Εργασία για το μάθημα «Στοχαστικές Μέθοδοι» του ΠΜΣ Στατιστική και Μοντελοποίηση

Ζυγογιάννης Κωνσταντίνος

Ιακωβίδης Ισίδωρος

Χριστόπουλος Γιώργος

Προσομοίωση στοιχείων ημιμαρκοβιανής αλυσίδας με χρήση συλλογής δεδομένων καιρού.

Σκοπός της εργασίας είναι η περιγραφή και παρουσίαση μιας ημιμαρκοβιανής αλυσίδας συλλέγοντας δεδομένα καιρού από την ιστοσελίδα http://meteosearch.meteo.gr/

<u>Περιγραφή δεδομένων.</u>

Επιλέξαμε την συλλογή πληροφοριών για δεδομένα βροχόπτωσης που αφορούν τις χρονικές περιόδους 1 Νοεμβρίου του 2017 έως 31 Μαΐου του 2018 και 1 Νοεμβρίου του 2018 έως 31 Μαΐου του 2019 στην περιοχή της Θεσσαλονίκης και συγκεκριμένα στον σταθμό της Καλαμαριάς μέσω του ιστότοπου http://meteosearch.meteo.gr/stationInfo.asp.

<u>Πληροφορίες σταθμού Καλαμαριάς (Θεσσαλονίκη (Καλαμαριά) (LGR9)).</u>

Υψόμετρο 5m, Θέση : Ναυτική Διοίκηση Βορείου Ελλάδας. Βρίσκεται σε χώμα. Ύψος αισθητήρων θερμ/υγρ 2m. Ύψος ανεμόμετρου 5m. Μέχρι 19/04/2016 βρισκόταν στον Ναυτικό Όμιλο Θεσσαλονίκης σε χώμα με αισθητήρες ύψους 2-5 m.

Σύμφωνα με την διεθνή βιβλιογραφία η κατηγοριοποίηση του ποσοστού βροχόπτωσης σε μια περιοχή ορίζεται ως εξής:

1) $Ασθενής\ αν\ < 2\ \ χιλιοστα/h$
2) Μέτρια αν $2-6\ \ χιλιοστα/h$
3) Ισχυρή αν $> 6\ \ χιλιοστα/h$

<u>Περιγραφή της διαχείρισης των δεδομένων.</u>

Θεωρούμε μια λίστα 212 θέσεων όπου κάθε της στοιχείο προκύπτει από τον μέσο όρο βροχόπτωσης της αντίστοιχης μέρας τον χρονικών περιόδων από 1 Νοεμβρίου του

2017 έως 31 Μαΐου του 2018 και 1 Νοεμβρίου του 2018 έως 31 Μαΐου του 2019. Παρατηρείται το φαινόμενο της έλλειψης βροχόπτωσης κατά την διάρκεια καλοκαιρινών μηνών, γεγονός που θα οδηγούσε ότι λίγες τιμές θα έχουν μεγάλους χρόνους παραμονής στην κατάσταση μηδενική βροχή. Για αυτόν τον λόγο επιλέχθηκε να εξαιρεθούν τα δεδομένα των καλοκαιρινών μηνών.

Αρχικά δίνεται το ιστόγραμμα των μέσω τιμών βροχής για το διάστημα της μελέτης.

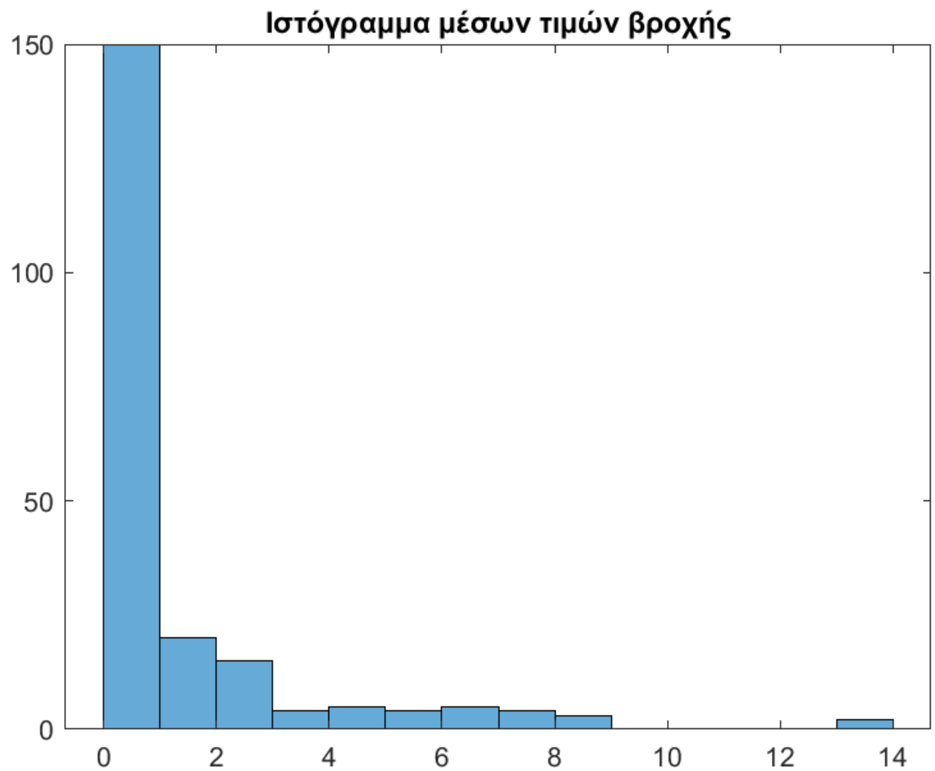

Figure 1 Ιστόγραμμα των μέσω τιμών βροχής για κάθε μέρα.

Παρατηρείται ότι δεν υπάρχουν πολλές τιμές για την μέση βροχόπτωση για τα διαστήματα [2,6] και $(6, +\infty)$, επομένως θα θεωρήσουμε τις δύο αυτές κατηγορίες ως μία. Δηλαδή οι καταστάσεις ορίζονται παρακάτω ως εξής.

- Μηδενική αν δεν παρατηρείτε βροχόπτωση (κατάσταση 1)
- $Ασθενής\ αν\ < 2 \quad χιλιοστα/h$ (κατάσταση 2)
- Ισχυρή αν $\geq 2 \quad χιλιοστα/h$ (κατάσταση 3)

Στη συνέχεια το ιστόγραμμα για τις καταστάσεις, δηλαδή των τιμών της μεταβλητής rain_state.

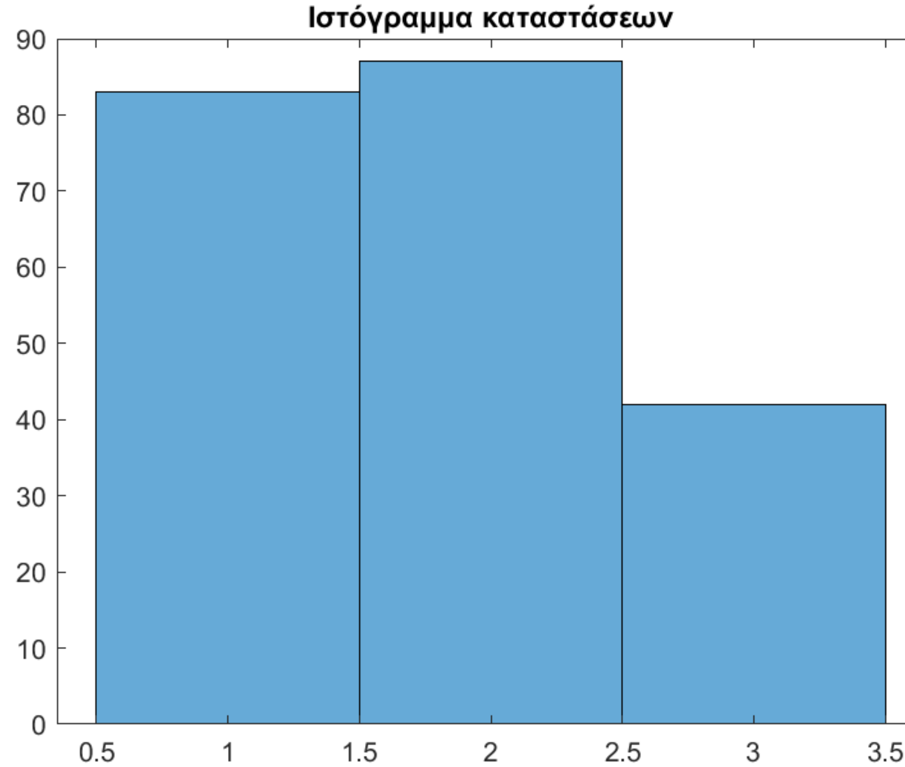

Figure 2 Ιστόγραμμα της μεταβλητής rain_state.

<u>Δημιουργία Ημι Μαρκοβιανής αλυσίδας.</u>

Επειδή οι μετρήσεις της βροχής λαμβάνονται σε σταθερά χρονικά διαστήματα (κάθε μέρα) έχουμε ως αποτέλεσμα ότι κάθε χρονική στιγμή είναι χρονική στιγμή μετάβασης. Επομένως ισχύει ότι $H(1) = J$, όπου J είναι ο πίνακας με όλα τα στοιχεία του μονάδες και $H(m) = 0$, για κάθε $m \neq 1$. Δηλαδή το μοντέλο μας είναι Μαρκοβιανό και ομογενές ως προς τον χρόνο.

Για να κάνουμε τη μελέτη με Ημι-Μαρκοβιανή αλυσίδα πρέπει να δημιουργήσουμε τις μεταβλητές rain_state2 και time με τον εξής περιορισμό: δεν επιτρέπεται η μετάβαση από την κατάσταση i στην κατάσταση i και η πληροφορία των συνεχόμενων ίδιων καταστάσεων στην μεταβλητή rain_state διατηρείται στην μεταβλητή time ως χρόνος παραμονής σε κάθε κατάσταση.

Η μεταβλητή rain_state2 είναι ένα διάνυσμα στήλη που περιλαμβάνει τις 118 αλλαγές κατάστασης και η μεταβλητή time τους 118 χρόνους παραμονής αντίστοιχα.

Τέλος αφού δεν επιτρέπεται η μετάβαση από την κατάσταση i στην κατάσταση i εξ ορισμού τα διαγώνια στοιχεία του πίνακα πιθανοτήτων μετάβασης P και των πινάκων πιθανοτήτων των χρόνων παραμονής $H(m)$ που θα εκτιμηθούν θα είναι 0.

Στη συνέχεια το ιστόγραμμα των καταστάσεων της HMA, δηλαδή των τιμών της μεταβλητής rain_state2.

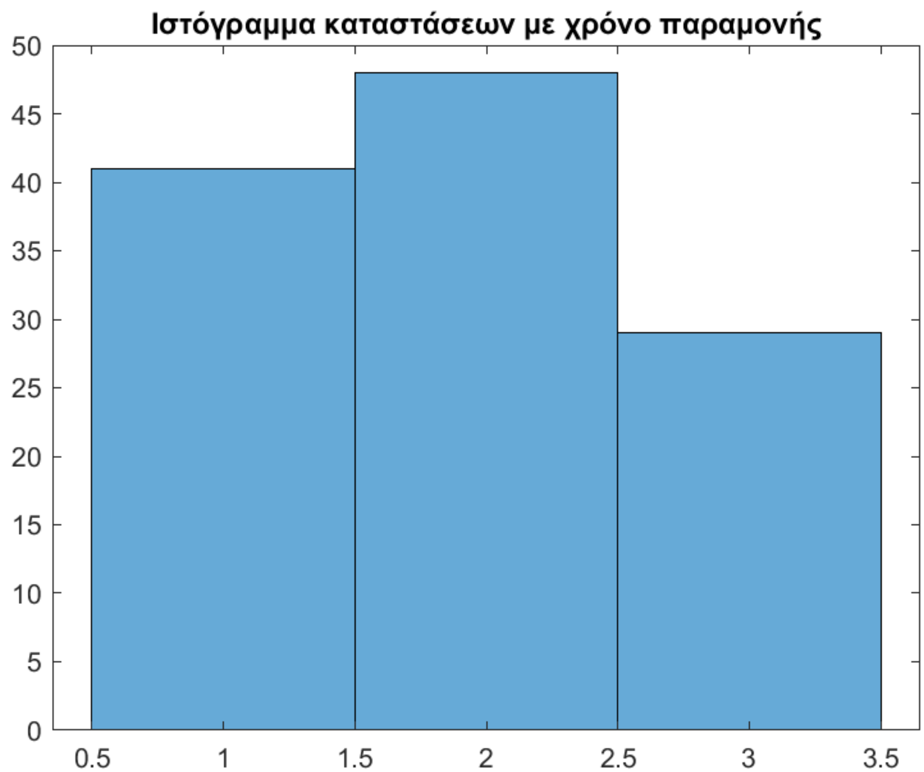

Figure 3 Ιστόγραμμα της μεταβλητής rain_state2

<u>Εκτίμηση πινάκων P και H(m).</u>

Θα ορίσουμε τον πίνακα N_P όπου το στοιχείο στην θέση (i, j) δηλώνει το πλήθος των μεταβάσεων από την κατάσταση i στην κατάσταση j. Επίσης η μέγιστη τιμή της μεταβλητής time δηλαδή ο μέγιστος χρόνος παραμονής σε μία κατάσταση είναι $m_{max} = 6$, επομένως θα ορίσουμε 6 πίνακες $N\_H(m), m = \{1,2,3,4,5,6\}$ όπου το στοιχείο στην θέση (i, j) δηλώνει το πλήθος που παρουσιάστηκε χρόνος παραμονής m στην κατάσταση i δεδομένου ότι η επόμενη μετάβαση έγινε στην κατάσταση j. Τέλος θα δημιουργήσουμε τις εμπειρικές κατανομές των χρόνων παραμονής υπό συνθήκη δηλαδή τις $T_{12}, T_{13}, T_{21}, T_{23}, T_{31}, T_{32}$, οι οποίες έχουν η κάθε μία N_P_{ij} στοιχεία.

Ο πίνακας $N\_P = \begin{pmatrix} 0 & 30 & 10 \\ 29 & 0 & 19 \\ 12 & 17 & 0 \end{pmatrix}$ και αντίστοιχα οι πίνακες $N\_H(1) = \begin{pmatrix} 0 & 16 & 8 \\ 18 & 0 & 9 \\ 6 & 12 & 0 \end{pmatrix}$, $N\_H(2) = \begin{pmatrix} 0 & 2 & 2 \\ 6 & 0 & 5 \\ 6 & 4 & 0 \end{pmatrix}$, $N\_H(3) = \begin{pmatrix} 0 & 4 & 0 \\ 3 & 0 & 3 \\ 0 & 0 & 0 \end{pmatrix}$, $N\_H(4) = \begin{pmatrix} 0 & 3 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$, $N\_H(5) = \begin{pmatrix} 0 & 4 & 0 \\ 2 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$, $N\_H(6) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Ο πίνακας πιθανοτήτων μετάβασης P της ΗΜΑ αλυσίδας προκύπτει ως εξής από τον πίνακα N_P

$$P_{ij} = \frac{N\_P_{ij}}{\sum_j N\_P_{ij}}$$

Ενώ οι πίνακες $H(m)$ υπολογίζονται ως εξής από τους πίνακες $N\_H(m)$ και N_P

$$H_{ij}(m) = \frac{N\_H_{ij}(m)}{N\_P_{ij}}$$

Ο πίνακας P είναι $P = \begin{pmatrix} 0 & 0.75 & 0.25 \\ 0.6042 & 0 & 0.3958 \\ 0.4138 & 0.5862 & 0 \end{pmatrix}$ και είναι στοχαστικός, και οι πίνακες $H(m), m = \{1,2,3,4,5,6\}$ είναι $H(1) = \begin{pmatrix} 0 & 0.5333 & 0.8 \\ 0.6207 & 0 & 0.4737 \\ 0.5 & 0.7059 & 0 \end{pmatrix}$, $H(2) = \begin{pmatrix} 0 & 0.0667 & 0.2 \\ 0.2069 & 0 & 0.2632 \\ 0.5 & 0.2353 & 0 \end{pmatrix}$, $H(3) = \begin{pmatrix} 0 & 0.1333 & 0 \\ 0.1034 & 0 & 0.1579 \\ 0 & 0 & 0 \end{pmatrix}$, $H(4) = \begin{pmatrix} 0 & 0.1 & 0 \\ 0 & 0 & 0.0526 \\ 0 & 0.0588 & 0 \end{pmatrix}$, $H(5) = \begin{pmatrix} 0 & 0.1333 & 0 \\ 0.0690 & 0 & 0.0526 \\ 0 & 0 & 0 \end{pmatrix}$ και $H(6) = \begin{pmatrix} 0 & 0.0333 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Υπολογισμός μέσης τιμής και διασποράς για την κατανομή του χρόνου παραμονής σε κάθε κατάσταση.

Τέλος, υπολογίστηκαν η μέση τιμή και η διασπορά για την κατανομή του χρόνου παραμονής σε κάθε κατάσταση T_1, T_2 και T_3 χρησιμοποιώντας τις μέσες τιμές και τις διασπορές των εμπειρικών κατανομών των χρόνων παραμονής υπό συνθήκη $T_{12}, T_{13}, T_{21}, T_{23}, T_{31}, T_{32}$.

$$E[T_i] = \sum_j p_{ij}\, E[T_{ij}]$$

$$Var[T_i] = -E^2[T_i] + \sum_j p_{ij} \left[Var[T_{ij}] + E^2[T_{ij}]\right]$$

Και επομένως έχουμε $E[T_1] = 2.05$, $E[T_2] = 1.7917$, $E[T_3] = 1.4483$, $Var[T_1] = 2.3715$, $Var[T_2] = 1.3457$, $Var[T_3] = 0.4854$.

Υπολογισμός πινάκων πιθανοτήτων μετάβασης σε διάστημα Q(m).

Στη συνέχεια υπολογίστηκαν κατά σειρά οι πίνακες $W(m)$, οι πίνακες πυρήνες $C(m)$, η συνάρτηση επιβίωσης ${}^{>}W(m)$ και οι πίνακες πιθανοτήτων μετάβασης σε διάστημα $Q(m)$ για $m = 1 \dots m_{\max}$, όπου $m_{\max}$ ο μέγιστος παρατηρούμενος χρόνος παραμονής σε μια κατάστασης, δηλαδή $m_{\max} = 6$.

$\mathrm{Q}(1) = \begin{pmatrix} 0.4000 & 0.4000 & 0.2000 \\ 0.3750 & 0.4375 & 0.1875 \\ 0.2069 & 0.4138 & 0.3793 \end{pmatrix}$, $\mathrm{Q}(2) = \begin{pmatrix} 0.4914 & 0.3078 & 0.2009 \\ 0.3138 & 0.4359 & 0.2503 \\ 0.4448 & 0.4017 & 0.1534 \end{pmatrix}$, $\mathrm{Q}(3) = \begin{pmatrix} 0.4436 & 0.3973 & 0.1591 \\ 0.4017 & 0.3672 & 0.2311 \\ 0.3660 & 0.3872 & 0.2468 \end{pmatrix}$, $\mathrm{Q}(4) = \begin{pmatrix} 0.4343 & 0.3849 & 0.1807 \\ 0.3807 & 0.4152 & 0.2041 \\ 0.4030 & 0.3924 & 0.2046 \end{pmatrix}$, $\mathrm{Q}(5) = \begin{pmatrix} 0.3557 & 0.4587 & 0.1856 \\ 0.4365 & 0.3609 & 0.2026 \\ 0.4075 & 0.3994 & 0.1931 \end{pmatrix}$,$\mathrm{Q}(6) = \begin{pmatrix} 0.3965 & 0.4028 & 0.2007 \\ 0.3869 & 0.4186 & 0.1945 \\ 0.4074 & 0.3962 & 0.1964 \end{pmatrix}$

Στη συνέχεια παραθέτονται γραφικά ανά τριάδες τα στοιχεία κάθε στήλης του πίνακα $Q(m)$.

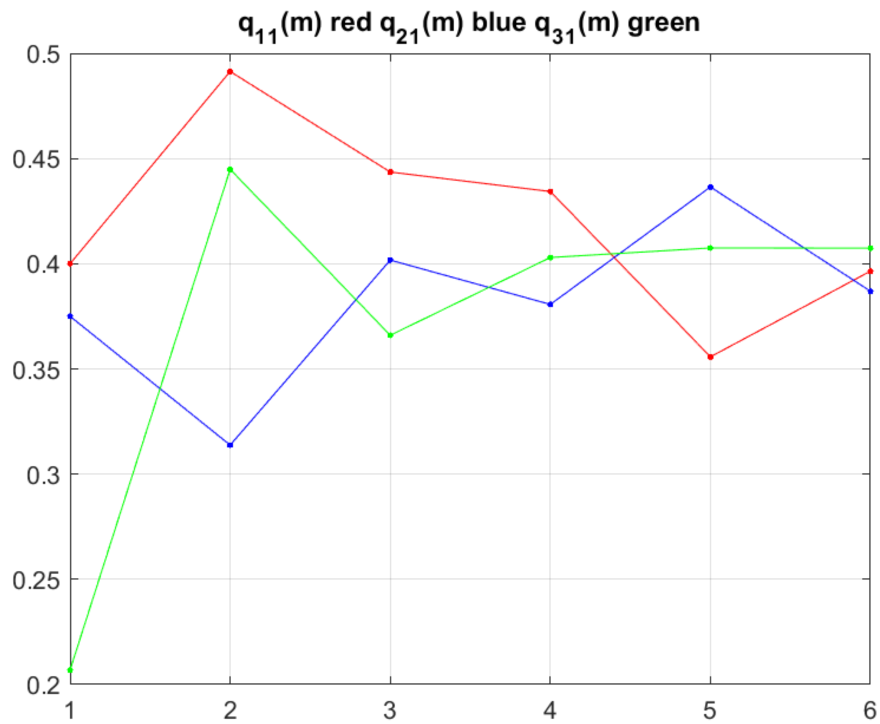

Figure 4 Εξέλιξη των πιθανοτήτων μετάβασης σε διάστημα $q_{11}(m)$, $q_{21}(m)$ και $q_{31}(m)$.

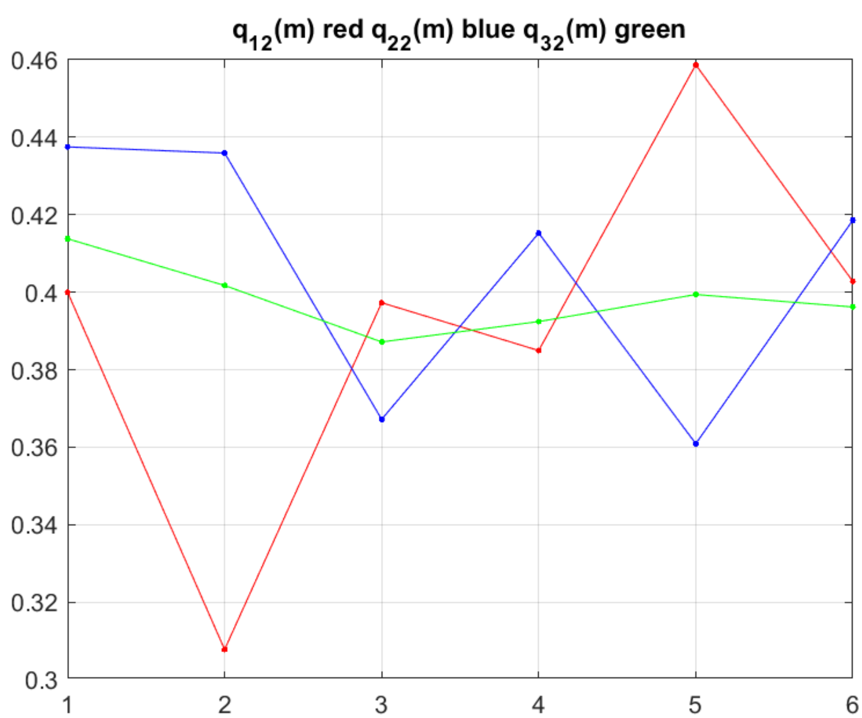

Figure 5 Εξέλιξη των πιθανοτήτων μετάβασης σε διάστημα $q_{12}(m)$, $q_{22}(m)$ και $q_{32}(m)$.

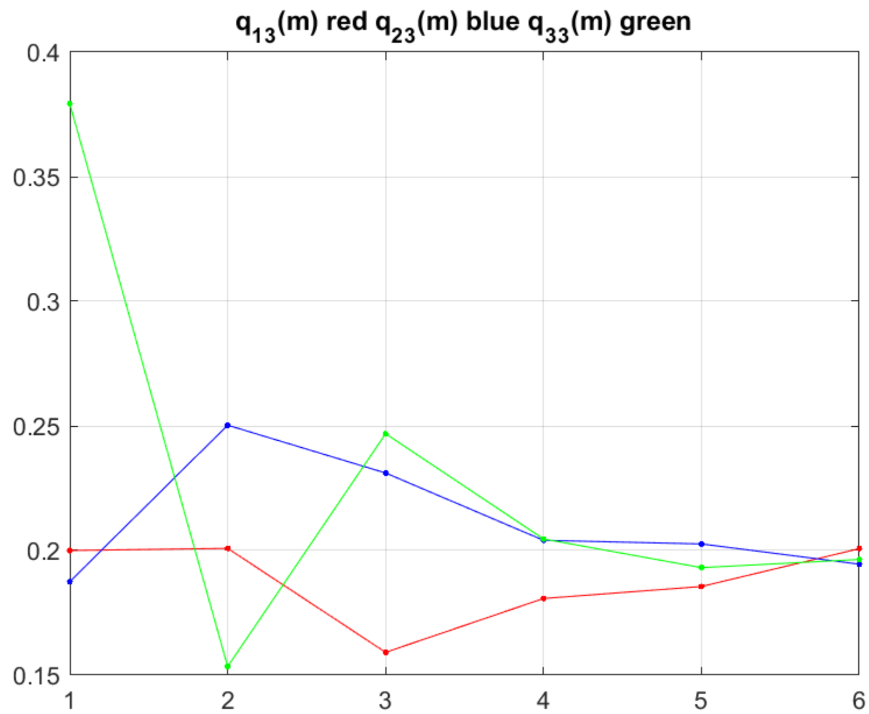

Figure 6 Εξέλιξη των πιθανοτήτων μετάβασης σε διάστημα $q_{13}(m), q_{23}(m)$ *και* $q_{33}(m)$.