131,5 - 110,25}$$

$$= 1,967$$

$$\tilde{b} = \bar{y} - \tilde{a}\bar{t}$$
$$= 213,8 - 1,967 \times 10,5$$
$$= 213,8 - 20,65$$
$$= 193,2$$

$$m(t) = 1,967t + 193,2$$

Μειονεκτήματα της μεθόδου

- Έλλειψη ταχείας προσαρμογής της μεθόδου σε αλλαγές κλίσης της τάσης (τις λαμβάνει υπ'όψη της μετά από $2m+1$ παρατηρήσεις).
- Απώλεια πληροφοριών ($2m$ παρατηρήσεις χάνονται, δηλαδή οι m πρώτες και οι m τελευταίες).