Εργασία για το μάθημα «Ειδικά Θέματα ΙΙ: Γενικευμένα Μαρκοβιανά Συστήματα» του ΠΜΣ Στατιστική και Μοντελοποίηση

Κρυφά Μαρκοβιανά μοντέλα

Διάρκεια παραμονής στις καταστάσεις

Ζυγογιάννης Κωνσταντίνος

Ιακωβίδης Ισίδωρος

Χριστόπουλος Γιώργος

ΕΙΣΑΓΩΓΗ

Σκοπός της παρακάτω εργασίας είναι η μελέτη των Κρυφών Μαρκοβιανών Μοντέλων, εστιάζοντας στο μεγαλύτερο μέρος στα μοντέλα εκείνα όπου ο χρόνος παραμονής στις καταστάσεις της Μαρκοβιανής αλυσίδας ορίζεται με σαφή τρόπο και είναι, εν γένει, διαφορετικός από εκείνον που ορίζει το απλό κρυφό μοντέλο. Αρχικά ορίζεται το απλό Κρυφό Μαρκοβιανό Μοντέλο, ενώ στη συνέχεια παρουσιάζονται τα τρία βασικά προβλήματα που συνδέονται με αυτό , καθώς και ο τρόπος αντιμετώπισής τους. Έπειτα γενικεύουμε τα προηγούμενα, ορίζοντας τα Κρυφά ημι-Μαρκοβιανά Μοντέλα, με την βοήθεια των οποίων καθορίζουμε με σαφή τρόπο τη διάρκεια παραμονής της αλυσίδας σε συγκεκριμένες καταστάσεις. Το νέο αυτό μοντέλο μπορεί πλέον να χρησιμοποιηθεί σε ένα μεγάλο πλήθος εφαρμογών. Αφού περιγραφεί η λειτουργεία τους ,διατυπώνονται και επιλύονται τα τρία βασικά προβλήματα στη γενικότερη μορφή τους. Αναφέρεται ως πρώτο παράδειγμα η Coxian κατανομή και έπειτα το σύνολο των κυρτά μονότονων κατανομών. Χρησιμοποιούμε την κατανομή Γάμμα για την μοντελοποίηση του προβλήματος του χρόνου παραμονής, ενώ μια ιδιότητα της κατανομής αυτής, μας επιτρέπει να την αντικαταστήσουμε με ένα ευρύ σύνολο γνωστών κατανομών.

ΠΕΡΙΕΧΟΜΕΝΑ

Διακριτές Μαρκοβιανές αλυσίδες

Σε αυτήν την παράγραφο θα γίνει μια σύντομη περιγραφή των διακριτών μαρκοβιανών αλυσίδων που γενικώς στο υπόλοιπο κομμάτι της παρούσας εργασίας θα θεωρούνται γνωστές. Ο λόγος της εισαγωγής αυτής είναι η παρουσίαση ορισμένων προβλημάτων που θα μας απασχολήσουν, όπως και οι λύσεις τους.

Έστω ένα σύστημα που αποτελείται από N διακριτές καταστάσεις $S_1, S_2, \dots, S_N$. Ανά τακτά χρονικά διαστήματα (διακριτού χρόνου) το σύστημα υποβάλλεται σε αλλαγή κατάστασης (πιθανώς παραμένει και στην ίδια κατάσταση) σύμφωνα με ένα σετ πιθανοτήτων που σχετίζονται με τις καταστάσεις. Οι χρονικές στιγμές που έχουν σχέση με τις αλλαγές κατάστασης του συστήματος συμβολίζονται με $t = 1, 2, \dots$ και η συγκεκριμένη κατάσταση σε χρόνο t συμβολίζεται με q_t. Για την ειδική περίπτωση μίας διακριτής , πρώτης τάξης μαρκοβιανής αλυσίδας η πιθανολογική περιγραφή της περιορίζεται μόνο στην παρούσα και την προηγούμενη κατάσταση, δηλαδή:

$$P[q_t = S_t \mid q_{t-1} = S_{t-1}, q_{t-2} = S_{t-2}, \dots] = P[q_t = S_t \mid q_{t-1} = S_{t-1}] \qquad (1)$$

Επιπλέον θεωρούμε μόνο τις διαδικασίες στις οποίες το δεξί μέλος της (1) είναι ανεξάρτητο του χρόνου, επομένως έχουμε ένα σετ πιθανοτήτων μετάβασης της μορφής

$$a_{ij} = P(q_t = S_j \mid q_{t-1} = S_i), \ \ 1 \leq i, j \leq N \qquad (2)$$

και υπόκεινται στους περιορισμούς $a_{ij} \geq 0$ και $\sum_{i=1}^{N} a_{ij} = 0$.

Η παραπάνω στοχαστική διαδικασία είναι ένα παρατηρούμενο μαρκοβιανό μοντέλο αφού το αποτέλεσμα της διαδικασίας είναι μία ομάδα καταστάσεων, μία για κάθε χρονική στιγμή, και κάθε κατάσταση αντιστοιχεί σε ένα φυσικό γεγονός (παρατηρούμενο).

Το παρακάτω παράδειγμα είναι ενδεικτικό των ερωτημάτων που μπορούν να προκύψουν σε μια αλυσίδα Markov.

Έστω ο πίνακας $A = \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.4 & 0.3 & 0.3 \\ 0.2 & 0.7 & 0.1 \end{pmatrix}$ των πιθανοτήτων μετάβασης που περιγράφει την μαρκοβιανή αλυσίδα m του παραδείγματος.

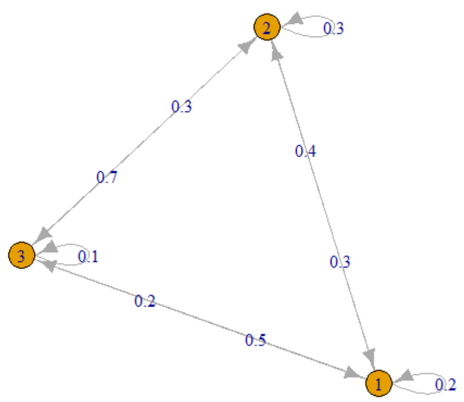

Figure1 Η Μαρκοβιανή αλυσίδα του παραδείγματος

Ερωτήσεις που άμεσα προκύπτουν είναι:
A) Πια η πιθανότητα εμφάνισης μιας συγκεκριμένης ακολουθίας καταστάσεων ;
B) Πια η πιθανότητα παραμονής σε μια συγκεκριμένη κατάσταση για χρόνο ακριβώς d;

Λύση του προβλήματος Α

Έστω για παράδειγμα μια ακολουθία καταστάσεων :
$O = S_3 \rightarrow S_3 \rightarrow S_3 \rightarrow S_1 \rightarrow S_2$,τότε :

$$P(O \mid m) = P(S_3)P(S_3 | S_3)P(S_3 \mid S_3)P(S_1 \mid S_3)P(S_2 \mid S_1) = 1\,(0.1)(0.1)(0.2)(0.3)$$
$$= 0.0006$$

Όπου $P\big(S_i \mid S_j\big)$ σημαίνει πιθανότητα μετάβασης από την κατάσταση j στην κατάσταση i.

Λύση του προβλήματος Β

Έστω η ακολουθία $O = S_i \rightarrow S_i \rightarrow S_i \ldots \rightarrow S_i \rightarrow S_j$όπου έχουμε παραμείνει στην κατάσταση S_i ακριβώς d φορές και η τελευταία κατάσταση S_j είναι διαφορετική. Τότε για τον υπολογισμό της πιθανότητας παραμονής έχουμε

$$P(0 \mid m) = P(S_i)P(S_i \mid S_i)P(S_i \mid S_i) \ldots P\big(S_j \mid S_i\big) = 1a_{ii}a_{ii} \ldots a_{ii}(1 - a_{ii})$$
$$= {a_{ii}}^{d-1}(1 - a_{ii})$$

Η παραπάνω πιθανότητα παραμονής στην κατάσταση i συμβολίζεται με $p_i(d)$. Χρησιμοποιώντας την παραπάνω πιθανότητα μπορούμε εύκολα να υπολογίσουμε το μέσο αναμενόμενο χρόνο παραμονής σε μία κατάσταση δηλαδή,

$$\bar{d}_i = E\big(p_i(d)\big) = \sum_{d=1}^{\infty} d p_i(d) = \sum_{d=1}^{\infty} d{a_{ii}}^{d-1}(1 - a_{ii}) = (1 - a_{ii})\sum_{d=1}^{\infty} d{a_{ii}}^{d-1}$$
$$= (1 - a_{ii})\left(\sum_{d=0}^{\infty} {a_{ii}}^{d}\right)' = (1 - a_{ii})(1/(1 - a_{ii})' = \frac{1}{1 - a_{ii}}$$

Κρυφά Μαρκοβιανά μοντέλα

Σε αυτήν την παράγραφο επεκτείνουμε την θεωρία των μαρκοβιανών μοντέλων στα κρυφά μαρκοβιανά μοντέλα. Σε πολλά προβλήματα που προκύπτουν παρατηρείτε το φαινόμενο να μην είναι εφικτή η παρατήρηση της κατάστασης της μαρκοβιανής αλυσίδας στον χρόνο t. Επομένως κρυφό μαρκοβιανό μοντέλο είναι μία διπλά ενσωματωμένη στοχαστική διαδικασία όπου η υποβόσκουσα στοχαστική διαδικασία δεν είναι παρατηρήσιμη (είναι κρυφή) και μπορεί να παρατηρηθεί μόνο μέσω μίας άλλης στοχαστικής διαδικασίας που παράγει τις παρατηρήσεις. Στο παρακάτω παράδειγμα φαίνεται μια τέτοια κατάσταση.

Έστω Ν δοχεία στα οποία βρίσκονται αρκετές μπάλες από Μ διαφορετικά χρώματα. Εκεί, μέσω μιας τυχαίας διαδικασίας επιλέγεται από ένα αρχικό δοχείο μια μπάλα, καταγράφεται το χρώμα της και επιστρέφει στο δοχείο. Στη συνέχεια επιλέγεται ξανά κάποιο δοχείο και η ίδια διαδικασία συνεχίζεται για πεπερασμένου πλήθους βήματα. Ενδιαφερόμαστε για την μοντελοποίηση της παραπάνω διαδικασίας που θα την ονομάζουμε κρυφό μαρκοβιανό μοντέλο.

Στοιχεία ενός κρυφού μαρκοβιανού μοντέλου.

Ένα κρυφό μαρκοβιανό σύστημα χαρακτηρίζεται από τα ακόλουθα:

1) N : Οι καταστάσεις του μοντέλου. Οι καταστάσεις μιας κρυφής μαρκοβιανής αλυσίδας είναι κρυφές, όμως ανάλογα την φύση του προβλήματος θα δίνεται κάποια φυσική ερμηνεία σε αυτές. Στο προηγούμενο παράδειγμα των δοχείων-μπαλών το πλήθος τον διακεκριμένων δοχείων είναι οι καταστάσεις του μοντέλου. Ένα απλό παράδειγμα μοντελοποίησης θα ήταν να επιλέξουμε το δοχείο να είναι για παράδειγμα μοναδικό. Εν γένει οι καταστάσεις ίσως επικοινωνούν με κάποια πιθανότητα μετά από κάποιο πεπερασμένο χρονικό βήμα (εργοδικές αλυσίδες) αλλά μπορεί και όχι. Συμβολίζουμε το σύνολο των N αυτών καταστάσεων με $\{S_1, S_2, \ldots, S_N\}$.

2) M : Το πλήθος των διαφορετικών παρατηρήσεων ανά κατάσταση που μπορεί να γνωρίζει ο παρατηρητής. Στο παράδειγμά μας των δοχείων-μπαλών, το πλήθος των διαφορετικών χρωμάτων είναι το παρατηρούμενο αντικείμενο. Θεωρούμε $V = \{v_1, v_2, \ldots, v_M\}$, τα διακεκριμένα παρατηρούμενα αντικείμενα.

3) Οι πιθανότητες μετάβασης των καταστάσεων, $A = \{a_{ij}\}$ δηλαδή $a_{ij} = P(q_t = S_j \mid q_{t-1} = S_i)$, $1 \le i, j \le N$. Για την ειδική περίπτωση όπου οποιαδήποτε κατάσταση είναι προσβάσιμη από όλες τις άλλες καταστάσεις με ένα και μόνο βήμα έχουμε ότι $a_{ij} > 0$ για όλα τα i, j. Για άλλους τύπους κρυφών μαρκοβιανών μοντέλων θα έχουμε $a_{ij} = 0$ για ένα ή περισσότερα ζεύγη (i, j).

4) Η πιθανότητα του παρατηρούμενου αντικειμένου στην κατάσταση j
$b_j(k) = P(v_k \text{ στο χρόνο } t \mid q_t = j)$, $i \in \{1,2, \ldots, N\}$και$k \in \{1,2, \ldots, M\}$.

5) Η αρχική κατάσταση $\pi = \{\pi_i\}$ δηλαδή $\pi_i = P(q_1 = S_i)$, $i \in \{1,2, \ldots, N\}$

Δεδομένων των τιμών των N, M, A, B, π κατασκευάζεται μια πεπερασμένη ακολουθία παρατηρήσεων $O = O_1 O_2 \ldots O_T$ με τον παρακάτω τρόπο:

1) Επιλέγοντας μια αρχική κατάσταση $q_1 = S_i$ με βάση την αρχική συνθήκη
2) Ορίζουμε $t = 2$
3) Επιλέγουμε $O_t = v_k$ με γνώμονα την κατανομή της πιθανότητας $b_i(k)$ για την κατάσταση S_i.
4) Μετακινούμαστε σε μια καινούργια κατάσταση q_{t+1} σύμφωνα πάντα με τις πιθανότητες μετάβασης a_{ij}.
5) Ορίζουμε $t = t + 1$, επιστρέφουμε στο βήμα 3) αν $t < T$
   Σε διαφορετική περίπτωση, η διαδικασία τερματίζεται.

Η παραπάνω διαδικασία μπορεί να χρησιμοποιηθεί και ως γεννήτορας παρατηρήσεων αλλά και ως μοντέλο για το πως παρήχθη μία δοσμένη ακολουθία παρατηρήσεων από ένα κατάλληλο κρυφό μαρκοβιανό μοντέλο. Τέλος χρησιμοποιούμε τον συμβολισμό $\lambda = (A, B, \pi)$ για να υποδείξουμε το σύνολο των παραμέτρων του μοντέλου.

Τα τρία βασικά προβλήματα για τα κρυφά μαρκοβιανά μοντέλα.

Πρόβλημα 1: Δεδομένης μιας ακολουθίας παρατηρήσεων $O = O_1 O_2 \dots O_T$ και ενός μοντέλου $\lambda = (A, B, \pi)$, πως υπολογίζεται αποτελεσματικά η
$P(O \mid \lambda) = \textit{πιθανότητα εμφάνισης της ακολουθίας } O;$

Η λύση του παραπάνω προβλήματος δημιουργεί αυτόματα ένα τρόπο αξιολόγησης κάθε μοντέλου με βάση τις πιθανότητες των παρατηρούμενων ακολουθιών. Έτσι ο ερευνητής μπορεί να επιλέξει ανάλογα την φύση του προβλήματος και των δεδομένων του το κατάλληλο για αυτόν μοντέλο.

Λύση του Προβλήματος1

Έστω συγκεκριμένη ακολουθία παρατηρήσεων $O = O_1 O_2 \dots O_T$ και κάποιο μοντέλολ. Αναζητάμε την$P(O \mid \lambda)$.

Ένας φυσιολογικός τρόπος είναι κάποιος να υπολογίσει όλους τους πιθανούς συνδυασμούς των καταστάσεων μήκους T. Έστω $Q = q_1 q_2 \dots q_T$, με q_1 αρχική κατάσταση, μια ακολουθία καταστάσεων. Η πιθανότητα εμφάνισης της ακολουθίας παρατηρήσεων δεδομένης της ακολουθίας καταστάσεων και θεωρώντας ανεξαρτησία των παρατηρήσεων είναι

$$P(O \mid Q, \lambda) = P(O_1 O_2 \dots O_T \mid Q, \lambda) = P(O_1 \mid Q, \lambda) P(O_2 \mid Q, \lambda) \dots P(O_T \mid Q, \lambda)$$
$$= P(O_1 \mid q_1, \lambda) P(O_2 \mid q_2, \lambda) \dots P(O_T \mid q_T, \lambda) = \prod_{t=1}^{T} b_{q_t}(O_t) \qquad (3)$$

Η πιθανότητα της ακολουθίας καταστάσεων είναι

$$P(Q \mid \lambda) = \pi_{q_\iota} a_{q_1 q_2} \dots a_{q_{T-1} q_T} \qquad (4)$$

Η από κοινού πιθανότητα να συμβούν ταυτόχρονα και τα δύο γεγονότα είναι

$$P(O, Q \mid \lambda) = P(O \mid Q, \lambda) * P(Q \mid \lambda) \qquad (5)$$

η οποία υπολογίζεται από τις (3) και (4). Η επιθυμητή πιθανότητα τελικά θα βρεθεί υπολογίζοντας το άθροισμα πάνω σε όλες τις δυνατές ακολουθίες καταστάσεων $Q = q_1 q_2 \dots q_T$.

Άρα

$$P(O \mid \lambda) = \sum_{\text{όλες οι δυνατές ακολουθίες } Q} P(O, Q \mid \lambda) * P(Q \mid \lambda)$$
$$= \sum_{q_1, q_2, \dots, q_T} \pi_{q_1} b_{q_1}(O_1)\, a_{q_1 q_2} b_{q_2}(O_2) \dots a_{q_{T-1} q_T} b_{q_T}(O_T) \quad (6)$$

Όπου από την εξίσωση (5) υπολογίζουμε για κάθε κατάσταση ξεχωριστά.

Η παραπάνω διαδικασία είναι αρκετά χρονοβόρα και δύσκολη υπολογιστικά καθώς υπολογίζει το πλήθος όλων των δυνατών καταστάσεων. Παρακάτω χρησιμοποιώντας μια δυναμική τεχνική μπορούμε να υπολογίσουμε την ζητούμενη πιθανότητα γρηγορότερα και πιο αποτελεσματικά. Πιο συγκεκριμένα, ο υπολογισμός της πιθανότητας $P(O \mid \lambda)$ σύμφωνα με τον παραπάνω τύπο απαιτεί $2TN^T$υπολογισμούς, αφού σε κάθε χρονική στιγμή $t = 1, 2, \dots, T$ υπάρχουν N πιθανές καταστάσεις (δηλαδή υπάρχουν N^T διαφορετικές ακολουθίες μήκους T) και για κάθε ακολουθία καταστάσεων απαιτούνται $2T$ υπολογισμοί για κάθε όρο του αθροίσματος της σχέσης (6).

Οι Εμπρός-Πίσω μέθοδοι

Έστω η βοηθητική εμπρός μεταβλητή$\alpha_t(i) = P(O_1 O_2 \dots O_t\ , q_t = S_i | \lambda)$,που είναι η πιθανότητα της μερικής ακολουθίας μέχρι τον χρόνο t, και στην χρονική στιγμή t να βρισκόμαστε στην κατάσταση S_i.

Η εμπρός μεταβλητή $\alpha_t(i)$ υπολογίζεται επαγωγικά ως εξής

1) $\alpha_1(i) = \pi_i b_i(O_1), 1 \le i \le N$
2) $\alpha_{t+1}(j) = \left(\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right) b_j(O_{t+1})\ , 1 \le j \le N, 1 \le t \le T-1$
3) $P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i)$

Το βήμα 1 υπολογίζει τις προς τα εμπρός πιθανότητες ως την κοινή πιθανότητας της κατάστασης S_iκαι της αρχικής παρατήρησης O_i.

Το βήμα 2 λαμβάνει υπόψιν πως η κατάσταση S_j είναι προσβάσιμη στον χρόνο $t+1$ από N διαφορετικές καταστάσεις,$S_i, 1 \le j \le N$, στον χρόνο t. Εφόσον $\alpha_t(i)$ είναι η κοινή πιθανότητα των γεγονότων ότι παρατηρείται η ακολουθία παρατηρήσεων $O_1 O_2 \dots O_t$, και η κατάσταση του συστήματος σε χρόνο t είναι S_i, τότε το γινόμενο $\alpha_t(i) a_{ij}$ είναι η κοινή πιθανότητα ότι η ακολουθία παρατηρήσεων $O_1 O_2 \dots O_t$ παρατηρείται και ότι το σύστημα σε χρόνο $t+1$ βρίσκεται στην κατάσταση S_j,ενώ την χρονική στιγμή t βρίσκονταν στην κατάσταση S_i. Αθροίζοντας αυτό το γινόμενο για όλες τις N πιθανές καταστάσεις S_i, $1 \le i \le N$,για χρόνο t, έχει ως αποτέλεσμα την πιθανότητα της κατάστασης S_j για χρόνο $t+1$ με όλες τις προηγούμενες παρατηρήσεις. Είναι εύκολο να παρατηρήσουμε ότι η πιθανότητα $\alpha_{t+1}(j)$ προκύπτει πολλαπλασιάζοντας το προηγούμενο άθροισμα με την πιθανότητα $b_j(O_{t+1})$. Ο υπολογισμός του βήματος 2 γίνεται για όλες τις καταστάσεις j, $1 \le j \le N$ για δεδομένο tκαι στη συνέχεια οι υπολογισμοί επαναλαμβάνονται για $t = 1,2, \dots, T-1$.

Τέλος στο βήμα 3 υπολογίζεται η επιθυμητή πιθανότητα $P(O \mid \lambda)$ ως το άθροισμα των προς τα εμπρός μεταβλητών $\alpha_T(i)$.

Αν εξετάσουμε τις υπολογιστικές απαιτήσεις για τους υπολογισμούς των $\alpha_t(j)$, $1 \le j \le N$, $1 \le t \le T$, βλέπουμε ότι απαιτούνται υπολογισμοί της τάξης N^2T αντί για $2TN^T$. Το πλεονέκτημα της παραπάνω δυναμικής διαδικασίας για την επίλυση του προβλήματος έγκειται στο επαγωγικό βήμα του υπολογισμού. Δηλαδή σε αντίθεση με την αρχική μέθοδο όσο μεγάλη και να είναι η ακολουθία παρατήρησης σε κάθε βήμα υπολογισμού μόνο το συνολικό πλήθος των καταστάσεων επηρεάζει την τελική τιμής της μεταβλητής $\alpha_t(i)$.

Με αντίστοιχο τρόπο ορίζουμε την πίσω μεταβλητή ως εξής. Έστω

$$\beta_t(i) = P(O_{t+1} O_{t+2} \dots O_T | q_t = S_i\ , \lambda)$$

Δηλαδή την πιθανότητα της μερικής ακολουθίας παρατηρήσεων από τον χρόνο $t+1$ έως T. Αντίστοιχα μπορεί να υπολογιστούν επαγωγικά οι τιμές $\beta_t(i)$ για κάθε κατάσταση.

1) $\beta_T(i) = 1, 1 \le i \le N$
2) $\beta_t(i) = \sum_{j=1}^{N} a_{ij} \beta_{t+1}(j)\, b_j(O_{t+1})\ , t = T-1, T-2, \dots\ 1,\ 1 \le i \le N$

Στο βήμα 1 ορίζονται οι πιθανότητες $\beta_T(i) = 1$ για όλα τα i. Στο βήμα 2 βλέπουμε ότι για να βρεθεί το σύστημα στην κατάσταση S_i σε χρόνο t και ,δεδομένης της ακολουθίας παρατηρήσεων από την χρονική στιγμή $t+1$ έως την χρονική στιγμή T, πρέπει να συνυπολογίσουμε όλες τις πιθανές καταστάσεις S_j για την χρονική στιγμή $t+1$ καθώς και την μετάβαση από την κατάσταση S_i στην κατάσταση S_j (δηλαδή η πιθανότητα μετάβασης a_{ij}) αλλά και την παρατήρηση O_{t+1} στην κατάσταση j (δηλαδή ο όρος $b(O_{t+1})$) και στη συνέχεια να υπολογίσουμε για την υπόλοιπη-μερική ακολουθία παρατηρήσεων από την κατάσταση j.

Πρόβλημα 2

Δοσμένης μίας ακολουθίας παρατηρήσεων $O = O_1O_2 \dots O_T$ και ενός μοντέλου $\lambda = (A, B, \pi)$, με ποιον τρόπο επιλέγουμε μία ακολουθία καταστάσεων $Q = q_1q_2 \dots q_T$ η οποία είναι βέλτιστη, δηλαδή ερμηνεύει με τον καλύτερο τρόπο τις παρατηρήσεις;

Λύνοντας το Πρόβλημα 2, προσπαθούμε να αποκαλύψουμε τα κρυφά μέρη του μοντέλου δηλαδή να βρούμε τη "σωστή" ακολουθία καταστάσεων. Πρέπει να σημειωθεί πως δεν υπάρχει "σωστή" ακολουθία καταστάσεων ,και για πρακτικούς λόγους χρησιμοποιείται ένα κριτήριο βελτιστοποίησης, ώστε το πρόβλημα να λυθεί όσο το δυνατόν καλύτερα. Υπάρχουν αρκετά διαφορετικά κριτήρια βελτιστοποίησης που είναι δυνατόν να χρησιμοποιήσουμε και, επομένως η επιλογή του κριτηρίου εξαρτάται σε μεγάλο βαθμό από τον σκοπό χρησιμοποίησης της ακολουθίας κρυφών καταστάσεων.

Λύση του Προβλήματος 2

Σε αντίθεση με το πρόβλημα 1, όπου μία ακριβής λύση μπορεί να δοθεί, υπάρχουν αρκετοί διαφορετικοί τρόποι για να λυθεί το πρόβλημα 2, δηλαδή να βρεθεί η "βέλτιστη" ακολουθία καταστάσεων, δεδομένης μίας ακολουθίας παρατηρήσεων. Η δυσκολία έγκειται στον ορισμό της βέλτιστης ακολουθίας καταστάσεων, δηλαδή, υπάρχουν αρκετά διαφορετικά κριτήρια. Για παράδειγμα, ένα κριτήριο βελτιστοποίησης είναι να επιλεγούν οι καταστάσεις q_t που είναι ξεχωριστά πιο πιθανές. Έτσι, μεγιστοποιείται ο αναμενόμενος αριθμός των σωστών ξεχωριστών καταστάσεων.

Ορίζουμε την συνάρτηση $\gamma_t(i) = P(q_t = S_i | O, \lambda)$, δηλαδή την πιθανότητα το σύστημα να βρίσκεται στην κατάσταση S_i την χρονική στιγμή t, δοσμένης της ακολουθίας παρατηρήσεων O και του μοντέλου λ.

Η συνάρτηση γ μπορεί να γραφεί χρησιμοποιώντας και τις προς τα εμπρός - προς τα πίσω μεταβλητές.

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O \mid \lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)}$$

Αφού το $\alpha_t(i)$ λαμβάνει υπόψιν την μερική ακολουθία παρατηρήσεων $O_1O_2 \dots O_t$ και την κατάσταση S_i την χρονική στιγμή t, ενώ το $\beta_t(i)$ την μερική ακολουθία παρατηρήσεων $O_tO_{t+1} \dots O_T$,δεδομένης της κατάστασης S_i την χρονική στιγμή t. Ο παράγοντας κανονικοποίησης $P(O \mid \lambda)$ κάνει την συνάρτηση γ ένα μέτρο πιθανότητας έτσι ώστε $\sum_{i=1}^{N} \gamma_t(i) = 1$.

Χρησιμοποιώντας την συνάρτηση γ μπορούμε να λύσουμε για τις ξεχωριστά πιο πιθανές καταστάσεις q_t

$$q_t = argmax[\gamma_t(i)], 1 \leq i \leq N, 1 \leq t \leq T$$

Αν και με αυτόν τον τρόπο μεγιστοποιείται ο αναμενόμενος αριθμός των σωστών καταστάσεων, δηλαδή διαλέγοντας για κάθε χρονική στιγμή t την πιο πιθανή κατάσταση, υπάρχουν κάποια προβλήματα όσον αφορά την ακολουθία καταστάσεων που προκύπτει. Για παράδειγμα, όταν ένα κρυφό μαρκοβιανό μοντέλο έχει πιθανότητες μετάβασης $a_{ij} = 0$, η "βέλτιστη" ακολουθία καταστάσεων μπορεί να μην είναι έγκυρη. Αυτό συμβαίνει γιατί με τον τρόπο που περιγράφηκε προηγουμένως,υπολογίζεται η πιο πιθανή κατάσταση για κάθε χρονική στιγμή χωρίς να λαμβάνουμε υπόψιν μας την πιθανότητα εμφάνισης ακολουθιών καταστάσεων.

Το πιο ευρέως διαδεδομένο κριτήριο είναι να βρούμε το μοναδικό καλύτερο μονοπάτι καταστάσεων, δηλαδή να μεγιστοποιήσουμε την $P(Q|O, \lambda)$ το οποίο είναι ισοδύναμο με το να μεγιστοποιήσουμε την $P(Q, O|\lambda)$. Ένας τέτοιος αλγόριθμος είναι ο αλγόριθμος Viterbi,που βασίζεται σε μεθόδους δυναμικού προγραμματισμού.

Για να βρούμε τη βέλτιστη ακολουθία καταστάσεων Q δοσμένης μίας ακολουθίας παρατηρήσεων O,ορίζουμε αρχικά την συνάρτηση δ ως εξής:

$$\delta_t(i) = maxP[q_1q_2 \dots q_t = i, O_1O_2 \dots, O_t|\lambda]$$

Δηλαδή $\delta_t(i)$ είναι η μέγιστη πιθανότητα για ένα μονοπάτι σε χρόνο t, το οποίο λαμβάνει υπόψιν τις πρώτες t παρατηρήσεις και τελειώνει στην κατάσταση S_i. Με επαγωγή έχουμε ότι: $\delta_{t+1}(i) = max[\delta_t(i)a_{ij}] * b_j(O_{t+1})$ (7)

Για να ανακτήσουμε την ακολουθία καταστάσεων πρέπει κρατάμε τα στοιχεία που μεγιστοποιούν την σχέση (7) για κάθε t και για κάθε j. Αυτό γίνεται με την διάταξη $\psi_t(j)$. Τα βήματα του αλγορίθμου είναι τα ακόλουθα:

1) $\delta_1(i) = \pi_i b_i(O_1), 1 \le i \le N$
   $\psi_1(i) = 0$
2) $\delta_t(i) = max[\delta_{t-1}(i)a_{ij}] * b_j(O_t), 1 \le j \le N, 2 \le t \le T$
   $\psi_t(j) = argmax[\delta_{t-1}(i)a_{ij}], 1 \le j \le N, 2 \le t \le T$
3) $P^* = max[\delta_T(i)]$
   $q^* = argmax[\delta_T(i)]$
4) $q_t^* = \psi_{t+1}(q_{t+1}^*), t = T - 1, T - 2, \dots, 1$

Πρόβλημα 3

Πως εκτιμώνται οι παράμετροι του μοντέλου $\lambda = (A, B, \pi)$ ώστε να μεγιστοποιηθεί η πιθανότητα $P(O \mid \lambda)$;

Στο πρόβλημα 3 προσπαθούμε να βελτιστοποιήσουμε τις παραμέτρους του μοντέλου έτσι ώστε να περιγράφουν με τον καλύτερο δυνατό τρόπο το πώς παρήχθη μία δοσμένη ακολουθία παρατηρήσεων. Η ακολουθία παρατηρήσεων που χρησιμοποιήθηκε για να προσαρμόσει τις παραμέτρους του μοντέλου ονομάζεται ακολουθία εκπαίδευσης, αφού χρησιμοποιείται για να "εκπαιδεύσει" το κρυφό μαρκοβιανό μοντέλο. Το πρόβλημα εκπαίδευσης είναι ιδιαίτερα κρίσιμο για τις εφαρμογές των κρυφών μαρκοβιανών μοντέλων, αφού μας επιτρέπει να προσαρμόσουμε τις παραμέτρους του μοντέλου με βέλτιστο τρόπο σε παρατηρούμενα δεδομένα, δηλαδή να δημιουργήσουμε το βέλτιστο μοντέλο για κάποιο φυσικό φαινόμενο.

Λύση του Προβλήματος 3

Το τρίτο και δυσκολότερο πρόβλημα σχετικά με τα κρυφά μαρκοβιανά μοντέλα είναι ο καθορισμός μίας μεθόδου προκειμένου να προσαρμοστούν οι παράμετροι $(\mathrm{A}, \mathrm{B}, \pi)$ώστε να μεγιστοποιηθεί η πιθανότητα της ακολουθίας παρατηρήσεων, δεδομένου του μοντέλου. Δεν υπάρχει γνωστή αναλυτική λύση για το μοντέλο που να μεγιστοποιεί την πιθανότητα της ακολουθίας παρατηρήσεων. Πρακτικά, δεδομένης οποιασδήποτε πεπερασμένης ακολουθίας παρατηρήσεων ως δεδομένων εκπαίδευσης, δεν υπάρχει βέλτιστος τρόπος για την εκτίμηση των παραμέτρων του μοντέλου. Μπορούμε ωστόσο να επιλέξουμε $\lambda = (\mathrm{A}, \mathrm{B}, \pi)$ τέτοια ώστε η πιθανότητα $\mathrm{P}(\mathrm{O}|\lambda)$ να μεγιστοποιηθεί τοπικά χρησιμοποιώντας μια επαναληπτική διαδικασία, όπως η μέθοδος Baum-Welch.

Για να περιγράψουμε τη διαδικασία επανεκτίμησης των παραμέτρων του κρυφού μαρκοβιανού μοντέλου, θα ορίσουμε αρχικά την

$$\xi_t(i,j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda)$$

δηλαδή την πιθανότητα να βρεθεί το μοντέλο στην κατάσταση $\mathrm{S_i}$ την χρονική στιγμή t και στην κατάσταση $\mathrm{S_j}$ τη χρονική στιγμή $\mathrm{t}+1$, δεδομένου του μοντέλου και της ακολουθίας παρατηρήσεων.

Σύμφωνα με τους ορισμούς των προς τα εμπρός και προς τα πίσω μεταβλητών, η $\xi_\mathrm{t}(\mathrm{i},\mathrm{j})$ μπορεί να γραφεί ως εξής:

$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(i)}{P(O|\lambda)} = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(i)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(i)}$$

Προηγουμένως ορίσαμε την συνάρτηση $\gamma_\mathrm{t}(\mathrm{i})$ ως την πιθανότητα να βρεθεί στην κατάσταση $\mathrm{S_i}$ την χρονική στιγμή t δεδομένης της ακολουθίας παρατηρήσεων και του μοντέλου, επομένως μπορούμε να γράψουμε την $\gamma_\mathrm{t}(\mathrm{i})$ χρησιμοποιώντας την $\xi_\mathrm{t}(\mathrm{i},\mathrm{j})$ αθροίζοντας ως προς j,δηλαδή $\gamma_\mathrm{t}(\mathrm{i}) = \sum_{\mathrm{j}=1}^{\mathrm{N}} \xi_\mathrm{t}(\mathrm{i},\mathrm{j})$.

Αν αθροίσουμε την $\gamma_\mathrm{t}(\mathrm{i})$ ως προς t έχουμε μία ποσότητα που μπορεί να ερμηνευτεί ως ο αναμενόμενος αριθμός των φορών που το σύστημα επισκέφτηκε την κατάσταση $\mathrm{S_i}$ ή ,ισοδύναμα, ο αναμενόμενος αριθμός μεταβάσεων που έγιναν από την κατάσταση $\mathrm{S_i}$. Όμοια αν αθροίσουμε την $\xi_\mathrm{t}(\mathrm{i},\mathrm{j})$ ως προς t (από $\mathrm{t}=1$ έως $\mathrm{t}=\mathrm{T}-1$), αυτό μπορεί να ερμηνευτεί ως ο αναμενόμενος αριθμός των μεταβάσεων από την κατάσταση $\mathrm{S_i}$ προς την κατάσταση $\mathrm{S_j}$.

$$\sum_{t=1}^{T-1} \gamma_t(i) = \text{αναμενόμενος αριθμός μεταβάσεων που έγιναν από την κατάσταση } S_i$$

$$\sum_{t=1}^{T-1} \xi_t(i,j) = \text{αναμενόμενος αριθμός μεταβάσεων από την κατάσταση } S_i \text{ προς την } S_j$$

Χρησιμοποιώντας τους παραπάνω τύπους, μπορεί να δοθεί μία μέθοδος για την επανεκτίμηση των παραμέτρων ενός κρυφού μαρκοβιανού συστήματος. Οι εκτιμήσεις για τις παραμέτρους είναι:

$$\overline{\pi_\iota} = αναμενόμενη\ συχνότητα\ για\ την\ κατάσταση\ S_i\ σε\ χρόνο\ (t=1) = \gamma_1(\iota)$$

$$\overline{a_{\iota J}} = \frac{αναμενόμενος\ αριθμός\ μεταβάσεων\ απο\ την\ κατάσταση\ S_i\ στην\ S_j}{αναμενόμενος\ αριθμός\ μεταβάσεων\ απο\ την\ κατάσταση\ S_i} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \xi_t(i,j)}$$

$$\overline{b_J}(k) = \frac{αναμενόμενος\ αριθμός\ φορών\ στην\ κατάσταση\ S_j\ και\ παρατήρησης\ του\ συμβόλου\ v_k}{αναμενόμενος\ αριθμός\ φορών\ στην\ κατάσταση\ S_j} = \frac{\sum_{t=1}^{T} \gamma_t(i),\ s.t. O_t = v_k}{\sum_{t=1}^{T} \gamma_t(i)}$$

Αν ορίσουμε το αρχικό μοντέλο $\lambda = (\mathrm{A}, \mathrm{B}, \pi)$ και $\bar{\lambda} = (\bar{\mathrm{A}}, \bar{\mathrm{B}}, \bar{\pi})$ να είναι το μοντέλο μετά την επανεκτίμηση των παραμέτρων, έχει αποδειχθεί από τον Baum ότι είτε το αρχικό μοντέλο λ ορίζει ένα κρίσιμο σημείο της συνάρτησης πιθανοφάνειας ,και σε αυτή την περίπτωση έχουμε $\lambda = \bar{\lambda}$, είτε το μοντέλο $\bar{\lambda}$ είναι πιο πιθανό από το μοντέλο λ με την έννοια ότι $\mathrm{P}(\mathrm{O}\,|\bar{\lambda}) > P(\mathrm{O} \mid \lambda)$, δηλαδή έχουμε βρει ένα καινούργιο μοντέλο $\bar{\lambda}$ από το οποίο η ακολουθία παρατηρήσεων είναι πιο πιθανό να έχει προκύψει.

Επαναλαμβάνοντας αυτή τη διαδικασία, με χρήση του $\bar{\lambda}$ στη θέση του λ, κάθε φορά βελτιώνουμε την πιθανότητα η ακολουθία παρατηρήσεων Ο να παραχθεί από το μοντέλο, μέχρι ένα οριακό σημείο. Το τελικό αποτέλεσμα της διαδικασίας επανεκτίμησης των παραμέτρων καλείται εκτίμηση μέγιστης πιθανοφάνειας του κρυφού μαρκοβιανού συστήματος. Πρέπει να σημειωθεί ότι ο αλγόριθμος οδηγεί σε τοπικό μέγιστο μόνο.

Μία σημαντική πτυχή της διαδικασίας επανεκτίμησης των παραμέτρων του μοντέλου είναι οι στοχαστικοί περιορισμοί στους οποίους υπόκεινται αυτές, δηλαδή

$$\sum_{i=1}^{N} \overline{\pi_\iota} = 1$$

$$\sum_{j=1}^{N} \overline{a_{\iota J}} = 1\,,\ \ 1 \le i \le N$$

$$\sum_{k=1}^{M} \overline{b_J}(k) = 1\,,\ \ 1 \le j \le N$$

Οι οποίοι ,όμως, ικανοποιούνται σε κάθε επανάληψη της μεθόδου.

Τύποι κρυφών μαρκοβιανών μοντέλων

Μέχρι τώρα λάβαμε υπόψιν μόνο τις ειδικές περιπτώσεις των εργοδικών ή πλήρως συνδεδεμένων κρυφών μαρκοβιανών μοντέλων στα οποία κάθε κατάσταση είναι προσβάσιμη από όλες τις υπόλοιπες καταστάσεις του μοντέλου. Ωστόσο, υπάρχουν άλλοι τύποι κρυφών μαρκοβιανών μοντέλων με ενδιαφέρουσες ιδιότητες που μοντελοποιούν καλύτερα τα δεδομένα σε διάφορες εφαρμογές.

Ένα τέτοιο μοντέλο είναι το αριστερά-δεξιά μοντέλο ή Bakis μοντέλο. Η κρυφή ακολουθία καταστάσεων που σχετίζεται με το μοντέλο αυτό έχει την ιδιότητα ότι όσο περνάει ο χρόνος ,ο δείκτης της κατάστασης μεγαλώνει (ή παραμένει ίδιος), δηλαδή η κατάσταση του μοντέλου εξελίσσεται από αριστερά προς τα δεξιά. Η βασική ιδιότητα όλων των αριστερά-δεξιά κρυφών μαρκοβιανών μοντέλων είναι ότι οι πιθανότητες μετάβασης για τις καταστάσεις έχουν την ακόλουθη ιδιότητα:

$$a_{ij} = 0, j < i$$

δηλαδή δεν επιτρέπεται η μετάβαση σε κάποια κατάσταση που έχει μικρότερο δείκτη από την κατάσταση που βρίσκεται εκείνη τη στιγμή το μοντέλο. Επιπλέον, η κατανομή πιθανότητας των αρχικών καταστάσεων έχει την εξής ιδιότητα:

$$\pi_i = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1' \end{cases}$$

αφού η ακολουθία καταστάσεων πρέπει να ξεκινήσει από την κατάσταση 1.

Εκτός από τους δομικούς περιορισμούς που αναπτύχθηκαν στην προηγούμενη παράγραφο ,ανάλογα με τη φύση του προβλήματος, μπορούμε να προσθέσουμε και άλλους περιορισμούς στο αριστερά-δεξιά μοντέλο. Ένας τέτοιος περιορισμός αφορά τις πιθανότητες μετάβασης των καταστάσεων του μοντέλου, έτσι ώστε να μην επιτρέπονται μεγάλες αλλαγές στον δείκτη των καταστάσεων. Ένας τέτοιος περιορισμός θα έχει τη μορφή:

$$a_{ij} = 0, j > i + \Delta$$

Για την τελευταία κατάσταση, στο αριστερά-δεξιά μοντέλο, οι πιθανότητες μετάβασης ορίζονται ως εξής:

$$a_{Nj} = \begin{cases} 0, & i < N \\ 1, & i = N \end{cases}$$

Τέλος, είναι σημαντικό να σημειωθεί ότι η εφαρμογή των περιορισμών στο αριστερά-δεξιά μοντέλο δεν επηρεάζουν την διαδικασία επανεκτίμησης των παραμέτρων. Αυτό συμβαίνει γιατί οποιαδήποτε παράμετρος του κρυφού μαρκοβιανού μοντέλου που έχει την τιμή 0 αρχικά θα παραμείνει 0 καθ’ όλη τη διαδικασία επανεκτίμησης.

Το βασικό πρόβλημα – Διάρκεια παραμονής στις καταστάσεις

Τα Κρυφά Μαρκοβιανά Μοντέλα (ΚΜΜ) είναι μια ιδιαίτερα σημαντική κατηγορία μοντέλων που συνδέονται με ένα ευρύ φάσμα εφαρμογών. Ωστόσο, είναι αρκετά περιοριστικό το γεγονός ότι ο χρόνος παραμονής του μοντέλου σε οποιαδήποτε κατάσταση ακολουθεί γεωμετρική κατανομή, όπως έχουμε αποδείξει και προηγουμένως :

$$p_i(d) = \textit{πιθανότητα } d \textit{ συνεχόμενων παρατηρήσεων στην κατάσταση } S_i = a_{ii}^{d-1}(1-a_{ii}) \quad (8)$$

Ακριβώς αυτός ο περιορισμός καθιστά τα ΚΜΜ ακατάλληλα να χρησιμοποιηθούν για να περιγράψουν ορισμένα μοντέλα. Για το σκοπό αυτό, θα χρειαστεί να επεκτείνουμε τα μοντέλα αυτά, ώστε να μην υπόκεινται στον παραπάνω περιορισμό και, ως εκ τούτου, να οδηγηθούμε σε μία ακριβέστερη και μαθηματικά ορθότερη περιγραφή για τις εφαρμογές εκείνες όπου τα ΚΜΜ υστερούν.

Οπότε εισάγουμε το Κρυφό ημι-Μαρκοβιανό Μοντέλο(ΚΗΜΜ), το οποίο διαφοροποιείται από το ΚΜΜ εφόσον επιτρέπουμε στην υποκείμενη διαδικασία να είναι μια ημι-Μαρκοβιανή αλυσίδα. Κάθε κατάσταση θα έχει μεταβλητή διάρκεια παραμονής, η οποία σχετίζεται με το πλήθος των παρατηρήσεων που παράγονται όσο παραμένουμε στην ίδια κατάσταση.

Τα ΚΗΜΜ έχουν εφαρμοστεί επιτυχώς σε διάφορες ερευνητικές περιοχές, με κυριότερη πεδίο εφαρμογής την αναγνώριση ομιλίας. Χαρακτηριστικά αναφέρουμε ότι χρησιμοποιήθηκαν στην αναγνώριση γονιδίων DNA, στα ηλεκτροκαρδιογραφήματα, στην αναγνώριση γραφικού χαρακτήρα, στην ανίχνευση θέσης πιθανών στόχων σε στρατιωτικές επιχειρήσεις, στην πρόβλεψη της πρωτεϊνικής δομής, στην ανίχνευση κίνησης κυτταρικών δικτύων, στην ανάλυση μοτίβων κατά τη διάρκεια άνθισης φυτών, στη μοντελοποίηση του Διαδικτύου, στη μοντελοποίηση πρόβλεψης του καιρού, στη μαγνητική τομογραφία MRI του εγκεφάλου, στη μοντελοποίηση χρηματοοικονομικών χρονοσειρών κ.α.

Οπότε, έχουμε γενικεύσει το ΚΜΜ σε ένα νέο μοντέλο μεταβλητής διάρκειας, ή χρόνου προσωρινής διαμονής, σε κάθε κατάσταση. Συνεπώς, η διάρκεια d είναι μία σαφώς ορισμένη τυχαία μεταβλητή για κάθε κατάσταση του ΚΗΜΜ. Η κύρια διαφοροποίηση μεταξύ των δύο μοντέλων είναι πως στο ΚΜΜ υποθέτουμε ότι λαμβάνουμε μία παρατήρηση σε κάθε κατάσταση, ενώ στο ΚΗΜΜ λαμβάνουμε μια ακολουθία παρατηρήσεων. Το πλήθος αυτών των παρατηρήσεων εξαρτάται από το χρόνο παραμονής στην κατάσταση, d.

Περιγράφουμε συνοπτικά, στα επόμενα 4 βήματα, τον τρόπο λειτουργίας του Κρυφού Ημι-Μαρκοβιανού Μοντέλου.

- Επιλέγεται μία αρχική κατάσταση, $\mathrm{q_1 = S_i}$, σύμφωνα με το αρχικό διάνυσμα $\mathrm{\pi_i}$.
- Επιλέγεται η διάρκεια $\mathrm{d_1}$, σύμφωνα με την συνάρτηση πυκνότητας της διάρκειας παραμονής $\mathrm{p_{q_1}(d_1)}$. (Υποθέτουμε ότι η συνάρτηση $\mathrm{p_q(d)}$ λαμβάνει μέγιστη διάρκεια παραμονής ίση με D).
- Η ακολουθία παρατηρήσεων $\mathrm{O_1 O_2 \dots O_{d_1}}$ επιλέγεται σύμφωνα με την από κοινού πυκνότητα $\mathrm{b_{q_1}(O_1 O_2 \dots O_{d_1})}$. Συνήθως υποθέτουμε ανεξαρτησία των παρατηρήσεων, οπότε:

$$b_{q_1}(O_1O_2 \dots O_{d_1}) = \prod_{t=1}^{d_1} b_{q_1}(O_t)$$

- Η επόμενη κατάσταση, $q_2 = S_j$, επιλέγεται σύμφωνα με τον πίνακα πιθανοτήτων μετάβασης, $a_{q_1q_2}$, με τον περιορισμό ότι $a_{q_2q_1} = 0$, δηλαδή δεν επιτρέπεται μετάβαση πίσω στην προηγούμενη κατάσταση. (Αυτός ο περιορισμός προκύπτει φυσιολογικά, εφόσον θεωρούμε ότι από την κατάσταση q_1 λαμβάνουμε ακριβώς d_1 παρατηρήσεις).

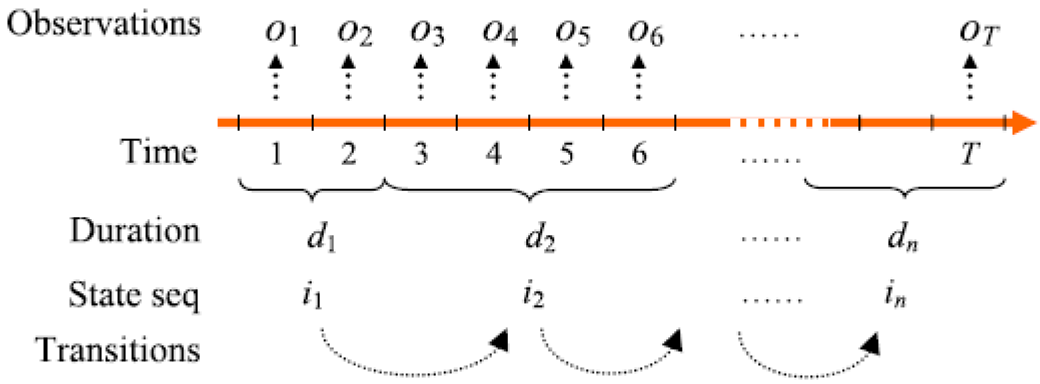

Figure2 Εποπτική περιγραφή του Κρυφού Ημι-Μαρκοβιανού Μοντέλου

<u>Σχόλιο</u>

Εύκολα μπορούμε να δούμε ότι το παραπάνω πιο σύνθετο μοντέλο ανάγεται στο κλασικό Κρυφό Μαρκοβιανό Μοντέλο που έχουμε περιγράψει ως τώρα, αν θεωρήσουμε ότι τα $p_q(d)$ κατανέμονται γεωμετρικά, όπως υποδηλώνεται από την (8).Στη συνέχεια, θα πρέπει να τροποποιηθούν οι τύποι που έχουν δοθεί ως τώρα, προκειμένου να υπολογιστεί η πιθανότητα $P(O|\lambda)$ καθώς και να επανεκτιμηθούν όλοι οι παράμετροι του νέου μοντέλου. Υποθέτουμε γενικά, ότι η πρώτη κατάσταση ξεκινά τη χρονική στιγμή $t = 1$, καθώς και ότι η τελευταία κατάσταση σταματά σε χρόνο $t = T$.

Οπότε ορίζουμε την προς τα εμπρός μεταβλητή ως

$$a_t(i) = P(O_1O_2 \dots O_t, (S_i \text{ λήγει στο χρόνο } t)|\lambda) \qquad (9)$$

Υποθέτουμε ότι επισκέπτονται συνολικά r καταστάσεις κατά τη διάρκεια των πρώτων t παρατηρήσεων, τις οποίες συμβολίζουμε με $q_1, q_2, \dots, q_r$ με αντίστοιχους χρόνους παραμονής $d_1, d_2, \dots, d_r$.

Συνεπώς, από την (9) προκύπτουν φυσιολογικά οι περιορισμοί:

$$q_r = S_i$$

$$\sum_{s=1}^{r} d_s = t$$

Οπότε η (9) μπορεί να γραφεί ως εξής:

$$a_t(i) = \sum_q \sum_d \pi_{q_1} p_{q_1}(d_1) P(O_1O_2 \dots O_{d_1}|q_1) a_{q_1q_2} p_{q_2}(d_2) P(O_{d_1+1}O_{d_1+2} \dots O_{d_1+d_2}|q_2) \dots$$

$$a_{q_{r-1}q_r} p_{q_r}(d_r) P(O_{d_1+d_2+\dots+d_{r-1}+1} \dots O_t|q_r)$$

Όπου η άθροιση γίνεται για όλες τις καταστάσεις q και όλες τις πιθανές διάρκειες d.

Οπότε επαγωγικά καταλήγουμε στην σχέση

$$a_t(j) = \sum_{i=1}^{N} \sum_{d=1}^{D} a_{t-d}(i)\, a_{ij} p_j(d) \prod_{s=t-d+1}^{t} b_j(O_s) \qquad (10)$$

Όπου D είναι η μέγιστη διάρκεια παραμονής σε οποιαδήποτε κατάσταση.
Για να δούμε αναλυτικότερα πώς λειτουργεί η αναδρομική σχέση (10), αρχικά υπολογίζουμε:

$$a_1(i) = \pi_i p_i(1) b_i(O_1)$$

$$a_2(i) = \pi_i p_i(2) \prod_{s=1}^{2} b_i(O_s) + \sum_{\substack{j=1 \\ j\neq i}}^{N} a_1(j) a_{ji}\, p_i(1) b_i(O_2)$$

$$a_3(i) = \pi_i p_i(3) \prod_{s=1}^{3} b_i(O_s) + \sum_{d=1}^{2} \sum_{\substack{j=1 \\ j\neq i}}^{N} a_{3-d}(j) a_{ji}\, p_i(d) \prod_{s=4-d}^{3} b_i(O_s)$$

και συνεχίζουμε ομοίως μέχρι να υπολογιστεί το $a_D(i)$. Έπειτα μπορούμε να χρησιμοποιήσουμε την (10) για όλα τα $t > D$. Ανάλογα με την περίπτωση των ΚΜΜ ,καταλήγουμε στην ιδιαίτερα απλή σχέση :

$$P(O|\lambda) = \sum_{i=1}^{N} a_T\,(i) \qquad (11)$$

Στο σημείο αυτό ,προκειμένου να κάνουμε επανεκτίμηση παραμέτρων, εισάγουμε 3 νέες μεταβλητές:

$$\alpha_t^*(i) = P(O_1 O_2 \ldots O_t, (S_i\ \xi\varepsilon\kappa\iota\nu\acute{\alpha}\ \sigma\varepsilon\ \chi\rho\acute{o}\nu o\ t+1)|\lambda)$$

$$\beta_t(i) = P(O_{t+1} O_{t+2} \ldots O_T |(S_i\ \lambda\acute{\eta}\gamma\varepsilon\iota\ \sigma\tau o\ \chi\rho\acute{o}\nu o\ t), \lambda)$$

$$\beta_t^*(i) = P(O_{t+1} O_{t+2} \ldots O_T |(S_i\ \xi\varepsilon\kappa\iota\nu\acute{\alpha}\ \sigma\varepsilon\ \chi\rho\acute{o}\nu o\ t+1)\ , \lambda)$$

Οι σχέσεις που συνδέονται οι παραπάνω μεταβλητές είναι οι εξής:

$$\alpha_t^*(j) = \sum_{i=1}^{N} a_t(i)\, a_{ij}$$

$$a_t(i) = \sum_{d=1}^{D} \alpha_{t-d}^*(i)\, p_i(d) \prod_{s=t-d+1}^{t} b_j(O_s)$$

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij}\, \beta_t^*(j)$$

$$\beta_t^*(i) = \sum_{d=1}^{D} \beta_{t+d}(i)\, p_i(d) \prod_{s=t+1}^{t+d} b_i(O_s)$$

Οπότε σύμφωνα με τις τελευταίες σχέσεις, οι τύποι που προκύπτουν για την επανεκτίμηση των παραμέτρων του μοντέλου, είναι :

$$\overline{\pi}_i = \frac{\pi_i \beta_0^*}{P(O|\lambda)} \quad (12a)$$

$$\overline{a_{ij}} = \frac{\sum_{t=1}^{T} a_t(i) a_{ij}\, \beta_t^*(j)}{\sum_{j=1}^{N} \sum_{t=1}^{T} a_t(i) a_{ij} \beta_t^*(j)} \quad (12b)$$

$$\overline{b_i}(k) = \frac{\sum^{T}_{\substack{t=1 \\ \varepsilon.\omega.O_t=k}} \left[\sum_{\tau<t} a_\tau^*(i)\beta_\tau^*(i) - \sum_{\tau<t} \alpha_\tau(i)\beta_\tau(i)\right]}{\sum_{k=1}^{M} \sum^{T}_{\substack{t=1 \\ \varepsilon.\omega.O_t=v_k}} \left[\sum_{\tau<t} a_\tau^*(i)\beta_\tau^*(i) - \sum_{\tau<t} \alpha_\tau(i)\beta_\tau(i)\right]} \quad (12c)$$

$$\overline{p_i}(d) = \frac{\sum_{t=1}^{T} \alpha_t^*(i) p_i(d) \beta_{t+d}(i) \prod_{s=t+1}^{t+d} b_i(O_s)}{\sum_{d=1}^{D} \left[\sum_{t=1}^{T} \alpha_t^*(i) p_i(d) \beta_{t+d}(i) \prod_{s=t+1}^{t+d} b_i(O_s)\right]} \quad (12d)$$

Η φυσική ερμηνεία των παραπάνω τύπων, και σε σύγκριση με τους αντίστοιχους στα ΚΜΜ, είναι :

- $\overline{\pi}_i$: η πιθανότητα ότι το μοντέλο ξεκινά με πρώτη κατάσταση την i, δεδομένης της ακολουθίας παρατηρήσεων O.
- $\overline{a_{ij}}$: ο αναμενόμενος αριθμός μεταβάσεων από την κατάσταση S_i στην S_j προς τον αναμενόμενο αριθμό συνολικών μεταβάσεων από την S_i, συμπεριλαμβάνοντας και τις περιπτώσεις όπου όταν οι $a -$ όροι για τους οποίους μία κατάσταση λήγει στο χρόνο t,
  συνδέονται με τους $\beta -$ όρους για τους οποίους μία νέα κατάσταση ξεκινά στο χρόνο $t + 1$.
- $\overline{b_i}(k)$: το αναμενόμενο πλήθος εμφανίσεων της παρατήρησης $O_t = v_k$ στην κατάσταση i,
  προς το αναμενόμενο πλήθος εμφάνισης οποιασδήποτε παρατήρησης στην κατάσταση i.
- $\overline{p_i}(d)$: το αναμενόμενο πλήθος των φορών που το σύστημα βρέθηκε στην κατάσταση i με διάρκεια παραμονής d, προς το αναμενόμενο πλήθος φορών που το σύστημα βρέθηκε στην κατάσταση i για οποιαδήποτε διάρκεια παραμονής.

Viterbi αλγόριθμος για ΚΗΜΜ

Ορίζουμε $S_{1:T} = S_1 S_2 \dots S_T$ όπου $S_t \in [1, N]$ είναι η κατάσταση στον χρόνο t. Αντίστοιχα ορίζουμε $O_{1:T} = O_1 O_2 \dots O_T$ όπου $O_t \in [1, M]$ είναι η παρατήρηση στον χρόνο t. Θέλουμε να βρούμε το μοναδικό καλύτερο μονοπάτι καταστάσεων $S_{1:T}$ που να αντιστοιχεί στην ακολουθία παρατηρήσεων $O_{1:T}$, δηλαδή μεγιστοποιεί την πιθανότητα $P[S_{1:T}|O_{1:T}, \lambda]$. Οι προς τα εμπρός πιθανότητες για τον αλγόριθμο Viterbi στα ΚΗΜΜ θα οριστούν ως εξής:

$$\delta_t(j,d) \equiv \begin{Bmatrix} max \\ S_{1:t-d} \end{Bmatrix} P[S_{1:t-d}, S_{t-d+1:t} = j, O_{1:T}|\lambda]$$

$$= \begin{Bmatrix} max \\ i \neq j \end{Bmatrix} \begin{Bmatrix} max \\ h \end{Bmatrix} \begin{Bmatrix} max \\ S_{1:t-d-h} \end{Bmatrix} P[S_{1:t-d-h}, S_{t-d-h+1,t-d} = i, S_{t-d+1:t} = j, O_{1:T}|\lambda]$$

$$= \begin{Bmatrix} max \\ i \neq j \end{Bmatrix} \begin{Bmatrix} max \\ h \end{Bmatrix} \delta_{t-d}(i,h) \alpha_{(i,h)(j,d)} b_{j,d}(O_{t-d+1:t})$$

για $1 \leq t \leq T$, $j \in S$, $d \in D$. Η ποσότητα $\delta_t(j,d)$ αντιπροσωπεύει την πιο πιθανή μερική ακολουθία καταστάσεων που τελειώνει την χρονική στιγμή t στη κατάσταση j και με διάρκεια d. Καταγράφουμε την προηγούμενη κατάσταση την οποία επιλέγει η $\delta_t(j,d)$ με $\psi(t,j,d) \equiv (t-d, i^*, h^*)$ όπου i^* είναι η προηγούμενη κατάσταση, h^* η χρονική της διάρκεια και $t-d$ ο χρόνος τερματισμού της. Η $\psi(t,j,d)$ βρίσκεται λαμβάνοντας υπ' όψιν

$$(i^*, h^*) = {}^{arg} \begin{Bmatrix} max \\ i \in S \backslash \{j\} \end{Bmatrix} \begin{Bmatrix} max \\ h \in D \end{Bmatrix} \{\delta_{t-d}(i,h) \alpha_{(i,h)(j,d)} b_{j,d}(O_{t-d+1:t})\}$$

Επομένως η ακολουθία καταστάσεων μπορεί να προσδιοριστεί βρίσκοντας την τελευταία κατάσταση που μεγιστοποιεί την πιθανοφάνεια. Για την γενική υπόθεση των οριακών συνθηκών η τελευταία κατάσταση είναι:

$$(t_1, j_1^*, d_1^*) = {}^{arg} \begin{Bmatrix} max \\ T+D-1 \geq t \geq T \end{Bmatrix} \begin{Bmatrix} max \\ j \in S \end{Bmatrix} \begin{Bmatrix} max \\ D \geq d \geq t-T+1 \end{Bmatrix} \delta_t(j,d)$$

ή για την απλοποιημένη υπόθεση των οριακών συνθηκών, $t_1 = T$ και

$$(j_1^*, d_1^*) = {}^{arg} \begin{Bmatrix} max \\ j \in S \end{Bmatrix} \begin{Bmatrix} max \\ d \in D \end{Bmatrix} \delta_T(j,d)$$

ιχνηλατούμε την ακολουθία καταστάσεων προς τα πίσω ως εξής:

$$(t_2, j_2^*, d_2^*) = \psi(t_1, j_1^*, d_1^*)$$

$$\dots$$

$$(t_n, j_n^*, d_n^*) = \psi(t_{n-1}, j_{n-1}^*, d_{n-1}^*)$$

μέχρι να οριστεί η πρώτη κατάσταση S_1, όπου $S_1 = j_n^*$ και $(j_n^*, d_n^*), \dots, (j_1^*, d_1^*)$ είναι η εκτιμώμενη ακολουθία καταστάσεων.

Μέχρι τώρα έχουμε δει τον τρόπο αντιμετώπισης του προβλήματος μεταβλητής διάρκειας παραμονής στις καταστάσεις του συστήματος. Η μοντελοποίηση διάφορων προβλημάτων διευκολύνεται σε σημαντικό βαθμό με τη χρήση των παραπάνω αποτελεσμάτων.

Ωστόσο, προκύπτουν και ορισμένα μειονεκτήματα αυτής της θεώρησης. Ένα μεγάλο τέτοιο μειονέκτημα, είναι πως αυξάνεται σημαντικά το πλήθος των υπολογισμών που χρειάζονται. Για παράδειγμα, μπορούμε να δούμε από τον τρόπο ορισμού της $\alpha_t(i)$, ότι δεσμεύουμε D φορές την μνήμη και χρειάζονται $\frac{D^2}{2}$ υπολογισμοί.

Ωστόσο, έχουν μελετηθεί διάφορες τεχνικές όπου μπορούμε να μειώσουμε το υπολογιστικό κόστος, και σε συγκεκριμένες περιπτώσεις ασυμπωτικά το κόστος υπολογισμού να συμπίπτει με εκείνο του συμβατικού ΚΜΜ. Οι μέθοδοι αυτοί ουσιαστικά εστιάζουν στον υπολογισμό των παραπάνω αθροισμάτων και γινομένων των αναδρομικών τύπων μόνο μια φορά στην αρχή του προβλήματος και στην αποθήκευση έπειτα για μελλοντική χρήση.

Ένα άλλο μειονέκτημα της παραπάνω μεθόδου είναι πως, για προκαθορισμένο πλήθος παρατηρήσεων T, το πλήθος των μεταβάσεων του συστήματος από μία κατάσταση σε οποιαδήποτε άλλη είναι κατά πολύ μικρότερο, γεγονός που συνεπάγεται ότι υπάρχουν λιγότερα δεδομένα για να εκτιμηθεί η πιθανότητα $\mathrm{p_i(d)}$. Αυτό σημαίνει πως γενικότερα το πρόβλημα της εκτίμησης των παραμέτρων του ΚΗΜΜ είναι πιο δύσκολο συγκριτικά με το συμβατικό ΚΜΜ.

Προκειμένου να εξαλειφθούν ή τουλάχιστον να περιοριστούν τα παραπάνω προβλήματα, θα αντιμετωπίσουμε το πρόβλημα της διάρκειας παραμονής στις καταστάσεις του ΚΗΜΜ όπως αυτό παρουσιάστηκε παραπάνω, χρησιμοποιώντας συγκεκριμένες παραμετρικές συναρτήσεις πυκνότητας της διάρκειας κατάστασης, σε αντίθεση με τις μη-παραμετρικές $\mathrm{p_i(d)}$ που χρησιμοποιήθηκαν ως τώρα.

Στη βιβλιογραφία προτείνονται διάφορες προσεγγίσεις για την αντιμετώπιση του συγκεκριμένου προβλήματος, ορισμένες από τις οποίες θα αναλύσουμε στις επόμενες σελίδες.

Διακριτή Coxian κατανομή

Ένα πρώτο παράδειγμα κρυφής ημιμαρκοβιανής αλυσίδας είναι η Coxian διακριτή κατανομή, όπου η πιθανότητα παραμονής $\mathrm{p_i(d)}$ακολουθεί $\mathrm{Cox(\mu,\theta)}$όπου μ,θ οι παράμετροι της κατανομής.

Η Coxian διακριτή κατανομή περιγράφεται μέσω μαρκοβιανής αλυσίδας όπου η πιθανότητα να γυρίσουμε μελλοντικά σε προηγούμενη κατάσταση είναι μηδενική και αν $\mu = \mu_1, \mu_2, \dots, \mu_N$ και $\theta = \theta_1, \theta_2, \dots, \theta_N$ όπου το πλήθος N εκφράζει τις φάσεις (καταστάσεις) της αλυσίδας. Ακόμα, η πιθανότητα σε κάθε φάση n να παραμείνουμε στην ίδια, δίνεται από τον τύπο $\mathrm{A_{nn}} = 1 - \theta_{\mathrm{n}}$με $\mathrm{n} = \{1,2,\dots,\mathrm{N}\}, \theta_{\mathrm{n}} \in (0,1]$.

Η χρονική διάρκεια παραμονής σε κάθε συγκεκριμένη φάση ακολουθεί γεωμετρική κατανομή $\mathrm{X_n} \sim \mathrm{Geom}(\theta_n)$. Επομένως, αν για παράδειγμα η μαρκοβιανή μας αλυσίδα ξεκινούσε από μια φάση (κατάσταση) κ με αρχική πιθανότητα μ_κ, τότε ο συνολικός χρόνος της μαρκοβιανής αλυσίδας θα δινόταν από την ποσότητα $X_\kappa + X_{\kappa+1} + \cdots X_N$ για κάποιο $\kappa \in \{1,2,\dots,N\}$ και για $0 \le \mu_\kappa \le 1$, $\sum_\kappa \mu_\kappa = 1$.

Στην δική μας περίπτωση θεωρούμε πως για κάθε κατάσταση i , οι πιθανότητες παραμονής $p_i(d)$ ακολουθούν $Cox(\mu^i, \theta^i)$ όπου μ^i, θ^i αποτελούν το σύνολο των παραμέτρων για κάθε κατάσταση $i \in \{1,2,\dots,N\}$. Όταν $\theta^i = 1$ για όλα τα n τότε ταυτίζεται με το συνήθη

Φυσιολογικά μπορούμε να θεωρήσουμε ότι το ζεύγος (i,n) είναι μία κατάσταση ΚΜΜ όπου $i \in \{1,2,\dots,N\}$ είναι η κατάσταση του ΚΗΜΜ και $n \in \{1,2,\dots,N\}$ είναι η φάση της Coxian κατανομής. Επομένως οι κλασικοί προς τα εμπρός-πίσω μέθοδοι μπορούν να χρησιμοποιηθούν για να απαντήσουμε στα γνωστά μας προβλήματα.
Για μια κατάσταση λοιπόν, (i,n) διακρίνουμε τις επόμενες περιπτώσεις:

1) είτε παραμένει στάσιμη με πιθανότητα ${A_{n,n}}^{(i)} = 1 - {\theta_n}^{(i)}$
2) είτε μετακινείται στην κατάσταση $(i, n+1)$με πιθανότητα

$$ {A_{n,n+1}}^{(i)} = {\theta_n}^{(i)}, \;\; n = 1,2, \dots, N-1 $$

3) είτε μετακινείται σε μια κατάσταση (j, n') αν $n = N$ με πιθανότητα $a_{ij}{\mu_{n'}}^{(j)}$

Μια περιγραφή της προς τα εμπρός μεταβλητής του αντίστοιχου αλγορίθμου, όταν το μοντέλο είναι όπως περιγράψαμε παραπάνω και ξεκινά πάντα από την φάση 1 και τελειώνει σε οποιαδήποτε φάση n ,είναι :

$$ \alpha_t(i,n,d) \begin{cases} \sum_{j\in S}\sum_{n'=1}^{N}\sum_{d'\in D} \alpha_{t-1}(j,n',d')p_j(d')a_{ij}b_{i1}(O_t) & n=1, d=1 \\ \alpha_{t-1}(i,1,d-1){A_{1,1}}^{(i)}b_{i1}(O_t) & n=1, d>1 \\ \sum_{n'=n-1}^{N} \alpha_{t-1}(i,n',d-1){A_{n,n}}^{(i)}b_{in}(O_t) & n>1, d>1 \\ \alpha_t(i,n,d) = 0 & n>1, d=1 \end{cases} $$

Όπου $p_i(d)$ είναι η πιθανότητα διάρκειας, $b_{in}(O_t)$ η πιθανότητα παρατήρησης, ${A_{n',n}}^{(i)}$ η πιθανότητα μετάβασης από την φάση (i,n') στη φάση (i,n) και $\alpha_t(i,n,d)$ η προς τα εμπρός μεταβλητή σε χρόνο t όταν η κατάσταση είναι η i, η φάση είναι n και η διάρκεια της κατάστασης i είναι d.

Σε αυτό το σημείο, μπορούμε να εκμεταλλευτούμε τον τροποποιημένο αλγόριθμο Viterbi, προκειμένου να μελετήσουμε μια ειδική κατηγορία κατανομών.

Κυρτά μονότονες κατανομές

Ο νέος διαφοροποιημένος αλγόριθμος Viterbi δίνει πως η νέα μέγιστη πιθανότητα ανά μονοπάτι δίνεται από τον τύπο :

$$ \delta_t(j) = \max_{d\in D} \max_{i\in S\setminus\{j\}} \left\{ \delta_{t-d}(i)a_{ij}p_j(d) \prod_{k=t-d+1}^{t} b_j(O_t) \right\} \qquad (13) $$

Έστω πως η κατανομή της πιθανότητας παραμονής $p_j(d)$ ικανοποιεί την συνθήκη της κοίλας μονοτονίας, δηλαδή :

$$ c_1p_j(d_1) \leq c_2p_j(d_2) \Rightarrow c_1p_j(d_1+\tau) \leq c_2p_j(d_2+\tau) \qquad \forall\tau \geq 0, $$

και σταθερές c_1, c_2 θετικές.

Αν λοιπόν στην δική μας περίπτωση ισχύει για $d_1 > d_2$

$$ \delta_{t-d_1}(i_1)a_{i_1j}p_j(d_1) \prod_{k=t-d_1+1}^{t} b_j(O_t) \leq \delta_{t-d_2}(i_2)a_{i_2j}p_j(d_2) \prod_{k=t-d_2+1}^{t} b_j(O_t) \; (*) \Rightarrow $$

$$ \delta_{t-d_1}(i_1)a_{i_1j}p_j(d_1+\tau) \prod_{k=t-d_1+1}^{t} b_j(O_t) $$

$$ \leq \delta_{t-d_2}(i_2)a_{i_2j}p_j(d_2+\tau) \prod_{k=t-d_2+1}^{t} b_j(O_t) \qquad (**) $$

Διαισθητικά αυτό σημαίνει πως για μια δεδομένη κατάσταση j, αν γνωρίζουμε πως οι παραμονές σε συγκεκριμένο χρόνο που διαρκούν περισσότερο $(d_1 > d_2)$ έχουν μικρότερη πιθανότητα$(*)$ τότε θα έχουν πάντα μικρότερη πιθανότητα όταν οι παραμονές προεκτείνονται σε διάρκεια $(**)$.

Το γεγονός αυτό μας παρέχει την δυνατότητα να μειώσουμε σημαντικά των αριθμό των στοιχείων της (13) που θα ελεγχθούν πραγματικά για να βρεθεί η ποσότητα $\delta_t(j)$. Με άλλα λόγια η ποσότητα $\delta_{t-d_1}(i_1)$ δε θα μπορούσε ποτέ να αποτελέσει λύση στο μέλλον, ανεξάρτητα της εξέλιξης της ακολουθίας μας.

Η προσέγγιση του Levinson μέσω της Γάμμα κατανομής

Στην παρούσα ενότητα, παρουσιάζεται η προσέγγιση κατά την οποία θα χρησιμοποιήσουμε συνεχή συνάρτηση πυκνότητας πιθανότητας για να περιγράψουμε το μοντέλο που θα έχει μεταβλητή διάρκεια παραμονής στην κάθε κατάσταση, σε χρόνο συνεχή. Συγκεκριμένα, θα χρησιμοποιηθεί η Γάμμα κατανομή, μία δι-παραμετρική κατανομή, η μέση τιμή και η διασπορά της οποίας προσδιορίζονται πλήρως από τις 2 παραμέτρους της.

Οπότε για τον χρόνο παραμονής $p_i(d)$ στην κατάσταση i, έχουμε ότι:

$$p_i(d) = \frac{\eta_i^{\nu_i}}{\Gamma(\nu_i)} d^{\nu_i - 1} e^{-\eta_i d} \ , \quad \tau > 0$$

Για τη μέση τιμή και διασπορά της Γάμμα γνωρίζουμε ότι :

$$\mu = \frac{\nu_i}{\eta_i}, \ \ \sigma^2 = \frac{\nu_i}{{\eta_i}^2}$$

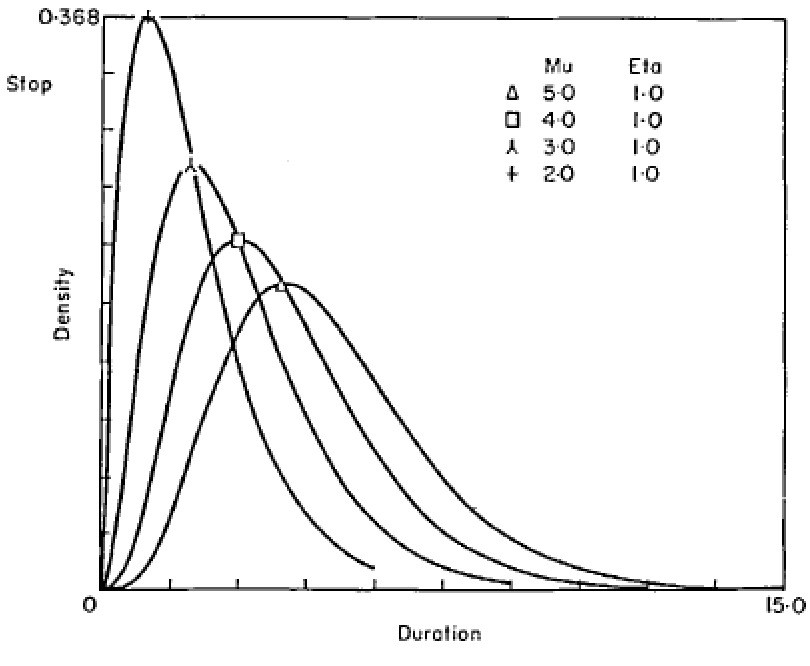

Figure3Η κατανομή Γάμμα G(ν,η) για σταθερή παράμετρο η=1

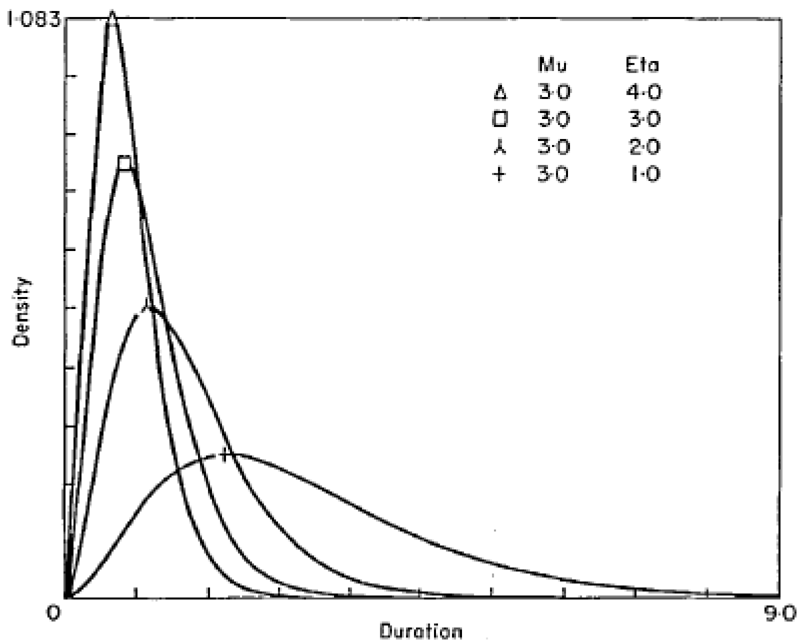

Figure4 Η Γάμμα κατανομή G(ν,η) για σταθερή παράμετρο ν=3

Από τον τρόπο κατασκευής του μοντέλου, για τον πίνακα πιθανοτήτων μετάβασης, a_{ij}, πρέπει να ισχύει ότι $a_{ii} = 0, \forall$ i $, 1 \leq i \leq n$, εφόσον ορίζεται σαφής χρόνος παραμονής στην κατάσταση i και δεν επιτρέπεται να μεταβούμε μελλοντικά σε προηγούμενη κατάσταση.

Ακόμη, υποθέτουμε πως τα $b_i(x)$, όπως έχουν οριστεί εξ αρχής, ακολουθούν πολυδιάστατες κανονικές κατανομές με μέση τιμή μ_i και πίνακα συνδιασπορών U_i, για κάθε $1 \leq i \leq n$ και κάθε $x \in \mathbb{R}^m$.

Συνεπώς, η πιθανότητα $P(O|\lambda)$ ορίζεται για κάθε ακολουθία παρατηρήσεων $O_1, O_2, \dots, O_T$, όπου $O_t \in \mathbb{R}^m$, για κάθε $1 \leq t \leq T$ και για κάθε λ τέτοιο ώστε:

$$\lambda = (A, \{\mu_i\}_{i=1}^n, \{U_i\}_{i=1}^n, \{\nu_i\}_{i=1}^n, \{\eta_i\}_{i=1}^n)$$

Στο σημείο αυτό, χρειαζόμαστε τις προς τα εμπρός και προς τα πίσω μεταβλητές που ορίστηκαν προηγουμένως, από όπου καταλήξαμε στη σχέση (11) :

$$P(O|\lambda) = \sum_{i=1}^{N} a_T\,(i)$$

Απ' όπου πλέον μπορούμε να υπολογίσουμε τα $a_T(i)$ με τον τρόπο που έχει περιγραφεί, μέσω της αναδρομικής σχέσης :

$$a_t(j) = \sum_{i=1}^{N}\sum_{d=1}^{D} a_{t-d}(i)\, a_{ij} p_j(d) \prod_{s=t-d+1}^{t} b_j(O_s)$$

Δηλαδή, δεδομένου του μοντέλου λ, του αρχικού διανύσματος π_i και αντικαθιστώντας την συνάρτηση $p_j(d)$ με τη συνάρτηση Γάμμα, μπορούμε πλέον να υπολογίσουμε το ζητούμενο $P(O|\lambda)$.

Εκτίμηση παραμέτρων

Η εκτίμηση των βασικών παραμέτρων του μοντέλου γίνεται ακριβώς με τον τρόπο που περιγράψαμε στις σχέσεις $(12a) - (12d)$

Οπότε απομένει η εκτίμηση των παραμέτρων της παρατηρούμενης διαδικασίας, δηλαδή των μ_j και U_j, καθώς των παραμέτρων της κατανομής, δηλαδή των ν_iκαι η_i. Στο σημείο αυτό, ξαναγράφουμε τους τύπους για τον υπολογισμό των προς τα εμπρός και προς τα πίσω μεταβλητών, που υπολογίστηκαν νωρίτερα, τροποποιώντας ελαφρώς τον συμβολισμό για λόγους διευκόλυνσης των πράξεων. Έτσι, θα έχουμε ότι :

$$a_t(j) = \sum_{\tau \le t} \sum_{\substack{i=1 \\ i \ne j}}^{n} a_{t-\tau}(i) a_{ij} p_j(\tau) \prod_{\theta=1}^{\tau} b_j(O_{t-\tau+\theta}), \quad 1 \le i \le n, 1 \le t \le T$$

$$\beta_t(i) = \sum_{\tau \le T-1} \sum_{\substack{j=1 \\ j \ne i}}^{n} a_{ij} p_j(\tau) \prod_{\theta=1}^{\tau} b_j(O_{t+\theta})\, \beta_{t+\tau}(j), \qquad 1 \le i \le n, 1 \le t \le T-1$$

Οριακές συνθήκες

Θεωρούμε ότι $a_0(1) = 1$ και $a_0(i) = 0$ για $i \ne 1$, δηλαδή ότι στον χρόνο $t = 0$ η διαδικασία ξεκινά από την κατάσταση q_1

Ομοίως, ο υπολογισμός των β ξεκινά θεωρώντας ότι $\beta_T(i) = 1$ για $1 \le i \le n$ Τελικά, συνδυάζοντας τις 2 παραπάνω σχέσεις καταλήγουμε στον υπολογισμό της πιθανότητας για οποιαδήποτε ακολουθία παρατηρήσεων $O = O_1, O_2, \dots, O_T$

$$P(O|\lambda) = \sum_{t=1}^{T} \sum_{\substack{j=1 \\ j \ne i}}^{n} \sum_{\tau \le t} a_{t-\tau}(i) a_{ij} p_j(\tau) \prod_{\theta=1}^{\tau} b_j(O_{t-\tau+\theta})\, \beta_t(j) \qquad (14)$$

Για την εκτίμηση της παραμέτρου μ_j, αρκεί να βρούμε πότε μηδενίζεται η $\frac{\partial P}{\partial \mu_j}$

$$\frac{\partial P}{\partial \mu_j} = \sum_{t=1}^{T} \sum_{\substack{i=1 \\ i \ne j}}^{n} \sum_{\tau \le t} a_{t-\tau}(i) a_{ij} p_j(\tau) \left(\frac{\partial}{\partial \mu_j} \left(\prod_{\theta=1}^{\tau} b_j(O_{t-\tau+\theta}) \right) \right) \beta_t(j) \qquad (15)$$

Λόγω της κανονικότητας που έχουμε υποθέσει, δηλαδή ότι $b_j(x) = N(\mu_j, U_j)$, παίρνουμε:

$$\frac{\partial}{\partial \mu_j} \left(\prod_{\theta=1}^{\tau} b_j(O_{t-\tau+\theta}) \right) = \sum_{r=1}^{\tau} \left(\prod_{\substack{\theta=1 \\ \theta \ne r}}^{\tau} b_j(O_{t-\tau+\theta}) \right) \frac{\partial N}{\partial \mu_j} \bigg| O_{t-\tau+\theta} =$$

$$\sum_{r=1}^{\tau} \left(\prod_{\substack{\theta=1 \\ \theta \ne r}}^{\tau} b_j(O_{t-\tau+\theta}) \right) b_j(O_{t-\tau+r}) U_j^{-1} (O_{t-\tau+r} - \mu_j)^T$$

$$= \left(\prod_{\theta=1}^{\tau} b_j(O_{t-\tau+\theta}) \right) \left(\sum_{\theta=1}^{\tau} U_j^{-1} (O_{t-\tau+\theta} - \mu_j)^T \right)$$

Οπότε αντικαθιστώντας στην (15), πολλαπλασιάζοντας με U_j^{-1} και θέτοντας το αποτέλεσμα ίσο με 0, έχουμε:

$$\sum_{t=1}^{T}\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{\tau\leq t} a_{t-\tau}(i)a_{ij}p_j(\tau)\left(\prod_{\theta=1}^{\tau} b_j(O_{t-\tau+\theta})\right)\left(\sum\nolimits_{\theta=1}^{\tau} O_{t-\tau+\theta}\right)\beta_t(j)=0$$

Όμως $\sum_{\theta=1}^{\tau} O_{t-\tau+\theta}=\tau\mu_j$, οπότε

$$\sum_{t=1}^{T}\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{\tau\leq t} a_{t-\tau}(i)a_{ij}p_j(\tau)\left(\prod_{\theta=1}^{\tau} b_j(O_{t-\tau+\theta})\right)(\tau\mu_j)\beta_t(j)=0\Rightarrow$$

$$\overline{\mu_j}=\frac{\sum_{t=1}^{T}\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{\tau\leq t} a_{t-\tau}(i)a_{ij}p_j(\tau)\left(\prod_{\theta=1}^{\tau} b_j(O_{t-\tau+\theta})\right)(\sum_{\theta=1}^{\tau} O_{t-\tau+\theta})\beta_t(j)}{\sum_{t=1}^{T}\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{\tau\leq t} \tau a_{t-\tau}(i)a_{ij}p_j(\tau)\left(\prod_{\theta=1}^{\tau} b_j(O_{t-\tau+\theta})\right)\beta_t(j)} \quad (16)$$

Για την εκτίμηση του U_j, αρχικά παρατηρούμε ότι :

$$\left.\frac{\partial N}{\partial U_j}\right| O_t=b_j(O_t)\left[-\frac{1}{2}U_j^{-1}-\frac{1}{2}(O_t-\mu_j)U_j^{-1}(U_j^{-1})^T(O_t-\mu_j)^T\right]$$

Στη συνέχεια δουλεύοντας ακριβώς με τον ίδιο τρόπο όπως και προηγουμένως μπορούμε να καταλήξουμε στο ζητούμενο, δηλαδή ότι :

$$\overline{U_j}=\frac{\sum_{t=1}^{T}\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{\tau\leq t} a_{t-\tau}(i)a_{ij}p_j(\tau)\left(\prod_{\theta=1}^{\tau} b_j(O_{t-\tau+\theta})\right)\left(\sum_{\theta=1}^{\tau}(O_{t-\tau+\theta}-\mu_j)(O_{t-\tau+\theta}-\mu_j)^T\right)\beta_t(j)}{\sum_{t=1}^{T}\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{\tau\leq t} \tau a_{t-\tau}(i)a_{ij}p_j(\tau)\left(\prod_{\theta=1}^{\tau} b_j(O_{t-\tau+\theta})\right)\beta_t(j)}$$

Για την εκτίμηση της παραμέτρου η_j της κατανομής, αρχικά παραγωγίζοντας την συνάρτηση πυκνότητας της Γάμμα, παίρνουμε:

$$\frac{\partial p_j(\tau)}{\partial \eta_j}=\frac{\partial}{\partial \eta_j}\left(\frac{\eta_j^{\nu_j}}{\Gamma(\nu_j)}\tau^{\nu_j-1}e^{-\eta_j\tau}\right)=\nu_j\frac{\eta_j^{\nu_j-1}}{\Gamma(\nu_j)}\tau^{\nu_j-1}e^{-\eta_j\tau}-\tau\frac{\eta_j^{\nu_j}}{\Gamma(\nu_j)}\tau^{\nu_j-1}e^{-\eta_j\tau}$$
$$=p_j(\tau)\left(\frac{\nu_j}{\eta_j}-\tau\right)$$

Ακολουθώντας άλλη μια φορά τα βήματα που ακολουθήσαμε στον υπολογισμό του $\overline{\mu_j}$, μπορούμε να καταλήξουμε στην παρακάτω σχέση εκτίμησης της η_j

$$\overline{\eta_j}=\frac{\nu_j\sum_{t=1}^{T} a_t(j)\beta_t(j)}{\sum_{t=1}^{T}\sum_{\tau\leq t}\tau\sum_{\substack{i=1\\i\neq j}}^{n} a_{t-\tau}(i)a_{ij}p_j(\tau)\left(\prod_{\theta=1}^{\tau} b_j(O_{t-\tau+\theta})\right)\beta_t(j)}$$

Τέλος, για την εκτίμηση της παραμέτρου ν_j έχει αποδειχθεί ότι, σε αντίθεση με τις υπόλοιπες παραμέτρους, δεν υπάρχει αναλυτικός τρόπος υπολογισμού εύρεσης ενός αντίστοιχου τύπου. Το πρόβλημα μπορεί να γίνει άμεσα αντιληπτό, από την παραγώγιση της $p_j(\tau)$ ως προς ν_j αυτή τη φορά. Για το λόγο αυτό, χρειάζεται να εισάγουμε μία νέα συνάρτηση, τη συνάρτηση ψ ή δίγαμμα συνάρτηση, όπως εμφανίζεται συχνά στη βιβλιογραφία

Έτσι, δουλεύοντας με παρόμοιο τρόπο όπως και για τις υπόλοιπες παραμέτρους, αποδεικνύεται ότι μπορούμε να φτάσουμε στο ακόλουθο αποτέλεσμα :

$$\psi(\nu_j) = \frac{\sum_{t=1}^{T}\sum_{\tau\leq 1} log\ (\eta_j\tau)\sum_{\substack{i=1\\ i\neq j}}^{n} a_{t-\tau}(i)a_{ij}p_j(\tau)\left(\prod_{\theta=1}^{\tau} b_j(O_{t-\tau+\theta})\right)\beta_t(j)}{\sum_{t=1}^{T} a_t(j)\beta_t(j)} \tag{17}$$

Η συνάρτηση ψ

Ορίζουμε την συνάρτηση ψ να είναι η λογαριθμική παράγωγος της Γάμμα, δηλαδή

$$\psi(x) = (log\ (\Gamma(x)\ )' = \frac{\Gamma'(x)}{\Gamma(x)}$$

Η ψ είναι συνεχώς διαφορίσιμη και γνησίως αύξουσα στο $\mathbb{R}^+$, με πραγματική ρίζα κοντά στο $x \approx 1.461$. Ακόμη ισχύουν οι εξής σχέσεις:

$$\psi(x+n) = \sum_{k=1}^{n}\frac{1}{x+k} + \psi(x+1), \quad x \in \mathbb{R}^+ \tag{18α}$$

$$\psi(x+1) = -\gamma + \sum_{k=1}^{\infty}(-1)^{k+1}\zeta(k+1)x^k\,, \quad x \in (-1,1) \tag{18β}$$

Όπου γ είναι η σταθερά των $Euler - Mascheroni$, $\gamma \approx 0.577$, και ζ είναι η συνάρτηση ζήτα του Riemann.

Οπότε θα προσεγγίσουμε την ζητούμενη εκτίμηση $\overline{\nu_j}$ με αριθμητική μέθοδο. Συγκεκριμένα εφαρμόζουμε την μέθοδο του Newton , δηλαδή

$$\nu_j^{(k+1)} = \nu_j^{(k)} - \frac{\psi\left(\nu_j^{(k)}\right) - C}{\psi'\left(\nu_j^{(k)}\right)}$$

Όπου με (k) συμβολίζουμε την k-οστή επανάληψη του αλγορίθμου, ενώ η σταθερά C είναι ίση με το δεξί μέλος της σχέσης (16). Τέλος, ο παρονομαστής μπορεί να υπολογιστεί παραγογίζοντας της σχέσεις (17α) και (17β). Η παραπάνω μέθοδος έχει αποδειχτεί ότι λόγω της καλής συμπεριφοράς της ψ στο $\mathbb{R}^+$,συγκλίνει πολύ γρήγορα με ιδιαίτερα μεγάλη ακρίβεια στη ζητούμενη εκτίμηση $\overline{\nu_j}$.

Σχόλιο

Προκειμένου να είναι πραγματικά χρήσιμοι οι τύποι που έχουμε δώσει για την εκτίμηση όλων των παραμέτρων παραπάνω, θα πρέπει να αποδειχθεί ότι όντως αυξάνουν την πιθανότητα $P(O|\lambda)$. Χωρίς να δώσουμε πλήρη απόδειξη, θα σκιαγραφήσουμε γιατί αυτό πράγματι συμβαίνει.

Αρχικά ορίζουμε τη συνάρτηση $g(\lambda,\overline{\lambda}) = P(O|\lambda)log\ (P(O|\overline{\lambda})$. Έχει αποδειχτεί ότι η ορθότητα της εκτίμησης προκύπτει εφόσον το $\overline{\lambda}$ είναι κρίσιμο σημείο για την g, και εφόσον η συνάρτηση πυκνότητας πιθανότητας f, εν προκειμένω η συνάρτηση πυκνότητας πιθανότητας της Γάμμα, είναι $log - κοίλη$ στο λ. Φυσικά, το $\overline{\lambda}$ είναι κρίσιμο σημείο γιατί έτσι το κατασκευάσαμε.

Κάνοντας τους απαραίτητους υπολογισμούς, για όλες τις παραμέτρους εκτός από τα ν_j, παρατηρούμε ότι

$\frac{\partial^2}{\partial\lambda^2}\left(log\left(f(\lambda)\right)\right) < 0$, δηλαδή ικανοποιείται το κριτήριο

Για $\lambda = v_j$,
$\frac{\partial^2}{\partial v_j^2} log\left(p_j(\tau)\right) = -\frac{\partial\psi}{\partial v_j} < 0$, αφού είδαμε ότι $\frac{\partial\psi}{\partial v_j} = \sum_{k=1}^{\infty} \frac{1}{(v_j+k)^2} > 0$, δηλαδή ικανοποιείται το κριτήριο.

Οικογένεια εκθετικών κατανομών

Στην παραπάνω ανάλυση, μπορεί να παρατηρήσει κανείς ότι δεν χρησιμοποιείται αυτή κάθε αυτή η συνάρτηση πυκνότητας της Γάμμα κατανομής. Ουσιαστικά, εκμεταλλευτήκαμε μόνο το γεγονός ότι η συνάρτηση πυκνότητας πιθανότητας της Γάμμα έχει την ιδιότητα να είναι log-κοίλη. Αυτή η ιδιότητα ισχύει για ένα σύνολο συναρτήσεων πυκνότητας πιθανότητας, το οποίο ονομάζουμε οικογένεια εκθετικών κατανομών, και στο οποίο ανήκουν πολλές από τις γνωστές μας κατανομές, όπως είναι η Bernoulli (με γνωστό πλήθος επαναλήψεων), η Poisson, η εκθετική, η κανονική, η χ-τετράγωνο ,η Γάμμα ,η Βήτα κλπ.

Ορισμός

Δεδομένου ενός μέτρου m, ορίζουμε την οικογένεια εκθετικών κατανομών να είναι εκείνες οι κατανομές των οποίων η συνάρτηση πυκνότητας πιθανότητας έχει την ακόλουθη μορφή:

$$p(x|\eta) = h(x)e^{\eta^T T(x) - A(\eta)}$$

Όπου η φυσική παράμετρος της κατανομής, h είναι μια συνάρτηση του x, $T(x)$ είναι μία ποσότητα που ονομάζεται επαρκές στατιστικό και η $A(\eta)$ είναι ο λογάριθμος της ροπογεννήτριας συνάρτησης

$$A(t) = log\big(E(e^{tX})\big)$$

Παραθέτουμε το ακόλουθο σημαντικό θεώρημα για τον χαρακτηρισμό ενός στατιστικού ως επαρκές.

Θεώρημα (Παραγοντοποίηση Fisher-Neyman)

Έστω η συνάρτηση πυκνότητας πιθανότητας $f_\theta(x)$. Το T θα είναι *επαρκές στατιστικό* για την παράμετρο θ αν και μόνο αν υπάρχουν μη αρνητικές συναρτήσεις g και h έτσι ώστε:

$$f_\theta(x) = h(x)g_\theta\big(T(x)\big)$$

Δηλαδή η $f_\theta(x)$ 'σπάει' σε ένα γινόμενο δύο συναρτήσεων, όπου η πρώτη δεν εξαρτάται από το θ, ενώ η εξάρτηση δεύτερη, η οποία εξαρτάται από το θ, εξαρτάται από το x μόνο μέσω του $T(x)$.

Παρατήρηση

Η φυσική παράμετρος η είναι συνάρτηση των παραμέτρων θ της κατανομής. Θα χρησιμοποιούμε η αντί για $\eta(\theta)$ για λόγους απλοποίησης του συμβολισμού.

Ακόμη, το πεδίο ορισμού της φυσικής παραμέτρου υπόκειται σε περιορισμούς, οποίοι προκύπτουν από το μέτρο m. Άναφέρουμε μόνο ότι το χωρίο Ω στο οποίο ορίζεται η η είναι:

$$\Omega = \left\{\eta \colon \int e^{\eta(\theta)T(x)}h(x)dm < \infty\right\}$$

όπου ,εν γένει, με το σύμβολο του ολοκληρώματος υπονοείται το ολοκλήρωμα $Lebesgue$

Χωρίς να επεκταθούμε στην θεωρητική ανάλυση που συνδέεται με τις παραπάνω έννοιες, δίνουμε στη συνέχεια παραδείγματα ορισμένων γνωστών κατανομών, υπολογίζοντας τις ποσότητες που αναφέρθηκαν παραπάνω, αλλά και επιβεβαιώνοντας τα θεωρητικά αποτελέσματα για την κατανομή Γάμμα που προηγήθηκαν.

- $Bernoulli$ κατανομή

Γνωρίζουμε ότι $p(x|\lambda) = \lambda^x(1-\lambda)^{1-x} = e^{log\left(\frac{\lambda}{1-\lambda}\right)x + log\,(1-\lambda)}$
Δηλαδή η κατανομή $Bernoulli$ ανήκει στην οικογένεια εκθετικών κατανομών με :
$\eta = log\left(\frac{\lambda}{1-\lambda}\right)$
$T(x) = x$
$A(\eta) = -\,log(1-\lambda) = log(1+e^{\eta})$
$h(x) = 1$

- $Poisson$ κατανομή

Γνωρίζουμε ότι $p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{1}{x!}e^{xlog\lambda - \lambda}$
Δηλαδή η κατανομή $Poisson$ ανήκει στην οικογένεια εκθετικών κατανομών με :
$\eta = log\lambda$
$T(x) = x$
$A(\eta) = \lambda = e^{\eta}$
$h(x) = \frac{1}{x!}$

- Κανονική κατανομή

Γνωρίζουμε ότι $p(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(x-\mu)^2} = \frac{1}{\sqrt{2\pi}}e^{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \frac{1}{2\sigma^2}\mu^2 - log\sigma}$
Δηλαδή η κανονική κατανομή ανήκει στην οικογένεια εκθετικών κατανομών με :
$\eta = \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix}$
$T(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$
$A(\eta) = \frac{\mu^2}{2\sigma^2} + log\sigma = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}log(-2\eta_2)$
όπου η_1 και η_2 τα στοιχεία της 1ης και 2ης γραμμής του πίνακα η αντίστοιχα.
$h(x) = \frac{1}{\sqrt{2\pi}}$

- Γάμμα κατανομή

Γνωρίζουμε ότι $p(x|\alpha,\beta) = \frac{\beta^\alpha x^{\alpha-1}e^{-\beta x}}{\Gamma(\alpha)} = e^{(\alpha-1)logx - \beta x - \left(log\left(\Gamma(\alpha)\right) - \alpha log\beta\right)}$
Δηλαδή η Γάμμα κατανομή ανήκει στην οικογένεια εκθετικών κατανομών με :

$$\eta = \begin{bmatrix} \alpha - 1 \\ -\beta \end{bmatrix}$$

$$T(x) = \begin{bmatrix} \log x \\ x \end{bmatrix}$$

$$A(\eta) = \log\big(\Gamma(\alpha)\big) - \alpha\log\beta = \log\big(\Gamma(\eta_1 + 1)\big) - (\eta_1 + 1)\log(-\eta_2)$$

$$h(x) = 1$$

<u>ΒΙΒΛΙΟΓΡΑΦΙΑ</u>

[1] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," in *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257-286, Feb. 1989, doi: 10.1109/5.18626.

[2] S. Levinson, "Continuously variable duration hidden Markov models for speech analysis," *ICASSP '86. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Tokyo, Japan, 1986, pp. 1241-1244, doi: 10.1109/ICASSP.1986.1168801.

[3] L. Liporace, "Maximum likelihood estimation for multivariate observations of Markov sources," in *IEEE Transactions on Information Theory*, vol. 28, no. 5, pp. 729-734, September 1982, doi: 10.1109/TIT.1982.1056544

[4] Yu, Shun-Zheng. (2010). Hidden semi-Markov models. Artif. Intell.. 174. 215-243. 10.1016/j.artint.2009.11.011.

[5] C. D. Mitchell and L. H. Jamieson, "Modeling duration in a hidden Markov model with the exponential family," *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Minneapolis, MN, USA, 1993, pp. 331-334 vol.2, doi: 10.1109/ICASSP.1993.319304.

[6] Shun-Zheng Yu. 2015. Hidden Semi-Markov Models: Theory, Algorithms and Applications (1st. ed.). Elsevier Science Publishers B. V., NLD.

[7]https://people.eecs.berkeley.edu/~jordan/courses/260-spring10/other-readings/chapter8.pdf (date of access 23/12/2020)