Θεματική εργασία για το μάθημα «Χρονικές Σειρές»

Εκτίμηση τάξης Μαρκοβιανής διαδικασίας, διαφορές και ομοιότητες με εκτίμηση τάξης AR διαδικασίας

Ζυγογιάννης Κωνσταντίνος
Τμήμα Μαθηματικών ΑΠΘ
Θεσσαλονίκη, 2021

Μαρκοβιανές Διαδικασίες

❖Μαρκοβιανή διαδικασία τάξης r, συνεχής χώρος καταστάσεων

$$P[a < X_t < \beta \mid X_{t_1} = s_1, X_{t_2} = s_2, \ldots, X_{t_n} = s_n] = P[a < X_t < \beta \mid X_{t_n} = s_n, X_{t_{n-1}} = s_{n-1}, \ldots, X_{t_{n-r+1}} = s_{n-r+1}]$$
$$\forall t_1 < t_2 < \ldots < t_n < t$$

❖Μαρκοβιανή αλυσίδα τάξης r, διακριτός χώρος καταστάσεων

$$P[X_{t_{n+1}} = s_{n+1} \mid X_{t_n} = s_n, \ldots, X_{t_0} = s_0] = P[X_{t_{n+1}} = s_{n+1} \mid X_{t_n} = s_n, X_{t_{n-1}} = s_{n-1}, \ldots, X_{t_{n-r+1}} = s_{n-r+1}]$$

❖Ομοιογένεια ως προς τον χρόνο

$$P[X_{n+1} = s_{n+1} \mid X_n = s_n, \ldots, X_0 = s_0] = P[X_{n+1} = s_{n+1} \mid X_n = s_n, X_{n-1} = s_{n-1}, \ldots, X_{n-r+1} = s_{n-r+1}]$$

Μαρκοβιανή αλυσίδα τάξης 1

Έστω χώρος καταστάσεων S={1,2}

	1	2
1	p_{11}	p_{12}
2	p_{21}	p_{22}

όπου $p_{ij} = P[X_1 = j \mid X_0 = i]$, πιθανότητα μετάβασης σε 1 βήμα

και $p_{ij}(n) = P[X_n = j \mid X_0 = i]$, πιθανότητα μετάβασης σε n βήματα

Σε πραγματικά δεδομένα, τα στοιχεία του P εκτιμώνται ως εξής:

$\overline{p_{ij}} = \dfrac{n_{ij}}{n_{ix}}$, πηλίκο του πλήθουστων μεταβάσεων της αλυσίδας από την i στην j, προς το πλήθος των μεταβάσεων από την i οπουδήποτε.

Goodman and Kruskal's Lambda

$$\lambda_{\mathrm{B}} = \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1}$$

ε_1:πιθανότητα να μην προβλέψουμε σαν επόμενη κατάσταση αυτήν με τη μεγαλύτερη πιθανότητα εμφάνισης

ε_2 : η ίδια πιθανότητα, δεσμεύοντας αυτή τη φορά ως προς την X_{t-1}

$$\overline{\lambda_{\mathrm{B}}} = L_B = \frac{\sum_{i=1}^{m} n_{M_i} - \max_j (C_j)}{N - 1 - \max_j (C_j)}$$

n_{M_i} : η υψηλότερη συχνότητα στην i-γραμμή

C_j : το άθροισμα των στοιχείων της j-στήλης

N : μέγεθος δείγματος

Κάνουμε τον έλεγχο υπόθεσης : $H_0 : \lambda_{\mathrm{B}} = 0$ vs $H_1 : \lambda_{\mathrm{B}} \neq 0$

Το στατιστικό του ελέγχου είναι: $z = \dfrac{L_B}{\sqrt{Var(L_B)}} \sim N(0,1)$

❖Γενικά $0 \leq L_B \leq 1$

❖ $L_B = 0$, όταν η γνώση της προηγούμενης κατάστασης δεν βελτιώνει καθόλου την πρόβλεψη της επόμενης, και $L_B = 1$ όταν έχουμε τέλεια πρόβλεψη της επόμενης κατάστασης, γνωρίζοντας την προηγούμενη.

Έλεγχος X^2

- Ο έλεγχος υπόθεσης που κάνουμε είναι :

H_0: οι παρατηρήσεις είναι ανεξάρτητες, δηλαδή $p(i,j) = p_i p_j$

vs

H_1 : οι παρατηρήσεις προέρχονται από 1ης τάξης μαρκοβιανή αλυσίδα

- Συμβολισμός: p_i είναι η πιθανότητα εμφάνισης του συμβόλου i

$p(i,j)$ Είναι η πιθανότητα εμφάνισης του ζεύγους (i,j)

Και ορίζουμε : $N_2 = \sum_{i,j} n_{ij} = N-1$, $n_{i+} = \sum_j n_{ij}$

Υπό τη μηδενική υπόθεση, η κατάσταση i θα ακολουθείται από την κατάσταση j σε συνολικό πλήθος ζευγών N2, $e_{ij} = N_2 p_i p_j$ φορές

- Εκτίμηση παραμέτρων : $\overline{p_i} = \dfrac{n_{i+}}{N_2}$, $\overline{p_j} = \dfrac{n_{+j}}{N_2}$ $\left(\Rightarrow \overline{e_{ij}} = \dfrac{n_{i+} n_{+j}}{N_2} \right)$

Το στατιστικό του ελέγχου είναι $X^2 = \sum_{i,j} \dfrac{(n_{ij} - \overline{e_{ij}})^2}{\overline{e_{ij}}} \approx X^2_{(m-1)^2}$

Εκτίμηση τάξης μέσω κριτηρίων πληροφορίας

- Τα κριτήρια που θα παρουσιαστούν, αποτελούν ορισμένους από τους τρόπους που διαθέτουμε για να εκτιμήσουμε πόση πληροφορία περιέχεται σε ένα μοντέλο καθορισμένης τάξης και να κάνουμε σύγκριση με άλλες τάξεις. Σκοπός είναι να βρεθεί μια ισορροπία μεταξύ της πληροφορίας που παρέχει το μοντέλο (μέγιστη δυνατή) και των παραμέτρων που πρέπει να εκτιμηθούν (ελάχιστες δυνατές).

- Αρχικά ορίζεται ένα σταθερό άνω φράγμα R για την τάξη : $r \leq R$

- Η εκτίμηση της τάξης της Μαρκοβιανής διαδικασίας θα είναι :

$$\bar{r} = \arg\min\{AIC(t), t = 0,1,\ldots,R\}$$

$$\bar{r} = \arg\min\{BIC(t), t = 0,1,\ldots,R\}$$

$$\bar{r} = \arg\min\{EDC(t), t = 0,1,\ldots,R\}$$

AIC – BIC - EDC

❖ Οι τιμές των κριτιρίων που θα ασχοληθούμε υπολογίζονται ως εξής

$$AIC(t) = -2\log(L(t)) + 2m^t(N-1)$$

$$BIC(t) = -2\log(L(t)) + m^t\log(N)$$

$$EDC(t) = -2\log(L(t)) + \gamma(t)c_N$$

όπου L(t) είναι η συνάρτηση πιθανοφάνειας για τις πιθανότητες μετάβασης του μαρκοβιανού μοντέλου τάξης t, ο λογάριθμος της οποίας ορίζεται ως εξής :

$$\log(L(t)) = \sum_{i_1 \ldots i_{t+1}} n_{i_1 \ldots i_{t+1}} \log \frac{n_{i_1 \ldots i_{t+1}}}{n_{i_1 \ldots i_t}}$$

(κάνουμε τη σύμβαση ότι $\frac{0}{0} = 0$ και $0(\infty) = 0$)

Όπου $n_{i_1 \ldots i_t}$ είναι το πλήθος εμφάνισης της ακολουθίας καταστάσεων $i_1 \ldots i_t$.

❖Για το κριτήριο EDC, γ(t) είναι μία γνησίως αύξουσα συνάρτηση του t, ενώ c_N είναι γενικά μια ακολουθία θετικών τ.μ. , η επιλογή της οποίας μπορεί να γίνει έτσι ώστε να έχει ορισμένες «επιθυμητές» ιδιότητες (π.χ. ισχυρή-ασθενής συνέπεια εκτίμησης).

AR(αυτοπαλινδρομούμενες) διαδικασίες

- Προσδιορισμός τάξης με το κριτήριο της μερικής αυτοσυσχέτισης

Αρχικά προσαρμόζουμε στη χρονοσειρά τα παρακάτω AR μοντέλα :

AR(1): $x_t = \phi_{11} x_{t-1} + z_t$

AR(2): $x_t = \phi_{12} x_{t-1} + \phi_{22} x_{t-2} + z_t$

… …

AR(k): $x_t = \phi_{1k} x_{t-1} + \phi_{2k} x_{t-2} + \ldots + \phi_{kk} x_{t-k} + z_t$

- Στη συνέχεια κάνουμε εκτίμηση των παραμέτρων για το κάθε μοντέλο, και ενδιαφερόμαστε για τους συντελεστές $\overline{\phi_{\tau\tau}}, \tau = 1,2,\ldots, k$
- Σύμφωνα με το κριτήριο, εάν η χρονοσειρά να προέρχεται από στοχαστική διαδικασία τύπου AR(p), για κάποιο $1 \leq p \leq k$, τότε θα ισχύει ότι $\overline{\phi_{pp}} \neq 0$ και $\overline{\phi_{\tau\tau}} = 0, \forall \tau > p$. (Πρακτικά δεν απαιτούμε να μηδενίζεται ο συντελεστής της μερικής αυτοσυσχέτισης, αλλά να «πέφτει» σε μία ζώνη του μηδενός, που ορίζεται π.χ. από το διάστημα $\pm \frac{z_{a/2}}{\sqrt{N}}$,για επίπεδο σημαντικότητας a.)

Κριτήρια πληροφορίας για AR

- Η εκτίμηση τάξης των αυτοπανδρομούμενων μοντέλων θα γίνει μέσω των κριτηρίων AIC, BIC, FPE και έπειτα θα δοθούν ορισμένες διορθώσεις των παραπάνω κριτηρίων.

- Όπως και στα Μαρκοβιανά, τα κριτήρια αυτά βασίζονται στην πιθανοφάνεια των δεδομένων , η οποία τώρα εκτιμάται από την διασπορά των σφαλμάτων s_z^2 του AR μοντέλου. Και πάλι , με τα κριτήρια αυτά προσπαθούμε να πετύχουμε μία ισορροπία μεταξύ της καλής προσαρμογής στα δεδομένα και της πολυπλοκότητας του μοντέλου.

- $AIC(t) = \log(s_z^2) + \dfrac{2t}{N}$

- $BIC(t) = \log(s_z^2) + \dfrac{t\log(N)}{N}$

- $FPE(t) = s_z^2 \dfrac{N+t}{N-t}$

Η εκτίμηση της τάξης p του μοντέλου είναι εκείνη για την οποία ελαχιστοποιείται η συνάρτηση του κριτηρίου που επιλέγουμε

Σχολιασμός κριτηρίων πληροφορίας - Διορθώσεις

- Όλα τα προηγούμενα κριτήρια έχουν έναν όρο ποινής (penalty term), δηλαδή μία συνάρτηση που λαμβάνει υπόψιν το μέγεθος δείγματος N. Ενώ έχει αποδειχθεί ότι και τα 3 κριτήρια συμπεριφέρονται «καλά» και τα αποτελέσματα που δίνουν ασυμπτωτικά μοιάζουν μεταξύ τους (για πολύ μεγάλα N), δεν συμβαίνει το ίδιο όταν περιοριζόμαστε σε μικρότερα δείγματα.

- Στη συνέχεια προτείνουμε δύο εναλλακτικές προσεγγίσεις για το AIC και μία για το FPE, οι οποίες δουλεύουν καλύτερα για μικρά δείγματα , με κριτήριο την μεροληψία εκτίμησης, ενώ προσεγγίζουν τα AIC και FPE αντίστοιχα, καθώς $N \to \propto$.

- $AICF(t) = \log(s_z^2) + \dfrac{2t}{N-2t}$

- $AICC(t) = \log(s_z^2) + \dfrac{2t+2}{N-t-2}$

- $FPER(t) = s_z^2 \dfrac{1+\dfrac{t}{N-t}}{1-\dfrac{t}{N-t}}$

Βιβλιογραφία

- On Determination of the Order of a Markov Chain, L. C. ZHAO, C. C. Y. DOREA and C. R. GONC¸ALVES , (2001)
- Estimating the Order of an Autoregressive Model Using Normalized Maximum Likelihood, Daniel F. Schmidt, Member, IEEE, and Enes Makalic, (2011)
- Finite sample criteria for autoregressive model order selection, M.Karimi, (2007)
- http://ikee.lib.auth.gr/record/101333/files/gri-2008-1266.pdf?fbclid=IwAR3EwzLxKlPBWxYrbnd3kv7woYdruVzAEzWQt3W6jUbVkKZ-TCxIDvHfLOs
- Στοχαστικές μέθοδοι στις επιχειρησιακές έρευνες, Π.-Χ.Γ. Βασιλείου, (1999)
- Σημειώσεις μαθήματος «Ανάλυση Χρονοσειρών», Κουγιουμτζής Δημήτρης, (2020-2021)