

Πιθανότητες και Στοχαστικές Διαδικασίες

Γ. Πάσχος

1 Οκτωβρίου 2011

Περιεχόμενα

1	Βασικές Έννοιες Πιθανοθεωρίας	7
1.1	Χώρος Πιθανοτήτων	7
1.2	Υπό συνθήκη πιθανότητα	10
1.3	Ιδιότητες Πιθανοτήτων	11
1.4	Τυχαία μεταβλητή	12
1.4.1	Κατανομή	12
1.4.2	Ροπές	13
1.5	Χαρακτηριστική Συνάρτηση	14
1.6	Διακριτές τ.μ.	14
1.6.1	Συνάρτηση δείκτης	17
1.6.2	Κατανομή Bernoulli	17
1.6.3	Κατανομή Binomial	17
1.6.4	Κατανομή Geometric	18
1.6.5	Κατανομή Poisson	19
1.7	Συνεχείς τ.μ.	20
1.7.1	Ομοιόμορφη Κατανομή (Uniform)	22
1.7.2	Κατανομή Gauss	23
1.7.3	Εκθετική κατανομή	24
1.7.4	Μίξη ΣΠΠ	25
1.7.5	Θετικές μεταβλητές	25
1.8	Συναρτήσεις τυχαίων μεταβλητών	25
1.8.1	Γραμμικός μετασχηματισμός τ.μ.	26
1.8.2	Μέθοδος Jacobian	26
1.8.3	Μέθοδος κατανομής	27
2	Πολλές μεταβλητές	29
2.1	Δύο τυχαίες μεταβλητές	29
2.1.1	Για τις διακριτές τ.μ.	30
2.1.2	Για τις συνεχείς τ.μ.	31
2.2	Κατανομές υπό συνθήκη	32
2.2.1	Για τις διακριτές τ.μ.	32
2.2.2	Για τις συνεχείς τ.μ.	34
2.2.3	Η περίπτωση μιας διακριτής και μιας συνεχούς τ.μ.	35
2.3	Πολλές τυχαίες μεταβλητές	35
2.3.1	Ανεξαρτησία	35
2.3.2	Ανάλυση με υπό συνθήκη κατανομές	36
2.3.3	Ροπές πολλών μεταβλητών	37

2.3.4	Συναρτήσεις πολλών μεταβλητών	38
2.4	Τυχαία διανύσματα	43
2.4.1	Γραμμικοί μετασχηματισμοί	44
2.4.2	Διαγωνοποίηση πινάκων	45
2.4.3	Το διάνυσμα Gauss	46
2.5	Μιγαδικές τυχαίες μεταβλητές	48
2.5.1	Διάνυσμα μιγαδικών μεταβλητών	48
2.5.2	Μιγαδικό διάνυσμα Gauss	49
2.6	Εξομοίωση τ.μ.	49
2.7	Προβλήματα στοχαστικής γεωμετρίας	50
2.7.1	Buffon's needle	50
2.7.2	Επιλογή σημείων στο δίσκο (Disk point picking)	51
3	Σύγκλιση ακολουθιών	53
3.1	Ανισότητες	53
3.1.1	Ανισότητα Markov	53
3.1.2	Ανισότητα Chebyshev	54
3.1.3	Φράγμα Chernoff	54
3.1.4	Ανισότητα Jensen	55
3.1.5	Ανισότητα Cauchy–Schwarz	55
3.2	Σύγκλιση τ.μ.	55
3.2.1	Σύγκλιση γενικά	56
3.2.2	Βέβαια σύγκλιση	56
3.2.3	Σύγκλιση με πιθανότητα 1	57
3.2.4	Σύγκλιση με την έννοια της τετραγωνικής μέσης τιμής	57
3.2.5	Σύγκλιση σε πιθανότητα	58
3.2.6	Σύγκλιση σε κατανομή	59
3.3	Δειγματική μέση τιμή	59
3.4	Νόμος των μεγάλων αριθμών	61
3.5	Κεντρικό οριακό θεώρημα	62
3.5.1	Προσέγγιση Gauss	62
3.6	Εκτίμηση	63
3.6.1	Μέσο τετραγωνικό σφάλμα	63
3.6.2	Εκτίμηση της X	63
3.6.3	Εκτίμηση της X δεδομένης της $Y = y$	63
3.6.4	Ιδιότητες σφάλματος εκτίμησης	64
3.6.5	Εκτίμηση της X δεδομένων πολλών μετρήσεων Y_1, Y_2, \dots, Y_n	65
4	Στοχαστικές διαδικασίες	67
4.1	Γενικά	67
4.1.1	Ροπές διαδικασιών	69
4.1.2	Πολλές διαδικασίες	70
4.2	Στασιμότητα	71
4.2.1	Στασιμότητα με την ευρεία έννοια (WSS)	72
4.2.2	Ιδιότητες αυτοσυσχέτισης για WSS διαδικασίες	72
4.2.3	Φασματική ισχύς	74
4.2.4	Κυκλική στασιμότητα	77
4.3	Εργοδικότητα	79

4.4	Μαρκοβιανές διαδικασίες	80
4.5	Διαδικασίες διακριτού χρόνου	80
4.5.1	i.i.d. διαδικασία	80
4.5.2	Διαδικασίες απαρίθμησης	81
4.6	Διαδικασίες συνεχούς χρόνου	85
4.6.1	i.i.d. Gauss	85
4.6.2	Γενική διαδικασία Gauss	85
4.6.3	Διαδικασία Wiener	86
5	Στοχαστική διαδικασία Poisson	89
5.1	Ορισμός της διαδικασίας Poisson	89
5.1.1	Πρώτος ορισμός ως διαδικασία σημείων	89
5.1.2	Δεύτερος ορισμός ως διαδικασία σημείων	90
5.1.3	Τρίτος ορισμός ως διαδικασία απαρίθμησης	90
5.2	Ιδιότητες της διαδικασίας Poisson	90
5.2.1	Βασικές ιδιότητες	90
5.2.2	Δέσμευση ως προς k	92
5.2.3	Εναπόθεση (superposition) δύο διαδικασιών Poisson	92
5.2.4	Λέπτυνση (thinning) διαδικασίας Poisson	93
5.2.5	Ιδιότητα PASTA	94
5.3	Άλλες μορφές της διαδικασίας Poisson	94
5.3.1	Μη ομογενής	94
5.3.2	Μείγμα Poisson	95
5.3.3	Χωρική διαδικασία Poisson	96
5.3.4	Τυχαίο σήμα τηλεγράφου	96
5.3.5	Shot noise	98
6	Διαδικασίες Markov	101
6.1	Αλυσίδες Markov διακριτού χρόνου	102
6.1.1	Εισαγωγή	102
6.1.2	Στάσιμη κατανομή - equilibrium	103
6.1.3	Κατηγοριοποίηση καταστάσεων και αλυσίδων	105
6.1.4	Βασικά θεωρήματα για τις αλυσίδες Markov διακριτού χρόνου	106
6.1.5	Στασιμη ανάλυση - πως λύνουμε μια εργοδική DTMC	107
6.1.6	Μεταβατική ανάλυση - Χρόνοι πρώτης άφιξης	108
6.2	Αλυσίδες Markov συνεχούς χρόνου	111
6.2.1	Εισαγωγή	111
6.2.2	Κατηγοριοποίηση CTMC	111
6.2.3	Εξισώσεις Chapman-Kolmogorov για τις CTMC	111
6.2.4	Στάσιμη ανάλυση	113
6.2.5	Μεταβατική ανάλυση	115
6.2.6	Διαδικασία Γεννήσεων-Θανάτων	115

Κεφάλαιο 1

Βασικές Έννοιες Πιθανοθεωρίας

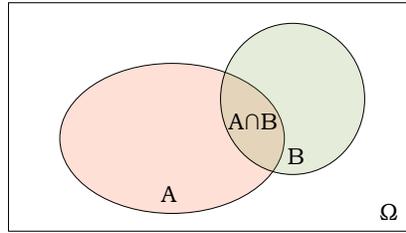
1.1 Χώρος Πιθανοτήτων

Ως **δειγματικό χώρο** (Sample space) Ω ορίζουμε το σύνολο με όλα τα πιθανά στοιχεία ω (δηλ. αποτελέσματα) ενός τυχαίου πειράματος· ο δειγματικός χώρος είναι ένα μη κενό σύνολο.

Παράδειγμα:

- Ρίχνω ένα κέρμα, $\Omega = \{Γ, Κ\}$.
- Ρίχνω ένα ζάρι, $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Αριθμός πελατών στην ουρά της τράπεζας, $\Omega = \{0, 1, \dots\}$.
- Διάρκεια μιας τηλεφωνικής κλήσης, $\Omega = \{\mathbb{R} : \omega \geq 0\}$.

Ο δειγματικός χώρος πρέπει να είναι *εξαντλητικός*, δηλαδή να καλύπτει όλα τα αποτελέσματα του πειράματος, και να έχει *αμοιβαία αποκλειόμενα στοιχεία*, με άλλα λόγια κάθε πείραμα να έχει ένα μοναδικό αποτέλεσμα. Παρόλα αυτά, ο δειγματικός χώρος για ένα πείραμα δεν είναι μοναδικός. Για παράδειγμα, για το πείραμα του ζαριού μπορούμε να χρησιμοποιήσουμε ως δειγματικό χώρο το συνεχές διάστημα $[0, 6]$ χωρισμένο σε έξι διαστήματα ίσου μήκους, ή το διακριτό σύνολο $\{1, 2, 3, 4, 5, 6\}$.



Σχήμα 1.1: Διάγραμμα Venn με γεγονότα-υποσύνολα του χώρου Ω .

Γεγονός (Event) A ονομάζουμε ένα υποσύνολο του Ω .

Παράδειγμα:

- Να έρθει το κέρμα γράμματα, $A = \{\Gamma\} \subset \Omega$.
- Να φέρω με το ζάρι άρτιο αριθμό, $A = \{2, 4, 6\}$.
- Οι πελάτες στην ουρά να είναι περισσότεροι από 5, $A = \{6, 7, \dots\}$.
- Η διάρκεια μιας τηλεφωνικής κλήσης να είναι αρνητική, $A = \{\omega \in \Omega : \omega < 0\} = \emptyset$.

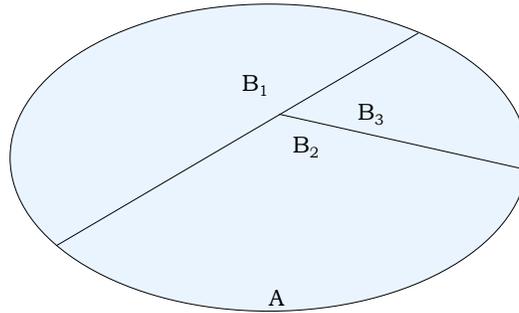
Το \emptyset συμβολίζει το κενό σύνολο, ή ισοδύναμα, το αδύνατο γεγονός. Το γεγονός Ω είναι το βέβαιο γεγονός. Ορίζουμε τα παρακάτω γεγονότα ως συναφή με τα γεγονότα A και B :

- Το συμπληρωματικό $A^c = \{\omega \in \Omega : \omega \notin A\}$,
- ένωση $A \cup B = \{\omega \in \Omega : \omega \in A \text{ ή } \omega \in B\}$,
- τομή $A \cap B = \{\omega \in \Omega : \omega \in A \text{ και } \omega \in B\}$.
- Αν τα γεγονότα, A και B είναι *ασυμβίβαστα* τότε $A \cap B = \emptyset$ και
- αν τα σύνολα $B_i, i = 1, 2, \dots$ αποτελούν διαμέριση του A τότε ισχύει $B_i \cap B_j = \emptyset$ και $\cup_i B_i = A$.

Πεδίο (Field) \mathcal{F} ονομάζουμε το σύνολο των γεγονότων-υποσυνόλων του Ω που είναι κλειστό σε πράξεις αριθμήσιμων συνόλων (συμπληρωματικό, ένωση, τομή). Το σύνολο αυτό πρέπει πάντα να περιέχει το αδύνατο και το βέβαιο γεγονός, \emptyset και Ω αντίστοιχα, και για κάθε δύο γεγονότα A και B , όλα τα προκύπτοντα γεγονότα $A^c, B^c, A \cap B, A \cup B$.

Μέτρο πιθανότητας (Probability Measure) μ είναι μια συνάρτηση που αναθέτει σε κάθε στοιχείο ενός συνόλου έναν αριθμό. Γνωστά παραδείγματα είναι το μήκος, η επιφάνεια κ.α. Εδώ μας ενδιαφέρει το μέτρο της πιθανότητας P που αναθέτει σε κάθε γεγονός έναν αριθμό στο διάστημα $[0, 1]$, άρα $P : \mathcal{F} \rightarrow [0, 1]$. Το τρίπτυχο Ω, \mathcal{F}, P ονομάζεται χώρος πιθανοτήτων (Probability space).

Το μέτρο πιθανότητας ορίζεται με τις παρακάτω ιδιότητες:



Σχήμα 1.2: Τα ασυμβίβαστα γεγονότα B_1, B_2, B_3 αποτελούν διαμέριση του A .

$$P(\emptyset) = 0$$

$$P(\Omega) = 1$$

$$P(A^c) = 1 - P(A)$$

$$A \subset B \Rightarrow P(A) \leq P(B)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C)$$

$$0 \leq P(A) \leq 1$$

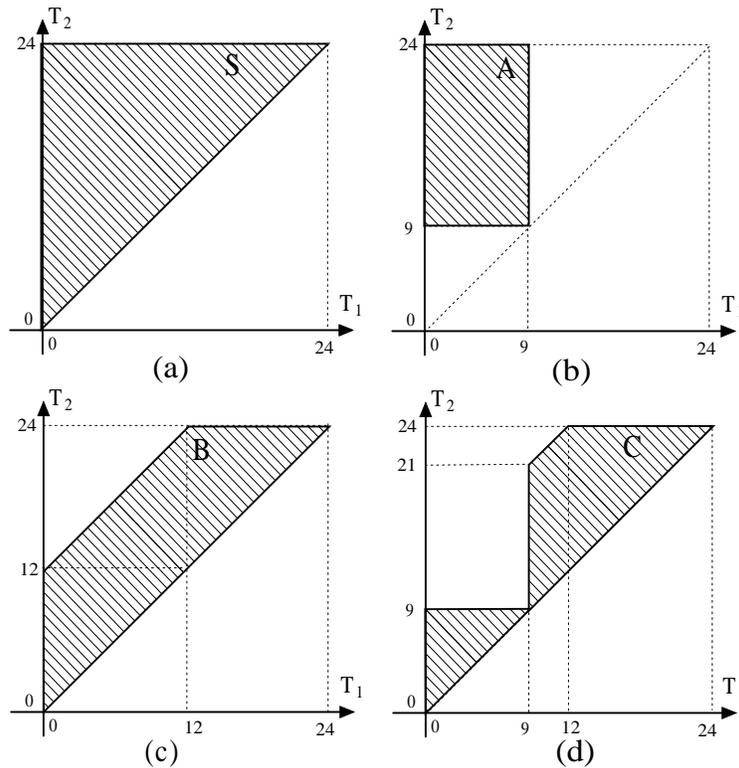
Ερώτηση: Αν A και B είναι ανεξάρτητα γεγονότα, τι συμβαίνει με την πιθανότητα να συμβούν ταυτόχρονα;

Παράδειγμα: (Φοιτητικό ωράριο) Σε ένα 24-ωρο, ένας φοιτητής ξυπνά τη χρονική στιγμή T_1 και κοιμάται τη χρονική στιγμή T_2 , όπου $0 \leq T_1 < T_2 \leq 24$.

1. Βρείτε το δειγματικό χώρο Ω και σχεδιάστε τον στο επίπεδο $x - y$, αν τα αποτελέσματα του πειράματος αποτελούνται από όλες τις πιθανές δυάδες (T_1, T_2) .
2. Προσδιορίστε το ενδεχόμενο A ο φοιτητής να είναι ξύπνιος στις 9 π.μ. και σχεδιάστε το στο επίπεδο.
3. Προσδιορίστε το ενδεχόμενο B ο φοιτητής να κοιμάται για περισσότερη ώρα από την ώρα που είναι ξύπνιος εντός του 24-ώρου, και σχεδιάστε το στο επίπεδο.
4. Σχεδιάστε την περιοχή που αντιστοιχεί στο $C = A^c \cap B$ και περιγράψτε το αντίστοιχο ενδεχόμενο με λόγια.

Λύση:

1. $\Omega = \{(T_1, T_2) : 0 \leq T_1 < T_2 \leq 24\}$. Ο δειγματικός χώρος απεικονίζεται στο Σχήμα 1.3(a) γραμμοσκιασμένος.



Σχήμα 1.3: Δειγματικός χώρος φοιτητικού ωραρίου.

2. $A = \{(T_1, T_2) : 0 \leq T_1 < T_2 \leq 24, T_1 \leq 9, T_2 > 9\}$. Το σύνολο A απεικονίζεται γραμμοσκιασμένο στο Σχήμα 1.3(b).
3. $B = \{(T_1, T_2) : 0 \leq T_1 < T_2 \leq 24, T_2 - T_1 \leq 12\}$. Το σύνολο B απεικονίζεται γραμμοσκιασμένο στο Σχήμα 1.3(c).
4. Το σύνολο $C = A \cap B$ σχεδιάζεται με γραμμοσκίαση στο Σχήμα 1.3(d) και αντιστοιχεί στο ενδεχόμενο ο φοιτητής να κοιμάται για περισσότερο από 12 ώρες, και επιπλέον να κοιμάται στις 9 π.μ.

■

1.2 Υπό συνθήκη πιθανότητα

Μερικές φορές χρειάζεται να αναρωτηθούμε για το συμβάν ενός γεγονότος γνωρίζοντας ότι ένα άλλο γεγονός, συσχετιζόμενο ή μη, έχει ήδη συμβεί. Όταν δεσμεύουμε ένα γεγονός τροποποιούμε τον δειγματικό χώρο με τον οποίο υπολογίζουμε τις πιθανότητες. Δεδομένου ότι $P(B) > 0$:

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

ονομάζεται υπό συνθήκη πιθανότητα. Προφανώς αν A, B είναι ανεξάρτητα, τότε $P(A|B) = P(A)$.

1.3 Ιδιότητες Πιθανοτήτων

Θεώρημα 1 (Ολική πιθανότητα) Έστω $\{B_i\}, i = 1, \dots, N$ μια διαμέριση του Ω . Ισχύει:

$$\begin{aligned} P(A) &= \sum_{i=1}^N P(A \cap B_i) = \\ &= \sum_{i=1}^N P(B_i) P(A|B_i). \end{aligned}$$

Παράδειγμα: Ρίχνοντας ένα ζάρι και φέρνοντας ζυγό έχω 0.3 πιθανότητα να κερδίσω ένα παιχνίδι, ενώ αν φέρω μόνο έχω 0.7. Ποιά είναι η πιθανότητα να κερδίσω πριν ρίξω το ζάρι;

Λύση:

$$\begin{aligned} P(\text{Νίκη}) &= \sum_{\text{ζυγοί}} \frac{1}{6} 0.3 + \sum_{\text{μονοί}} \frac{1}{6} 0.7 = \\ &= 0.15 + 0.35 = 0.5. \end{aligned}$$

■

Θεώρημα 2 (Bayes) Έστω $P(B) > 0$ και $\{A_j\}$ μια διαμέριση του A . Τότε:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\sum_j P(B|A_j)P(A_j)}. \quad (1.1)$$

όπου $P(A)$ ονομάζεται και *a priori* πιθανότητα (δε λαμβάνει γνώση για το B) και $P(A|B)$ ονομάζεται και *a posteriori* πιθανότητα.

Παράδειγμα: (Τεστ πολλαπλών επιλογών) Ένας μαθητής απαντά μια ερώτηση πολλαπλών επιλογών, με N δυνατές επιλογές. Αν ο μαθητής δεν ξέρει την απάντηση, επιλέγει μια από τις δυνατές επιλογές στην τύχη, χωρίς κάποια προτίμηση. Αν η πιθανότητα να ξέρει την απάντηση είναι p , ποιά είναι η δεσμευμένη πιθανότητα να ξέρει την απάντηση με δεδομένο ότι επέλεξε σωστά;

Λύση: Ορίζουμε τα ενδεχόμενα $A =$ 'Ο μαθητής επέλεξε σωστά' και $C =$ 'Ο μαθητής γνωρίζει την απάντηση'.

Ουσιαστικά ψάχνουμε την πιθανότητα $P(C|A)$. Έχουμε:

$$P(C|A) = \frac{P(A|C)P(C)}{P(A|C)P(C) + P(A|C^c)P(C^c)} = \frac{1 \times p}{1 \times p + \frac{1}{N} \times (1 - p)} = \frac{p}{p + \frac{1-p}{N}}.$$

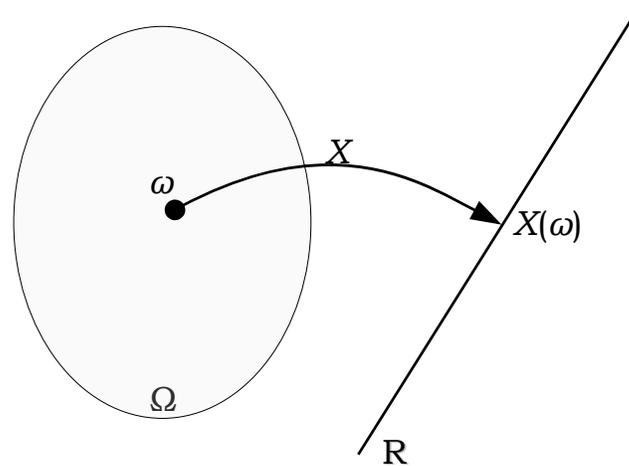
■

Θεώρημα 3 (De Morgan) $(A \cap B)^c = A^c \cup B^c$ και $(A \cup B)^c = A^c \cap B^c$.

Ερώτηση: Αποδείξτε τα δύο παραπάνω θεωρήματα. Υπόδειξη 1: χρησιμοποιήστε τις υπό συνθήκη πιθανότητες. Υπόδειξη 2: θεωρήστε ένα στοιχείο ℓ του χώρου που ικανοποιεί το $(A \cap B)^c$.

Παρατηρήστε τις παρακάτω αντιστοιχίες σε μεγέθη που θα εμφανίζονται στη συνέχεια:

- $\bar{A} \equiv A^c$.
- $P(A \cap B) \equiv P(AB) \equiv P(A, B)$.
- $P(A \cup B) \equiv P(A + B)$.



Σχήμα 1.4: Μια τ.μ. X που είναι συνάρτηση από τον δειγματικό χώρο Ω στο πεδίο των πραγματικών αριθμών. ω είναι στοιχεία του Ω και $X(\omega)$ είναι τιμές της τ.μ. ή ισοδύναμα δείγματα (realizations).

1.4 Τυχαία μεταβλητή

Τυχαία μεταβλητή X ονομάζεται μια συνάρτηση που αντιστοιχίζει τον δειγματικό χώρο στο πεδίο \mathbb{R} ή σε κάποιο υποσύνολο του.

Παράδειγμα:

- X : Το άθροισμα δύο ζαριών. Πεδίο τιμών: $\{2, 3, \dots, 12\}$.
- U : Ο αριθμός συνεχόμενων ρίψεων κέρματος που έρχονται Κ. Πεδίο τιμών: \mathbb{N} .
- V : Η διαφορά της μετρούμενης τάσης RMS στο δίκτυο από τα 220Volt. Πεδίο τιμών: $\{\mathbb{R} : V \geq -220\}$.
- T : Η διάρκεια μιας τηλεφωνικής κλήσης. Πεδίο τιμών: \mathbb{R}^+ .

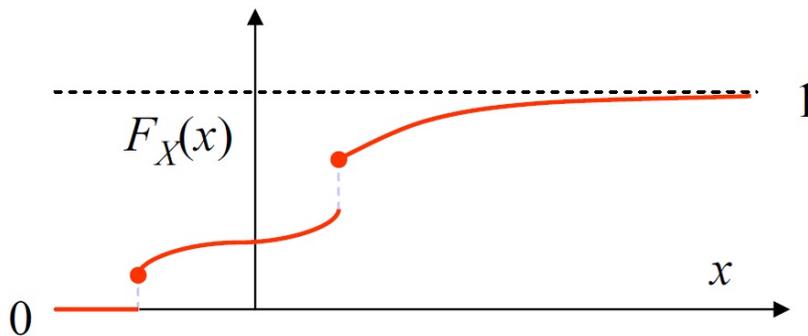
Είναι φανερό ότι υπάρχουν διακριτές τυχαίες μεταβλητές (τ.μ.) (για παράδειγμα X, Y) και συνεχείς τ.μ. (για παράδειγμα V, T) οι οποίες τυγχάνουν διαφορετικής αντιμετώπισης κατά περίπτωση αλλά έχουν και κοινές ιδιότητες.

1.4.1 Κατανομή

Εξ' ορισμού ισχύει ότι η ποσότητα $X(\omega) \leq x$ είναι ένα γεγονός που ανήκει στο πεδίο \mathcal{F} και πάντα μπορούμε να υπολογίσουμε το

$$\{X(\omega) \leq x\} = \{\omega \in \Omega : X(\omega) \leq x\} \subseteq \Omega.$$

Άρα μπορούμε πάντα να ορίσουμε την πιθανότητα $P(X(\omega) \leq x)$, η οποία ονομάζεται **κατανομή** (distribution) και 'συνοδεύει' κάθε τ.μ.



Σχήμα 1.5: Ένα παράδειγμα κατανομής όπου φαίνονται χαρακτηριστικά οι ιδιότητές της.

Μια τ.μ. μπορεί να μελετηθεί πριν από κάποιο πείραμα (για παράδειγμα μπορούμε να υπολογίσουμε τη μέση τιμή της). Επίσης, μπορούμε να υπολογίσουμε μια τιμή της τ.μ. όπως ακριβώς με μια συνάρτηση (για παράδειγμα $X(\omega_1) = 7$). Καθώς, όμως, το ω διαλέγεται στην τύχη κάθε φορά, η γνώση αυτή δεν μας βοηθάει. Αντίθετα, η κατανομή μιας τ.μ. καθορίζει πλήρως μια τ.μ. και για αυτό το λόγο ονομάζεται **πλήρες μοντέλο**. Επίσης, γράφουμε $F_X(x) \equiv P(X \leq x)$, που μερικές φορές για συντομία γράφεται ως $F(x)$, εφόσον βέβαια δεν υπάρχει περίπτωση σύγχυσης.

Η κατανομή $F_X(x)$ έχει τις παρακάτω ιδιότητες (δες σχήμα 1.5):

- Είναι αύξουσα.
- Είναι συνεχής από δεξιά.
- Ισχύει πάντα $F(-\infty) = 0$, $F(\infty) = 1$.

Ερώτηση: Ποιά είναι η κατανομή της τ.μ. X : το αποτέλεσμα ενός δίκαιου ζαριού;

1.4.2 Ροπές

Οι ροπές (moments) είναι στατιστικοί τελεστές επί μιας τ.μ. Η πιο γνωστή ροπή είναι η πρώτη ροπή ή αλλιώς **αναμενόμενη τιμή** (expectation), η οποία συμβολίζεται ως $E[X]$. Ιδιότητες της αναμενόμενης τιμής:

- $E[aX + bY + c] = aE[X] + bE[Y] + c$ (γραμμικός τελεστής)
- $E[XY] \stackrel{\text{ανεξ.}}{=} E[X]E[Y]$

Καθ'ομοιο τρόπο ορίζουμε τη n -οστή ροπή ως $E[X^n]$. Επίσης, ορίζουμε την n -οστή κεντρική ροπή ως $E[(X - E[X])^n]$ (ονομάζεται κεντρική επειδή αφαιρούμε απο τις τιμές της τ.μ. την αναμενόμενη τιμή). Η δεύτερη κεντρική ροπή ονομάζεται **μεταβλητότητα** (variance) και συμβολίζεται με $\text{Var}(X)$,

$$\begin{aligned} \text{Var}(X) &\doteq E[(X - E[X])^2] = \\ &= E[X^2] - E[X]^2. \end{aligned}$$

Ιδιότητες της μεταβλητότητας:

- $\text{Var}(aX + c) = \text{Var}(aX) = a^2\text{Var}(X)$

- $\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y) \stackrel{\text{ανεξ.}}{=} a^2\text{Var}(X) + b^2\text{Var}(Y)$
- $\text{Var}(X) \geq 0$.

όπου $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$ είναι η συμμεταβλητότητα δύο τ.μ.

Η τετραγωνική ρίζα της μεταβλητότητας χρησιμοποιείται κατά κόρον στην στατιστική και ονομάζεται τυπική απόκλιση (standard deviation) $\sigma = \sqrt{\text{Var}(X)}$. Άλλες ονομαστές ροπές είναι η τρίτη κεντρική ροπή ή λοξότητα (skewness) και η τέταρτη κεντρική ροπή ή κύρτωση (kurtosis).

Οι ροπές είναι ενδεικτικές της μορφής της κατανομής μιας τ.μ. (πράγμα που εξηγεί και τις ονομασίες τους). Η γνώση όλων των ροπών μιας τ.μ. ισοδυναμεί με τη γνώση της κατανομής (μοντέλο ροπών). Επίσης, υπάρχουν συγκεκριμένες κατανομές όπου οι ροπές τους είναι κατά πλειοψηφία μηδενικές ή εκφράζονται ως συνάρτηση των δύο πρώτων ροπών. Ένα τυπικό παράδειγμα είναι η κανονική κατανομή όπου η γνώση των δύο πρώτων ροπών είναι αρκετή για να περιγράψει ολόκληρη την τ.μ. (ισοδυναμεί με την γνώση της κατανομής).

1.5 Χαρακτηριστική Συνάρτηση

Η **χαρακτηριστική συνάρτηση** καθορίζει μια τ.μ. με τον ίδιο τρόπο που το κάνει η κατανομή. Δηλαδή, η γνώση της κατανομής είναι ισοδύναμη με τη γνώση της χαρακτηριστικής συνάρτησης. Υπάρχει μάλιστα πάντα 1 προς 1 αντιστοιχία μεταξύ της χαρακτηριστικής συνάρτησης και της κατανομής για κάθε τ.μ. Ορίζεται ως

$$\Phi(\omega) \doteq E[e^{j\omega X}],$$

όπου $j \equiv \sqrt{-1}$. Η χαρακτηριστική συνάρτηση είναι ένας τελεστής με ιδιότητες Fourier: έστω ανεξάρτητες τ.μ. X_1, X_2 με χαρακτηριστικές συναρτήσεις $\Phi_{X_1}(\omega)$ και $\Phi_{X_2}(\omega)$ αντίστοιχα. Τότε:

$$\Phi_{X_1+X_2}(\omega) = \Phi_{X_1}(\omega)\Phi_{X_2}(\omega).$$

Αναπτύσσοντας το $e^{j\omega X}$ κατά Taylor¹ γύρω από το 0 έχουμε:

$$\Phi(\omega) = E \left[\sum_{k=0}^{\infty} \frac{(j\omega X)^k}{k!} \right] = \sum_{k=0}^{\infty} \frac{(j\omega)^k}{k!} E[X^k].$$

Συνεπώς, από τη χαρακτηριστική συνάρτηση μπορούμε να υπολογίσουμε οποιαδήποτε ροπή:

$$E[X^n] = (-j)^n \Phi_X^{(n)}(\omega) \Big|_{\omega=0},$$

όπου $\Phi_X^{(n)}(\omega) \doteq \frac{d^n \Phi(\omega)}{d\omega^n}$. Για τον ίδιο λόγο μπορούμε να χρησιμοποιήσουμε και τη συνάρτηση γεννήτρια ροπών (moment generator function ή m.g.f.).

$$M_X(t) \doteq E[e^{\omega X}].$$

1.6 Διακριτές τ.μ.

Οι τ.μ. που αντιστοιχίζουν το δειγματικό χώρο σε ένα διακριτό σύνολο \mathcal{S}_d ονομάζονται διακριτές τ.μ. Για αυτές ισχύει $P(X \in \mathcal{S}_d) = 1$. Η κατανομή μιας διακριτής τ.μ. καθορίζεται από τις πιθανότητες $p_i \doteq P(X = x_i)$, $x_i \in \mathcal{S}_d$. Συγκεκριμένα:

¹Ισχύει ότι $e^x \doteq 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$

$$P(X \leq x) = \sum_{i: x_i \leq x} p_i, \quad \forall x_i \in \mathcal{S}_d.$$

Επιπρόσθετα, μπορούμε να ορίσουμε την **συνάρτηση μάζας -πιθανότητας** ή ΣΜΠ (probability mass function ή p.m.f.) ως

$$p_X(x) = P(X = x) = \begin{cases} p_i & x \in \mathcal{S}_d \\ 0 & x \notin \mathcal{S}_d. \end{cases}$$

Κάθε ΣΜΠ πρέπει να τηρεί τις παρακάτω συνθήκες:

- $\sum_{x_i \in \mathcal{S}_d} p_X(x_i) = 1.$
- $p_X(x_i) > 0, \quad \forall x_i \in \mathcal{S}_d.$

Επίσης, κάθε διάνυσμα με θετικές τιμές που έχει άθροισμα στοιχείων 1, μπορεί να θεωρηθεί ως ΣΜΠ.

Παράδειγμα: (Κατανομή ελαχίστου) Ρίχνουμε το ζάρι δύο φορές, έστω X το πρώτο αποτέλεσμα και Y το δεύτερο. Να βρεθεί η ΣΜΠ του $\min\{X, Y\}$.

Λύση: Θα βρω όλα τα ενδεχόμενα όπου το X είναι το ελάχιστο και αντίστοιχα τις πιθανότητές τους:

$$\begin{aligned} X = 1 & \quad \text{και } Y = 1, 2, 3, 4, 5, 6 \text{ με πιθανότητα } \frac{6}{36} \\ X = 2 & \quad \text{και } Y = 2, 3, 4, 5, 6 \text{ με πιθανότητα } \frac{5}{36} \\ X = 3 & \quad \text{και } Y = 3, 4, 5, 6 \text{ με πιθανότητα } \frac{4}{36} \\ X = 4 & \quad \text{και } Y = 4, 5, 6 \text{ με πιθανότητα } \frac{3}{36} \\ X = 5 & \quad \text{και } Y = 5, 6 \text{ με πιθανότητα } \frac{2}{36} \\ X = 6 & \quad \text{και } Y = 6 \text{ με πιθανότητα } \frac{1}{36}. \end{aligned}$$

Λόγω συμμετρίας το ίδιο θα συμβαίνει όταν το Y είναι το ελάχιστο. Προσοχή, όμως, πρέπει να αποφύγουμε να μετρήσουμε τις περιπτώσεις $X = Y$ δύο φορές. Άρα, τελικά έχουμε

$$\begin{aligned} \min\{X, Y\} = 1 & \quad \text{με πιθανότητα } \frac{11}{36} \\ \min\{X, Y\} = 2 & \quad \text{με πιθανότητα } \frac{9}{36} \\ \min\{X, Y\} = 3 & \quad \text{με πιθανότητα } \frac{7}{36} \\ \min\{X, Y\} = 4 & \quad \text{με πιθανότητα } \frac{5}{36} \\ \min\{X, Y\} = 5 & \quad \text{με πιθανότητα } \frac{3}{36} \\ \min\{X, Y\} = 6 & \quad \text{με πιθανότητα } \frac{1}{36}. \end{aligned}$$

Παρατηρήστε ότι η πιθανότητα του ενδεχομένου $\{\min\{X, Y\} = 1\}$ είναι ίδια με την πιθανότητα να φέρουμε τουλάχιστον έναν άσο. (χρήσιμο για όσους παίζουν τάβλι!) ■

Η αναμενόμενη τιμή διακριτής τ.μ. βρίσκεται ως

$$E[X] = \sum_{x \in \mathcal{S}_d} P(X = x) x = \sum_i p_i x_i.$$

Παράδειγμα: Η αναμενόμενη τιμή της X : το αποτέλεσμα ενός δίκαιου ζαριού, είναι:

$$E[X] = \sum_{i=1}^6 p_i x_i = \left(\frac{1}{6}1 + \frac{1}{6}2 + \frac{1}{6}3 + \frac{1}{6}4 + \frac{1}{6}5 + \frac{1}{6}6\right) = 3.5.$$

Άρα η αναμενόμενη τιμή μπορεί να μην ανήκει καν στο \mathcal{S}_d .

Η n -οστή ροπή υπολογίζεται ως

$$E[X^n] = \sum_{x \in \mathcal{S}_d} p_i x_i^n.$$

Παράδειγμα: Η μεταβλητότητα της X : το αποτέλεσμα ενός δίκαιου ζαριού, είναι:

$$\text{Var}(X) = \sum_{i=1}^6 p_i x_i^2 - E[X]^2 = \left(\frac{1}{6}1 + \frac{1}{6}4 + \frac{1}{6}9 + \frac{1}{6}16 + \frac{1}{6}25 + \frac{1}{6}36\right) - 3.5^2 = \frac{35}{12}.$$

Για τη χαρακτηριστική συνάρτηση μιας διακριτής τ.μ. μπορούμε να γράψουμε:

$$\Phi_X(\omega) = E[e^{j\omega X}] = \sum_i p_i e^{j\omega x_i}.$$

Επίσης, για διακριτές τ.μ. που παίρνουν μη αρνητικές ακέραιες τιμές ($x_i \equiv i \geq 0$), μπορούμε να ορίσουμε τον ισοδύναμο μετασχηματισμό Z :

$$G_X(z) \doteq E[z^X] = \sum_{i=0}^{\infty} p_i z^i, |z| \leq 1,$$

ο οποίος ονομάζεται συνάρτηση-γεννήτρια πιθανοτήτων (probability-generating function ή p.g.f.). Έχουμε:

$$p_i = P(X = i) = \frac{1}{i!} G_X(z)^{(i)} \Big|_{z=0}.$$

Ερώτηση: Δείξτε ότι για τ.μ. που παίρνουν μη αρνητικές ακέραιες τιμές ισχύει $\text{Var}(X) = G_X''(1) + G_X'(1) - (G_X'(1))^2$.

1.6.1 Συνάρτηση δείκτης

Η **συνάρτηση δείκτης** (Indicator function) είναι η πιο πρώιμη μορφή τ.μ. Λαμβάνει δυαδικές τιμές $\mathcal{S}_d = \{0, 1\}$ και περιγράφεται από το παρακάτω:

$$\mathbb{1}_{\{A\}}(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A, \end{cases}$$

όπου A είναι ένα γεγονός. Προφανώς ισχύει:

$$P(\mathbb{1}_{\{A\}} = 1) = P(A)$$

και

$$P(\mathbb{1}_{\{A\}} = 0) = P(A^c) = 1 - P(A).$$

Επίσης ισχύει

$$E[\mathbb{1}_{\{A\}}] = P(\mathbb{1}_{\{A\}} = 1) \cdot 1 + P(\mathbb{1}_{\{A\}} = 0) \cdot 0 = P(A).$$

1.6.2 Κατανομή Bernoulli

Χρησιμοποιείται για να περιγράψει πειράματα που μπορεί να πετύχουν ή να αποτύχουν (Bernoulli trials). Αν $X \sim \text{Bernoulli}(p)$, $0 \leq p \leq 1$, τότε

$$\begin{aligned} P(X = 1) &= p \\ P(X = 0) &= 1 - p \\ \Phi(\omega) &= 1 - p + pe^{j\omega}, \omega \in \mathbb{R} \\ G_X(z) &= 1 - p + pz, |z| \leq 1 \\ E[X] &= p \\ \text{Var}(X) &= p(1 - p). \end{aligned}$$

1.6.3 Κατανομή Binomial

Χρησιμοποιείται για να περιγράψει τον αριθμό επιτυχιών σε n διαδοχικά πειράματα Bernoulli που πετυχαίνουν κάθε ένα με πιθανότητα p . Αν $X \sim B(p, n)$, $0 \leq p \leq 1$, $n \in \mathbb{N}^*$ τότε

$$\begin{aligned} P(X = i) &= \binom{n}{i} p^i (1 - p)^{n-i}, 0 \leq i \leq n \\ \Phi(\omega) &= (1 - p + pe^{j\omega})^n, \omega \in \mathbb{R} \\ G_X(z) &= (1 - p + pz)^n, |z| \leq 1 \\ E[X] &= np \\ \text{Var}(X) &= np(1 - p), \end{aligned}$$

όπου $\binom{n}{i} \doteq \frac{n!}{i!(n-i)!}$. Επίσης ισχύει ότι $X = \sum_{i=1}^n Y_i$, όπου $Y_i \sim \text{Bernoulli}(p)$ και ανεξάρτητες μεταξύ τους.

Παράδειγμα: Ρίχνουμε ένα δίκαιο κέρμα 10 φορές. (α) Ποιά είναι η πιθανότητα να έρθει 6 φορές Γ; (β) Πόσες φορές είναι αναμενόμενο να έρθει Γ;

Λύση:

(α) Θεωρούμε ως επιτυχία να έρθει Γ. Το πείραμά μας περιγράφεται από μια τ.μ. $X \sim B(0.5, 10)$. Κατά συνέπεια

$$P(X = 6) = \binom{10}{6} 0.5^6 (0.5)^4 = 0.2051.$$

(β) $E[X] = 10 \cdot 0.5 = 5$. ■

Η μέση τιμή της Binomial μπορεί να υπολογιστεί ως εξής

$$\begin{aligned} E[X] &= \sum_{i=0}^n i \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=1}^n i \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \\ &= p \sum_{i=1}^n \frac{n!}{(i-1)!(n-i)!} p^{i-1} (1-p)^{n-i} = p \sum_{k=0}^{n-1} \frac{n!}{k!(n-k-1)!} p^k (1-p)^{n-k-1} \\ &= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-1-k)!} p^k (1-p)^{n-1-k} = np \sum_{k=0}^{\nu} \frac{\nu!}{k!(\nu-k)!} p^k (1-p)^{\nu-k} = np. \end{aligned}$$

1.6.4 Κατανομή Geometric

Χρησιμοποιείται για να περιγράψει τον αριθμό των διαδοχικά αποτυχημένων πειραμάτων Bernoulli μέχρι την πρώτη επιτυχία. Ανάλογα με την χρήση μπορούμε να έχουμε διαφορετικές εκδοχές της γεωμετρικής κατανομής.

Αρχικά θεωρούμε την περίπτωση που η τ.μ. παίρνει τιμές στο σύνολο $\{0, 1, \dots\}$. Τότε $X \sim G_0(p)$, $0 < p \leq 1$ και

$$\begin{aligned} P(X = i) &= p(1-p)^i, i \geq 0. \\ \Phi(\omega) &= \frac{p}{1 - (1-p)e^{j\omega}}, \omega \in \mathbb{R} \\ G_X(z) &= \frac{p}{1 - (1-p)z}, |z| \leq 1 \\ E[X] &= \sum_i i \cdot p(1-p)^i = \frac{1-p}{p} \\ \text{Var}(X) &= \frac{1-p}{p^2}, \end{aligned}$$

όπου χρησιμοποιήσαμε τις ταυτότητες $\sum_{i=0}^{\infty} ip^i \doteq \frac{p}{(1-p)^2}$ και $\sum_{i=0}^{\infty} i^2 p^i \doteq \frac{p(1+p)}{(1-p)^3}$.

Αντίθετα, αν οι προσπάθειες μετράνε από το 1, δηλαδή η τ.μ. παίρνει τιμές στο σύνολο $\{1, 2, \dots\}$. Τότε $X \sim G_1(p)$, $0 < p \leq 1$ και

$$\begin{aligned} P(X = i) &= p(1-p)^{i-1}, i \geq 1. \\ \Phi(\omega) &= \frac{pe^{j\omega}}{1 - (1-p)e^{j\omega}}, \omega \in \mathbb{R} \\ G_X(z) &= \frac{pz}{1 - (1-p)z}, |z| \leq 1 \\ E[X] &= \sum_i i \cdot p(1-p)^{i-1} = \frac{1}{p} \\ \text{Var}(X) &= \frac{1-p}{p^2}, \end{aligned}$$

Αντίστοιχα ορίζονται και οι γεωμετρικές μεταβλητές που αφορούν στην πρώτη αποτυχία μετρώντας από το 0, $G_0(1-p)$ και μετρώντας από το 1, $G_1(1-p)$.

Παράδειγμα: Σε ένα κανάλι επικοινωνιών μεταδίδουμε 1 bit κάθε δευτερόλεπτο. Ο δέκτης λαμβάνει κάθε bit με πιθανότητα p . Βρείτε τον αναμενόμενο ωφέλιμο ρυθμό μετάδοσης (throughput) και την αναμενόμενη καθυστέρηση για κάθε bit.

Λύση: Για τον ωφέλιμο ρυθμό μετάδοσης έχουμε

$$E[T] = p.$$

Για την καθυστέρηση, θεωρούμε την τ.μ. $X \sim G_1(p)$ που περιγράφει τον αριθμό συνεχόμενων λαθών (χωρίς λάθος καθυστερώ 1 δευτερόλεπτο, με ένα λάθος δύο δευτερόλεπτα κ.ο.κ.). Προφανώς έχουμε

$$E[D] = \frac{1}{p}.$$

■

Η γεωμετρική κατανομή έχει την ιδιότητα της απώλειας μνήμης (memoryless property), η οποία εκφράζεται ως εξής:

$$P(X \geq i + j | X \geq i) = P(X \geq j).$$

Για την γεωμετρική κατανομή αυτό αποδεικνύεται ως εξής:

$$\begin{aligned} P(X \geq j) &= 1 - P(X < j) = 1 - \sum_{k=0}^{j-1} p(1-p)^k \\ &= 1 - p \sum_{k=0}^{j-1} (1-p)^k = 1 - p \frac{1 - (1-p)^j}{1 - 1 + p} = (1-p)^j. \end{aligned}$$

Επίσης,

$$\begin{aligned} P(X \geq i + j | X \geq i) &= \frac{P(X \geq i + j, X \geq i)}{P(X \geq i)} \\ &= \frac{P(X \geq i + j)}{P(X \geq i)} = \frac{(1-p)^{i+j}}{(1-p)^i} = (1-p)^j. \end{aligned}$$

1.6.5 Κατανομή Poisson

Χρησιμοποιείται για να περιγράψει τον αριθμό συμβάντων σε ένα δεδομένο χρονικό πλαίσιο αν τα συμβάντα γίνονται ανεξάρτητα το ένα από το άλλο και με γνωστό ρυθμό λ . Αν $X \sim \text{Poisson}(\lambda)$, $\lambda > 0$, τότε

$$\begin{aligned} P(X = i) &= \frac{\lambda^i e^{-\lambda}}{i!}, i \geq 0. \\ \Phi(\omega) &= e^{\lambda(e^{j\omega} - 1)}, \omega \in \mathbb{R} \\ G_X(z) &= e^{\lambda(z-1)}, |z| \leq 1 \\ E[X] &= \lambda \\ \text{Var}(X) &= \lambda. \end{aligned}$$

Παράδειγμα: Είναι γνωστό ότι στο τηλεφωνικό κέντρο του ΟΤΕ οι τηλεφωνικές κλήσεις καταφτάνουν με ρυθμό 10 το δευτερόλεπτο. Ποιά η πιθανότητα σε ένα δευτερόλεπτο να έρθουν 20 κλήσεις;

Λύση: Θεωρούμε ότι οι τηλεφωνικές κλήσεις γεννιούνται με ανεξάρτητο τρόπο (λόγω της ιδιωτικότητας των χρηστών)². Άρα, θεωρούμε ότι ο αριθμός κλήσεων σε ένα δευτερόλεπτο είναι μια τ.μ. $X \sim \text{Poisson}(10)$

$$P(X = 20) = \frac{10^{20} e^{-10}}{20!} = e^{\log 20^{20} + \log e^{-10} - \sum_{i=2}^{20} \log i} = e^{-6.2839} = 0.0019,$$

όπου \log είναι ο νεπέριος λογάριθμος (βάση το e). ■

Ερώτηση: Υπολογίστε την αναμενόμενη τιμή χρησιμοποιώντας την p.m.f. της Poisson.

Έστω η τ.μ. $X \sim B\left(\frac{\lambda}{n}, n\right)$ με ΣΜ-Π:

$$\begin{aligned} p_X(k) &= \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \\ &= \frac{n(n-1)\dots(n-k+1)}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n}{n} \frac{n-1}{n} \dots \frac{n-k+1}{n} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k}. \end{aligned}$$

Αν λάβουμε το όριο της συνάρτησης αυτής $n \rightarrow \infty$ τότε όλα τα κλάσματα με παρανομαστή n τείνουν στη μονάδα, και:

$$\left(1 - \frac{\lambda}{n}\right)^{n-k} \rightarrow e^{-\lambda}.$$

Τελικά λαμβάνουμε

$$\lim_{n \rightarrow \infty} p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

που είναι η ΣΜ-Π Poisson(λ).

1.7 Συνεχείς τ.μ.

Οι τ.μ. που αντιστοιχίζουν τον δειγματικό χώρο σε ένα συνεχές σύνολο \mathcal{S}_c ονομάζονται συνεχείς τ.μ. Όταν η κατανομή μιας συνεχούς τ.μ. X είναι παραγωγίσιμη (σύνηθες), μπορούμε να ορίσουμε την **συνάρτηση πυκνότητας πιθανότητας** (probability density function ή p.d.f.) $f_X(x)$:

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

Κατ' αντιστοιχία ισχύει

$$F_X(x) = \int_{-\infty}^x f_X(u) du.$$

Κάθε p.d.f. πρέπει να τηρεί τις παρακάτω συνθήκες:

² Αυτή η παραδοχή δεν ισχύει βέβαια σε περίπτωση συσχέτισης, όπως μετά από έναν σεισμό ή σε μια μεγάλη γιορτή.

- $\int_{-\infty}^{\infty} f_X(x)dx = 1$.
- Αφού η $F_X(x)$ είναι αύξουσα, θα πρέπει $f_X(x) \geq 0, \forall x$.

Οι συνεχείς τ.μ. έχουν τις παρακάτω ιδιότητες:

- $P(X = x) = \sigma_{\text{υν}} \cdot 0$
- $P(a < X < b) = \sigma_{\text{υν}} \cdot P(a \leq X \leq b) = \int_a^b f_X(x)dx$
- $P(X \in A) = \int_A f_X(x)dx$
- $P(X \in \mathcal{S}_c) = \int_{\mathcal{S}_c} f_X(x)dx = 1$
- $E[X] = \int_{\mathcal{S}_c} x dF_X(x) = \text{παραγ.} \int_{\mathcal{S}_c} x f_X(x)dx$
- $E[X^n] = \int_{\mathcal{S}_c} x^n dF_X(x) = \text{παραγ.} \int_{\mathcal{S}_c} x^n f_X(x)dx$.

Επίσης, όταν $\mathcal{S}_c = \mathbb{R}$, τότε

- $P(X \in \Omega) = \int_{-\infty}^{\infty} f_X(x)dx = 1$
- $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$
- $E[X] = \int_{-\infty}^{\infty} x dF_X(x) = \text{παραγ.} \int_{-\infty}^{\infty} x f_X(x)dx$
- $E[X^n] = \int_{-\infty}^{\infty} x^n dF_X(x) = \text{παραγ.} \int_{-\infty}^{\infty} x^n f_X(x)dx$.

Παράδειγμα: (Μια p.d.f. μπορεί να είναι αυθαίρετα μεγάλη) Έστω η τ.μ. X με συνάρτηση πυκνότητας πιθανότητας

$$f_X(x) = \begin{cases} \frac{1}{2\sqrt{x}} & \text{αν } x \in (0, 1] \\ 0 & \text{αλλού.} \end{cases}$$

Τότε για τιμές $x \rightarrow 0^+$ η p.d.f. λαμβάνει πολύ μεγάλες τιμές. Όμως ισχύει πάντα ότι $\int_A f_X(x)dx \leq 1$.

Για τη χαρακτηριστική συνάρτηση ισχύει:

$$\Phi(\omega) = E[e^{j\omega X}] = \int_{-\infty}^{\infty} e^{j\omega x} dF_X(x) = \text{παραγ.} \int_{-\infty}^{\infty} e^{j\omega x} f_X(x)dx.$$

Παρατηρούμε λοιπόν ότι η χαρακτηριστική συνάρτηση είναι ουσιαστικά ο μετασχηματισμός Fourier της συνάρτησης πυκνότητας πιθανότητας. Κατά αντιστοιχία μπορούμε να ορίσουμε και τον αντίστοιχο μετασχηματισμό Laplace, που στη συγκεκριμένη περίπτωση ονομάζεται μετασχηματισμός Laplace-Stieltjes:

$$\tilde{F}_X(s) \doteq E[e^{-sX}] = \int_0^{\infty} e^{-sx} dF_X(x), s \in \mathbb{C}.$$

Ερώτηση: Πως μπορούμε να υπολογίσουμε την n -οστή ροπή γνωρίζοντας τον μετασχηματισμό $\tilde{F}_X(s)$;

1.7.1 Ομοιόμορφη Κατανομή (Uniform)

Η ομοιόμορφη κατανομή είναι η πιο απλή στη σύλληψη κατανομή. Αν $X \sim U(\alpha, b)$, $\alpha, b \in \mathbb{R}$ και $\alpha < b$, τότε

$$\begin{aligned} F(x) &= \begin{cases} 0 & x < \alpha \\ \frac{x-\alpha}{b-\alpha} & x \in [\alpha, b] \\ 1 & x > b \end{cases} \\ f(x) &= \begin{cases} \frac{1}{b-\alpha} & x \in [\alpha, b] \\ 0 & \text{αλλού} \end{cases} \\ \Phi(\omega) &= \frac{e^{j\omega b} - e^{j\omega\alpha}}{j\omega(b-\alpha)}, \omega \in \mathbb{R} \\ \tilde{F}(s) &= \frac{e^{-sb} - e^{-s\alpha}}{s(\alpha-b)}, s \in \mathbb{C} \\ E[X] &= \frac{\alpha + b}{2} \\ \text{Var}(X) &= \frac{(b-\alpha)^2}{12}. \end{aligned}$$

Βασίζεται στην ομογένεια της πιθανότητας και κατά συνέπεια η πιθανότητα εξαρτάται μόνο από το εύρος υπολογισμού και όχι από τις οριακές τιμές. Έστω $\tau < b - \alpha$:

$$P(\alpha \leq X \leq \alpha + \tau) = P(b - \tau \leq X \leq b).$$

Παράδειγμα: Έστω τυχαία μεταβλητή $U \sim U(-1, 1)$. Η ομοιομορφία σημαίνει ότι η πιθανότητα να εμφανιστεί η U σε κάποιο υποδιάστημα εξαρτάται αποκλειστικά από το μήκος του υποδιαστήματος, και είναι ανάλογη με αυτό. Προσδιορίστε την κατανομή πιθανότητας της U και προσδιορίστε τις πιθανότητες: $P(U > 0)$, $P(|U| < 1/3)$, $P(|U| \leq 3/4)$, $P(U < 5)$, $P(1/3 < U < 1/2)$.

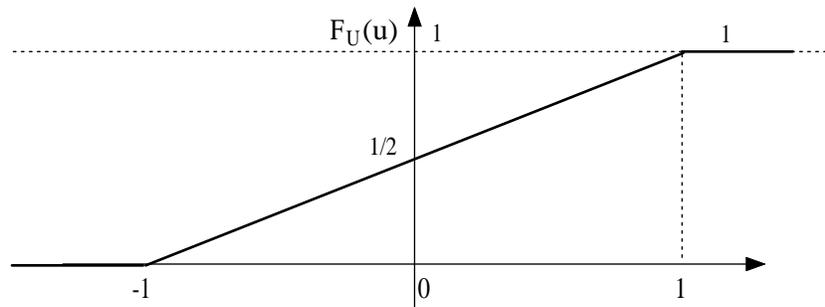
Λύση: Η κατανομή πιθανότητας που ικανοποιεί την ιδιότητα της ομοιομορφίας είναι η

$$F_U(u) = \begin{cases} 1, & u > 1, \\ \frac{u+1}{2}, & -1 \leq u \leq 1, \\ 0, & u < -1. \end{cases}$$

και σχεδιάζεται στο Σχήμα 1.6. Χρησιμοποιώντας την κατανομή πιθανότητας, έχουμε:

$$\begin{aligned} P(U > 0) &= 1 - P(U \leq 0) = 1 - F_U(0) = \frac{1}{2}, \\ P(|U| < \frac{1}{3}) &= P(-\frac{1}{3} < U < \frac{1}{3}) = F_U((\frac{1}{3})^-) - F_U(-\frac{1}{3}) = \frac{1}{3}, \\ P(|U| \geq \frac{3}{4}) &= 1 - P(|U| < \frac{3}{4}) = 1 - P(-\frac{3}{4} < U < \frac{3}{4}) = 1 - F_U((\frac{3}{4})^-) + F_U(-\frac{3}{4}) = \frac{1}{4}, \\ P(U < 5) &= F_U(5^-) = 1, \\ P(\frac{1}{3} < U < \frac{1}{2}) &= F_U((\frac{1}{2})^-) - F_U(\frac{1}{3}) = \frac{1}{12}, \end{aligned}$$

λόγω της συνέχειας χρησιμοποιήσαμε το όριο $F(x^-) \doteq \lim_{t \rightarrow x^-} F(t)$. ■

Σχήμα 1.6: Κατανομή $U(-1, 1)$.

1.7.2 Κατανομή Gauss

Χρησιμοποιείται σε πολλά μοντέλα στη φυσική και όχι μόνο. Για παράδειγμα, χρησιμοποιείται για να περιγράψει διακυμάνσεις που προκύπτουν από μεγάλο αριθμό αστάθμητων παραγόντων. Για τον ίδιο λόγο αποτελεί βάση της στατιστικής. Επίσης, το πιο διάσημο μοντέλο καναλιού επικοινωνιών ονομάζεται Gaussian channel καθώς υποθέτει προσθετικό θόρυβο υπο τη μορφή μιας τ.μ. Gauss.

Η Gaussian κατανομή περιγράφεται πλήρως από τη μέση τιμή (συμβολίζεται με μ) και τη μεταβλητότητα (συμβολίζεται με σ^2). Αν $X \sim \mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}_+$, τότε

$$\begin{aligned} F(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x-\mu}{\sqrt{2\sigma^2}} \right) \right], x \in \mathbb{R} \\ f(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R} \\ \Phi(\omega) &= e^{j\omega\mu - \frac{1}{2}\omega^2\sigma^2}, \omega \in \mathbb{R} \\ \tilde{F}(s) &= e^{-s\mu + \frac{1}{2}s^2\sigma^2}, s \in \mathbb{C} \\ E[X] &= \mu \\ \operatorname{Var}(X) &= \sigma^2. \end{aligned}$$

όπου $\operatorname{erf}(\cdot)$ είναι η συνάρτηση σφάλματος (error function). Μια μεταβλητή $X \sim \mathcal{N}(0, 1)$ ονομάζεται κανονική (Normal).

Παράδειγμα: (Η κατανομή Gaussian διατηρεί την κανονικότητά της) Έστω ανεξάρτητες τ.μ. $X, Y \sim \mathcal{N}(\mu, \sigma^2)$ και $Z = X + Y$. Αρχικά μπορούμε εύκολα να δείξουμε ότι

$$E[Z] = E[X + Y] = E[X] + E[Y] = 2\mu$$

και

$$\operatorname{Var}(Z) = \operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) = 2\sigma^2.$$

Κατόπιν υπολογίζουμε την χαρακτηριστική συνάρτηση της Z

$$\begin{aligned} \Phi_Z(\omega) &= \Phi_{X+Y}(\omega) = \Phi_X(\omega)\Phi_Y(\omega) \\ &= e^{j\omega\mu - \frac{1}{2}\omega^2\sigma^2} e^{j\omega\mu - \frac{1}{2}\omega^2\sigma^2} \\ &= e^{j\omega(2\mu) - \frac{1}{2}\omega^2(2\sigma^2)}. \end{aligned}$$

Άρα, παρατηρούμε ότι $Z \sim \mathcal{N}(2\mu, 2\sigma^2)$. Επίσης, η ιδιότητα αυτή ισχύει για οποιονδήποτε αριθμό από ανεξάρτητες Gaussian μεταβλητές που αθροίζουμε και ακόμα και αν έχουν διαφορετικές μέσες τιμές και μεταβλητότητες. Επίσης, η τυχαία μεταβλητή $W = aX + b$ όπου $a \neq 0, b$ σταθερές είναι επίσης Gaussian και συγκεκριμένα $W \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$. Άρα, συμπεραίνουμε ότι η Gaussian κατανομή παραμένει αναλλοίωτη σε γραμμικούς μετασχηματισμούς.

Παράδειγμα: Έστω ένα Gaussian διάνυσμα $\{X, Y\}$ όπου $X, Y \sim \mathcal{N}(0, \sigma^2)$ ανεξάρτητες τ.μ. Το μέτρο του διανύσματος αυτού αναπαριστά την απόσταση που έχουν ρίψεις με κάποιο στατιστικό σφάλμα από το κέντρο ενός στόχου. Το ίδιο φαινόμενο συμβαίνει και σε έναν τηλεπικοινωνιακό δέκτη. Ο πομπός στέλνει ένα σήμα-διάνυσμα στο οποίο προστίθεται Gaussian θόρυβος στην τετμημένη και στην τεταγμένη. Η κατανομή που περιγράφει το μέτρο του διανύσματος και καταδεικνύει την πιθανότητα λάθους ονομάζεται Rayleigh, $R = \sqrt{X^2 + Y^2} \Rightarrow R \sim \text{Rayleigh}(\sigma)$.

1.7.3 Εκθετική κατανομή

Σχετίζεται άμεσα με τη στοχαστική διαδικασία Poisson και με τις διαδικασίες Markov και χρησιμοποιείται ευρύτατα στη μοντελοποίηση για να περιγράψει μεγέθη πακέτων πληροφορίας, χρόνους μεταξύ δύο συμβάντων κ.α.

Η εκθετική κατανομή περιγράφεται πλήρως από το ρυθμό (συμβολίζεται με λ). Αν $X \sim \text{Exp}(\lambda)$, $\lambda > 0$, τότε

$$\begin{aligned} F(x) &= 1 - e^{-\lambda x}, x \geq 0 \\ f(x) &= \lambda e^{-\lambda x}, x \geq 0 \\ \Phi(\omega) &= \frac{\lambda}{\lambda - j\omega}, \omega \in \mathbb{R} \\ \tilde{F}(s) &= \frac{\lambda}{\lambda + s}, s \in \mathbb{C} \\ E[X] &= \frac{1}{\lambda} \\ \text{Var}(X) &= \frac{1}{\lambda^2}. \end{aligned}$$

Το άθροισμα δύο ή περισσότερων εκθετικών μεταβλητών με ίδιο ρυθμο λ έχει κατανομή Gamma. Επίσης, αν $X \sim \text{Exp}(\lambda)$ και $Y = \sqrt{2X\sigma^2\lambda}$, τότε $Y \sim \text{Rayleigh}(\sigma)$.

Η εκθετική κατανομή έχει την ιδιότητα απώλειας μνήμης:

$$\begin{aligned} P(X \geq s + t | X \geq s) &= \frac{P(X \geq s+t, X \geq s)}{P(X \geq s)} = \frac{P(X \geq s+t)}{P(X \geq s)} = \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X \geq t). \end{aligned}$$

Επίσης, η εκθετική κατανομή μπορεί να αποτελέσει τη βάση για πιο περίπλοκα μοντέλα κατανομών που προκύπτουν από αλγεβρικές πράξεις μεταξύ εκθετικών μεταβλητών. Είναι γενικά ένα πολύτιμο εργαλείο. Για παράδειγμα, η κατανομή μίξη-εκθετικών χρησιμοποιείται για την μοντελοποίηση αφίξεων πακέτων στο Internet. Αν $X \sim \text{hyperexponential}(\mathbf{q}, \boldsymbol{\lambda})$, τότε

$$P(X > x) = \sum_i q_i e^{-\lambda_i x}.$$

1.7.4 Μίξη ΣΠΠ

Έστω οι συναρτήσεις $f_1(x)$ και $f_2(x)$ που είναι ΣΠΠ. Τότε και η συνάρτηση $f_3(x) = w_1 f_1(x) + w_2 f_2(x)$ με $w_1 + w_2 = 1, w_1, w_2 \geq 0$ είναι ΣΠΠ. Πράγματι,

$$\int_{-\infty}^{\infty} f_3(x) dx = w_1 \int_{-\infty}^{\infty} f_1(x) dx + w_2 \int_{-\infty}^{\infty} f_2(x) dx = w_1 + w_2 = 1,$$

ενώ μπορούμε να δείξουμε ότι για κάθε x , ισχύει $f_3(x) \geq 0$.

1.7.5 Θετικές μεταβλητές

Έστω X μια συνεχής θετική μεταβλητή, δηλαδή $X(\omega) \geq 0, \forall \omega \in \Omega$. Τότε

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x dF_X(x) = \\ &= \int_0^{\infty} x dP(X \leq x) = \\ &= \int_0^{\infty} -x [P(X > x)]' dx = \\ &= -x P(X > x) \Big|_0^{\infty} + \int_0^{\infty} x' P(X > x) dx = \\ &= \int_0^{\infty} P(X > x) dx, \end{aligned}$$

όπου η πιθανότητα $P(X > x)$ ονομάζεται **πιθανότητα ουράς** (tail probability).

Επίσης, αν $P(g(X) \geq a) = 1$ τότε

$$E[g(X)] = g(a) + \int_a^{\infty} g'(x) P(X > x) dx$$

και για τις διακριτές τ.μ.

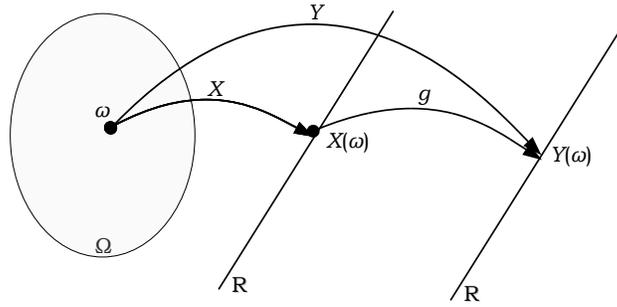
$$\begin{aligned} \sum_{i=1}^{\infty} P(X \geq i) &= \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} P(X = j) \\ &= \sum_{j=1}^{\infty} \sum_{i=1}^j P(X = j) \\ &= \sum_{j=1}^{\infty} j P(X = j) \\ &= E[X]. \end{aligned}$$

1.8 Συναρτήσεις τυχαίων μεταβλητών

Αν X με πεδίο τιμών \mathcal{S} είναι τ.μ. και $g : \mathcal{S} \rightarrow \hat{\mathcal{S}}$ συνάρτηση, τότε $Y = g(X)$ είναι τυχαία μεταβλητή, $Y : \Omega \rightarrow \hat{\mathcal{S}}$. Δείτε το σχήμα 1.8.

Αν η X είναι διακριτή, ισχύει

$$E[g(X)] = \sum_i p_i g(x_i)$$



Σχήμα 1.7: Η συνάρτηση τ.μ. είναι και αυτή τ.μ.

και αν είναι συνεχής,

$$E[g(X)] = \int_{S \cap \mathcal{S}} g(x) dF_X(x).$$

Με βάση τα παραπάνω μπορούμε να υπολογίσουμε και τις υπόλοιπες σχέσεις για τις συναρτήσεις τ.μ.

1.8.1 Γραμμικός μετασχηματισμός τ.μ.

Αν X είναι τ.μ. και a, b σταθερές, το $Y = aX + b$ λέγεται γραμμικός μετασχηματισμός. Επίσης, λέγεται και μετασχηματισμός θέσης και κλίμακας (location and scale transformation). Όπως είδαμε, μια γκαουσιανή τ.μ. μένει αναλλοίωτη σε τέτοιο μετασχηματισμό. Γενικά, όμως, ισχύει

$$E[Y] = aE[X] + b \text{ και } \text{Var}(Y) = a^2 \text{Var}(X).$$

Για την ΣΠΠ και με $a \neq 0$ ισχύει

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

1.8.2 Μέθοδος Jacobian

Με τη βοήθεια της μεθόδου αυτής μπορούμε να υπολογίσουμε την πυκνότητα πιθανότητας μιας συνάρτησης τ.μ. (όταν μπορούμε να ορίσουμε την αντίστροφη συνάρτηση). Έστω τ.μ. X με πυκνότητα $f_X(x)$, συνάρτηση g με ρίζες στην εξίσωση $y = g(x)$ στο πεδίο τιμών της Y τις $\{x_i\}$. Αν $Y = g(X)$, τότε:

$$f_Y(y) = \sum_{x_i} \frac{f_X(x)}{|g'(x)|} \Big|_{x=x_i}.$$

Παράδειγμα: Έστω $X \sim U(0, 1)$ και $Y = X^2$.

Λύση: Η Y έχει πεδίο τιμών το $[0, 1]$ και η εξίσωση $y = g(x)$ έχει ρίζες $x_1 = \sqrt{y}$ και $x_2 = -\sqrt{y}$, όπου κρατάμε μόνο την πρώτη καθώς η δεύτερη είναι εκτός πεδίου τιμών. Επίσης $|g'(x)| = 2|x|$ και $f_X(x)$ δίνεται παραπάνω. Έτσι, έχουμε

$$f_Y(y) = \left. \frac{f_X(x)}{|g'(x)|} \right|_{x=x_1} = \frac{1}{2\sqrt{y}}, y \in (0, 1].$$

Δοκιμάστε στο matlab αυτό: `hist((unifrnd(0,1,10000,1)).^2 , 50).`

1.8.3 Μέθοδος κατανομής

Έστω η τ.μ. X και η $Y = g(X)$.

Για τις διακριτές:

$$p_Y(y) = \sum_{\{x:g(x)=y\}} p_X(x).$$

Για τις συνεχείς:

$$F_Y(y) = P(g(X) \leq y) = \int_{\{x:g(x) \leq y\}} f_X(x) dx.$$

Παράδειγμα: Θα επαναλάβουμε το πρόβλημα της προηγούμενης υποενότητας όπου $X \sim U(0, 1)$ και $Y = X^2$.

Λύση: Για $y \in (0, 1]$

$$F_Y(y) = P(X^2 \leq y) = \stackrel{X \geq 0}{=} P(X \leq \sqrt{y}) = F_X(\sqrt{y}).$$

Συνεπώς

$$f_Y(y) = \frac{dF_Y(y)}{dy} = F'_X(\sqrt{y}) = (\sqrt{y})' f_X(\sqrt{y}) = \frac{1}{2\sqrt{y}}.$$

■

Κεφάλαιο 2

Πολλες μεταβλητές

2.1 Δύο τυχαίες μεταβλητές

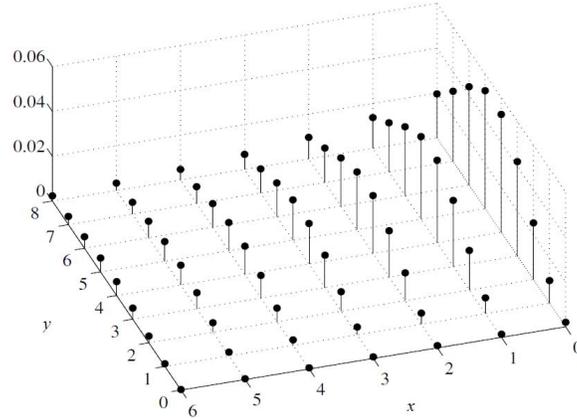
Σε ένα πείραμα είναι δυνατόν να μας ενδιαφέρει η γνώση δύο τυχαίων μεταβλητών και η αλληλεπίδρασή τους. Για παράδειγμα, αναρωτιόμαστε ποιά είναι η πιθανότητα να βρέξει αύριο δεδομένου ότι έβρεξε σήμερα. Η απάντηση μπορεί να δωθεί εφόσον γνωρίζουμε πως οι δύο τ.μ. συμπεριφέρονται από κοινού. Σε αυτήν την περίπτωση το πλήρες μοντέλο (δηλ. η πλήρης γνώση για τις δύο τ.μ.) απαιτεί τη γνώση της **από κοινού κατανομής** (joint distribution):

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y), \quad (2.1)$$

ή κατ' αντιστοιχία της κοινής χαρακτηριστικής συνάρτησης:

$$\Phi_{X,Y}(\omega_1, \omega_2) \doteq E[e^{j\omega_1 X} e^{j\omega_2 Y}] = E[e^{j(\omega_1 X + \omega_2 Y)}].$$

Εφόσον γνωρίζουμε το πλήρες μοντέλο, μπορούμε να υπολογίσουμε κάθε τυχαία μεταβλητή χωριστά αν



Σχήμα 2.1: Παράδειγμα από κοινού ΣΜΠ για δύο διακριτές τ.μ.

επιθυμούμε. Αυτό γίνεται χρησιμοποιώντας την ταυτότητα $A = A \cap \Omega$ ως εξής:

$$F_X(x) = \lim_{y \rightarrow +\infty} F_{XY}(x, y) =^{\text{συμβ.}} F_{XY}(x, +\infty)$$

$$F_Y(y) = \lim_{x \rightarrow +\infty} F_{XY}(x, y) =^{\text{συμβ.}} F_{XY}(+\infty, y)$$

$$\Phi_X(\omega) = \Phi_{X,Y}(\omega, 0)$$

$$\Phi_Y(\omega) = \Phi_{X,Y}(0, \omega).$$

Επίσης, είναι δυνατόν να ορίσουμε ροπές που αφορούν στις δύο τ.μ. ταυτόχρονα, π.χ. $E[X^n Y^m]$.

2.1.1 Για τις διακριτές τ.μ.

Στο σχήμα 2.1 εικονίζεται μια από κοινού ΣΜΠ δύο διακριτών τ.μ. $p_{X,Y}(x, y)$. Προφανώς ισχύει $\sum_x \sum_y p_{X,Y}(x, y) = 1$ και $p_{X,Y} \geq 0, \forall x, y$. Επίσης, μπορούμε να υπολογίσουμε άμεσα τις ΣΜΠ του X και του Y ξεχωριστά, οι οποίες ονομάζονται **οριακές ΣΜΠ** (marginal p.m.f.)

$$p_X(x) = \sum_y p_{X,Y}(x, y),$$

$$p_Y(y) = \sum_x p_{X,Y}(x, y).$$

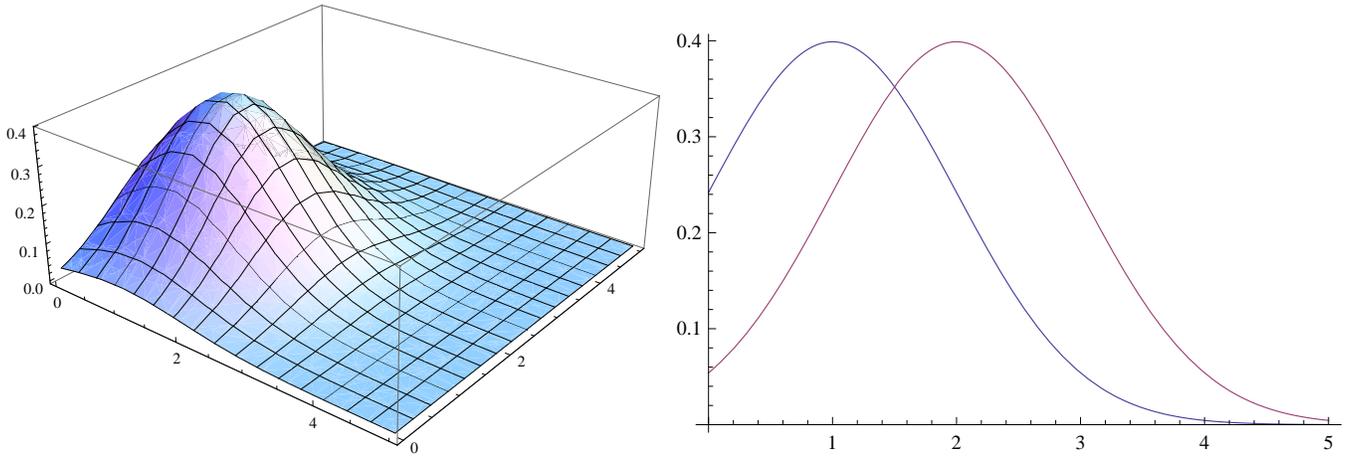
Για να υπολογίσουμε πιθανότητες ενδεχομένων χρειάζεται απλά να αθροίσουμε την από κοινού ΣΜΠ στο σωστό σύνολο τιμών. Για παράδειγμα, $P(X + Y = 3) = \sum_{x \in S_X} \sum_{y=3-x: y \in S_Y} p_{XY}(x, y)$.

Για την από κοινού χαρακτηριστική συνάρτηση ισχύει

$$\Phi_{X,Y}(\omega_1, \omega_2) = \sum_x \sum_y e^{j(\omega_1 x + \omega_2 y)} p_{X,Y}(x, y).$$

Για οποιαδήποτε συνάρτηση $g(X, Y)$ έχουμε

$$E[g(X, Y)] = \sum_x \sum_y g(x, y) p_{X,Y}(x, y)$$



Σχήμα 2.2: Παράδειγμα από κοινού ΣΠΠ για δύο συνεχείς τ.μ. (αριστερά) και οι αντίστοιχες οριακές ΣΠΠ (δεξιά).

και άρα

$$E[X^n Y^m] = \sum_x \sum_y x^n y^m p_{X,Y}(x, y).$$

2.1.2 Για τις συνεχείς τ.μ.

Εφόσον η από κοινού κατανομή είναι παραγωγίσιμη, ορίζουμε την **από κοινού ΣΠΠ** (joint p.d.f.) ως

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y},$$

η οποία έχει παρεμφερείς ιδιότητες με ΣΠΠ ($\int_{S_X} \int_{S_Y} f_{X,Y}(x, y) dy dx = 1$ και $f_{X,Y}(x, y) \geq 0, \forall x, y$). Στο σχήμα 2.2 εικονίζεται μια από κοινού ΣΠΠ για δύο συνεχείς τ.μ. Η γνώση της συμπεριφοράς κάθε τ.μ. δίνεται από τις οριακές ΣΠΠ που υπολογίζονται ως

$$\begin{aligned} f_X(x) &= \int_{S_Y} f_{X,Y}(x, y) dy, \\ f_Y(y) &= \int_{S_X} f_{X,Y}(x, y) dx. \end{aligned}$$

Για να υπολογίσουμε πιθανότητες ενδεχομένων θα πρέπει απλά να ολοκληρώσουμε στο σωστό διάστημα

$$P((X, Y) \in A) = \int \int_{(x,y) \in A} f_{XY}(x, y) dx dy.$$

Θα πρέπει πάντα να ισχύει $\int_{S_X} \int_{S_Y} f_{X,Y}(x, y) dy dx = 1$.

Για οποιαδήποτε συνάρτηση $g(X, Y)$ έχουμε

$$E[g(X, Y)] = \int_{S_X} \int_{S_Y} g(x, y) f_{X,Y}(x, y) dy dx$$

και άρα

$$E[X^n Y^m] = \int_{S_X} \int_{S_Y} x^n y^m f_{X,Y}(x, y) dy dx.$$

2.2 Κατανομές υπό συνθήκη

Προσοχή: όταν δεσμεύουμε μεταβλητές αναφερόμαστε σε διαφορετικό δειγματικό χώρο. Η κατανομή της X υπό συνθήκη της Y ορίζεται ως

$$F_{X|Y}(x, y) = \frac{P(X \leq x, Y = y)}{P(Y = y)},$$

όπου αντί για το $P(X \leq x, Y = y)$ μερικές φορές γράφουμε $P(X \leq x|Y = y)$. Η κατανομή ορίζεται πάντα για $P(Y = y) > 0$, δηλαδή το $\{X \leq x, Y = y\}$ είναι ένα ενδεχόμενο αν το $\{Y = y\}$ να μπορεί να συμβεί. Σε αντίθετη περίπτωση, δεν έχει νόημα να μιλάμε για την κατανομή αυτή. Στη συνέχεια, θα συναντήσουμε πολλές περιπτώσεις μεγεθών υπό συνθήκη, όπου παντού υπονοείται ότι ο παρανομαστής θα πρέπει πάντα να είναι μη μηδενική ποσότητα. Θα παραλήψουμε την αναφορά για πρακτικούς λόγους έχοντας υπόψη ότι τα αντίστοιχα μεγέθη δεν ορίζονται για τις περιπτώσεις που ο παρανομαστής είναι μηδεν.

2.2.1 Για τις διακριτές τ.μ.

Μια διακριτή τ.μ. X μπορεί να δεσμευτεί από ένα γεγονός και μπορούμε να μιλάμε για την ΣΜΠ υπό συνθήκη ενός γεγονότος

$$p_{X|A}(x) = P(X = x|A) = \frac{P(\{X = x\} \cap A)}{P(A)}.$$

Παρατηρήστε ότι για κάθε γεγονός έχουμε $\sum_x P(\{X = x\} \cap A) = P(A)$, το οποίο επαληθεύει ότι η $p_{X|A}$ είναι συνάρτηση μάζας πιθανότητας.

Παράδειγμα: Έστω X μια τ.μ.: το αποτέλεσμα ενός δίκαιου ζαριού και A το γεγονός ότι προκύπτει άρτιος αριθμός. Βρείτε την υπό συνθήκη συνάρτηση μάζας πιθανότητας $p_{X|A}(k)$.

Λύση:

$$\begin{aligned} p_{X|A}(k) &= P(X = k|A) = \\ &= \frac{P(X = k \cap A)}{P(A)} = \\ &= \begin{cases} 1/3 & k = 2, 4, 6 \\ 0 & \text{αλλιώς.} \end{cases} \end{aligned}$$

■

Κατ' επέκταση, μπορούμε να ορίσουμε την ΣΜΠ υπό συνθήκη του γεγονότος $\{Y = y\}$ για κάθε y που ικανοποιεί $P(Y = y) > 0$ λαμβάνοντας έτσι την υπό συνθήκη ΣΜΠ μεταξύ δύο τ.μ.:

$$p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)}.$$

Παρατηρήστε ότι η συνάρτηση $p_{X|Y}(x|y)$ είναι ΣΜΠ για δεδομένο y , δηλαδή είναι μια παραμετροποιημένη ΣΜΠ. Ισοδύναμα μπορούμε να γράψουμε τον γνώστό **κανόνα αλυσίδας** (chain rule)

$$p_{X,Y}(x, y) = p_Y(y)p_{X|Y}(x|y).$$

που ισχύει πάντα για όλα τα y ώστε $p_Y(y) > 0$. Ο κανόνας της αλυσίδας μπορεί να επεκταθεί για n τ.μ. με τη λογική ότι η από κοινού γνώση ισούτε με τη γνώση μιας μεταβλητής επί τη γνώση μιας δεύτερης μεταβλητής δεδομένης της πρώτης, κ.ο.κ.

Παράδειγμα: Έστω η από κοινού ΣΜΠ $P(X = 1, Y = 1) = \frac{1}{3}, P(X = 2, Y = 1) = \frac{1}{3}, P(X = 1, Y = 2) = \frac{1}{3}, P(X = 2, Y = 2) = 0$. Τότε μπορούμε να υπολογίσουμε ότι

$$\begin{aligned} p_{X|Y}(x|1) &= \left\{ \frac{1}{2}, \frac{1}{2} \right\} \\ p_{X|Y}(x|2) &= \{1, 0\} \\ E[X|Y = 1] &= \frac{3}{2} \\ E[X|Y = 2] &= 1 \\ E[X] &= P(Y = 1) E[X|Y = 1] + P(Y = 2) E[X|Y = 2] = \frac{2}{3} \cdot \frac{3}{2} + \frac{1}{3} \cdot 1 = \frac{4}{3}. \end{aligned}$$

Είναι φανερό από το παράδειγμα ότι το μέγεθος $E[X|Y = y] = \sum_x x p_{X|Y}(x|y)$ έχει μεγάλη σημασία: ονομάζεται υπό συνθήκη αναμενόμενη τιμή (conditional expectation) και είναι συνάρτηση του y . Χρησιμοποιώντας αυτό το μέγεθος έχουμε

$$\begin{aligned} E[X] &= \sum_x x p_X(x) \\ &= \sum_x x \sum_y p_{X,Y}(x, y) \\ &= \sum_x x \sum_y p_Y(y) p_{X|Y}(x|y) \\ &= \sum_y p_Y(y) \sum_x x p_{X|Y}(x|y) \\ &= \sum_y p_Y(y) E[X|Y = y], \end{aligned}$$

που είναι ένας πολύ χρήσιμος τρόπος υπολογισμού της μέσης τιμής της X όταν γνωρίζουμε την υπο συνθήκη αναμενόμενη τιμή και την κατανομή της Y . Το ίδιο μπορούμε να κάνουμε και με μια οποιαδήποτε διαμέριση του δειγματικού χώρου A_1, \dots, A_n :

$$E[X] = \sum_{k=1}^n P(A_k) E[X|A_k].$$

Παράδειγμα: Ένας κρατούμενος ζει σε ένα κελί με τρεις πόρτες. Μια πόρτα οδηγεί στην ελευθερία, μια άλλη οδηγεί μέσα από ένα τούνελ μιας ημέρας ξανά στο κελί και η τρίτη κάνει το ίδιο μέσα από ένα τούνελ τριών ημερών. Αν υποθέσουμε ότι ο κρατούμενος ξεχνάει τι επιλογές έκανε τις προηγούμενες μέρες, πόσο καιρό κατά μέσο όρο θα χρειαστεί για να απελευθερωθεί;

Λύση: Έστω X η πόρτα που επιλέγεται κατά τυχή, $S_d = \{1, 2, 3\}$ και $P(X = i) = \frac{1}{3}, i = 1, 2, 3$. Έστω Y ο χρόνος σε μέρες για τη σωτηρία. Τότε λόγω της αμνησίας του κρατουμένου έχουμε

$$\begin{aligned} E[Y|X = 1] &= 0 \\ E[Y|X = 2] &= E[Y] + 1 \\ E[Y|X = 3] &= E[Y] + 3. \end{aligned}$$

Άρα

$$EY = E_X[E[Y|X]] = \sum_i p_i E[Y|X = i] = \frac{1}{3}0 + \frac{1}{3}(EY + 1) + \frac{1}{3}(EY + 3),$$

από το οποίο καταλήγουμε ότι

$$EY = 4.$$

■

2.2.2 Για τις συνεχείς τ.μ.

Μια συνεχής τ.μ. X μπορεί επίσης να δεσμευτεί από ένα γεγονός

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{P(X \in A)} & \text{αν } x \in A, \\ 0 & \text{αλλού.} \end{cases}$$

Παρατηρήστε ότι

$$P(X \in B|X \in A) = \frac{P(X \in B, X \in A)}{P(X \in A)} = \frac{\int_{A \cap B} f_X(x) dx}{P(X \in A)} = \int_B f_{X|A}(x) dx,$$

που σημαίνει ότι μπορούμε να χρησιμοποιήσουμε την $f_{X|A}$ ως ΣΠΠ για να υπολογίσουμε πιθανότητες. Υποθέτοντας ότι οι συναρτήσεις είναι παραγωγίσιμες, άρα και συνεχείς, μπορούμε να ακολουθήσουμε τον παρακάτω συλλογισμό

$$\begin{aligned} f_{X,Y}(x, y) &= \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y} \\ &= \frac{P(x \leq X \leq x + \partial x, y \leq Y \leq y + \partial y)}{\partial x \partial y} \\ &= \frac{P(x \leq X \leq x + \partial x | y \leq Y \leq y + \partial y) P(y \leq Y \leq y + \partial y)}{\partial x \partial y} \\ &= \frac{P(x \leq X \leq x + \partial x | y \leq Y \leq y + \partial y) f_Y(y)}{\partial x} \\ &\approx \frac{P(x \leq X \leq x + \partial x | Y = y) f_Y(y)}{\partial x} \\ &= f_{X|Y}(x|y) f_Y(y), \end{aligned}$$

που είναι ο κανόνας της αλυσίδας για τις συνεχείς τ.μ. Άρα, καταλήγουμε στον ορισμό της υπό συνθήκη ΣΠΠ που είναι

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

Παρατηρήστε ότι η $f_{X|Y}(x|y)$ είναι ΣΠΠ για σταθερό y , είναι δηλαδή μια παραμετροποιημένη ΣΠΠ. Μπορούμε να την χρησιμοποιήσουμε για να υπολογίσουμε πιθανότητες

$$P(X \in A|Y = y) = \int_A f_{X|Y}(x|y) dx$$

και μέσες τιμές

$$\begin{aligned} E[g(X)|Y = y] &= \int g(x) f_{X|Y}(x|y) dx \\ E[g(X)] &= \int E[g(X)|Y = y] f_Y(y) dy. \end{aligned}$$

2.2.3 Η περίπτωση μιας διακριτής και μιας συνεχούς τ.μ.

Αν υποθέσουμε ότι η X είναι διακριτή τ.μ. ενώ η Y συνεχής, τότε για το ζευγάρι (X, Y) μπορούμε να ορίσουμε την από κοινού κατανομή (2.1) κανονικά. Αντίθετα, τόσο η από κοινού ΣΜΠ όσο και η από κοινού ΣΠΠ δεν έχουν νόημα, η πρώτη γιατί θα έδινε μηδενικές πιθανότητες και η δεύτερη γιατί η συνάρτηση F_{XY} δεν είναι παραγωγίσιμη. Χρησιμοποιώντας, όμως, επιχειρήματα από τις προηγούμενες παραγράφους, μπορούμε να καταλήξουμε στο συμπέρασμα ότι θα ισχύει

$$\frac{f_{Y|X}(y|x)}{f_Y(y)} = \frac{p_{X|Y}(x|y)}{p_X(x)},$$

για όλα τα (x, y) έτσι ώστε $p_X(x) > 0, f_Y(y) > 0$.

Παράδειγμα: Εκπέμπεται ένα ψηφιακό σήμα X με $P(X = 1) = p$ και $P(X = 0) = 1 - p$. Στο λαμβανόμενο σήμα $Y = X + N$ προστίθεται θόρυβος $N \sim \mathcal{N}(0, 1)$. Υπολογίστε την πιθανότητα $P(X = 1)$ σαν συνάρτηση της τιμής y του δέκτη.

Λύση: Αρχικά παρατηρούμε ότι δεσμευμένου του $X = x$, η έξοδος είναι το άθροισμα γκαουσιανής και σταθεράς, δηλαδή $Y|X \sim \mathcal{N}(x, 1)$. Επίσης, ισχύει $f_Y(y) = P(X = 1) f_Y(y|x = 1) + P(X = 0) f_Y(y|x = 0)$. Στη συνέχεια έχουμε

$$\begin{aligned} P(X = 1|Y = y) &= \frac{p_X(1)f_{Y|X}(y|1)}{f_Y(y)} = \frac{p_X(1)f_{Y|X}(y|1)}{p_X(1)f_{Y|X}(y|1) + p_X(0)f_{Y|X}(y|0)} \\ &= \frac{\frac{p}{\sqrt{2\pi}}e^{-(y-1)^2/2}}{\frac{p}{\sqrt{2\pi}}e^{-(y-1)^2/2} + \frac{1-p}{\sqrt{2\pi}}e^{-y^2/2}} \\ &= \frac{p}{p + (1-p)e^{\frac{1}{2}-y}}. \end{aligned}$$

■

2.3 Πολλές τυχαίες μεταβλητές

Η γνώση της συμπεριφοράς πολλών μεταβλητών X_1, X_2, \dots, X_n σε ένα πείραμα περιγράφεται από το πλήρες μοντέλο:

$$F_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n),$$

το οποίο είναι ισοδύναμο με τη γνώση της χαρακτηριστικής συνάρτησης

$$\Phi_{X_1 X_2 \dots X_n}(\omega_1, \omega_2, \dots, \omega_n) = E[e^{j(\omega_1 X_1 + \omega_2 X_2 + \dots + \omega_n X_n)}]$$

ή τη γνώση όλων των ροπών (όλων των πιθανών συνδυασμών).

2.3.1 Ανεξαρτησία

Οι τ.μ. X_1, X_2, \dots, X_n είναι από κοινού ανεξάρτητες (mutually independent) αν και μόνο αν για κάθε διάλυμα τιμών (x_1, x_2, \dots, x_n) ισχύει ένα από τα

$$\begin{aligned} F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) &= F_{X_1}(x_1)F_{X_2}(x_2) \dots F_{X_n}(x_n) \\ p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) &= p_{X_1}(x_1)p_{X_2}(x_2) \dots p_{X_n}(x_n) \\ f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) &= f_{X_1}(x_1)f_{X_2}(x_2) \dots f_{X_n}(x_n). \end{aligned}$$

Μια από τις παραπάνω ισότητες υπονοεί και τις υπόλοιπες. Αυτό σημαίνει ότι αν αποδείξουμε ότι για δύο τ.μ. ισχύει $f_{X,Y}(x,y) = f_X(x)f_Y(y), \forall x,y$ τότε αυτές οι τ.μ. είναι ανεξάρτητες μεταξύ τους. Όταν δύο τ.μ. είναι ανεξάρτητες μεταξύ τους λέμε ότι είναι **ανεξάρτητες ανά ζεύγη** (pairwise independent). Είναι δυνατόν τρεις τ.μ. να είναι όλες ανεξάρτητες ανά ζεύγη αλλά να μην είναι από κοινού ανεξάρτητες. Όταν n τ.μ. είναι από κοινού ανεξάρτητες, τότε οποιοσδήποτε συνδυασμός αυτών είναι επίσης ανεξάρτητος (άρα και ανά ζεύγη), κάτι που αποδεικνύεται εύκολα λαμβάνοντας οριακές κατανομές ως προς τις μεταβλητές που μας ενδιαφέρει. Για την ανεξαρτησία των τ.μ. γράφουμε $X_1 \perp X_2 \perp \dots \perp X_n$.

Αν $X_1 \perp X_2 \perp \dots \perp X_n$, τότε

$$\begin{aligned} E[X_1 X_2 \dots X_n] &= E[X_1]E[X_2] \dots E[X_n] \\ \Phi_{X_1, X_2, \dots, X_n}(\omega_1, \omega_2, \dots, \omega_n) &= \Phi_{X_1}(\omega_1)\Phi_{X_2}(\omega_2) \dots \Phi_{X_n}(\omega_n). \end{aligned}$$

Για παράδειγμα, έστω X, Y συνεχείς, ανεξάρτητες τ.μ. Τότε ισχύει

$$\begin{aligned} E[XY] &= \int_{S_Y} \int_{S_X} xy f_{X,Y}(x,y) dx dy =^{\text{ανεξ.}} \int_{S_Y} \int_{S_X} xy f_X(x) f_Y(y) dx dy \\ &= \int_{S_X} x f_X(x) dx \int_{S_Y} y f_Y(y) dy = E[X]E[Y]. \end{aligned}$$

Είναι, επίσης, δυνατόν να ορίσουμε την ανεξαρτησία δύο τ.μ. X_1, X_2 δεδομένου μιας άλλης τ.μ. X_3 , δηλαδή $X_1 \perp X_2 | X_3$. Αυτό δεν σημαίνει ότι οι X_1, X_2 είναι ανεξάρτητες, αλλά ότι η ανεξαρτησία προκύπτει όταν γνωρίζουμε το X_3 . Όταν ισχύει κάτι τέτοιο, λέμε ότι οι τ.μ. ορίζουν μια αλυσίδα Markov $X_1 \rightarrow X_2 \rightarrow X_3$. Σε επόμενο κεφάλαιο θα δούμε και άλλους ορισμούς για τις αλυσίδες Markov.

Η από κοινού ανεξαρτησία μιας συλλογής τ.μ. υποδηλώνει ότι το κοινό πείραμα των τ.μ. έχει το ίδιο αποτέλεσμα με τη συλλογή πειραμάτων για κάθε τ.μ. ξεχωριστά. Όταν δύο τ.μ. συνδέονται με κάποιο ντετερμινιστικό νόμο, για παράδειγμα $Y = f(X)$, αναμένουμε ότι δε θα είναι ανεξάρτητες¹.

2.3.2 Ανάλυση με υπό συνθήκη κατανομές

Χρησιμοποιώντας τον κανόνα της αλυσίδας μπορούμε να γράψουμε

$$p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = p_{X_1}(x_1)p_{X_2|X_1}(x_2|x_1) \dots p_{X_n|X_{n-1}, \dots, X_1}(x_n|x_{n-1}, \dots, x_1)$$

και

$$f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = f_{X_1}(x_1)f_{X_2|X_1}(x_2|x_1) \dots f_{X_n|X_{n-1}, \dots, X_1}(x_n|x_{n-1}, \dots, x_1).$$

Επιπλέον, σε περίπτωση που οι τ.μ. X_1, \dots, X_n έχουν την ιδιότητα Markov θα ισχύει και

$$p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = p_{X_1}(x_1)p_{X_2|X_1}(x_2|x_1) \dots p_{X_n|X_{n-1}}(x_n|x_{n-1})$$

και

$$f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = f_{X_1}(x_1)f_{X_2|X_1}(x_2|x_1) \dots f_{X_n|X_{n-1}}(x_n|x_{n-1}).$$

¹Πράγματι, για να είναι Y, X ανεξάρτητες θα πρέπει $f(x) = c$.

2.3.3 Ροπές πολλών μεταβλητών

Για το μοντέλο των ροπών απαιτείται η γνώση όλων των ροπών $E[X_1^{\xi_1} X_2^{\xi_2} \dots X_n^{\xi_n}]$ (όλων των πιθανών συνδυασμών $\xi_i = 0, 1, \dots, \forall i$). Το άθροισμα των ξ_i καθορίζει την **τάξη** της ροπής ενώ, αν υπάρχουν παραπάνω από ένα $i : \xi_i > 0$, τότε ονομάζεται **ετεροροπή**. Άρα το $E[X_1 X_2^2 X_3]$ είναι μια ετεροροπή τέταρτης τάξης. Μοντέλα ροπών μεγαλύτερα της τέταρτης τάξης χρησιμοποιούνται σπάνια λόγω της δυσκολίας υπολογισμού.

Οι κεντρικές ροπές ονομάζονται και **σωρείτες** (cumulants).

Συσχέτιση

Η κεντρική ετεροροπή δεύτερης τάξης ονομάζεται συσχέτιση (covariance), και γράφεται

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y].$$

Η συσχέτιση λαμβάνεται ως αποτέλεσμα των τιμών των τ.μ. και δεν πρέπει να εκληφθεί ως σχέση εξάρτησης (η εξάρτηση δύο μεταβλητών υπονοεί την ύπαρξη κάποιου ντετερμινιστικού νόμου που εμπλέκει τις δύο τ.μ. έστω μερικώς). Λέμε ότι δύο τ.μ. X, Y είναι **ασυσχέτιστες** όταν $\text{Cov}(X, Y) = 0$. Προφανώς από τον ορισμό έχουμε ότι δύο ανεξάρτητες τ.μ. είναι και ασυσχέτιστες μεταξύ τους (αφού θα ισχύει $E[XY] = E[X]E[Y]$ και άρα $\text{Cov}(X, Y) = 0$). Το αντίθετο, όμως, δεν ισχύει πάντα.

Παράδειγμα: (Δύο ασυσχέτιστες μεταβλητές μπορεί είναι εξαρτημένες) Έστω η τ.μ. $\Theta \sim U(0, 2\pi)$ και $X = \cos \Theta$, $Y = \sin \Theta$. Το σημείο (X, Y) βρίσκεται πάντοτε πάνω στον μοναδιαίο κύκλο. Κάνοντας μετασχηματισμούς (π.χ. χρησιμοποιώντας τη μέθοδο Jacobian) μπορούμε να υπολογίσουμε τις οριακές κατανομές των X, Y που είναι μη μηδενικές στο διάστημα $(-1, 1)$. Συνεπώς, εφόσον η από κοινού κατανομή πρέπει να είναι 0 εντός του μοναδιαίου δίσκου, οι X, Y είναι εξαρτημένες². Όμως,

$$\begin{aligned} E[XY] &= E[\sin \Theta \cos \Theta] = \frac{1}{2\pi} \int_0^{2\pi} \sin \theta \cos \theta d\theta \\ &= \frac{1}{4\pi} \int_0^{2\pi} \sin 2\theta d\theta = 0 \\ E[X] &= \int_0^{2\pi} \cos \theta d\theta = 0 \\ E[Y] &= \int_0^{2\pi} \sin \theta d\theta = 0, \end{aligned}$$

δηλαδή $\text{Cov}(X, Y) = 0$. Άρα, οι X, Y είναι ασυσχέτιστες, αλλά όχι ανεξάρτητες.

Όταν διαχειριζόμαστε στατιστικά στοιχεία, μελετούμε συνήθως τη συσχέτιση σε μια κανονικοποιημένη μορφή της. Συγκεκριμένα ορίζουμε ως **συντελεστή συσχέτισης** (correlation coefficient) το

$$\rho_{X,Y} \doteq \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} = \frac{E[XY] - E[X]E[Y]}{\sqrt{E[X^2] - E[X]^2} \sqrt{E[Y^2] - E[Y]^2}}.$$

Πολλές φορές, για συντομία και εφόσον δεν υπάρχει πιθανότητα λάθους, ο συντελεστής αυτός παρουσιάζεται ως ρ . Επίσης, η μεταβλητότητα μπορεί να συμβολίζεται με $\sigma_X^2 \equiv \text{Var}(X)$. Ισχύει πάντα

$$-1 \leq \rho \leq 1,$$

όπου το θετικό πρόσημο δείχνει την τάση οι δύο μεταβλητές να λαμβάνουν ομόσημες τιμές. Επίσης, το $\rho = 0$ σημαίνει ότι οι δύο τ.μ. είναι ασυσχέτιστες (προσοχή όχι απαραίτητα ανεξάρτητες!). Όταν μάλιστα έχουμε $\rho = \pm 1$, τότε οι δύο τ.μ. είναι γραμμικά συσχετισμένες, δηλ. $Y = aX + b$.

²Η εξάρτηση είναι φανερή αφού θα πρέπει πάντα $X^2 + Y^2 = 1$.

2.3.4 Συναρτήσεις πολλών μεταβλητών

Οι συναρτήσεις πολλών μεταβλητών είναι και αυτές τ.μ.

Μέθοδος Jacobian για πολλές μεταβλητές

Είναι δυνατόν να επεκτείνουμε τη μέθοδο που περιγράφηκε στο προηγούμενο κεφάλαιο για την περίπτωση συνάρτησης πολλών μεταβλητών. Έστω τ.μ. X, Y με πυκνότητα $f_{XY}(x, y)$ και συναρτήσεις g, h με ρίζες στο σύστημα $\begin{cases} u = g(x, y) \\ w = h(x, y) \end{cases}$ τις $\{u_i, w_i\}$ εντός του πεδίου τιμών των U, W . Αν $U = g(X, Y)$ και $W = h(X, Y)$, τότε:

$$f_{UW}(u, w) = \sum_i \frac{f_{XY}(x, y)}{|J(x, y)|} \Bigg|_{(x=u_i, y=w_i)},$$

όπου

$$J(x, y) = \begin{vmatrix} \frac{\partial g(x, y)}{\partial x} & \frac{\partial g(x, y)}{\partial y} \\ \frac{\partial h(x, y)}{\partial x} & \frac{\partial h(x, y)}{\partial y} \end{vmatrix},$$

ονομάζεται Ιακωβιανή (Jacobian) ορίζουσα. Η μέθοδος γενικεύεται για περισσότερες μεταβλητές αλλά γίνεται αρκετά περίπλοκη. Επίσης, πρέπει πάντα να αναφερόμαστε σε συναρτήσεις που κάνουν 1-1 αντιστοίχιση (δηλ. που μπορούμε να λύσουμε το αντίστοιχο σύστημα και να τις αντιστρέψουμε).

Παράδειγμα: Έστω οι τ.μ. X, Y και $U = X + Y$. Βρείτε την ΣΠΠ της U .

Λύση: Θεωρούμε τις εξισώσεις $g(x, y) = x + y$ και $h(x, y) = x - y$. Προφανώς έχουμε ότι $U = g(X, Y)$, και θεωρούμε και την βοηθητική τ.μ. $W = h(X, Y)$.

Το σύστημα $\begin{cases} u = x + y \\ w = x - y \end{cases}$ έχει τη μοναδική λύση $(x, y) = (\frac{u+w}{2}, \frac{u-w}{2})$. Άρα μπορούμε να υπολογίσουμε την από κοινού ΣΠΠ χρησιμοποιώντας τη μέθοδο Jacobian. Έχουμε

$$J(x, y) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$$

και

$$f_{UW}(u, w) = \frac{f_{XY}(x, y)}{2} \Bigg|_{(x, y) = (\frac{u+w}{2}, \frac{u-w}{2})} = \frac{1}{2} f_{XY} \left(\frac{u+w}{2}, \frac{u-w}{2} \right).$$

Στη συνέχεια μπορούμε να υπολογίσουμε την οριακή ΣΠΠ που ζητούμε

$$\begin{aligned} f_U(u) &= \int_{-\infty}^{\infty} \frac{1}{2} f_{XY} \left(\frac{u+w}{2}, \frac{u-w}{2} \right) dw \\ &\stackrel{\text{ανεξ.}}{=} \int_{-\infty}^{\infty} \frac{1}{2} f_X \left(\frac{u+w}{2} \right) f_Y \left(\frac{u-w}{2} \right) dw \\ &= \int_{-\infty}^{\infty} f_X(w) f_Y(u-w) dw. \end{aligned}$$

Αυτό είναι βέβαια ανεμενόμενο γνωρίζοντας ότι η χαρακτηριστική συνάρτηση του αθροίσματος είναι το γινόμενο των χαρακτηριστικών συναρτήσεων (από ιδιότητες ΜΣ Fourier η ΣΠΠ του αθροίσματος θα πρέπει να είναι η συνέλιξη των αντίστοιχων ΣΠΠ). ■

Μέθοδος κατανομής για πολλές μεταβλητές

Για να υπολογίσουμε την ΣΠΠ της συνάρτησης U πολλών μεταβλητών χρησιμοποιούμε την κατανομή, η οποία υπολογίζεται ως εξής (παράδειγμα για δύο μεταβλητές):

$$F_U(u) = P(U \leq u) = P(g(X, Y) \leq u) = \iint_{(x,y):g(x,y) \leq u} f_{X,Y}(x, y) dx dy.$$

Για εφαρμογή αυτής της μεθόδου δες το παράδειγμα της ενότητας 2.7.1. Στο επόμενο παράδειγμα παρουσιάζεται μια παραπλήσια μέθοδος.

Παράδειγμα: (Το πλάτος ενός ανεξάρτητου διανύσματος Gauss) Έστω $X, Y \sim \mathcal{N}(0, \sigma^2)$ ανεξάρτητες μεταξύ τους και $R = \sqrt{X^2 + Y^2}$. Αποδείξτε ότι η τ.μ. R ακολουθεί την κατανομή Rayleigh.

Λύση: Γνωρίζουμε ότι

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}.$$

Για μια απειροστή επιφάνεια ισχύει

$$dA = dx dy = r dr d\theta,$$

όπου dr είναι η στοιχειώδης ακτίνα και $d\theta$ η στοιχειώδης γωνία (και $r = \sqrt{x^2 + y^2}$, $\theta = \arctan(\frac{y}{x})$). Θεωρούμε μια συγκεκριμένη επιφάνεια A όπου θα πρέπει

$$\begin{aligned} P((R, \Theta) \in A) &= P((X, Y) \in A) = \frac{1}{2\pi\sigma^2} \iint_{(x,y) \in A} e^{-\frac{x^2+y^2}{2\sigma^2}} dx dy \\ &= \frac{1}{2\pi\sigma^2} \iint_{(r,\theta) \in A} e^{-\frac{r^2}{2\sigma^2}} r dr d\theta. \end{aligned}$$

Αυτό σημαίνει ότι θα πρέπει

$$f_{R\Theta}(r, \theta) = \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}}.$$

Τελικά παίρνουμε

$$f_R(r) = \int_0^{2\pi} \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} d\theta = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}, r \geq 0.$$

και

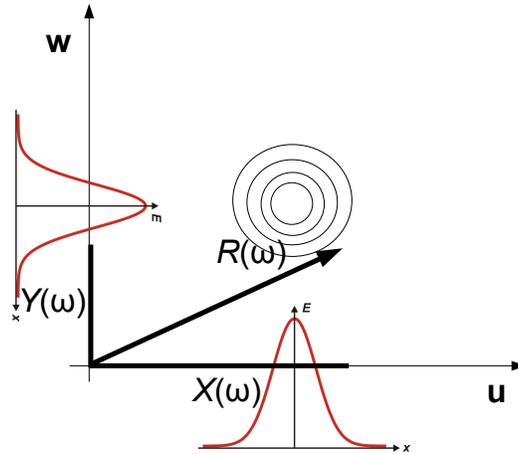
$$f_\Theta(\theta) = \int_0^\infty \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} dr = \frac{1}{2\pi}, \theta \in [0, 2\pi].$$

Άρα το μήκος του διανύσματος είναι κατανεμημένο κατά Rayleigh και η γωνία είναι ομοιόμορφα κατανεμημένη.

Στο σχήμα φαίνεται το διάνυσμα Gauss (δηλ. το τυχαίο διάνυσμα $\mathbf{v} = X\mathbf{u} + Y\mathbf{w}$). Το διάνυσμα αυτό έχει τυχαίο μήκος που όπως αποδείξαμε ακολουθεί την κατανομή Rayleigh όταν $X, Y \sim \mathcal{N}(0, \sigma)$. ■

Ερώτηση: Επαναλάβετε την απόδειξη του παραπάνω παραδείγματος χρησιμοποιώντας τη μέθοδο Jacobian.

Δοκιμάστε στο matlab αυτό: `rayleigh.m`



Σχήμα 2.3: Ένα realization ενός διανύσματος gaussian. Οι ομόκεντροι κύκλοι (ελλείψεις όταν η μεταβλητότητα δεν είναι ίδια για τις δύο διαστάσεις) απεικονίζουν σημεία με ίση πυκνότητα πιθανότητας.

Άθροισμα πολλών τ.μ.

Το άθροισμα τ.μ. είναι ειδική περίπτωση συνάρτησης. Όπως είδαμε στο προηγούμενο παράδειγμα, το άθροισμα δύο ανεξάρτητων τ.μ. έχει ΣΠΠ (ή ΣΜΠ) που δίνεται από τη συνέλιξη των αντίστοιχων ΣΠΠ (ή ΣΜΠ). Για την περίπτωση των διακριτών X, Y και $W = X + Y$, έχουμε

$$\begin{aligned}
 p_W(w) &= P(X + Y = w) \\
 &= \sum_{\{(x,y):x+y=w\}} P(X = x, Y = y) \\
 &= \sum_x P(X = x, Y = w - x) \\
 &\stackrel{\text{ανεξ.}}{=} \sum_x p_X(x) p_Y(w - x).
 \end{aligned}$$

Η ιδιότητα αυτή ισχύει και για περισσότερες τ.μ. Έστω X_1, X_2, \dots, X_n ανεξάρτητες τ.μ. με χαρακτηριστικές συναρτήσεις $\Phi_{X_1}(\omega), \Phi_{X_2}(\omega), \dots, \Phi_{X_n}(\omega)$ αντίστοιχα, και $Y = \sum_{i=1}^n X_i$. Ισχύει

$$\Phi_Y(\omega) \stackrel{\text{ανεξ.}}{=} \prod_{i=1}^n \Phi_{X_i}(\omega).$$

Το άθροισμα τ.μ. έχει μια ιδιότητα σύγκλισης που θα δούμε στο επόμενο κεφάλαιο. Επίσης, για τις πρώτες ροπές ισχύει (χωρίς να απαιτούμε οι X_i να είναι ανεξάρτητες)

$$E\left[\sum_i a_i X_i\right] = \sum_i a_i E[X_i]$$

και

$$\begin{aligned}
 \text{Var}\left(\sum_i a_i X_i\right) &= \sum_i a_i^2 \text{Var}(X_i) + \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2a_i a_j \text{Cov}(X_i, X_j) \\
 &\stackrel{\text{ανεξ.}}{=} \sum_i a_i^2 \text{Var}(X_i).
 \end{aligned}$$

Παράδειγμα: Έστω $X_i, i = 1, \dots, N$ ανεξάρτητες τ.μ. κατανομημένες με τον ίδιο τρόπο (i.i.d ή independent identically distributed) με μέση τιμή $E[X]$ και μεταβλητότητα $\text{Var}(X)$ και N διακριτή τυχαία μεταβλητή, $\mathcal{S}_d = \{1, 2, \dots\}$ με μέση τιμή $E[N]$ και μεταβλητότητα $\text{Var}(N)$. (α) Βρείτε τη μέση τιμή του $S = \sum_{i=1}^N X_i$. (β) Βρείτε την μεταβλητότητα S . (γ) Βρείτε τη χαρακτηριστική συνάρτηση $\Phi_S(\omega)$ δεδομένου ότι $\Phi_X(\omega) = \frac{\lambda}{\lambda + j\omega}$ και $G_N(z) = e^{\lambda(z-1)}$.

Λύση:

(α) Υπολογίζουμε με δέσμευση ως προς το N

$$\begin{aligned} E[S] &= E_N[E[S|N]] \\ &= \sum_i p_N(i) E\left[\sum_{k=1}^i X_k | N = i\right] \\ &= \sum_i p_N(i) \sum_{k=1}^i E[X_k] \\ &= \sum_i p_N(i) i E[X] \\ &= E[N]E[X]. \end{aligned}$$

(β) Υπολογίζουμε πρώτα

$$\begin{aligned} \text{Var}(S) &= \text{Var}\left(\sum_{i=1}^N X_i\right) \\ &= E\left[\left(\sum_{i=1}^N X_i\right)^2 - E\left[\sum_{i=1}^N X_i\right]^2\right] \\ &= E_N\left[E\left[\left(\sum_{i=1}^N X_i\right)^2 | N\right]\right] - (E[N]E[X])^2. \end{aligned}$$

Κατόπιν, θεωρούμε ότι $N = k$, δηλ. μια σταθερά, τότε

$$\begin{aligned} E\left[\left(\sum_{i=1}^N X_i\right)^2 | N = k\right] &= E\left[\sum_{i=1}^k X_i \sum_{j=1}^k X_j | N = k\right] \\ &= \sum_{i=1}^k \sum_{j=1}^k E[X_i X_j] \\ &= kE[X^2] + k(k-1)E[X]^2, \end{aligned}$$

όπου το τελευταίο προκύπτει αν αναλογιστούμε ότι $E[X_i X_j] = \overset{\text{ανεξ.}}{E[X]^2}, i \neq j$ και $E[X_i X_j] = E[X^2], i = j$. Στη συνέχεια μπορούμε να υπολογίσουμε το ζητούμενο

$$\begin{aligned}
\text{Var}(S) &= E_N \left[E \left[\left(\sum_{i=1}^N X_i \right)^2 \middle| N \right] \right] - (E[N]E[X])^2 \\
&= E_N [NE[X^2] + N(N-1)E[X]^2] - (E[N]E[X])^2 \\
&= E[N]E[X^2] + E[N^2]E[X]^2 - E[N]E[X]^2 - (E[N]E[X])^2 \\
&= E[N]\text{Var}(X) + E[X]^2\text{Var}(N).
\end{aligned}$$

(γ) Γενικά ισχύει

$$\begin{aligned}
\Phi_S(\omega) &= E[e^{j\omega S}] = E_N[E[e^{j\omega \sum_{i=1}^N X_i} | N]] \\
&= E_N[E[\prod_{i=1}^N e^{j\omega X_i}]] \stackrel{\text{ανεξ.}}{=} E_N[\prod_{i=1}^N E[e^{j\omega X_i}]] \\
&= E_N[\prod_{i=1}^N \Phi_X(\omega)] = \sum_{k=1}^{\infty} P(N=k) \Phi_X^k(\omega) \\
&= G_N(\Phi_X(\omega)),
\end{aligned}$$

αφού $G_N(z) \doteq \sum_{k=1}^{\infty} P(N=k) z^k$. Ειδικά για τους δοθέντες μετασχηματισμούς έχουμε:

$$\begin{aligned}
\Phi_S(\omega) &= G_N\left(\frac{\lambda}{\lambda + j\omega}\right) = e^{\lambda\left(\frac{\lambda}{\lambda + j\omega} - 1\right)} \\
&= e^{-\frac{1}{1 + j\omega/\lambda}}.
\end{aligned}$$

■

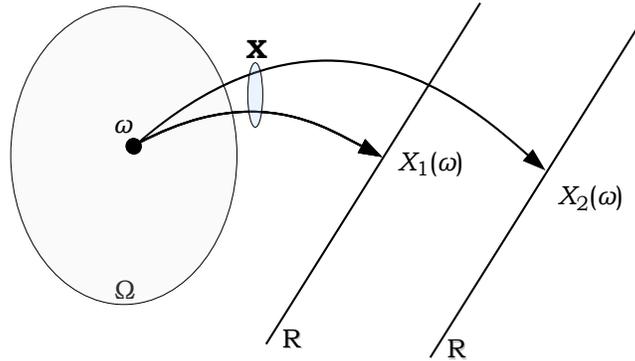
Μέγιστο και ελάχιστο πολλών τ.μ.

Έστω X_1, \dots, X_n τ.μ. και $Y = \max\{X_1, \dots, X_n\}$. Τότε μπορούμε να υπολογίσουμε την κατανομή του Y

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) = P(\max\{X_1, \dots, X_n\} \leq y) \\
&= P(X_1 \leq y, \dots, X_n \leq y) = F_{X_1, \dots, X_n}(y, \dots, y) \\
&\stackrel{\text{ανεξ.}}{=} \prod_{i=1}^n F_{X_i}(y).
\end{aligned}$$

Αντίστοιχα για $Y = \min\{X_1, \dots, X_n\}$ έχουμε

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) = P(\min\{X_1, \dots, X_n\} \leq y) \\
&= 1 - P(\min\{X_1, \dots, X_n\} > y) \\
&= 1 - P(X_1 > y, \dots, X_n > y) \stackrel{\text{ανεξ.}}{=} 1 - \prod_{i=1}^n (1 - F_{X_i}(y)).
\end{aligned}$$



Σχήμα 2.4: Ένα δείγμα $\mathbf{X}(\omega)$ ενός τυχαίου διανύσματος. Ένα στοιχείο του δειγματικού χώρου ω απεικονίζεται μέσω του τυχαίου διανύσματος σε δύο τιμές $X_1(\omega), X_2(\omega)$ που είναι τα στοιχεία του δισδιάστατου διανύσματος.

2.4 Τυχαία διανύσματα

Ένα πεπερασμένο, διατεταγμένο σύνολο τ.μ. ονομάζεται και τυχαίο διάνυσμα, $\mathbf{X} = [X_1, X_2, \dots, X_n]^T$. Για το τυχαίο διάνυσμα ισχύουν όλα όσα ισχύουν για την κοινή αντιμετώπιση τ.μ. Το τυχαίο διάνυσμα έχει κοινό δειγματικό χώρο με την έννοια ότι σε ένα μοναδικό πείραμα όπου έρχεται το αποτέλεσμα ω , παράγεται το δείγμα τυχαίου διανύσματος $\mathbf{X}(\omega)$, όπως στο σχήμα 2.4.

Οι μέσες τιμές των διανυσματικών στοιχείων (λόγω του πλήθους τους) μπορούν να αναπαρασταθούν σε ένα διάνυσμα ίσου μήκους που ονομάζεται **μέσο διάνυσμα** (mean vector)

$$\mathbf{m}_{\mathbf{X}} \doteq E[\mathbf{X}] = [E[X_1], E[X_2], \dots, E[X_n]]^T.$$

Ομοίως, οι δεύτερες ροπές μπορούν να αναπαρασταθούν σε έναν τετραγωνικό πίνακα που ονομάζεται **πίνακας συσχέτισης** (correlation matrix)

$$\mathbf{R}_{\mathbf{X}} = \begin{bmatrix} E[X_1^2] & E[X_1 X_2] & \dots & E[X_1 X_n] \\ E[X_2 X_1] & E[X_2^2] & \dots & E[X_2 X_n] \\ \vdots & \vdots & \ddots & \vdots \\ E[X_n X_1] & E[X_n X_2] & \dots & E[X_n^2] \end{bmatrix},$$

ενώ οι κεντρικές δεύτερες ροπές με τον τετραγωνικό πίνακα που ονομάζεται **πίνακας συμμεταβλητότητας** (covariance matrix)

$$\mathbf{K}_{\mathbf{X}} = \begin{bmatrix} E[(X_1 - m_1)^2] & E[(X_1 - m_1)(X_2 - m_2)] & \dots & E[(X_1 - m_1)(X_n - m_n)] \\ E[(X_2 - m_2)(X_1 - m_1)] & E[(X_2 - m_2)^2] & \dots & E[(X_2 - m_2)(X_n - m_n)] \\ \vdots & \vdots & \ddots & \vdots \\ E[(X_n - m_n)(X_1 - m_1)] & E[(X_n - m_n)(X_2 - m_2)] & \dots & E[(X_n - m_n)^2] \end{bmatrix},$$

όπου $m_i = E[X_i]$. Παρατηρήστε ότι τα στοιχεία της διαγωνίου του $\mathbf{K}_{\mathbf{X}}$ είναι τα $\text{Var}(X_i)$ και τα υπόλοιπα στοιχεία είναι της μορφής $\text{Cov}(X_i, X_j)$.

Άλλες παρατηρήσεις που μπορούμε να κάνουμε είναι:

- Οι πίνακες $\mathbf{R}_{\mathbf{X}}, \mathbf{K}_{\mathbf{X}}$ είναι συμμετρικοί $n \times n$ πίνακες.

- Αν και μόνο αν το διάνυσμα έχει ασυσχέτιστα στοιχεία X_1, X_2, \dots, X_n τότε ο πίνακας \mathbf{K}_X είναι διαγώνιος.
- Αν τα στοιχεία είναι ανεξάρτητα, τότε είναι και ασυσχέτιστα και άρα ο πίνακας \mathbf{K}_X είναι διαγώνιος (δεν ισχύει απαραίτητα το αντίθετο).
- Αν $\mathbf{m}_X = \mathbf{0}$ τότε $\mathbf{R}_X = \mathbf{K}_X$.
- Ισχύει πάντα $\mathbf{R}_X = E[\mathbf{X}\mathbf{X}^T]$ και
- $\mathbf{K}_X = E[(\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^T] = \mathbf{R}_X - \mathbf{m}_X\mathbf{m}_X^T$ ³.
- Ο πίνακας ετερομεταβλητότητας μεταξύ δύο τυχαίων διανυσμάτων ορίζεται ως

$$\mathbf{K}_{XY} \doteq E[(\mathbf{X} - \mathbf{m}_X)(\mathbf{Y} - \mathbf{m}_Y)^T] = \mathbf{R}_{XY} - \mathbf{m}_X\mathbf{m}_Y^T.$$

2.4.1 Γραμμικοί μετασχηματισμοί

Ο γραμμικός μετασχηματισμός έχει τη μορφή $y = ax + b$. Στο διανυσματικό χώρο, ο γραμμικός μετασχηματισμός γράφεται ως $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$, όπου με μικρό συμβολίζονται τα διανύσματα και με κεφαλαίο οι πίνακες. Παρακάτω, ο συμβολισμός θα τροποποιηθεί καθώς τα τυχαία διανύσματα (όπως και οι τ.μ.) συμβολίζονται με κεφαλαία γράμματα.

Έστω το τυχαίο διάνυσμα \mathbf{Y} που είναι γραμμικός μετασχηματισμός του \mathbf{X} και γράφουμε:

$$\mathbf{Y} = \mathbf{A}\mathbf{X}.$$

Έχουμε λοιπόν θεωρήσει ότι $\mathbf{b} = \mathbf{0}$. Για το τυχαίο διάνυσμα \mathbf{Y} μπορούμε να υπολογίσουμε τις δύο πρώτες ροπές συναρτήσει των ροπών του τυχαίου διανύσματος \mathbf{X} και των γραμμικών παραμέτρων \mathbf{A} . Έχουμε:

$$\begin{aligned} \mathbf{m}_Y &= E[\mathbf{A}\mathbf{X}] = \begin{bmatrix} \sum_{j=1}^n a_{1j}E[X_j] \\ \sum_{j=1}^n a_{2j}E[X_j] \\ \vdots \\ \sum_{j=1}^n a_{nj}E[X_j] \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \ddots & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix} \\ &= \mathbf{A}E[\mathbf{X}] = \mathbf{A}\mathbf{m}_X, \end{aligned}$$

$$\begin{aligned} \mathbf{K}_Y &= E[(\mathbf{Y} - \mathbf{m}_Y)(\mathbf{Y} - \mathbf{m}_Y)^T] = E[(\mathbf{A}\mathbf{X} - \mathbf{A}\mathbf{m}_X)(\mathbf{A}\mathbf{X} - \mathbf{A}\mathbf{m}_X)^T] \\ &= E[\mathbf{A}(\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^T\mathbf{A}^T] = \mathbf{A}E[(\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^T]\mathbf{A}^T \\ &= \mathbf{A}\mathbf{K}_X\mathbf{A}^T \end{aligned}$$

και⁴

$$\begin{aligned} \mathbf{K}_{XY} &= E[(\mathbf{X} - \mathbf{m}_X)(\mathbf{Y} - \mathbf{m}_Y)^T] = E[(\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^T\mathbf{A}^T] \\ &= \mathbf{K}_X\mathbf{A}^T. \end{aligned}$$

³ $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$.

⁴ $(\mathbf{A}\mathbf{B})^T = \mathbf{B}^T\mathbf{A}^T$.

Παράδειγμα: (Μετασχηματισμός ασυσχέτιστου διανύσματος): Έστω \mathbf{X} ένα ασυσχέτιστο τυχαίο διάνυσμα που οι συντεταγμένες του έχουν μεταβλητότητες σ . Λαμβάνουμε το γραμμικό μετασχηματισμό $\mathbf{Y} = \mathbf{A}\mathbf{X}$. Τότε για τον πίνακα συμμεταβλητότητας του \mathbf{Y} έχουμε:

$$\mathbf{K}_Y = \mathbf{A}\mathbf{K}_X\mathbf{A}^T = \mathbf{A}\sigma\mathbf{I}\mathbf{A}^T = \sigma\mathbf{A}\mathbf{A}^T.$$

Η πρώτη παρατήρηση είναι ότι ο γραμμικός μετασχηματισμός ενός ασυσχέτιστου διανύσματος μπορεί να είναι ένα συσχετισμένο διάνυσμα. Μάλιστα, στην επόμενη ενότητα θα προσπαθήσουμε να μετασχηματίσουμε ένα συσχετισμένο διάνυσμα σε ασυσχέτιστο. Μια άλλη παρατήρηση είναι ότι για να παραμείνει το διάνυσμα \mathbf{Y} ασυσχέτιστο μετά από έναν γραμμικό μετασχηματισμό, θα πρέπει ο πίνακας \mathbf{A} να είναι ορθογώνιος (δηλ. $\mathbf{A}\mathbf{A}^T = \mathbf{I}$) ή κάποιος πίνακας που ο πολλαπλασιασμός αυτός δίνει διαγώνιο πίνακα, δηλαδή να ισχύει $\sum_j \prod_i a_{ij} = 0$.

2.4.2 Διαγωνοποίηση πινάκων

Στην πράξη μας ενδιαφέρει να μπορούμε από ένα συσχετισμένο διάνυσμα \mathbf{X} (που για παράδειγμα είναι το διάνυσμα δεδομένων μας) να μπορούμε να το μετατρέψουμε σε ένα ασυσχέτιστο διάνυσμα \mathbf{Y} που είναι ευκολότερο να μελετηθεί. Σε μαθηματικούς όρους, κάνουμε την ερώτηση: υπάρχει πάντα πίνακας \mathbf{A} έτσι ώστε για οποιοδήποτε \mathbf{X} , και $\mathbf{Y} = \mathbf{A}\mathbf{X}$, ο πίνακας \mathbf{K}_Y να είναι διαγώνιος πίνακας;

Η απάντηση δίνεται από την γραμμική άλγεβρα και είναι, ναι! Συγκεκριμένα, η απάντηση λέει ότι κάθε συμμετρικός πίνακας με πραγματικές τιμές (ο \mathbf{K}_X είναι τέτοιος) μπορεί να διαγωνοποιηθεί από έναν πίνακα \mathbf{P} έτσι ώστε να ισχύει:

$$\mathbf{P}^T\mathbf{K}_X\mathbf{P} = \mathbf{\Lambda} \text{ και } \mathbf{P}^T\mathbf{P} = \mathbf{I}$$

και $\mathbf{\Lambda}$ είναι κάποιος διαγώνιος πίνακας.

Παράδειγμα: (Υπολογισμός του \mathbf{P}): Αρχικά υπολογίζουμε τις ιδιοτιμές και τα ιδιοδιανύσματα του \mathbf{K}_X με την εξίσωση:

$$\mathbf{K}_X\mathbf{e}_i = \lambda_i\mathbf{e}_i,$$

όπου \mathbf{e}_i είναι $n \times 1$ διανύσματα στήλης. Αν κανονικοποιήσουμε τα διανύσματα έτσι ώστε το άθροισμα τετραγώνων των στοιχείων είναι μονάδα, λαμβάνουμε μια ορθοκανονική βάση, δηλαδή:

$$\mathbf{e}_i^T\mathbf{e}_j = \delta_{i,j} = \begin{cases} 1 & \text{αν } i = j \\ 0 & \text{αν } i \neq j. \end{cases}$$

Ορίζουμε τώρα τον πίνακα \mathbf{P} ως τον πίνακα που έχει στήλες τα κανονικοποιημένα ιδιοδιανύσματα του \mathbf{K}_X και τον πίνακα $\mathbf{\Lambda}$ ως τον διαγώνιο πίνακα των ιδιοτιμών:

$$\mathbf{P} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n] \text{ και } \mathbf{\Lambda} = \text{diag}[\lambda_i].$$

Τότε ισχύει ότι

$$\begin{aligned} \mathbf{K}_X\mathbf{P} &= \mathbf{K}_X[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n] = [\mathbf{K}_X\mathbf{e}_1, \mathbf{K}_X\mathbf{e}_2, \dots, \mathbf{K}_X\mathbf{e}_n] \\ &= [\lambda_1\mathbf{e}_1, \lambda_2\mathbf{e}_2, \dots, \lambda_n\mathbf{e}_n] = \mathbf{P}\mathbf{\Lambda}. \end{aligned}$$

Τελικά, λόγω του ότι $\mathbf{P}^T\mathbf{P} = \mathbf{I}$ ισχύει για μια ορθοκανονική βάση, λαμβάνουμε

$$\mathbf{P}^T\mathbf{K}_X\mathbf{P} = \mathbf{P}^T\mathbf{P}\mathbf{\Lambda} = \mathbf{\Lambda}.$$

Συμπερασματικά, μπορούμε να διαγωνοποιήσουμε έναν οποιοδήποτε πίνακα συμμεταβλητότητας χρησιμοποιώντας έναν γραμμικό μετασχηματισμό. Με άλλα λόγια, κάθε συσχετισμένο τυχαίο διάνυσμα, μπορεί να μετατραπεί σε ασυσχέτιστο αν το εισάγουμε σε ένα γραμμικό σύστημα το οποίο έχει ως παραμέτρους τον ανάστροφο πίνακα των ιδιοδιανυσμάτων του πίνακα συμμεταβλητότητας του αρχικού τυχαίου διανύσματος.

Δοκιμάστε στο mathematica αυτό: `diagonalization_symmetric.nb`

Παράδειγμα: (Karhunen-Loeve expansion): Σε συνέχεια των παραπάνω, το τυχαίο διάνυσμα \mathbf{X} μπορεί να γραφεί ως

$$\mathbf{X} = \mathbf{I}\mathbf{X} = \mathbf{P}\mathbf{P}^T\mathbf{X} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n] \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \sum_{k=1}^n Y_k \mathbf{e}_k.$$

που ονομάζεται Karhunen-Loeve expansion και έχει πολλές εφαρμογές σε επεξεργασία στοχαστικών σημάτων.

2.4.3 Το διάνυσμα Gauss

Έστω $\mathbf{X} = [X_1, X_2]$ ένα διάνυσμα με ανεξάρτητες κανονικές τ.μ. (ανεξάρτητο διάνυσμα Gauss), δηλ. $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ και $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ με $\text{Cov}(X_1, X_2) = 0$ και $E[X_1 X_2] = \mu_1 \mu_2$. Στη συνέχεια ορίζουμε τον γραμμικό μετασχηματισμό $\mathbf{Y} = \mathbf{A}\mathbf{X}$ έτσι ώστε:

$$\begin{aligned} Y_1 &= a_{11}X_1 + a_{12}X_2 \\ Y_2 &= a_{21}X_1 + a_{22}X_2. \end{aligned}$$

Παρατηρήστε ότι το διάνυσμα \mathbf{Y} είναι επίσης ένα διάνυσμα Gauss όχι απαραίτητα ανεξάρτητο. Αυτό συμβαίνει γιατί κάθε γραμμικός συνδυασμός Gauss είναι πάλι Gauss (όπως αποδείξαμε παραπάνω) άρα και τα Y_1, Y_2 είναι Gauss. Μπορούμε να υπολογίσουμε τις ροπές τους

$$\begin{aligned} E[Y_1] &= E[a_{11}X_1 + a_{12}X_2] = a_{11}\mu_1 + a_{12}\mu_2 \\ E[Y_2] &= E[a_{21}X_1 + a_{22}X_2] = a_{21}\mu_1 + a_{22}\mu_2 \\ E[Y_1^2] &= E[(a_{11}X_1 + a_{12}X_2)^2] = a_{11}^2 E[X_1^2] + a_{12}^2 E[X_2^2] = a_{11}^2(\sigma_1^2 + \mu_1^2) + a_{12}^2(\sigma_2^2 + \mu_2^2) \\ E[Y_2^2] &= E[(a_{21}X_1 + a_{22}X_2)^2] = a_{21}^2 E[X_1^2] + a_{22}^2 E[X_2^2] = a_{21}^2(\sigma_1^2 + \mu_1^2) + a_{22}^2(\sigma_2^2 + \mu_2^2). \end{aligned}$$

Ο συντελεστής συσχέτισης μπορεί να υπολογιστεί ως εξής

$$\begin{aligned} \rho_{Y_1, Y_2} &= \frac{E[Y_1 Y_2]}{\sigma_{Y_1} \sigma_{Y_2}} \\ &= \frac{a_{11}a_{21}(\sigma_1^2 + \mu_1^2) + a_{12}a_{22}(\sigma_2^2 + \mu_2^2)}{\sqrt{a_{11}^2\sigma_1^2 + a_{12}^2\sigma_2^2 - 2a_{11}a_{12}\mu_1\mu_2} \sqrt{a_{21}^2\sigma_1^2 + a_{22}^2\sigma_2^2 - 2a_{21}a_{22}\mu_1\mu_2}}. \end{aligned}$$

Αποδεικνύοντας ότι και η υπό συνθήκη κατανομή $f_{Y_1|Y_2}(y_1|y_2)$ είναι κανονική και λύνοντας την μια μεταβλητότητα ως προς την άλλη, λαμβάνουμε την από κοινού κατανομή του συσχετισμένου διανύσματος Gauss:

$$f_{Y_1, Y_2}(y_1, y_2) = \frac{1}{2\pi\sqrt{1 - \rho^2}\sigma_{Y_1}\sigma_{Y_2}} e^{-\frac{\frac{(y_1 - m_1)^2}{\sigma_{Y_1}^2} - 2\rho\frac{(y_1 - m_1)(y_2 - m_2)}{\sigma_{Y_1}\sigma_{Y_2}} + \frac{(y_2 - m_2)^2}{\sigma_{Y_2}^2}}{2(1 - \rho^2)}},$$

όπου $m_1 \equiv E[Y_1]$, $m_2 \equiv E[Y_2]$, $\rho \equiv \rho_{Y_1, Y_2}$ και $|\rho| < 1$, δηλαδή δεν επιτρέπουμε στις Y_1, Y_2 να είναι η ίδια μεταβλητή, ή αλλιώς δεν επιτρέπουμε $a_{11} = a_{21}$ και $a_{12} = a_{22}$ ταυτόχρονα. Παρατηρούμε ότι για $\rho = 0$ προκύπτει το γινόμενο των κατανομών Y_1 και Y_2 , δηλαδή αν οι δύο μεταβλητές είναι ασυσχέτιστες τότε είναι και ανεξάρτητες.

Παράδειγμα: (Όταν δύο κανονικές τ.μ. είναι ασυσχέτιστες, τότε είναι και ανεξάρτητες)

Λύση: Για να αποδείξουμε την ιδιότητα αυτή που έχουν οι κανονικές τ.μ. θεωρούμε δύο ασυσχέτιστες κανονικές τ.μ. $X_1 \sim \mathcal{N}(0, \sigma_1^2)$, $X_2 \sim \mathcal{N}(0, \sigma_2^2)$ και την τ.μ. $Y_1 = a_{11}X_1 + a_{12}X_2$ η οποία είναι επίσης κανονική. Όπως ήδη γνωρίζουμε $Y_1 \sim \mathcal{N}(0, a_{11}^2\sigma_1^2 + a_{12}^2\sigma_2^2)$, αφού ο γραμμικός συνδυασμός κανονικών είναι κανονική με μέση τιμή το άθροισμα και μεταβλητότητα το άθροισμα μείον τη συμμεταβλητότητα (η οποία λόγω έλλειψης συσχέτισης είναι 0). Αρχικά, παρατηρούμε ότι ο μετασχηματισμός *Laplace – Stieltjes* της Y είναι

$$\tilde{F}_Y(s) = E[e^{sY}] = e^{s^2\sigma_Y^2/2} = e^{s^2(a_{11}^2\sigma_1^2 + a_{12}^2\sigma_2^2)/2}$$

Στη συνέχεια, υπολογίζουμε τον μετασχηματισμό *Laplace – Stieltjes* της από κοινού κατανομής (X_1, X_2) :

$$\tilde{F}_{X_1, X_2}(s_1, s_2) = E[e^{s_1X_1 + s_2X_2}] = e^{(s_1^2\sigma_1^2 + s_2^2\sigma_2^2)/2},$$

όπου η δεύτερη ισότητα ισχύει παρατηρώντας ότι το άθροισμα $s_1X_1 + s_2X_2$ είναι ουσιαστικά μια μεταβλητή σαν την Y_1 ενώ προφανώς ισχύει $E[e^{Y_1}] = \tilde{F}_{Y_1}(1)$. Όμως, γνωρίζουμε ότι για δύο ανεξάρτητες κανονικές μεταβλητές θα ισχύει

$$\tilde{F}_{X_1, X_2}(s_1, s_2) = \tilde{F}_{X_1}(s_1)\tilde{F}_{X_2}(s_2) = e^{(s_1^2\sigma_1^2 + s_2^2\sigma_2^2)/2}.$$

Εφόσον υπάρχει 1-1 σχέση μετασχηματισμών με κατανομές, καταλήγουμε στο συμπέρασμα ότι οι από κοινού κατανομές δύο ασυσχέτιστων κανονικών τ.μ. και δύο ανεξάρτητων κανονικών τ.μ. είναι ίδιες. ■

Δοκιμάστε στο mathematica αυτό: `joint_pdf_gauss.nb`

Επίσης παρατηρούμε ότι η από κοινού κατανομή δύο συσχετισμένων κανονικών τ.μ. καθορίζεται πλήρως από τις ροπές $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$. Αυτό ισχύει και για το γενικό διάνυσμα Gauss.

$$f_{\mathbf{X}}(\mathbf{x}) \doteq f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \frac{e^{-\frac{(\mathbf{x}-\mathbf{m})^T \mathbf{K}^{-1} (\mathbf{x}-\mathbf{m})}{2}}}{(2\pi)^{n/2} |\mathbf{K}|^{1/2}},$$

όπου $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ είναι το διάνυσμα στήλη με τις μεταβλητές της συνάρτησης, $\mathbf{m} = [E[X_1], E[X_2], \dots, E[X_n]]^T$ είναι το μέσο διάνυσμα και

$$\mathbf{K} = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \dots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \dots & \text{Var}(X_n) \end{bmatrix}$$

είναι ο πίνακας συμμεταβλητότητας.

2.5 Μιγαδικές τυχαίες μεταβλητές

Έστω δύο τ.μ. X, Y που λαμβάνουν πραγματικές τιμές, δηλαδή $\mathcal{S}_X, \mathcal{S}_Y = \mathbb{R}$. Τότε ορίζουμε την μιγαδική τ.μ. ως

$$Z(\omega) = X(\omega) + jY(\omega),$$

όπου $\omega \in \Omega$ και η $Z(\omega)$ παίρνει τιμές στο \mathbb{C} . Από τις ιδιότητες των πράξεων στο πεδίο των μιγαδικών αριθμών, γνωρίζουμε ότι το συζυγές $Z^*(\omega)$ είναι επίσης μιγαδική τ.μ., ενώ το μέτρο $|Z(\omega)|$ και η γωνία $\arg\{Z(\omega)\}$ είναι πραγματικές τ.μ. Επίσης, ισχύει $\mathcal{R}\{Z(\omega)\} \equiv X(\omega)$ και $\mathcal{I}\{Z(\omega)\} \equiv Y(\omega)$ που είναι αντίστοιχα το πραγματικό και το φανταστικό μέρος της $Z(\omega)$, επίσης πραγματικές τ.μ.

Καθώς το πεδίο \mathbb{C} και ο διανυσματικός χώρος \mathbb{R}^2 είναι ισοδύναμα, μπορούμε να σκεφτούμε τις μιγαδικές τ.μ. σαν τυχαία διανύσματα στο \mathbb{R}^2 . Πιο συγκεκριμένα, η κατανομή της Z ορίζεται ως η από κοινού κατανομή των X, Y

$$F_Z(\mathbf{z}) = F_{X,Y}(z_1, z_2), z_1, z_2 \in \mathbb{R},$$

όπου \mathbf{z} το διάνυσμα πραγματικών αριθμών (z_1, z_2) . Αντίστοιχα ορίζονται οι ΣΜΠ και ΣΠΠ. Η μέση τιμή της Z θα είναι

$$m_Z \doteq E[Z] = E[X] + jE[Y],$$

και είναι μιγαδικός αριθμός, ενώ η συμμεταβλητότητα (covariance)

$$\text{Cov}(Z) = E[(Z - m_Z)(Z - m_Z)^*] = E[(X - m_X)^2] + E[(Y - m_Y)^2],$$

είναι πραγματικός αριθμός και σε όρους σημάτων εκφράζει την ισχύ της $Z(\omega)$. Τέλος, ορίζεται και η ψευδοσυμμεταβλητότητα (pseudocovariance)

$$\text{Pse}(Z) = E[(Z - m_Z)^2] = E[(X - m_X)^2] - E[(Y - m_Y)^2] + 2jE[(X - m_X)(Y - m_Y)],$$

Όταν η ψευδοσυμμεταβλητότητα είναι ίση με 0, τότε η $Z(\omega)$ λέγεται **κανονική μιγαδική μεταβλητή** (proper complex random variable). Αυτό προϋποθέτει ότι τα X, Y είναι κανονικοποιημένα ως προς τη διασπορά (δηλαδή $E[(X - m_X)^2] = E[(Y - m_Y)^2]$) και ότι είναι ασυσχέτιστα (δηλαδή $E[(X - m_X)(Y - m_Y)] = 0$).

2.5.1 Διάνυσμα μιγαδικών μεταβλητών

Θεωρούμε τα τυχαία διανύσματα \mathbf{X}, \mathbf{Y} διάστασης N με μέσες τιμές $\mathbf{m}_X, \mathbf{m}_Y$ και πίνακες συμμεταβλητότητας $\mathbf{K}_X, \mathbf{K}_Y$. Ορίζουμε το μιγαδικό διάνυσμα

$$\mathbf{Z} = \mathbf{X} + j\mathbf{Y}$$

που λαμβάνει τιμές στο \mathbb{C}^N . Η κατανομή του τυχαίου μιγαδικού διανύσματος \mathbf{Z} καθορίζεται από την από κοινού κατανομή των τ.μ. $X_1, \dots, X_N, Y_1, \dots, Y_N$.

Το διάνυσμα μέσης τιμής κατασκευάζεται από τα αντίστοιχα διανύσματα των \mathbf{X}, \mathbf{Y}

$$\mathbf{m}_Z = \mathbf{m}_X + j\mathbf{m}_Y$$

Αντίστοιχα και ο πίνακας συμμεταβλητότητας

$$\mathbf{K}_Z = E[(\mathbf{Z} - \mathbf{m}_Z)(\mathbf{Z} - \mathbf{m}_Z)^\dagger] = \mathbf{K}_X + \mathbf{K}_Y + j(\mathbf{K}_{XY} - \mathbf{K}_{YX}),$$

όπου το σύμβολο \dagger αναφέρεται στον συζυγοανάστροφο πίνακα, δηλαδή $\mathbf{A}^\dagger = (\mathbf{A}^T)^* = (\mathbf{A}^*)^T$. $\mathbf{K}_X = E[(\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^T]$, $\mathbf{K}_{XY} = E[(\mathbf{X} - \mathbf{m}_X)(\mathbf{Y} - \mathbf{m}_Y)^T]$ και αντίστοιχα για τα υπόλοιπα.

Κατ' αντιστοιχία ορίζεται και ο πίνακας ψευδοσυμμεταβλητότητας

$$\hat{\mathbf{K}}_Z = E[(\mathbf{Z} - \mathbf{m}_Z)(\mathbf{Z} - \mathbf{m}_Z)^T] = \mathbf{K}_X - \mathbf{K}_Y + j(\mathbf{K}_{XY} + \mathbf{K}_{YX}),$$

Ένα μιγαδικό τυχαίο διάνυσμα λέγεται **κανονικό** αν $\hat{\mathbf{K}}_{\mathbf{Z}} = \mathbf{0}$ ή ισοδύναμα αν

$$K_{\mathbf{X}} = K_{\mathbf{Y}} \text{ και } K_{\mathbf{X}\mathbf{Y}} = -K_{\mathbf{Y}\mathbf{X}}.$$

Συνεπώς, για ένα κανονικό μιγαδικό τυχαίο διάνυσμα θα ισχύει

$$K_{\mathbf{Z}} = 2K_{\mathbf{X}} + 2jK_{\mathbf{X}\mathbf{Y}}.$$

2.5.2 Μιγαδικό διάνυσμα Gauss

Το μιγαδικό διάνυσμα Gauss χρησιμοποιείται ευρύτατα στις επικοινωνίες καθώς περιγράφει το κοινό μοντέλο δύο διαδικασιών χρόνου της συνιστώσας φάσης (in-phase component) και της ορθογωνικής (quadrature component) ενώ ταυτόχρονα είναι το πιο απλό στοχαστικό μοντέλο που προσφέρει αυτή την μοντελοποίηση. Η απλότητα έχει να κάνει με το γεγονός ότι όπως γνωρίζουμε, η κατανομή της τ.μ. Gauss καθορίζεται από τις δύο πρώτες ροπές της (μέση τιμή και διασπορά), ενώ η κατανομή του διανύσματος Gauss καθορίζεται αντίστοιχα από το διάνυσμα μέσης τιμής και τον πίνακα συμμεταβλητότητας.

Πιο συγκεκριμένα, μας ενδιαφέρει το κανονικό μιγαδικό διάνυσμα Gauss που ονομάζεται κυκλικά συμμετρικό μιγαδικό διάνυσμα Gauss (circularly symmetric complex Gaussian vector), συμβολίζεται με $\mathcal{CN}(\mathbf{m}, \mathbf{K})$ και όπως γίνεται κατανοητό καθορίζεται πλήρως από τα \mathbf{m}, \mathbf{K} . Η p.d.f. είναι

$$f_{\mathbf{Z}}(\mathbf{z}) \doteq f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = \frac{e^{-(\mathbf{Z}-\mathbf{m})^\dagger \mathbf{K}^{-1}(\mathbf{Z}-\mathbf{m})}}{\pi^N |\mathbf{K}|},$$

Η ψευδοσυμμεταβλητότητα είναι μηδενικός πίνακας καθώς το κυκλικά συμμετρικό μιγαδικό διάνυσμα Gauss είναι κανονικό. Περισσότερες πληροφορίες μπορείτε να βρείτε στο [1].

2.6 Εξομοίωση τ.μ.

Κατά τη διάρκεια εξομοιώσεων, συχνά απαιτείται η παραγωγή τυχαίων τιμών που ακολουθούν κάποια συγκεκριμένη κατανομή, είναι δηλαδή δείγματα (realizations) μιας τ.μ. Το πρώτο ζήτημα που μας απασχολεί είναι πως θα παράγουμε τυχαίους αριθμούς γενικά. Στην πράξη είναι δυνατόν να παράγουμε ψευδοτυχαίους αριθμούς παρατηρώντας γεγονότα που εξαρτώνται από πολλούς παράγοντες⁵. Το τελευταίο ψηφίο της τιμής του δείκτη χρηματιστηρίου ή το τελευταίο ψηφίο της δυαδικής απεικόνισης του ρολογιού του επεξεργαστή, είναι τέτοια παραδείγματα. Όμως, παρατηρείται ότι οι μέθοδοι αυτές δεν έχουν τόσο καλές ιδιότητες. Για αυτό το λόγο, έχουν εξελιχθεί ντετερμινιστικοί αλγόριθμοι⁶ που παράγουν ακολουθίες ψευδοτυχαίων αριθμών με καλές ιδιότητες τυχειότητας, δηλαδή ακολουθίες τυχαίων αριθμών που είναι σχεδόν ασυσχέτιστες. Οι αλγόριθμοι αυτοί αντιστοιχίζουν μια ψευδοτυχαία ακολουθία σε έναν αριθμό που ονομάζεται σπόρος (seed). Χρησιμοποιώντας το ίδιο seed είναι δυνατόν να επαναλάβουμε το ίδιο τυχαίο πείραμα δεύτερη φορά έχοντας ακριβώς την ίδια ακολουθία συμβάντων.

Το δεύτερο ζήτημα είναι πως παράγουμε τυχαίους αριθμούς από μια συγκεκριμένη κατανομή. Δεχόμενοι ότι οι παραπάνω μέθοδοι παράγουν ομοιόμορφα κατανομημένους αριθμούς, θα πρέπει να επέμβουμε στους τυχαίους αριθμούς έτσι ώστε να έχουμε το επιθυμητό αποτέλεσμα. Σε περίπτωση που χρησιμοποιούμε κάποιο μαθηματικό πακέτο (όπως το matlab), υπάρχουν έτοιμες ρουτίνες που εξυπηρετούν στην παραγωγή τυχαίων αριθμών για ένα μεγάλο αριθμό γνωστών κατανομών. Είναι χρήσιμο όμως να αναφερθούμε σε μερικές πρακτικές μεθόδους που μπορούμε να χρησιμοποιήσουμε στις περιπτώσεις που δεν έχουμε στη διάθεσή μας τέτοιες ρουτίνες.

⁵Ο καθ. D.H. Lehmer, πρωτοπόρος στην υπολογιστική θεωρία αριθμών, όρισε την ακολουθία τυχαίων αριθμών το 1951 ως εξής “.. a random sequence is a vague notion in which each term is unpredictable to the uninitiated and whose digits pass a certain number of tests traditional with statisticians..”

⁶Το matlab, για παράδειγμα, χρησιμοποιεί τον αλγόριθμο Mersenne Twister [10].

Η πρώτη μέθοδος είναι να εφαρμόσουμε μια συνάρτηση στα δεδομένα. Αν για παράδειγμα γνωρίζουμε ότι για $X \sim U(0, 1)$ και $Y = X^2$, το Y έχει κατανομή ρίζας και μας ενδιαφέρει να παράγουμε τιμές από αυτήν την κατανομή, τότε μπορούμε να παράγουμε τους τυχαίους αριθμούς με την ομοιόμορφη γεννήτρια και να τους υψώνουμε στο τετράγωνο, ενώ είμαστε σίγουροι ότι οι προκύπτοντες αριθμοί θα ανήκουν στην ζητούμενη κατανομή. Για να γενικεύσουμε αυτή τη μέθοδο ακολουθούμε την εξής πρακτική. Υποθέτουμε ότι μπορούμε να παράγουμε αριθμούς δείγματα της $X \sim U(0, 1)$ ενώ μας ενδιαφέρει να παράγουμε δείγματα της Y με κατανομή $F_Y(y)$. Έστω ότι $Z = F_Y^{-1}(X)$, τότε ισχύει

$$P(Z \leq y) = P(F_Y^{-1}(X) \leq y) = P(X \leq F_Y(y)) = F_Y(y),$$

όπου η τελευταία ισότητα προκύπτει από το ότι $P(X \leq h) = h$ όταν η X είναι ομοιόμορφη στο $[0, 1]$ και $0 < h < 1$. Συνεπώς, αρκεί να ορίζεται η αντίστροφη της κατανομής που μας ενδιαφέρει. Τότε, εφαρμόζουμε την αντίστροφη αυτή σε όλους τους τυχαίους αριθμούς και πετυχαίνουμε το ζητούμενο.

Όταν μας ενδιαφέρει να παράγουμε ένα γεγονός με μια πιθανότητα $p = 0.6$ έχοντας στη διάθεσή μας την $X \sim U(0, 1)$, μπορούμε να συγκρίνουμε τον τυχαίο αριθμό με το κατώφλι 0.6, έτσι ώστε όταν δεν το ξεπερνάει να πραγματοποιούμε το γεγονός. Τέλος, όταν θέλουμε να παράγουμε αριθμούς από μίξη κατανομών, τότε μπορούμε να εκτελούμε πρώτα ένα ομοιόμορφο πείραμα και να διαλέγουμε μια από τις κατανομές (με πιθανότητα ίση με το βάρος της στη μίξη) και στη συνέχεια να παράγουμε τον τυχαίο αριθμό από την αντίστοιχη κατανομή.

2.7 Προβλήματα στοχαστικής γεωμετρίας

2.7.1 Buffon's needle

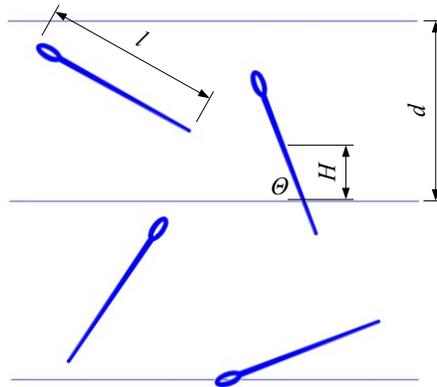
Η βελόνα του Buffon είναι ένα διάσημο παράδειγμα που μελετά γεωμετρικά σχήματα που παράγονται από τυχαία τοποθετημένα αντικείμενα. Δύο παράλληλες ευθείες σε απόσταση d ορίζουν μια επιφάνεια πάνω στην οποία πετάμε μια βελόνα μήκους $l < d$. Ποιά είναι η πιθανότητα η βελόνα να τέμνει μια από τις ευθείες;

Αρχικά, θα θεωρήσουμε σαν σημείο τοποθέτησης της βελόνας το κέντρο της και συγκεκριμένα την απόστασή του από την κάτω ευθεία (έστω τ.μ. H) ενώ θα επιτρέψουμε στη βελόνα να έχει μια τυχαία γωνία με την κάτω ευθεία (έστω τ.μ. Θ). Τόσο η H όσο και η Θ είναι ομοιόμορφα κατανομημένες και ανεξάρτητες μεταξύ τους (από την εκφώνηση του προβλήματος). Παρατηρήστε ότι θα έχουμε πειράματα όπου το κέντρο της βελόνας πέφτει πιο κοντά στην πάνω ευθεία ($H > d/2$) και άλλα που πέφτει κοντά στην κάτω ευθεία ($H < d/2$). Τα πειράματα αυτά περιγράφονται από την ίδια κατανομή λόγω συμμετρίας. Διαλέγουμε να μελετήσουμε τα δεύτερα. Η από κοινού ΣΠΠ θα είναι

$$f_{H,\Theta}(h, \theta) = \begin{cases} \frac{4}{\pi d} & \text{αν } h \in [0, \frac{d}{2}], \theta \in [0, \frac{\pi}{2}] \\ 0 & \text{αλλού.} \end{cases}$$

Για να τέμνει λοιπόν η βελόνα την κάτω γραμμή, θα πρέπει να ισχύει $H \leq \frac{l}{2} \sin \Theta$. Άρα η ζητούμενη πιθανότητα είναι

$$\begin{aligned} P\left(H \leq \frac{l}{2} \sin \Theta\right) &= \int_{[0, \frac{\pi}{2}]} \int_{[0, \frac{l}{2} \sin \theta]} f_{H,\Theta}(h, \theta) dh d\theta \\ &= \frac{4}{\pi d} \int_0^{\frac{\pi}{2}} \int_0^{\frac{l}{2} \sin \theta} dh d\theta \\ &= \frac{4}{\pi d} \int_0^{\frac{\pi}{2}} \frac{l}{2} \sin \theta d\theta \\ &= \frac{2l}{\pi d}. \end{aligned}$$



Σχήμα 2.5: Η βελόνα του Buffon πετάγεται στην τύχη μεταξύ δύο ευθειών. Ζητείται η πιθανότητα η βελόνα να τέμνει μια από τις δύο ευθείες.

Η μέθοδος αυτή είναι μια πειραματική μέθοδος υπολογισμού του π . Μπορεί κανείς να επαναλάβει το πείραμα της βελόνας n φορές και να μετρήσει πόσες φορές η βελόνα άγγιζε μια από τις δύο γραμμές, A_n . Τότε

$$\pi = \frac{2l}{d \lim_{n \rightarrow \infty} \frac{A_n}{n}}.$$

2.7.2 Επιλογή σημείων στο δίσκο (Disk point picking)

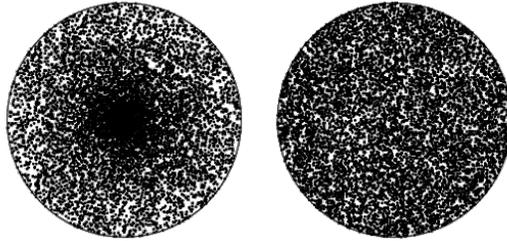
Θέλουμε να βρούμε την κατανομή των σημείων πάνω σε έναν κυκλικό δίσκο με ακτίνα r_0 έτσι ώστε να έχουν ομοιόμορφη πυκνότητα πιθανότητας. Παρατηρούμε ότι αν προσπαθήσουμε να παράγουμε τα σημεία στο καρτεσιανό σύστημα συντεταγμένων τότε μας ενδιαφέρει να παράγουμε τα σημεία (X, Y) όπου X και Y , ανεξάρτητες, ομοιόμορφα κατανεμημένες τ.μ. στο $[-1, 1]$. Όμως, με τη μέθοδο αυτή παράγουμε και σημεία εκτός του δίσκου (για παράδειγμα το σημείο $(0.85, 0.9)$). Υπάρχουν δύο τρόποι να αποφύγουμε αυτό το τεχνικό πρόβλημα.

(α) Μπορούμε να παράγουμε τα σημεία στο σύστημα πολικών συντεταγμένων όπου προφανώς τα όρια της ακτίνας και της γωνίας είναι σαφώς ορισμένα. Συγκεκριμένα παράγουμε σημεία (R, Θ) με $R \in [0, r_0]$ και $\Theta \in [0, 2\pi]$. Αν όμως θεωρήσουμε ομοιόμορφη κατανομή για την ακτίνα, καταλήγουμε σε λάθος (δες σχήμα 2.6) αφού οι ακτίνες απομακρύνονται μεταξύ τους καθώς απομακρυνόμαστε από το κέντρο του δίσκου. Για να υπολογίσουμε σωστά την κατανομή της τυχαίας ακτίνας έχουμε

$$\begin{aligned} f_{R\Theta}(r, \theta) dr d\theta &= P(r \leq R \leq r + dr, \theta \leq \Theta \leq \theta + d\theta) \\ &= \frac{1}{\pi r_0^2} r dr d\theta. \end{aligned}$$

Άρα $f_{R\Theta}(r, \theta) = \frac{r}{\pi r_0^2}$. Συνεπώς η οριακή κατανομή της ακτίνας θα είναι

$$f_R(r) = \int_0^{2\pi} \frac{r}{\pi r_0^2} d\theta = \frac{2r}{r_0^2}, r \in [0, r_0].$$



Σχήμα 2.6: Αν μας ζητηθεί να διαλέξουμε ένα σημείο σε κυκλικό δίσκο με ομοιόμορφο τρόπο, λαμβάνοντας ομοιόμορφα την ακτίνα και την γωνία καταλήγουμε σε λάθος. Αριστερά φαίνεται η λανθασμένη μέθοδος και δεξιά η ορθή.

(β) Εναλλακτικά, μπορούμε να παράγουμε τα σημεία στο καρτεσιανό σύστημα και κάθε φορά που παράγουμε ένα σημείο εκτός του δίσκου να το παραβλέπουμε και να συνεχίζουμε το πείραμά μας. Η αρχή αυτή που μας επιτρέπει να παραβλέπουμε τα σημεία ονομάζεται κανόνας απόρριψης (rejection rule). Η μέθοδος αυτή είναι πιο αργή καθώς το πείραμά μας πρέπει να διαρκέσει περισσότερο χρόνο ώστε να λάβουμε τον ίδιο αριθμό σημείων με την μέθοδο (α).

Δοκιμάστε στο matlab αυτό: `disk_point_picking.m`

Κεφάλαιο 3

Σύγκλιση ακολουθιών

3.1 Ανισότητες

Μερικές φορές γνωρίζουμε μόνο κάποιες από τις ροπές για μια τ.μ. και όχι την κατανομή της. Σε αυτές τις περιπτώσεις, μπορούμε να χρησιμοποιήσουμε κάποιες σημαντικές ανισότητες για να υπολογίσουμε φράγματα για πιθανότητες.

3.1.1 Ανισότητα Markov

Έστω μια θετική τ.μ. (δηλ. παίρνει μόνο θετικές τιμές ή $\mathcal{S}_X \subseteq \mathbb{R}_+$) και $a > 0$. Τότε

$$\begin{aligned} E[X] &= \int_0^{\infty} x f_X(x) dx \\ &= \int_0^a x f_X(x) dx + \int_a^{\infty} x f_X(x) dx \\ &\geq \int_a^{\infty} x f_X(x) dx \\ &\geq \int_a^{\infty} a f_X(x) dx \\ &= aP(X \geq a). \end{aligned}$$

Άρα

$$P(X \geq a) \leq \frac{E[X]}{a}, \text{ για κάθε } a > 0.$$

Παράδειγμα: Έστω η τ.μ. $X \sim \text{Exp}(1)$ και η $Y \sim U(0, 1)$. Πόσο αυστηρό είναι το φράγμα Markov για τις πιθανότητες $P(X \leq 1)$ και $P(Y \geq \frac{1}{2})$;

Λύση: Και οι δύο τ.μ. είναι θετικές. Για την πρώτη, έχουμε

$$P(X \leq 1) = 1 - P(X \geq 1) \geq 1 - \frac{E[X]}{1} = 0,$$

και η πραγματική τιμή της πιθανότητας είναι $1 - \frac{1}{e}$.

Για την δεύτερη έχουμε

$$P\left(Y \geq \frac{1}{2}\right) \leq \frac{E[Y]}{1/2} = 1,$$

και η πραγματική τιμή της πιθανότητας είναι $\frac{1}{2}$. Άρα και στις δύο περιπτώσεις τα φράγματα είναι πολύ χαλαρά (για την ακρίβεια δεν μας δίνουν καμία πληροφορία). ■

3.1.2 Ανισότητα Chebyshev

Για οποιαδήποτε τ.μ. με $E[X] = \mu$ και $\text{Var}(X) = \sigma^2$ έχουμε

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}, \text{ για κάθε } a > 0.$$

Η ανισότητα Chebyshev προκύπτει από την ανισότητα Markov αν αντικαταστήσουμε $X \sim (X - \mu)^2$ και $a \sim a^2$.

Παράδειγμα: Ένας server έχει καθυστέρηση αντίδρασης 15sec κατά μ.ο. με τυπική απόκλιση 3sec. Ποιά η πιθανότητα η καθυστέρηση σε ένα request να απέχει περισσότερο από 5sec από το μ.ο.

Λύση: Θα χρησιμοποιήσουμε την ανισότητα Chebyshev για $\mu = 15\text{sec}$, $\sigma = 3\text{sec}$ και $a = 5\text{sec}$.

$$P(|X - 15| \geq 5) \leq \frac{9}{25} = 0.36.$$

Άρα, γνωρίζουμε ότι είναι το πολύ 36%. ■

3.1.3 Φράγμα Chernoff

Αν γνωρίζουμε περισσότερη πληροφορία για μια τ.μ. από τη μέση τιμή και τη μεταβλητότητά της, μπορούμε να χρησιμοποιήσουμε ένα πιο αυστηρό φράγμα, το φράγμα Chernoff.

$$\begin{aligned}
P(X \geq a) &= \int_0^{\infty} \mathbb{1}_{\{x \geq a\}} f_X(x) dx \\
&\leq \int_0^{\infty} e^{s(x-a)} f_X(x) dx \\
&= e^{-sa} \int_0^{\infty} e^{sx} f_X(x) dx = e^{-sa} E[e^{sX}] \\
&= e^{-sa} \tilde{F}_X(s), \forall s > 0.
\end{aligned}$$

Αν γνωρίζουμε, λοιπόν, τον μετασχηματισμό Laplace-Stieltjes της X μπορούμε να βελτιστοποιήσουμε το παραπάνω φράγμα ως προς s για κάθε ζητούμενο a , δηλαδή να βρούμε το s που δίνει το πιο αυστηρό φράγμα για την πιθανότητα που μας ενδιαφέρει.

3.1.4 Ανισότητα Jensen

Η ανισότητα του Jensen είναι ένα σημαντικό αποτέλεσμα που αφορά σε κυρτές συναρτήσεις. Μια συνάρτηση $f: \mathcal{X} \rightarrow \mathcal{S}$ λέγεται κυρτή στο πεδίο \mathcal{X} όταν κάθε δύο σημεία $x_1, x_2 \in \mathcal{X}$ και οποιοδήποτε $\lambda \in [0, 1]$ ισχύει

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2).$$

Αυτό πρακτικά σημαίνει ότι το ημιεπίπεδο που ορίζεται πάνω από τη συνάρτηση είναι κυρτό, δηλαδή το ευθύγραμμο τμήμα που ενώνει δύο σημεία του ημιεπιπέδου είναι πάντα εντός του ημιεπιπέδου. Η ανισότητα του Jensen λέει ότι, αν X είναι τ.μ. και f κυρτή συνάρτηση, τότε

$$f(E[X]) \leq E[f(X)].$$

3.1.5 Ανισότητα Cauchy–Schwarz

Στη γραμμική άλγεβρα η ανισότητα Cauchy–Schwarz λέει ότι το εσωτερικό γινόμενο δύο διανυσμάτων είναι πάντα μικρότερο από το γινόμενο των μέτρων τους. Εφαρμοσμένη στη θεωρία πιθανοτήτων, η ανισότητα αυτή μας λέει ότι

$$E[XY]^2 \leq E[X^2]E[Y^2].$$

Συνέπεια αυτού είναι ότι

$$\text{Cov}(X, Y)^2 \leq \text{Var}(X) \text{Var}(Y).$$

3.2 Σύγκλιση τ.μ.

Μια ακολουθία τ.μ. είναι ένα διάνυσμα τ.μ. με άπειρες αλλά αριθμήσιμες διαστάσεις (countably infinite). Με άλλα λόγια, έχουμε μια συλλογή τ.μ. που συμβολίζεται ως $\{X_n\}$ όπου $n = 1, 2, \dots$. Επίσης, όπως ακριβώς και με τα τυχαία διανύσματα, μια τυχαία ακολουθία αντιστοιχίζει μια ακολουθία τιμών σε κάθε τυχαίο αποτέλεσμα ω , το οποίο ονομάζεται δείγμα ακολουθίας (realization of random sequence) ή αλλιώς δειγματική ακολουθία (sample sequence) και συμβολίζεται με $\{X_1(\omega), X_2(\omega), \dots\}$. Πιο εύκολα, μπορεί κανείς να σκέφτεται το sample sequence σαν μια ακολουθία πραγματικών αριθμών x_1, x_2, \dots που προκύπτουν από τις τ.μ. μόλις το αποτέλεσμα του πειράματος ω καθοριστεί.

Μέχρι τώρα ασχοληθήκαμε κυρίως με σύνολα πειραμάτων (γεγονότα). Στο κεφάλαιο αυτό, όμως, θα μελετήσουμε και τη συμπεριφορά τ.μ. ως προς λίγα πειράματα που έχουν μηδενική πιθανότητα να συμβούν, ή ακόμα και ένα από αυτά.

3.2.1 Σύγκλιση γενικά

Σύμφωνα με το κριτήριο Cauchy, μια ακολουθία $\{x_n\}$ συγκλίνει όταν: για κάθε $\epsilon > 0$ μπορούμε να βρούμε $n'(\epsilon)$ τέτοιο ώστε έχουμε $|x_n - x_m| < \epsilon$ για οποιαδήποτε $n, m > n'$. Με άλλα λόγια, αν η ακολουθία συγκλίνει, από ένα σημείο (που μπορεί φυσικά να εξαρτάται από τις απαιτήσεις σύγκλισης, δηλαδή το ϵ) και μετά όλες οι τιμές της ακολουθίας βρίσκονται σε ένα μικρό διάστημα $[-\epsilon, \epsilon]$. Αν μπορεί κανείς να δείξει μια τέτοια ιδιότητα για μια τυχαία ακολουθία, τότε λέμε ότι η ακολουθία συγκλίνει σε αριθμό και γράφουμε

$$X_n(\omega) \rightarrow c \text{ όταν } n \rightarrow \infty \quad \forall \omega \in \Omega.$$

Αυτή η σχέση είναι συνήθως αυστηρή για μια τυχαία ακολουθία καθώς απαιτούμε όλα τα τυχαία πειράματα να αντιστοιχούν σε ακολουθίες που καταλήγουν σε τιμές μέσα σε ένα μικρό διάστημα. Απαιτούμε δηλαδή να μην υπάρχει κανένα πείραμα που να μπορεί να ξεφύγει από το διάστημα αυτό.

Παράδειγμα: Έστω η ακολουθία με πειράματα $X_i \sim \text{Bernoulli}(p)$, $X_n = \frac{1}{n} \sum_i^n X_i$. Έχει βέβαια σύγκλιση;

Λύση: Η απάντηση είναι όχι. Για να το αποδείξουμε, προσέξτε ότι για οσοδήποτε μεγάλο n'

$$\exists \omega_0 : X_{n'}(\omega_0) = 0 \text{ και}$$

$$\exists \omega_1 : X_{n'}(\omega_1) = 1,$$

δηλαδή υπάρχουν πειράματα που είναι διασκορπισμένα στο διάστημα $[0, 1]$. Άρα, ο παραπάνω συλλογισμός δεν μπορεί να επαληθευτεί για μικρά ϵ . ■

Παρότι η ακολουθία του παραδείγματος δεν συγκλίνει πάντα στη μέση τιμή, από πειράματα μπορεί κανείς να παρατηρήσει ότι η συντριπτική πλειοψηφία των πειραμάτων καταλήγει κοντά σε αυτήν την τιμή. Μας ενδιαφέρει, λοιπόν, και η σύγκλιση υπό όρους. Παρακάτω αναλύουμε τύπους σύγκλισης για τυχαίες ακολουθίες. Οι τύποι αυτοί παρουσιάζονται σε φθίνουσα σειρά αυστηρότητας με την ιδιότητα ότι αν μια ακολουθία συγκλίνει με αυστηρούς όρους, τότε θα συγκλίνει και με λιγότερο αυστηρούς όρους.

3.2.2 Βέβαια σύγκλιση

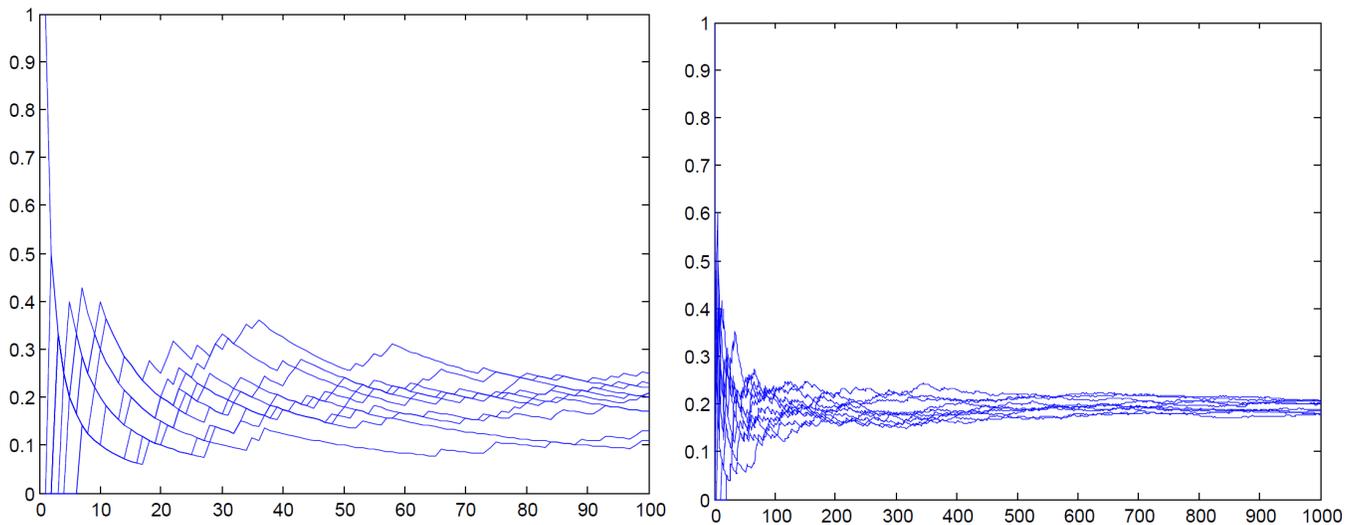
Όταν μια τυχαία ακολουθία X_n συγκλίνει σε μια τ.μ. X για κάθε τυχαίο αποτέλεσμα ω σημαίνει ότι μπορώ να βρω μεγάλα n' τέτοια ώστε η τιμή της ακολουθίας για $n > n'$ είναι για όλα τα ενδεχόμενα ω γύρω από την τιμή $X(\omega)$, δηλαδή στο διάστημα $[X(\omega) - \epsilon, X(\omega) + \epsilon]$. Γράφουμε

$$X_n(\omega) \rightarrow X(\omega) \text{ όταν } n \rightarrow \infty \quad \forall \omega \in \Omega.$$

Παράδειγμα: Έστω η ακολουθία $X_n = e^{-1/n} X$, όπου X μια τ.μ. Προφανώς, έχουμε

$$\lim_{n \rightarrow \infty} X_n = X,$$

που σημαίνει ότι για κάθε πείραμα ω η ακολουθία X_n συγκλίνει στην τ.μ. X , έχουμε δηλαδή βέβαια σύγκλιση.



Σχήμα 3.1: Σύγκλιση του αθροίσματος ανεξάρτητων Bernoulli με $p = 0.2$. Στην εικόνα φαίνονται 10 sample sequences, αριστερά για $n < 100$ και δεξιά για $n < 1000$. Όπως θα δείξουμε παρακάτω, το άθροισμα i.i.d. ακολουθιών συγκλίνει σε κατανομή στην κατανομή Gauss με μέση τιμή p και μεταβλητότητα $\frac{p(1-p)}{n}$. Επίσης συγκλίνει με πιθανότητα 1 στην αναμενόμενη τιμή (που είναι 0.2).

3.2.3 Σύγκλιση με πιθανότητα 1

Είναι δυνατόν η σύγκλιση να ισχύει για τα περισσότερα πειράματα: όλα εκτός από λίγα. Αν μπορούμε να αποδείξουμε ότι η πιθανότητα αυτών των λίγων πειραμάτων είναι 0, τότε μιλάμε για σύγκλιση με πιθανότητα 1, ή αλλιώς σχεδόν βέβαια σύγκλιση (convergence with probability one or almost sure convergence). Θα πρέπει να ισχύει $P(A) = 1$, όπου $A = \{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}$ και τότε γράφουμε

$$X_n(\omega) \rightarrow X(\omega) \text{ a.s.} \quad \text{όταν } n \rightarrow \infty$$

ή

$$P\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1.$$

Προσέξτε τη διαφορά με την βέβαια σύγκλιση: έχουμε ότι $A^c \neq \emptyset$ αλλά επίσης ισχύει $P(A^c) = 0$. Με άλλα λόγια, υπάρχουν πειράματα στα οποία η σύγκλιση αποτυγχάνει, η μάζα πιθανότητας που συγκεντρώνουν όμως αυτά, είναι μηδενική. Το παράδειγμα του αθροίσματος τ.μ. Bernoulli είναι μια τέτοια περίπτωση.

3.2.4 Σύγκλιση με την έννοια της τετραγωνικής μέσης τιμής

Μια ασθενέστερη περίπτωση σύγκλισης είναι αυτή της σύγκλισης με την έννοια της τετραγωνικής μέσης τιμής (mean square sense convergence), όπου ζητούμε από την δεύτερη ροπή της τυχαίας ακολουθίας να συγκλίνει στη δεύτερη ροπή μιας τ.μ. Λέμε ότι μια τυχαία ακολουθία συγκλίνει σε μια τ.μ. με την έννοια της τετραγωνικής μέσης τιμής όταν ισχύει

$$\lim_{n \rightarrow \infty} E[(X_n - X)^2] = 0.$$

Παράδειγμα: Έστω ω μια τ.μ. ομοιόμορφα κατανομημένη στο $[0, 1]$. Συγκλίνει η ακολουθία $V_n(\omega) = \omega \left(1 - \frac{1}{n}\right)$ με την έννοια της τετραγωνικής μέσης τιμής;

Λύση: Από προηγούμενο παράδειγμα γνωρίζουμε ήδη ότι η $V_n(\omega)$ συγκλίνει βέβαια στην ω . Άρα εξετάζουμε το

$$E[(V_n(\omega) - \omega)^2] = E\left[\left(-\frac{\omega}{n}\right)^2\right] = \int_0^1 \left(\frac{\omega}{n}\right)^2 d\omega = \frac{1}{n^2},$$

όπου χρησιμοποιήσαμε την κατανομή της ω . Άρα ισχύει $\lim_{n \rightarrow \infty} E[(V_n(\omega) - \omega)^2] = 0$, που σημαίνει ότι η V_n συγκλίνει στην ω με την έννοια της τετραγωνικής μέσης τιμής. ■

3.2.5 Σύγκλιση σε πιθανότητα

Σε περιπτώσεις που τα δείγματα ω , για τα οποία η σύγκλιση αποτυγχάνει, συγκεντρώνουν μάζα πιθανότητας μη μηδενική (άρα δεν ισχύουν οι παραπάνω τύποι σύγκλισης), είναι δυνατόν να αναρωτηθούμε αν η πιθανότητα του κριτηρίου του Cauchy συγκλίνει στο μηδέν, αν γίνεται δηλαδή αυθαιρέτως μικρή για αρκετά μεγάλο n . Σε αυτήν την περίπτωση (που είναι και η ασθενέστερη μορφή σύγκλισης) γράφουμε

$$X_n(\omega) \rightarrow X(\omega) \text{ in prob.} \quad \text{όταν } n \rightarrow \infty$$

ή

$$\lim_{n \rightarrow \infty} P(X_n = X) = 1$$

ή

$$P(|X_n - X| > \epsilon) \rightarrow 0 \text{ όταν } n \rightarrow \infty.$$

Οι παραπάνω τύποι σύγκλισης σε τ.μ. έχουν την παρακάτω σχέση

$$\text{βέβαια} \Rightarrow \text{με πιθανότητα } 1 \Rightarrow \text{σε πιθανότητα}$$

Παράδειγμα: Έστω ο δειγματικός χώρος $\Omega = [0, 1]$ και το ω διαλέγεται με ομοιόμορφο τρόπο. Για $n = 1, 2, \dots$ ορίζουμε τις παρακάτω ακολουθίες

$$\begin{aligned} U_n(\omega) &= \frac{\omega}{n} \\ V_n(\omega) &= \omega \left(1 - \frac{1}{n}\right) \\ W_n(\omega) &= \omega e^n \\ Y_n(\omega) &= \cos(2\pi n\omega) \\ Z_n(\omega) &= e^{-n(n\omega-1)}. \end{aligned}$$

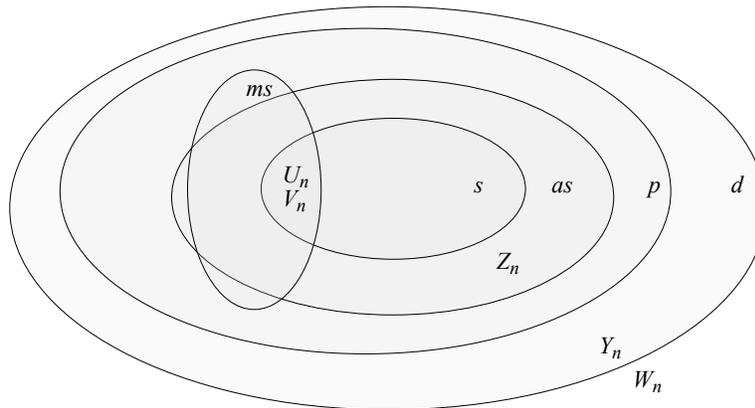
Εξετάστε τις ακολουθίες αυτές ως προς τη σύγκλισή τους.

Λύση: (α) Μπορούμε να δείξουμε ότι $U_n \rightarrow 0$ για κάθε ω . Πράγματι, για κάθε ω έχουμε $U_n \leq \frac{1}{n}$ που μπορεί ναδειχθεί ότι συγκλίνει στο 0. Άρα, η U_n συγκλίνει βέβαια σε αριθμό.

(β) Για την V_n παρατηρούμε ότι $\lim_{n \rightarrow \infty} V_n = \omega$ για κάθε ω . Όμως, η $X(\omega) = \omega$ είναι η ομοιόμορφη τ.μ. στο $[0, 1]$. Άρα, η V_n συγκλίνει βέβαια στην X .

(γ) Για $\omega = 0$ έχουμε ότι η W_n συγκλίνει στο μηδέν, αλλά για $\omega \neq 0$ αποκλίνει. Οπότε, η W_n αποκλίνει.

(δ) Για $\omega \neq 0$ η Y_n ταλαντώνεται μεταξύ -1 και 1 , οπότε δε συγκλίνει.



Σχήμα 3.2: Σχέση μεταξύ τύπων σύγκλισης: s: βέβαια σύγκλιση, as: σχεδόν βέβαια σύγκλιση (με πιθανότητα 1), p: σύγκλιση σε πιθανότητα, d: σύγκλιση σε κατανομή και ms: σύγκλιση με την έννοια της μέσης τετραγωνικής τιμής. Στο σχήμα ταξινομούνται και τυχαίες ακολουθίες των παραδειγμάτων.

(ε) Για $\omega = 0$, η Z_n αποκλίνει. Όταν, $\omega \neq 0$ συγκλίνει στο 0 καθώς μπορούμε πάντα να βρούμε n αρκετά μεγάλο ώστε $n > 1/\omega$. Όμως $P(\omega = 0) = 0$, άρα η Z_n συγκλίνει με πιθανότητα 1 στο 0. ■

Δοκιμάστε στο matlab αυτό: `convergence_exercise.m`

3.2.6 Σύγκλιση σε κατανομή

Για μια τυχαία ακολουθία $\{X_n\}$ με κατανομές $\{F_n(x)\}$, λέμε ότι συγκλίνει σε κατανομή όταν υπάρχει κατανομή F έτσι ώστε

$$F_n(x) \rightarrow F(x) \text{ in prob.} \quad \text{όταν } n \rightarrow \infty$$

για όλα τα x που η F είναι συνεχής. Η σύγκλιση σε κατανομή δεν πρέπει να συγχέεται με τους τύπους σύγκλισης σε τ.μ.

3.3 Δειγματική μέση τιμή

Έστω X_1, \dots, X_n ανεξάρτητες τ.μ. με την ίδια κατανομή (independent and identically distributed ή i.i.d.) και έστω η τ.μ.

$$S_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Η τ.μ. αυτή ονομάζεται δειγματική μέση τιμή (sample mean). Μπορεί κανείς να σκεφτεί την S_n ως το εξής τυχαίο πείραμα: λαμβάνουμε n τιμές από τη μεταβλητή X και προσπαθούμε να εκτιμήσουμε τον μέσο όρο της. Περιμένουμε, φυσικά, η S_n να έχει μεγάλη σχέση με την $E[X]$ αλλά όχι απαραίτητα να είναι ίση.

Η τ.μ. S_n είναι ένας **εκτιμητής** μέσης τιμής. Ο εκτιμητής έχει δύο ιδιότητες:

1. θα πρέπει η αναμενόμενη τιμή του να είναι κοντά στην αναμενόμενη τιμή της τ.μ. που θέλουμε να εκτιμήσουμε. Όταν είναι ακριβώς ίση, ο εκτιμητής λέγεται μη πολωμένος (unbiased estimator).

2. Η μεταβλητότητά του θα πρέπει να είναι μικρή.

Για τη δειγματική μέση τιμή μπορούμε εύκολα να υπολογίσουμε:

$$E[S_n] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = E[X]$$

και

$$\text{Var}(S_n) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\text{Var}(X)}{n}.$$

Άρα, η δειγματική μέση τιμή είναι ένας μη πολωμένος εκτιμητής με καλές ιδιότητες για n μεγάλο.

Παράδειγμα: Θέλουμε να εκτιμήσουμε το $J = \int_0^1 g(x)dx$ με $0 \leq g(x) \leq 1$ για όλα τα x . Έχουμε στη διάθεσή μας τους παρακάτω εκτιμητές

(α) $U = \mathbb{1}_{\{Y \leq g(X)\}}$

(β) $V = g(X)$

(γ) $W = \frac{1}{2}[g(X) + g(1 - X)]$

όπου X, Y ανεξάρτητες τ.μ. και $X, Y \sim U(0, 1)$. Εξετάστε την ποιότητά τους ως εκτιμητές.

Λύση:

(α) Έχουμε

$$\begin{aligned} E[U] &= E[\mathbb{1}_{\{Y \leq g(X)\}}] = P(Y \leq g(X)) = \int_0^1 \int_0^{g(x)} dy dx = \int_0^1 g(x) dx = J \\ E[U^2] &= E[\mathbb{1}_{\{Y \leq g(X)\}}^2] = E[\mathbb{1}_{\{Y \leq g(X)\}}] = J. \end{aligned}$$

(β) Έχουμε

$$\begin{aligned} E[V] &= E[g(X)] = \int_0^1 g(x) dx = J \\ E[V^2] &= E[g^2(X)] = \int_0^1 g^2(x) dx \leq \int_0^1 g(x) dx = J, \end{aligned}$$

αφού $g(x) \leq 1$.

(γ) Έχουμε

$$\begin{aligned} E[W] &= E\left[\frac{1}{2}[g(X) + g(1 - X)]\right] = \frac{1}{2} \left(\int_0^1 g(x) dx + \int_0^1 g(1 - x) dx \right) = J \\ E[W^2] &= \frac{1}{4} E[g^2(X) + 2g(X)g(1 - X) + g^2(1 - X)] = \frac{1}{2} E[V^2] + \frac{1}{2} \int_0^1 g(x)g(1 - x) dx \\ &= E[V^2] + \frac{1}{2} \int_0^1 g(x)[g(1 - x) - g(x)] dx = E[V^2] - \frac{1}{4} \int_0^1 [-2g(x)g(1 - x) + 2g^2(x)] dx \\ &= E[V^2] - \frac{1}{4} \int_0^1 [g(x) + g(1 - x)]^2 dx \leq E[V^2], \end{aligned}$$

αφού το ολοκλήρωμα είναι μια θετική ποσότητα. Συμπερασματικά, όλοι οι εκτιμητές είναι μη πολωμένοι και έχουμε $\text{Var}(W) \leq \text{Var}(V) \leq \text{Var}(U)$, δηλαδή ο W είναι ο καλύτερος εκτιμητής. ■

3.4 Νόμος των μεγάλων αριθμών

Αν χρησιμοποιήσουμε την ανισότητα Chebyshev για την δειγματική μέση τιμή έχουμε:

$$\begin{aligned} P(|S_n - E[S_n]| \geq \epsilon) &\leq \frac{\text{Var}(S_n)}{\epsilon^2} \\ &\text{ή} \\ P(|S_n - E[X]| \geq \epsilon) &\leq \frac{\text{Var}(X)}{n\epsilon^2} \\ &\text{ή} \\ P(|S_n - E[X]| < \epsilon) &\geq 1 - \frac{\text{Var}(X)}{n\epsilon^2}. \end{aligned}$$

Με άλλα λόγια, η δειγματική μέση τιμή συγκλίνει σε πιθανότητα στην αναμενόμενη τιμή της X . Αυτό είναι ένα πολύ σημαντικό οριακό θεώρημα για τη θεωρία πιθανοτήτων. Μας λέει ότι για μεγάλο n , η μάζα πιθανότητας της δειγματικής μέσης τιμής συγκεντρώνεται σε μια μικρή περιοχή γύρω από το $E[X]$. Επαληθεύει, ουσιαστικά, την διαίσθηση ότι λαμβάνοντας μέσους όρους με αρκετά δεδομένα, μπορούμε να υπολογίσουμε την αναμενόμενη τιμή.

Η παραπάνω σύγκλιση σε πιθανότητα ονομάζεται ‘Ασθενής νόμος των μεγάλων αριθμών’. Υπάρχει, επίσης, ο ‘Ισχυρός νόμος των μεγάλων αριθμών’ ο οποίος λέει ότι η δειγματική μέση τιμή συγκλίνει με πιθανότητα 1 στην αναμενόμενη τιμή, δηλαδή

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \rightarrow X \text{ a.s.}$$

ή

$$P\left(\lim_{n \rightarrow \infty} \frac{X_1 + X_2 + \cdots + X_n}{n} \rightarrow X\right) = 1.$$

Παράδειγμα: (Δημοσκόπηση): Έστω p το ποσοστό των ψηφοφόρων που υποστηρίζουν κάποιον υποψήφιο. Θέλουμε να κάνουμε μια δημοσκόπηση ρωτώντας τυχαίους περαστικούς τι θα ψηφίσουν. Πόσους περαστικούς πρέπει να ρωτήσουμε για να έχουμε σφάλμα 0.01 με βεβαιότητα 95% στα αποτελέσματά μας;

Λύση: Θα θεωρήσουμε ότι κάθε περαστικός είναι μια Bernoulli τ.μ. που υποστηρίζει τον υποψήφιο μας με πιθανότητα p . Άρα, ρωτώντας n περαστικούς, το άθροισμα των θετικών απαντήσεων θα είναι η δειγματική μέση τιμή της τ.μ. αυτής. Γνωρίζουμε ότι

$$P(|S_n - p| \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2}.$$

Δυστυχώς, η τιμή της p είναι άγνωστη. Γνωρίζουμε όμως ότι $p(1-p) \leq \frac{1}{4}$ για κάθε πιθανή τιμή της p . Συνεπώς,

$$P(|S_n - p| \geq \epsilon) \leq \frac{1}{4n\epsilon^2}.$$

Για $\epsilon = 0.01$ και πιθανότητα 0.05 χρειαζόμαστε, λοιπόν $n \geq 50,000$ περαστικούς. Το αποτέλεσμα αυτό, πάντως, είναι λίγο απαισιόδοξο τόσο λόγω του p (όταν αυτό είναι κοντά στο 1 ή στο 0), όσο και λόγω της ανισότητας Chebyshev που δεν είναι τόσο αυστηρή σε αυτήν την περίπτωση. ■

3.5 Κεντρικό οριακό θεώρημα

Σύμφωνα με το κεντρικό οριακό θεώρημα, για κάθε τ.μ. με μέση τιμή μ , μεταβλητότητα σ^2 και δειγματική μέση τιμή S_n , η τ.μ. $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$ (η οποία έχει αναμενόμενη τιμή μηδέν και μεταβλητότητα μονάδα) συγκλίνει σε κατανομή στην $\mathcal{N}(0, 1)$. Ισχύει δηλαδή

$$\lim_{n \rightarrow \infty} P(Z_n \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-x^2/2} dx.$$

Απόδειξη: Έστω η τ.μ. $Y_i = \frac{X_i - \mu}{\sqrt{n}\sigma}$. Προφανώς θα έχουμε $E[Y_i] = 0$ και $\text{Var}(Y_i) = E[Y_i^2] = \frac{1}{n}$. Για την χαρακτηριστική συνάρτηση της Y_i μπορούμε να γράψουμε

$$\Phi_{Y_i}(\omega) = 1 + j \frac{\omega E[Y_i]}{1!} - \frac{\omega^2 E[Y_i^2]}{2!} - j \frac{\omega^3 E[Y_i^3]}{3!} + \dots = 1 - \frac{\omega^2}{2n} + O\left(\frac{\omega^2}{n}\right).$$

Ως $O\left(\frac{\omega^2}{n}\right)$ ορίζουμε μια συνάρτηση του n που πηγαίνει ταχύτερα στο μηδέν από ότι η συνάρτηση $\frac{\omega^2}{n}$. Στη συνέχεια υπολογίζουμε την χαρακτηριστική συνάρτηση του Z_i

$$\Phi_{Z_i}(\omega) = \Phi_{\sum Y_i}(\omega) = \prod_{i=1}^n \Phi_{Y_i}(\omega) = \Phi_{Y_i}^n(\omega).$$

Άρα

$$\lim_{n \rightarrow \infty} \Phi_{Z_i}(\omega) = \lim_{n \rightarrow \infty} \left(1 - \frac{\omega^2}{2n} + O\left(\frac{\omega^2}{n}\right)\right)^n = e^{-\frac{\omega^2}{2}}.$$

Δηλαδή, η Z_i συγκλίνει σε χαρακτηριστική συνάρτηση στην $\mathcal{N}(0, 1)$. Από το θεώρημα συνέχειας του Levy προκύπτει ότι συγκλίνει και σε κατανομή στην $\mathcal{N}(0, 1)$. ■

Το κεντρικό οριακό θεώρημα είναι πολύ σημαντικό καθώς λέει ότι το άθροισμα πολλών ανεξάρτητων τυχαίων παραγόντων είναι κατανομημένο κατά Gauss ανεξάρτητα από τις αρχικές κατανομές, συνεχείς ή διακριτές. Αυτό το θεμελιώδες αποτέλεσμα εξηγεί διαισθητικά γιατί ο θόρυβος στις μετρήσεις μπορεί να μοντελοποιηθεί ως Gauss.

Ανάλογα με την κατανομή του X , απαιτείται διαφορετικό n έτσι ώστε η σύγκλιση να έχει ικανή ακρίβεια. Για παράδειγμα, σε ομοιόμορφες κατανομές $n = 8$ είναι αρκετό, ενώ σε εκθετικές απαιτείται αρκετά μεγαλύτερο n .

Δοκιμάστε στο matlab αυτό: `central_limit_theorem.m`

Ερώτηση: Αποδείξτε την ισχύ του θεωρήματος κάνοντας τις εξής πράξεις:

$$\Phi_{Z_i}(\omega) = E[e^{j\omega Z_i}] = \dots = e^{-\frac{\omega^2}{2}}.$$

3.5.1 Προσέγγιση Gauss

Χρησιμοποιώντας το ΚΟΘ μπορούμε να προβλέψουμε με μεγάλη ακρίβεια πιθανότητες που αφορούν ένα μεγάλο άθροισμα από ανεξάρτητες μεταβλητές.

Παράδειγμα: (Δημοσκόπηση 2): Στο παραπάνω παράδειγμα της δημοσκόπησης, παρατηρούμε ότι για αρκετά μεγάλο n , το $S_n - p$ είναι κατανομημένο κατά Gauss με αναμενόμενη τιμή 0 και μεταβλητότητα $\frac{p(1-p)}{n}$.

Μπορούμε να ισχυριστούμε, δηλαδή, ότι κατά προσέγγιση ισχύει $S_n - p \approx Y \sim \mathcal{N}(0, \frac{p(1-p)}{n})$. Άρα, η ζητούμενη πιθανότητα είναι

$$P(|S_n - p| \geq \epsilon) \approx P(Y \leq -\epsilon \cup Y \geq \epsilon) = 2(1 - F_Y(\epsilon)),$$

λόγω συμμετρίας. Καθώς αυξανόμενης της μεταβλητότητας (η καμπάνα απλώνει) η πιθανότητα αυξάνει, χρησιμοποιώντας τη μέγιστη μεταβλητότητα, που είναι $\frac{1}{4n}$, βρίσκουμε ένα πάνω φράγμα για την πιθανότητα:

$$2(1 - F_Y(\epsilon)) \leq 2(1 - F_Z(\epsilon)),$$

όπου $Z \sim \mathcal{N}(0, \frac{1}{4n})$. Στο σημείο αυτό πρέπει να ανατρέξουμε σε πίνακες της κανονικής κατανομής ώστε να βρούμε το μικρότερο n που για $\epsilon = 0.01$ εγγυάται $2(1 - F_Z(\epsilon)) \leq 0.05$. Αυτό είναι το $n = 9604$ που είναι πολύ πιο ενθαρρυντικό από το αποτέλεσμα που λάβαμε χρησιμοποιώντας την ανισότητα Chebyshev. Μπορεί να βελτιωθεί ακόμα περισσότερο αν γνωρίζουμε από πριν μια εκτίμηση του p , π.χ. ο υποψήφιος αποκλείεται να έχει δημοτικότητα περισσότερο από 10%. Στην πράξη, η εμπιστοσύνη στις δημοσκοπήσεις προκύπτει μετά την εξέταση των δεδομένων. Το εύρος των τιμών που περιλαμβάνει ένα ποσοστό των δεδομένων ονομάζεται διάστημα εμπιστοσύνης (confidence interval).

Στη προσέγγιση Gauss δεν βρίσκουμε αυστηρό όριο, όπως στο προηγούμενο παράδειγμα, παρά μόνο μια προσέγγιση της ζητούμενης πιθανότητας, οπότε θα ήταν χρήσιμο να γνωρίζουμε πόσο καλή είναι αυτή η προσέγγιση. Γενικά, όταν το n είναι τόσο μεγάλο όσο στο παραπάνω παράδειγμα, η προσέγγιση Gauss είναι εξαιρετικά καλή.

3.6 Εκτίμηση

Σε πολλές πρακτικές εφαρμογές θέλουμε να εκτιμήσουμε την τιμή μιας τυχαίας μεταβλητής, πολλές φορές έχοντας κάποια μερική γνώση. Για παράδειγμα, μπορεί να θέλουμε να εκτιμήσουμε την τιμή μιας μετοχής γνωρίζοντας το παρελθόν της, ή τη στάθμη του σήματος γνωρίζοντας μια ενθόρυβη μέτρηση.

3.6.1 Μέσο τετραγωνικό σφάλμα

Το πρώτο πρόβλημα που ανακύπτει είναι πως θα αποτιμήσουμε την ποιότητα της εκτίμησης, δηλαδή το σφάλμα της. Η πιο διαδεδομένη μετρική της ποιότητας ενός σφάλματος είναι το **μέσο τετραγωνικό σφάλμα**. Έστω η τ.μ. που προσπαθούμε να εκτιμήσουμε X και η εκτίμηση που διαθέτουμε c . Το μέσο τετραγωνικό σφάλμα ορίζεται ως $E[(X - c)^2]$. Η εκτίμηση μπορεί να είναι και αυτή μια τυχαία μεταβλητή, ή μια σταθερά.

Μια ενδελεχής ανάλυση για μετρικές σφάλματος μπορείτε να βρείτε στο [2].

3.6.2 Εκτίμηση της X

Όταν δεν έχουμε άλλη γνώση, η καλύτερη δυνατή εκτίμηση της X με το κριτήριο του μέσου τετραγωνικού σφάλματος είναι η $c^* = E[X]$. Έχουμε

$$E[(X - c)^2] = \text{Var}(X - c) + E[X - c]^2 = \text{Var}(X) + (E[X] - c)^2, \quad (3.1)$$

άρα το $c = E[X]$ ελαχιστοποιεί αυτόν τον όρο. Επιπλέον, ο εκτιμητής $E[X]$ ονομάζεται και μη πολωμένος αφού για το μέσο σφάλμα έχουμε $E[X - E[X]] = 0$.

3.6.3 Εκτίμηση της X δεδομένης της $Y = y$

Συχνά μας δίνεται η γνώση μιας τ.μ. Y συναφούς με την X (για παράδειγμα μπορεί να μας δίνεται η γνώση του θορύβου ενώ προσπαθούμε να εκτιμήσουμε το σήμα λήψης). Αυτό σημαίνει ότι πριν εκτιμήσουμε την τιμή

X (*smega*) μαθαίνουμε με κάποιο τρόπο ότι $Y(\omega) = y$. Κατά συνέπεια, πρέπει να βρούμε το c που ελαχιστοποιεί το μέγεθος $E[(X - c)^2|Y = y]$. Επαναλαμβάνοντας το (3.1), βρίσκουμε ότι

$$c^* = E[X|Y = y],$$

που είναι συνάρτηση της τιμής y όπως αναμενόταν.

Παραπάνω είδαμε ότι η εκτίμηση της X είναι συνάρτηση των τιμών της Y . Κατέπεκταση, μπορούμε να θεωρήσουμε την πιο γενική μορφή εκτιμητή που είναι μια συνάρτηση $g(Y)$. Θα δούμε ότι η καλύτερη τέτοια συνάρτηση (πάντα με κριτήριο το μέσο τετραγωνικό σφάλμα είναι η $E[X|Y]$). Από το παραπάνω έχουμε

$$\begin{aligned} E[(X - E[X|Y = y])^2|Y = y] &\leq E[(X - c)^2|Y = y] \Rightarrow \\ E[(X - E[X|Y = y])^2|Y = y] &\leq E[(X - g(y))^2|Y = y] \Rightarrow \\ E[(X - E[X|Y])^2|Y] &\leq E[(X - g(Y))^2|Y] \Rightarrow \\ E[E[(X - E[X|Y])^2|Y]] &\leq E[E[(X - g(Y))^2|Y]] \Rightarrow \\ E[(X - E[X|Y])^2] &\leq E[(X - g(Y))^2] \end{aligned}$$

3.6.4 Ιδιότητες σφάλματος εκτίμησης

Θα χρησιμοποιήσουμε τον συμβολισμό $\hat{X} = E[X|Y]$ για τον εκτιμητή και $\tilde{X} = X - \hat{X}$ για το σφάλμα εκτίμησης (και τα δύο είναι τ.μ.). Αρχικά παρατηρούμε ότι

$$E[\tilde{X}|Y] = E[X - \hat{X}|Y] = E[X|Y] - E[\hat{X}|Y] = \hat{X} - \hat{X} = 0.$$

Άρα, η δεσμευμένη μέση τιμή του σφάλματος είναι 0. Επίσης, σύμφωνα με τον νόμο της υπό συνθήκη μέσης τιμής θα έχουμε

$$E[\tilde{X}] = E[E[\tilde{X}|Y]] = 0.$$

Άρα και η μέση τιμή του σφάλματος είναι 0. Στη συνέχεια υπολογίζουμε

$$\text{Cov}(\hat{X}, \tilde{X}) = E[(\hat{X} - E[X])\tilde{X}] = E[E[(\hat{X} - E[X])\tilde{X}|Y]] = E[(\hat{X} - E[X])E[\tilde{X}|Y]] = 0$$

που σημαίνει ότι ο εκτιμητής και το σφάλμα είναι ασυσχέτιστα. Τέλος,

$$\text{Var}(X) = \text{Var}(\tilde{X} + \hat{X}) \stackrel{\text{ανεξ.}}{=} \text{Var}(\tilde{X}) + \text{Var}(\hat{X}),$$

δηλαδή, η διασπορά της X μπορεί να αναλυθεί στη διασπορά του εκτιμητή και του σφάλματος.

Παράδειγμα: Έστω ότι η παρατηρούμενη τ.μ. Y δεν παρέχει πληροφορία για την X επειδή είναι ανεξάρτητες. Τότε βλέπουμε ότι $\hat{X} = E[X|Y] = E[X]$ και θα ισχύει ότι ο εκτιμητής είναι μια σταθερά. Επιπλέον, θα ισχύει $\text{Var}(\hat{X}) = 0$ αφού είναι σταθερά. Άρα, καταλήγουμε ότι αν X, Y ανεξάρτητες τότε

$$\text{Var}(X) = \text{Var}(\tilde{X}).$$

Αυτή είναι μια συνθήκη που μπορεί να χρησιμοποιηθεί για να χαρακτηρίσει όλες τις περιπτώσεις που η μέτρηση της Y δεν μας βοηθάει να εκτιμήσουμε καλύτερα την X .

3.6.5 Εκτίμηση της X δεδομένων πολλών μετρήσεων Y_1, Y_2, \dots, Y_n

Επεκτείνοντας τα παραπάνω μπορούμε να δώσουμε μια γενική λύση στο πρόβλημα της εκτίμησης με κριτήριο το μέσο τετραγωνικό σφάλμα. Θα ισχύει δηλαδή

$$E[(X - E[X|Y_1, Y_2, \dots, Y_n])^2] \leq E[(X - g(Y_1, Y_2, \dots, Y_n))^2].$$

που σημαίνει ότι ο εκτιμητής $E[X|Y_1, Y_2, \dots, Y_n]$ ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα. Καθώς, όμως, δεν είναι πάντα εύκολο να κατασκευάσουμε έναν τέτοιο εκτιμητή, έχει ενδιαφέρον να γνωρίζουμε την αποδοτικότητα άλλων εκτιμητών. Εδώ θα εξετάσουμε την πιο απλή μορφή εκτιμητή, του **γραμμικού εκτιμητή**

$$g(Y_1, Y_2, \dots, Y_n) = a_1 Y_1 + \dots + a_n Y_n + b$$

Κατά συνέπεια, το γενικό πρόβλημα της εκτίμησης με γραμμικό εκτιμητή είναι

$$\min E[(X - a_1 Y_1 - \dots - a_n Y_n - b)^2].$$

Παράδειγμα: Λύστε το πρόβλημα εκτίμησης με γραμμικό εκτιμητή για $n = 1$.

Λύση: Θέλουμε να βρούμε τα a, b που ελαχιστοποιούν το $E[(X - aY - b)^2]$. Όμως παρατηρήστε ότι η τ.μ. $X - aY$ έχει ως βέλτιστο εκτιμητή το $b^* = E[X - aY]$. Άρα, μένει να βρούμε το a που ελαχιστοποιεί το $E[(X - aY - b)^2] = E[(X - aY - E[X - aY])^2] = \text{Var}(X - aY)$. Όμως

$$\text{Var}(X - aY) = \text{Var}(X) + a^2 \text{Var}(Y) + 2\text{Cov}(X, -aY) = \text{Var}(X) + a^2 \text{Var}(Y) - 2a \text{Cov}(X, Y)$$

Για να ελαχιστοποιήσουμε το παραπάνω, λύνουμε την εξίσωση $f'(a) = 0$ και βρίσκουμε

$$a^* = \frac{\text{Cov}(X, Y)}{\text{Var}(Y)} = \rho \sqrt{\frac{\text{Var}(X)}{\text{Var}(Y)}}.$$

Τελικά μπορούμε να υπολογίσουμε το καλύτερο δυνατό σφάλμα που θα είναι

$$E[(X - a^*Y - b^*)^2] = (1 - \rho^2) \text{Var}(X).$$

Αυτό είναι λογικό καθώς η ποιότητα του γραμμικού εκτιμητή σχετίζεται άμεσα με τον συντελεστή συσχέτισης των X, Y . Για παράδειγμα, αν όταν το X είναι μεγάλο τότε το Y είναι μεγάλο, έχουμε μεγάλη συσχέτιση (κοντά στο 1) και το σφάλμα εκτίμησης του γραμμικού εκτιμητή θα είναι μικρό. ■

Κεφάλαιο 4

Στοχαστικές διαδικασίες

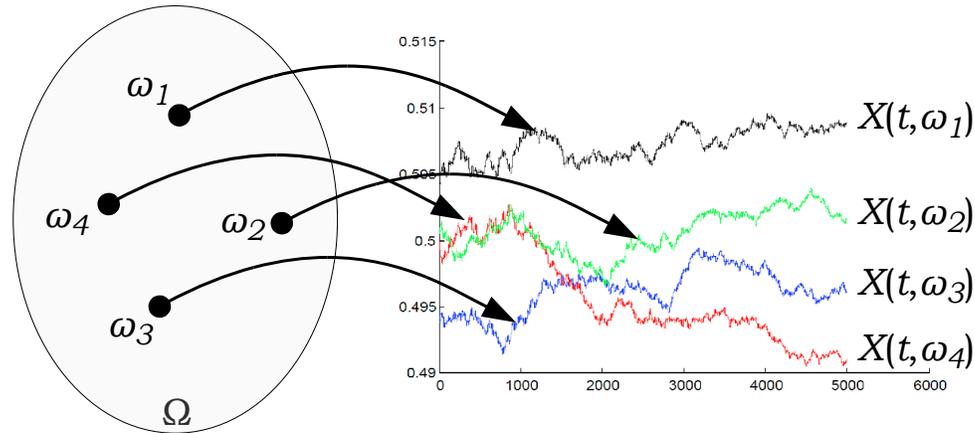
4.1 Γενικά

Η στοχαστική διαδικασία είναι μια αντιστοίχιση τυχαίων στοιχείων του δειγματικού χώρου $\omega \in \Omega$ σε συναρτήσεις. Γράφουμε $X(t, \omega)$, $X : \mathbb{R} \times \Omega \rightarrow \mathbb{R}$ όπου $t \in \mathbb{R}$ η παράμετρος της στοχαστικής διαδικασίας (η παράμετρος των τυχαίων συναρτήσεων) που μπορεί να είναι συνεχής (π.χ. χρόνος) ή διακριτή (π.χ. ένας ακέραιος). Θα περιοριστούμε στη μελέτη διαδικασιών με αιτιατότητα, δηλαδή με παράμετρο που λαμβάνει τιμές σε μονοδιάστατο σύνολο και άρα δύο τιμές μπορούν να έχουν μια σαφή σχέση σύγκρισης (partial order relation)¹. Τέτοιες διαδικασίες ονομάζονται διαδικασίες χρόνου και συμβολίζονται με X_t αν έχουμε συνεχή χρόνο και X_n αν έχουμε διακριτό.

Για σταθερό t_0 , το $X(t_0, \omega)$ είναι μια τυχαία μεταβλητή. Για σταθερό ω_0 , το $X(t, \omega_0)$ είναι μια συνάρτηση του t που ονομάζεται δειγματικό μονοπάτι (sample path). Άρα, η στοχαστική διαδικασία μπορεί να γίνει αντιληπτή και ως ένα σύνολο από τυχαίες μεταβλητές, μια λογική επέκταση, δηλαδή, των τυχαίων διανυσμάτων και των τυχαίων ακολουθιών.

Το πεδίο τιμών της παραμέτρου των συναρτήσεων λέγεται και χώρος παραμέτρου (parameter space), ενώ το πεδίο τιμών των τ.μ. λέγεται χώρος καταστάσεων (state space). Συνολικά έχουμε τέσσερις βασικές κατηγορίες στοχαστικών διαδικασιών που προκύπτουν ως ζευγάρια διαλέγοντας διακριτό ή συνεχή χώρο παραμέτρου t και

¹Τα τυχαία πεδία (random fields) είναι ένα αντίθετο παράδειγμα όπου η παράμετρος παίρνει τιμές σε μια σφαίρα n διαστάσεων



Σχήμα 4.1: Η στοχαστική διαδικασία αντιστοιχίζει σε κάθε δείγμα ω του δειγματικού χώρου ένα δειγματικό μονοπάτι.

διακριτό ή συνεχή χώρο καταστάσεων για τις τ.μ.

Επεκτείνοντας, λοιπόν, τη λογική που ακολουθήσαμε σε προηγούμενα κεφάλαια, μπορούμε να υποθέσουμε ότι γνωρίζοντας το πλήρες μοντέλο, δηλαδή την από κοινού κατανομή όλων των τυχαίων μεταβλητών, θα μπορούσαμε να υπολογίσουμε τη πιθανότητα οποιουδήποτε ζητούμενου γεγονότος. Το πρόβλημα που προκύπτει, όμως, είναι ότι μια στοχαστική διαδικασία έχει πιθανώς άπειρες (και μάλιστα μη μετρήσιμα άπειρες) τυχαίες μεταβλητές. Άρα και το πλήρες μοντέλο θα απαιτεί την από κοινού γνώση απείρων τ.μ. Επιπρόσθετα, πολύ συχνά δεν γνωρίζουμε τη μορφή του πλήρους μοντέλου ή δεν μπορούμε να αναγνωρίσουμε τον χώρο πιθανοτήτων για μια στοχαστική διαδικασία. Συνεπώς, οι μέθοδοι που χρησιμοποιούμε στις στοχαστικές διαδικασίες είναι συνήθως διαφορετικές και λιγότερο αισιόδοξες.

Το βασικότερο βήμα απώλειας γνώσης για την στοχαστική διαδικασία, που είναι όμως σχεδόν πάντα απαραίτητο στις συνεχείς διαδικασίες είναι η δειγματοληψία k σημείων της διαδικασίας

$$\{X_k\} : X_1 = X(t_1, \omega), X_2 = X(t_2, \omega), \dots, X_k = X(t_k, \omega).$$

Με άλλα λόγια, διαλέγουμε k χρονικές στιγμές που αντιστοιχούν σε k τ.μ. και μελετούμε τη συμπεριφορά τους. Η συμπεριφορά αυτή μπορεί να καθοριστεί από την απο κοινού κατανομή k -οστής τάξης:

$$F_{X_1, X_2, \dots, X_k}(x_1, x_2, \dots, x_k) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_k \leq x_k).$$

Αντίστοιχα ορίζονται και οι k -οστής τάξης ΣΠΠ και ΣΜΠ για συνεχείς και διακριτές τ.μ. Αυτές οι συναρτήσεις δεν διαφέρουν σε τίποτα στη λογική τους από τις αντίστοιχες από κοινού συναρτήσεις που συναντήσαμε σε προηγούμενα κεφάλαια. Η γνώση τους δε, δεν καθορίζει πλήρως μια διαδικασία, μπορεί όμως να μας πληροφορήσει για σημαντικές ιδιότητες, όπως είναι η στασιμότητα, η εργοδικότητα, η ιδιότητα *Markov*, η ιδιότητα στασίμων αυξήσεων κ.α.

Παράδειγμα:

- Η στάθμη του λαμβανόμενου σήματος σε ένα κινητό είναι διαδικασία συνεχούς χρόνου και συνεχούς χώρου καταστάσεων.

- Ο αριθμός πελατών στην ουρά μιας τράπεζας είναι διαδικασία συνεχούς χρόνου και διακριτού χώρου καταστάσεων.
- Οι μετρήσεις της τάσης του ηλεκτρικού δικτύου κάθε μια ώρα είναι μια διαδικασία διακριτού χρόνου και συνεχούς χώρου καταστάσεων.
- Τα φύλλα που μοιράζονται σε μια παρτίδα ποκερ από τον υπολογιστή είναι μια διαδικασία διακριτού χρόνου και διακριτού χώρου καταστάσεων.

4.1.1 Ροπές διαδικασιών

Ανάλογα με την τάξη του δείγματος της διαδικασίας που χρησιμοποιούμε μπορούμε να ορίσουμε αντίστοιχης τάξης ετεροροπές. Λόγω της περιπλοκότητας που ανακύπτει, συνήθως ασχολούμαστε με χαμηλές τάξεις, μέχρι 4. Αν λάβουμε την πρώτη τάξη σε κάθε σημείο της παραμέτρου t , μπορούμε να ορίσουμε μεγέθη όπως η **συνάρτηση αναμενόμενης τιμής** (mean function) και η **συνάρτηση μεταβλητότητας** variance function:

$$m_X(t) \equiv E[X(t)] = \int_{-\infty}^{\infty} x f_{X(t)}(x) dx$$

και

$$\sigma_X^2(t) \equiv \text{Var}(X(t)) = \int_{-\infty}^{\infty} (x - m_X(t))^2 f_{X(t)}(x) dx,$$

οι οποίες είναι συναρτήσεις του χρόνου. Προφανώς, για $t = t_0$, η τιμή $m_X(t_0)$ είναι η αναμενόμενη τιμή της τυχαίας μεταβλητής $X(t_0, \omega)$.

Κατόπιν, μπορούμε να εξετάσουμε τη δεύτερη τάξη, όπου από την διαδικασία ξεχωρίζουμε δύο διαφορετικές τ.μ. σε δύο διαφορετικές τιμές της παραμέτρου t . Ορίζουμε την **συνάρτηση αυτοσυσχέτισης** (autocorrelation function):

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X(t_1)X(t_2)}(x, y) dx dy$$

και τη **συνάρτηση αυτοσυμμεταβλητότητας** (autocovariance function)

$$C_X(t_1, t_2) = E[(X(t_1) - m_X(t_1))(X(t_2) - m_X(t_2))] = R_X(t_1, t_2) - m_X(t_1)m_X(t_2)$$

που είναι αντίστοιχα η ετεροροπή και η κεντρική ετεροροπή δεύτερης τάξης μεταξύ δύο τ.μ. της ίδιας στοχαστικής διαδικασίας (γιαυτό και το πρώτο συνθετικό -αυτο). Οι συναρτήσεις αυτές, αν και παρέχουν περιορισμένη γνώση για μια διαδικασία, είναι πολύ σημαντικές καθώς δείχνουν την συσχέτιση που έχουν δύο σημεία της διαδικασίας μεταξύ τους. Συγκεκριμένα, αυτές οι συναρτήσεις περιγράφουν τη σχέση μεταξύ οποιονδήποτε δύο τ.μ. της διαδικασίας, χωρίς να εξηγούν όμως τι συμβαίνει μεταξύ τριών τ.μ.

Μπορούμε να παρατηρήσουμε ότι $\text{Var}(X(t)) = C_X(t, t)$ και να ορίσουμε το συντελεστή συσχέτισης (correlation coefficient) μιας διαδικασίας ως

$$\rho_X(t_1, t_2) = \frac{C_X(t_1, t_2)}{\sqrt{C_X(t_1, t_1)}\sqrt{C_X(t_2, t_2)}}.$$

Παράδειγμα: (Ημιτονοειδές σήμα με τυχαίο πλάτος): Έστω η διαδικασία συνεχούς χρόνου $X(t) = A \cos 2\pi t$ όπου A μια τ.μ. Η συνάρτηση μέσης τιμής θα είναι:

$$m_X(t) = E[A \cos 2\pi t] = E[A] \cos 2\pi t.$$

Παρατηρήστε ότι για $t : \cos 2\pi t = 0$ η διαδικασία λαμβάνει μόνο μηδενικές τιμές. Η συνάρτηση αυτοσυσχέτισης θα είναι:

$$R_X(t_1, t_2) = E[A \cos 2\pi t_1 A \cos 2\pi t_2] = E[A^2] \cos 2\pi t_1 \cos 2\pi t_2$$

και η συνάρτηση αυτοσυμμεταβλητότητας:

$$\begin{aligned} C_X(t_1, t_2) &= R_X(t_1, t_2) - m_X(t_1)m_X(t_2) \\ &= E[A^2] \cos 2\pi t_1 \cos 2\pi t_2 - E[A]^2 \cos 2\pi t_1 \cos 2\pi t_2 \\ &= \text{Var}(A) \cos 2\pi t_1 \cos 2\pi t_2. \end{aligned}$$

Στην περίπτωση διακριτού χώρου καταστάσεων, τα ολοκληρώματα θα είναι αθροίσματα. Επίσης, στην περίπτωση διακριτού χρόνου διαδικασιών, οι παραπάνω ροπές θα έχουν ως παράμετρο το n αντί για το t . Οι αντίστοιχοι ορισμοί, όμως, μπορούν να προκύψουν φυσιολογικά από τα παραπάνω.

4.1.2 Πολλές διαδικασίες

Σε αναλογία με τις τυχαίες μεταβλητές, μπορούμε να θεωρήσουμε κοινά μοντέλα πιθανότητας για δύο ή περισσότερες τυχαίες διαδικασίες. Χαρακτηριστικό παράδειγμα που κάτι τέτοιο απαιτείται, είναι η περίπτωση που έχουμε μια τυχαία διαδικασία σαν είσοδο ενός συστήματος και ενδιαφερόμαστε να μελετήσουμε την έξοδο (που είναι συνάρτηση της εισόδου) μαζί με την είσοδο και πιθανόν κάποιο ανεξάρτητο θόρυβο.

Σε αντιστοιχία, μπορούμε να ορίσουμε από κοινού μοντέλα k -οστής- j -οστής τάξης λαμβάνοντας k δείγματα από την $X(t)$ και j δείγματα από την $Y(t)$. Ορίζοντας $\mathbf{X} = X(t_1), \dots, X(t_k)$, $\mathbf{Y} = Y(t'_1), \dots, Y(t'_j)$, έχουμε:

$$F_{\mathbf{XY}}(x_1, \dots, x_k, y_1, \dots, y_j) \doteq P(X(t_1) \leq x_1, \dots, X(t_k) \leq x_k, Y(t'_1) \leq y_1, \dots, Y(t'_j) \leq y_j).$$

Δύο στοχαστικές διαδικασίες είναι ανεξάρτητες **αν και μόνο αν** για κάθε επιλογή k, j και κάθε επιλογή των $t_1, \dots, t_k, t'_1, \dots, t'_j$ ισχύει

$$F_{\mathbf{XY}}(x_1, \dots, x_k, y_1, \dots, y_j) = F_{\mathbf{X}}(x_1, \dots, x_k)F_{\mathbf{Y}}(y_1, \dots, y_j).$$

Το πιο απλό μοντέλο (που όμως θα χρησιμοποιήσουμε εκτενέστερα) είναι η περίπτωση $k = j = 1$:

$$F_{X(t_1)Y(t_2)}(x, y) \doteq P(X(t_1) \leq x, Y(t_2) \leq y).$$

Προφανώς, για κάθε επιλογή t_1, t_2 (δύο χρονικών στιγμών στις διαδικασίες X, Y αντίστοιχα) είναι δυνατόν να έχουμε μια διαφορετική τέτοια συνάρτηση. Κάθε τέτοια συνάρτηση, όμως, είναι συνάρτηση κατανομής και τηρεί τις συνθήκες που περιγράψαμε στο 1^ο κεφάλαιο. Αντίστοιχα, μπορούμε να ορίσουμε τη **συνάρτηση ετεροσυσχέτισης** (cross-correlation)

$$R_{X,Y}(t_1, t_2) = E[X(t_1)Y(t_2)].$$

Δύο διαδικασίες $X(t), Y(t)$ λέγονται **ορθογώνιες** αν $R_{X,Y}(t_1, t_2) = 0$ για όλα τα t_1, t_2 . Επίσης, ορίζουμε την **ετεροσυμμεταβλητότητα** (cross-covariance)

$$C_{X,Y}(t_1, t_2) = E[(X(t_1) - m_X(t_1))(Y(t_2) - m_Y(t_2))] = R_{X,Y}(t_1, t_2) - m_X(t_1)m_Y(t_2).$$

Δύο διαδικασίες $X(t), Y(t)$ λέγονται **ασυσχέτιστες** αν $C_{X,Y}(t_1, t_2) = 0$ για όλα τα t_1, t_2 .

Παράδειγμα: (Λήψη σήματος σε κανάλι προσθετικού ανεξάρτητου θορύβου.) Έστω το σήμα λήψεως $Y(t) = X(t) + N(t)$, όπου $N(t)$ είναι ανεξάρτητος θόρυβος. Βρείτε την ετεροσυσχέτιση μεταξύ της

εξόδου και της εισόδου.

Λύση: Θα είναι:

$$\begin{aligned} R_{XY}(t_1, t_2) &= E[X(t_1)Y(t_2)] \\ &= E[X(t_1)X(t_2) + X(t_1)N(t_2)] \\ &= R_X(t_1, t_2) + E[X(t_1)N(t_2)] \\ &\stackrel{\text{ανεξ.}}{=} R_X(t_1, t_2) + m_X(t_1)m_N(t_2). \end{aligned}$$

Αν υποθέσουμε ότι ο θόρυβος έχει μέση τιμή μηδέν, τότε παρατηρούμε ότι η ετεροσυσχέτιση ταυτίζεται με την αυτοσυσχέτιση του σήματος εισόδου. Αυτό είναι λογικό καθώς η έξοδος είναι συσχετισμένη με την είσοδο αλλά όχι με τον θόρυβο. ■

4.2 Στασιμότητα

Κάποιες στοχαστικές διαδικασίες έχουν την ιδιότητα ότι η φύση της τυχαιότητάς τους παραμένει σταθερή με το χρόνο. Δηλαδή, η παρατηρούμενη συμπεριφορά της διαδικασίας δεν εξαρτάται από την αρχική στιγμή δειγματοληψίας παρά μόνο από τις χρονικές διαφορές μεταξύ των δειγμάτων. Μια στοχαστική διαδικασία λέγεται **στάσιμη με την αυστηρή έννοια** όταν για οποιοδήποτε k , οποιαδήποτε δείγματα t_1, \dots, t_k και οποιαδήποτε χρονική διαφορά $\tau \in \mathbb{R}$ ισχύει:

$$F_{X(t_1), \dots, X(t_k)}(x_1, \dots, x_k) = F_{X(t_1+\tau), \dots, X(t_k+\tau)}(x_1, \dots, x_k).$$

Αυτό σημαίνει ότι αν μετακινήσουμε το σημείο του πρώτου δείγματος (t_1) κατά τ φροντίζοντας να διατηρήσουμε την σχετική θέση όλων των δειγμάτων ίδια, τότε η συμπεριφορά των δειγμάτων θα είναι η ίδια.

Δύο διαδικασίες $X(t), Y(t)$ λέγονται **από κοινού στάσιμες** αν ο παραπάνω ορισμός ισχύει για την από κοινού κατανομή των δύο διαδικασιών.

Η 'αυστηρότητα' του παραπάνω ορισμού έγκειται στο γεγονός ότι βάση του ορισμού θα πρέπει τα μοντέλα όλων των τάξεων να παρουσιάζουν στασιμότητα. Κατά συνέπεια, για την κατανομή πρώτης τάξης βρίσκουμε ότι

$$F_X(t)(x) = F_X(t+\tau)(x) = F_X(x).$$

Δηλαδή, η κατανομή μιας στάσιμης διαδικασίας παραμένει σταθερή για όλα τα t . Αυτό σημαίνει ότι οι αντίστοιχες συναρτήσεις μέσης τιμής και μεταβλητότητας θα είναι σταθερές

$$m_X(t) = {}^{\sigma\tau} m \quad \text{και} \quad \sigma_X^2(t) = {}^{\sigma\tau} \sigma^2.$$

Για την κατανομή δεύτερης τάξης βρίσκουμε ότι

$$F_{X(t_1), X(t_2)}(x_1, x_2) = F_{X(t_1+\tau), X(t_2+\tau)}(x_1, x_2) = F_{X(t_1), X(t_1+\tau)}(x_1, x_2) = F_{X(0), X(\tau)}(x_1, x_2).$$

Δηλαδή, η από κοινού κατανομή δύο τ.μ. της διαδικασίας είναι ίδια για όλες τις μεταβλητές που απέχουν $\tau = t_2 - t_1$ μεταξύ τους. Αντίστοιχα, οι ροπές δεύτερης τάξης θα είναι συναρτήσεις μόνο του τ :

$$R_X(t_1, t_2) = {}^{\sigma\tau} R_X(\tau) \quad \text{και} \quad C_X(t_1, t_2) = {}^{\sigma\tau} C_X(\tau).$$

Παράδειγμα: Ορίζουμε ως διαδικασία αθροίσματος την διακριτού χρόνου διαδικασία $S_n = X_1 + \dots + X_n$, $n \in \mathbb{N}^*$ και X_i i.i.d. τυχαίες μεταβλητές με μέση τιμή m_X και μεταβλητότητα σ_X^2 . Είναι η διαδικασία αυτή

στάσιμη;

Λύση: Η απάντηση είναι όχι. Όπως γνωρίζουμε από τα προηγούμενα κεφάλαια ισχύει

$$m_S(n) = nm_X \quad \text{και} \quad \sigma_S^2(n) =^{\text{ανεξ.}} n\sigma_X^2.$$

Άρα, τόσο η μέση τιμή όσο και η μεταβλητότητα είναι συναρτήσεις του χρόνου², δηλαδή η διαδικασία S_n δεν είναι στάσιμη. ■

4.2.1 Στασιμότητα με την ευρεία έννοια (WSS)

Μια στοχαστική διαδικασία X_t λέγεται στάσιμη με την ευρεία έννοια (wide-sense stationary ή WSS) όταν για κάθε t ισχύει:

$$m_X(t) = m \quad \text{και} \quad C_X(t_1, t_2) = C_X(\tau)$$

δηλαδή, η μέση τιμή είναι σταθερή και η αυτοσυμμεταβλητότητα εξαρτάται μόνο από τη διαφορά $\tau = t_2 - t_1$. Ισοδύναμα, η αυτοσυσχέτιση θα είναι συνάρτηση μόνο του τ ($R_X(\tau)$). Όμοια, μπορούμε να ορίσουμε και την από κοινού στασιμότητα με την ευρεία έννοια για δύο στοχαστικές διαδικασίες.

Παράδειγμα: (Μια WSS διαδικασία δεν είναι απαραίτητα στάσιμη με την αυστηρή έννοια)

Έστω η διαδικασία διακριτού χρόνου X_n που είναι ίση με την τ.μ. Y σε περιττούς χρόνους και ίση με την τ.μ. Z σε άρτιους, όπου

$$P(Y = y) = \begin{cases} \frac{1}{2} & y = -1 \\ \frac{1}{2} & y = 1 \\ 0 & \text{αλλού} \end{cases} \quad \text{και} \quad P(Z = z) = \begin{cases} \frac{1}{10} & z = -3 \\ \frac{9}{10} & z = \frac{1}{3} \\ 0 & \text{αλλού} \end{cases}$$

Είναι η X_n στάσιμη;

Λύση: Με την αυστηρή έννοια, η X_n δεν είναι στάσιμη καθώς η ΣΜΠ δεν είναι ίδια για όλα τα n . Για τα περιττά n είναι ίση με p_Y και για άρτια είναι ίση με p_Z που εμφανώς είναι διαφορετικές. Όμως, τόσο για n άρτιο, όσο και για n περιττό βρίσκουμε

$$m_X(n) = 0 \quad \text{και} \quad C_X(n_1, n_2) = \begin{cases} E[X(n_1)]E[X(n_2)] = 0 & n_1 \neq n_2 \\ E[X(n_1)]^2 = 1 & n_1 = n_2. \end{cases}$$

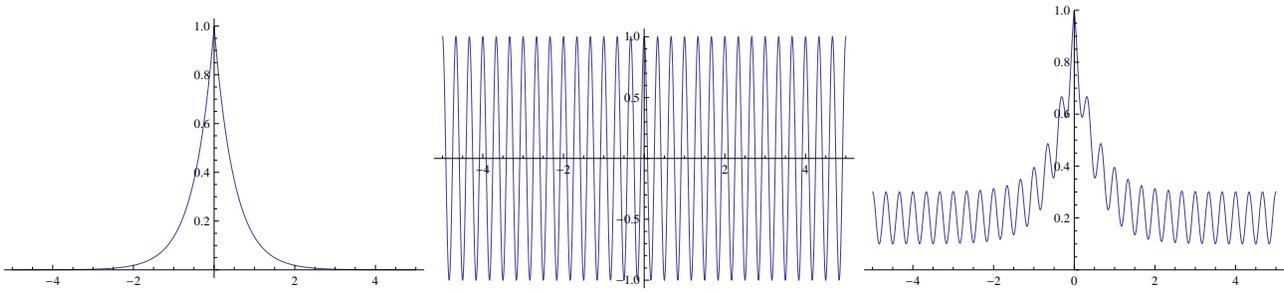
Δηλαδή, η X_n είναι WSS. ■

4.2.2 Ιδιότητες αυτοσυσχέτισης για WSS διαδικασίες

Για διαδικασίες στάσιμες με την ευρεία έννοια (άρα και για τις στάσιμες με την αυστηρή έννοια), η συνάρτηση αυτοσυσχέτισης παρουσιάζει τις παρακάτω ιδιότητες:

- Για $\tau = 0$ εκφράζει την μέση ισχύ του στοχαστικού σήματος, $R_X(0) = E[X(t)^2]$.
- Είναι συμμετρική γύρω από το μηδέν, $R_X(-\tau) = R_X(\tau)$.

²Εδώ οι χρονικές στιγμές είναι οι διακριτές στιγμές n .



Σχήμα 4.2: Συνάρτηση αυτοσυσχέτισης για (α) σήμα τηλέγραφου, (β) ημιτονοειδές σήμα με τυχαία φάση και (γ) το άθροισμα σταθεράς $m = \sqrt{0.2}$ και των (α-β), όλα κανονικοποιημένα με μοναδιαία μεταβλητότητα και μέση τιμή (και άρα $R_X(\tau) = \rho_X \leq 1$).

- Είναι μέτρο του ρυθμού μεταβολής μια διαδικασίας με την έννοια ότι ισχύει:

$$\begin{aligned} P(|X(t+\tau) - X(t)| > \epsilon) &= P((X(t+\tau) - X(t))^2 > \epsilon^2) \\ &\leq \frac{E[(X(t+\tau) - X(t))^2]}{\epsilon^2} \\ &= \frac{2(R_X(0) - R_X(\tau))}{\epsilon^2}, \end{aligned}$$

δηλαδή, αν η διαφορά $R_X(0) - R_X(\tau)$ είναι μικρή, η πιθανότητα δύο τιμές $X(t, \omega), X(t+\tau, \omega)$ να διαφέρουν περισσότερο από ϵ είναι και αυτή μικρή.

- Λόγω της ανισότητας Cauchy-Schwarz ($E[XY]^2 \leq E[X^2]E[Y^2]$) ισχύει $R_X(\tau) \leq R_X(0)$ για όλα τα τ .
- Αν ισχύει $R_X(\tau) = R_X(d)$, τότε η $R_X(\tau)$ είναι περιοδική με περίοδο d και η αντίστοιχη διαδικασία $X(t)$ είναι περιοδική με την μέση τετραγωνική έννοια, δηλαδή ισχύει

$$E[(X(t+d) - X(t))^2] = 0.$$

- Αν $X(t) = m + N(t)$, όπου $N(t)$ είναι διαδικασία με μηδενική μέση τιμή και αυτοσυσχέτιση τέτοια ώστε $\lim_{\tau \rightarrow \infty} R_N(\tau) = 0$, τότε για την αυτοσυσχέτιση του X θα ισχύει:

$$\begin{aligned} R_X(\tau) &= E[X(t)X(t+\tau)] = E[(m + N(t))(m + N(t+\tau))] \\ &= m^2 + 2mE[N(t)] + R_N(\tau) = m^2 + R_N(\tau), \end{aligned}$$

Άρα θα ισχύει $\lim_{\tau \rightarrow \infty} R_X(\tau) = m^2$ που σημαίνει ότι η αυτοσυσχέτιση του X θα τείνει στο τετράγωνο της μέσης τιμής της $X(t)$.

Συμπερασματικά, η συνάρτηση αυτοσυσχέτισης μιας στάσιμης διαδικασίας έχει τρεις συνιστώσες, (1) μια συνιστώσα που τείνει στο μηδέν για μεγάλες τιμές του τ , (2) μια περιοδική συνιστώσα και (3) μια συνιστώσα που οφείλεται στη μέση τιμή. Στο σχήμα 4.2 εικονίζονται παραδείγματα συναρτήσεων αυτοσυσχέτισης με μερικές ή όλες τις ιδιότητες.

4.2.3 Φασματική ισχύς

Αναφερόμενοι πάντα σε διαδικασίες WSS ή στάσιμες με την αυστηρή έννοια, μπορούμε να ορίσουμε τον ΜΣ Fourier της συνάρτησης αυτοσυσχέτισης

$$S_X(f) = \mathcal{F}\{R_X(\tau)\} = \int_{-\infty}^{\infty} R_X(\tau)e^{-j2\pi f\tau} d\tau,$$

που ονομάζεται **πυκνότητα φασματικής ισχύος** (power spectral density). Επίσης, αφού $R_X(\tau)$ και $S_X(f)$ είναι ζεύγος, ισχύει ότι ο αντίστροφος ΜΣ Fourier της πυκνότητας φασματικής ισχύος θα ισούτε με τη συνάρτηση αυτοσυσχέτισης

$$R_X(\tau) = \mathcal{F}^{-1}\{S_X(f)\} = \int_{-\infty}^{\infty} S_X(f)e^{j2\pi f\tau} df.$$

Στην επιστήμη του ηλεκτρολόγου μηχανικού, η δεύτερη ροπή της $X(t)$ αναφέρεται και ως μέση ισχύς του στοχαστικού σήματος $X(t)$. Πράγματι,

$$E[X^2(t)] = R_X(0) = \int_{-\infty}^{\infty} S_X(f)df,$$

γεγονός που είναι συνεπές με την απόδοση του $S_X(f)$ ως πυκνότητα φασματικής ισχύος. Για διαδικασίες που παίρνουν πραγματικές τιμές, ισχύει ότι η αυτοσυσχέτιση είναι άρτια συνάρτηση $R_X(\tau) = R_X(-\tau)$ και άρα

$$\begin{aligned} S_X(f) &= \int_{-\infty}^{\infty} R_X(\tau)e^{-j2\pi f\tau} d\tau \\ &= \int_{-\infty}^{\infty} R_X(\tau)(\cos 2\pi f\tau - j \sin 2\pi f\tau) d\tau \\ &= \int_{-\infty}^{\infty} R_X(\tau) \cos 2\pi f\tau d\tau, \end{aligned}$$

αφού η $\sin(x)$ είναι περιττή³.

Παρατηρούμε ότι, αφού $R_X(\tau) = C_X(\tau) + m_X^2$, τότε

$$\begin{aligned} S_X(f) &= \mathcal{F}\{C_X(\tau) + m_X^2\} \\ &= \mathcal{F}\{C_X(\tau)\} + m_X^2\delta(f), \end{aligned}$$

όπου $\delta(x)$ είναι η συνάρτηση δέλτα Dirac και έχουμε χρησιμοποιήσει το ζεύγος Fourier: $c\delta(f) = \mathcal{F}\{c\}$. Λόγω της θέσης της σταθεράς στο φάσμα, η συνιστώσα αυτή ονομάζεται και συνιστώσα “DC”. Γενικεύοντας την πυκνότητα φασματικής ισχύος για δύο από κοινού στάσιμες τυχαίες διαδικασίες, ορίζουμε την **ετερο-πυκνότητα φασματικής ισχύος** (cross-power spectral density):

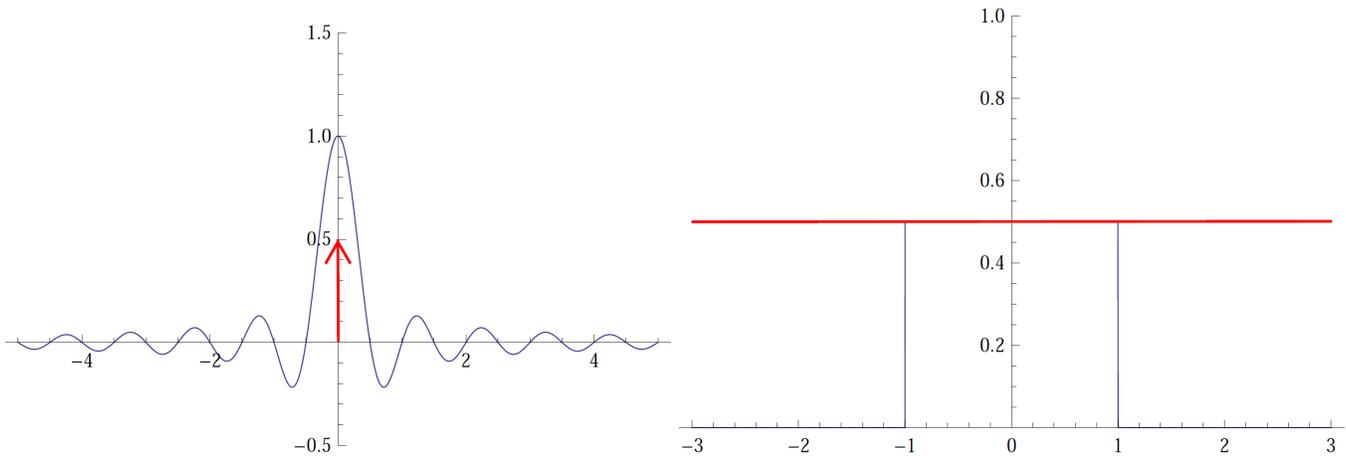
$$S_{X,Y}(f) \doteq \mathcal{F}\{R_{X,Y}(\tau)\}.$$

Γενικά, η $S_{X,Y}(f)$ είναι μιγαδική συνάρτηση του f ακόμα και όταν οι δύο διαδικασίες λαμβάνουν μόνο πραγματικές τιμές.

Παράδειγμα: (Λευκός θόρυβος) Έστω η στοχαστική διαδικασία θορύβου $X(t)$ με την παρακάτω συνάρτηση πυκνότητας φασματικής ισχύος:

$$S_X(f) = \begin{cases} \frac{N_0}{2} & f \in [-W, W] \\ 0 & \text{αλλού.} \end{cases}$$

³Το γινόμενο άρτιας και περιττής συνάρτησης είναι περιττή συνάρτηση και το ολοκλήρωμα περιττής συνάρτησης σε όλο το \mathbb{R} είναι ίσο με μηδέν.



Σχήμα 4.3: Συνάρτηση αυτοσυσχέτισης (αριστερά) και πυκνότητα φασματικής ισχύος (δεξιά) του λευκού θορύβου στάθμης $N_0 = 1$, με μπλε για $W = 1$ και με κόκκινο για $W \rightarrow \infty$.

Η διαδικασία $X(t)$ ονομάζεται και λευκός θόρυβος καθώς το φασματικό περιεχόμενο είναι σταθερό σε όλο το εύρος συχνοτήτων (όπως το λευκό χρώμα). Υπολογίστε την αυτοσυσχέτιση και τη μέση ισχύ. Επίσης, ερευνήστε την περίπτωση που το εύρος ζώνης απλώνεται στο άπειρο.

Λύση: Η μέση ισχύς υπολογίζεται εύκολα:

$$E[X^2(t)] = \int_{-W}^W \frac{N_0}{2} df = N_0 W.$$

Η συνάρτηση αυτοσυσχέτισης θα είναι

$$\begin{aligned} R_X(\tau) &= \frac{N_0}{2} \int_{-W}^W e^{j2\pi f\tau} df \\ &= \frac{N_0}{2} \frac{e^{-j2\pi W\tau} - e^{j2\pi W\tau}}{-j2\pi\tau} \\ &= N_0 \frac{\sin(2\pi W\tau)}{2\pi\tau} \\ &= N_0 W \operatorname{sinc}(2W\tau), \end{aligned}$$

όπου $\operatorname{sinc}(t) \doteq \frac{\sin(\pi t)}{\pi t}$, και το αποτέλεσμα είναι αναμενόμενο αφού πρόκειται για ένα επίσης γνωστό ζεύγος $\operatorname{rect}(f) = \mathcal{F}\{\operatorname{sinc}(t)\}$, όπου $\operatorname{rect}(x)$ η συνάρτηση παλμού. Μπορούμε, επίσης, να επαληθεύσουμε το $E[X^2(t)] = R_X(0)$.

Όταν το εύρος ζώνης απλώνεται στο άπειρο, δηλαδή $W \rightarrow \infty$, η μέση ισχύς είναι επίσης άπειρη. Για την συνάρτηση αυτοσυσχέτισης έχουμε

$$\lim_{W \rightarrow \infty} R_X(\tau) = \frac{N_0}{2} \delta(\tau).$$

Η συνάρτηση αυτοσυσχέτισης και η πυκνότητα φασματικής ισχύος για τις δύο περιπτώσεις φαίνονται στο σχήμα 4.3. ■

Παράδειγμα: (Ημιτονοειδές σήμα με τυχαία φάση) Έστω η στοχαστική διαδικασία $X(t) = a \cos(2\pi f_0 t + \Theta)$, όπου Θ είναι τ.μ. ομοιόμορφα κατανομημένη στο $[0, 2\pi]$. Βρείτε την πυκνότητα φασματικής ισχύος.

Λύση: θα βρούμε πρώτα τη συνάρτηση αυτοσυσχέτισης

$$\begin{aligned} R_X(t_1, t_2) &= E[a \cos(2\pi f_0 t_1 + \Theta) a \cos(2\pi f_0 t_2 + \Theta)] \\ &= \frac{a^2}{2\pi} \int_0^{2\pi} \frac{1}{2} [\cos(2\pi f_0(t_1 - t_2)) + \cos(2\pi f_0(t_1 + t_2) + 2\theta)] d\theta \\ &= \frac{a^2}{4\pi} \cos(2\pi f_0(t_1 - t_2)) \int_0^{2\pi} d\theta + \frac{a^2}{4\pi} \int_0^{2\pi} \cos(2\pi f_0(t_1 + t_2) + 2\theta) d\theta \\ &= \frac{a^2}{4\pi} \cos(2\pi f_0(t_1 - t_2)) \int_0^{2\pi} d\theta \\ &= \frac{a^2}{2} \cos(2\pi f_0(t_1 - t_2)). \end{aligned}$$

Δηλαδή, η διαδικασία είναι στάσιμη και έχουμε $R_X(\tau) = \frac{a^2}{2} \cos(2\pi f_0 \tau)$. Άρα η πυκνότητα φασματικής ισχύος θα είναι

$$\begin{aligned} S_X(f) &= \frac{a^2}{2} \mathcal{F}\{\cos(2\pi f_0 \tau)\} \\ &= \frac{a^2}{4} [\delta(f - f_0) + \delta(f + f_0)]. \end{aligned}$$

■

Φασματική ισχύς στις διαδικασίες διακριτού χρόνου

Στις διαδικασίες διακριτού χρόνου, η αυτοσυσχέτιση $R_X(n)$ είναι μια διακριτή ακολουθία. Κατά συνέπεια, η πυκνότητα φασματικής ισχύος ορίζεται ως η σειρά Fourier:

$$S_X(f) \doteq \mathcal{F}\{R_X(n)\} = \sum_{n=-\infty}^{\infty} R_X(n) e^{-j2\pi f n}.$$

Εφόσον η $S_X(f)$ είναι περιοδική με περίοδο 1, αρκεί να την μελετήσουμε στο διάστημα $f \in [-\frac{1}{2}, \frac{1}{2}]$. Επίσης, αν γνωρίζουμε την πυκνότητα φασματικής ισχύος μπορούμε να υπολογίσουμε την συνάρτηση αυτοσυσχέτισης (που είναι ακολουθία)

$$R_X(n) \doteq \int_{-1/2}^{1/2} S_X(f) e^{j2\pi f n} df, n = \dots, 0, 1, \dots$$

που ονομάζονται και συντελεστές Fourier της $S_X(f)$.

Παράδειγμα: Διαδικασία κινητού μέσου όρου (moving average process) Έστω η διαδικασία Y_n που ορίζεται ως

$$Y_n = X_n + aX_{n-1},$$

⁴Στην λύση χρησιμοποιήσαμε την τριγωνομετρική ταυτότητα $\cos a \cos b \doteq \frac{1}{2} [\cos(a + b) + \cos(a - b)]$.

όπου η X_n είναι μια διαδικασία λευκού θορύβου διακριτού χρόνου. Η διαδικασία αυτή ονομάζεται διαδικασία κινητού μέσου όρου (moving average process). Βρείτε τη μέση τιμή της, την αυτοσυσχέτιση και την πυκνότητα φασματικής ισχύος.

Λύση: Η διαδικασία λευκού θορύβου διακριτού χρόνου είναι μια ακολουθία ασυσχέτιστων τ.μ. με μέση τιμή μηδέν και μεταβλητότητα σ_X^2 . Για την συνάρτηση μέσης τιμής της Y_n έχουμε

$$E[Y_n] = E[X_n + aX_{n-1}] = E[X_n] + aE[X_{n-1}] = 0.$$

Επίσης ισχύει

$$E[Y_n Y_{n+k}] = E[X_n X_{n+k}] + aE[X_n X_{n+k-1}] + aE[X_{n-1} X_{n+k}] + a^2 E[X_{n-1} X_{n+k-1}] = \begin{cases} (1+a^2)\sigma_X^2 & k=0 \\ a\sigma_X^2 & k=\pm 1 \\ 0 & \text{αλλιώς,} \end{cases}$$

που είναι και η συνάρτηση αυτοσυσχέτισης $R_Y(k)$ αφού η μέση τιμή είναι πάντα μηδέν. Τέλος η πυκνότητα φασματικής ισχύος θα είναι

$$\begin{aligned} S_Y(f) &= \sum_{k=-\infty}^{\infty} R_Y(k) e^{-j2\pi f k} \\ &= \sum_{k=-1}^1 R_Y(k) e^{-j2\pi f k} \\ &= (1+a^2)\sigma_X^2 + a\sigma_X^2 (e^{j2\pi f} + e^{-j2\pi f}) \\ &= \sigma_X^2 (1+a^2 + 2a \cos(2\pi f)). \end{aligned}$$

■

4.2.4 Κυκλική στασιμότητα

Πολλές στοχαστικές διαδικασίες δημιουργούνται από την κυκλική επανάληψη διαδικασιών ανά χρονικές περιόδους T . Ένα παράδειγμα είναι το σήμα που παράγει το modem. Μια διαδικασία $X(t)$ λέγεται **κυκλικά στάσιμη** (cyclostationary) αν η από κοινού κατανομή των δειγμάτων είναι ίδια αν προσθέσουμε T σε κάθε δείγμα, δηλαδή αν για κάθε k, m και κάθε t_1, \dots, t_k ισχύει:

$$F_{X_1(t_1), \dots, X_k(t_k)}(x_1, \dots, x_k) = F_{X_1(t_1+mT), \dots, X_k(t_k+mT)}(x_1, \dots, x_k),$$

όπου $m \in \mathbb{N}$. Επίσης, όμοια με πριν, ορίζουμε την κυκλικά στάσιμη διαδικασία με την ευρεία έννοια, ως την διαδικασία που οι συναρτήσεις μέσης τιμής και αυτοσυμμεταβλητότητας έχουν τις ιδιότητες:

$$m_X(t+mT) = m_X(t) \quad \text{και} \quad C_X(t_1+mT, t_2+mT) = C_X(t_1, t_2)$$

Ερώτηση: Απόδειξε ότι αν η $X(t)$ είναι κυκλικά στάσιμη, τότε είναι και κυκλικά στάσιμη με την ευρεία έννοια.

Παράδειγμα: Ένα modem μεταδίδει μια ακολουθία δεδομένων που είναι δυαδική i.i.d. και ισοπίθανη: για να μεταδώσει το 1, το modem μεταδίδει έναν τετραγωνικό παλμό $p(t)$ με διάρκεια T και πλάτος 1, ενώ για να

μεταδώσει το 0, μεταδίδει έναν παλμό ίσης διάρκειας αλλά πλάτους -1. Είναι η προκύπτουσα διαδικασία $X(t)$ κυκλικά στάσιμη;

Λύση: Για την $X(t)$ μπορούμε να γράψουμε:

$$X(t) = \sum_{n=-\infty}^{\infty} A_n p(t - nT),$$

όπου A_n είναι μια ακολουθία από i.i.d. δυαδικές μεταβλητές που παίρνουν τις τιμές ± 1 ισοπίθανα και $p(t)$ είναι η συνάρτηση παλμού, $p(t) = 1, 0 \leq t < T$ και μηδέν αλλού. Η μέση τιμή είναι

$$m_X(t) = E \left[\sum_{n=-\infty}^{\infty} A_n p(t - nT) \right] = \sum_{n=-\infty}^{\infty} E[A_n] p(t - nT) = 0.$$

Η συνάρτηση αυτοσυμμεταβλητότητας είναι

$$C_X(t_1, t_2) = E[X(t_1)X(t_2)] - 0 = \begin{cases} E[X(t_1)^2] = 1 & nT \leq t_1, t_2 < (n+1)T \\ E[X(t_1)X(t_2)] = 0 & \text{αλλού} \end{cases}$$

το οποίο προκύπτει επειδή το διπλό άθροισμα μετατρέπεται σε μονό ώστε $n = k$ για να είναι οι A_n, A_k συσχετισμένες και στη συνέχεια να πρέπει να βρούμε τα t_1, t_2 έτσι ώστε να ισχύει $0 < t_1 - nT < T$ και $0 < t_2 - nT < T$ ταυτόχρονα. Άρα η $X(t)$ είναι κυκλικά στάσιμη με την ευρεία έννοια. Παρατηρήστε, όμως, ότι $C_X(0, \frac{T}{2}) \neq C_X(\frac{3T}{4}, \frac{5T}{4})$, άρα η $X(t)$ δεν είναι WSS, άρα ούτε και στάσιμη με την αυστηρή έννοια. ■

Οι κυκλικά στάσιμες διαδικασίες είναι χρήσιμες γιατί μπορούν να μετατραπούν σε στάσιμες. Έστω μια κυκλικά στάσιμη διαδικασία $X(t)$, τότε θα δείξουμε ότι η $X_s(t)$ είναι στάσιμη διαδικασία, όπου $X_s(t) = X(t + \Theta)$, και Θ είναι τ.μ. ομοιόμορφα κατανομημένη στο $[0, T]$ ανεξάρτητη από τη διαδικασία. Για την από κοινού κατανομή της $X_s(t)$ σε ένα αυθαίρετο σύνολο δειγμάτων έχουμε:

$$\begin{aligned} F_{X_s(t_1), \dots, X_s(t_k)}(x_1, \dots, x_k) &= P(X_s(t_1) \leq x_1, \dots, X_s(t_k) \leq x_k) \\ &= P(X(t_1 + \Theta) \leq x_1, \dots, X(t_k + \Theta) \leq x_k) \\ &= \int_0^T P(X(t_1 + \theta) \leq x_1, \dots, X(t_k + \theta) \leq x_k | \Theta = \theta) f_{\Theta}(\theta) d\theta \\ &= \frac{1}{T} \int_0^T P(X(t_1 + \theta) \leq x_1, \dots, X(t_k + \theta) \leq x_k | \Theta = \theta) d\theta \\ &= \frac{1}{T} \int_0^T P(X(t_1 + \theta) \leq x_1, \dots, X(t_k + \theta) \leq x_k) d\theta, \end{aligned}$$

όπου χρησιμοποιήσαμε την υπό συνθήκη ΣΠΠ των δειγμάτων ως προς την Θ και την ανεξαρτησία της γωνίας από τη διαδικασία. Επίσης,

$$\begin{aligned} F_{X_s(t_1+\tau), \dots, X_s(t_k+\tau)}(x_1, \dots, x_k) &= \frac{1}{T} \int_0^T P(X(t_1 + \tau + \theta) \leq x_1, \dots, X(t_k + \tau + \theta) \leq x_k) d\theta \\ &= \frac{1}{T} \int_{\tau}^{\tau+T} P(X(t_1 + \phi) \leq x_1, \dots, X(t_k + \phi) \leq x_k) d\phi \\ &= \frac{1}{T} \int_0^T P(X(t_1 + \theta) \leq x_1, \dots, X(t_k + \theta) \leq x_k) d\theta, \end{aligned}$$

όπου χρησιμοποιήσαμε ότι η $X(t)$ είναι κυκλικά στάσιμη, άρα η κατανομή της είναι ίδια σε ακέραια πολλαπλάσια του T , άρα περιοδική. Το ολοκλήρωμα περιοδικής συνάρτησης στο διάστημα $[0, T]$ θα είναι ίδιο με το ολοκλήρωμα στο διάστημα $[\tau, \tau + T]$. Χρησιμοποιώντας την υπό συνθήκη μέση τιμή (conditional expectation) μπορούμε να δείξουμε, ομοίως, ότι αν η $X(t)$ είναι κυκλικά στάσιμη με την ευρεία έννοια, τότε η $X_s(t)$ θα είναι WSS. Τότε αποδεικνύεται ότι ισχύει

$$m_{X_s} = \frac{1}{T} \int_0^T m_X(t) dt$$

και

$$R_{X_s}(\tau) = \frac{1}{T} \int_0^T R_X(t, t + \tau) dt.$$

4.3 Εργοδικότητα

Όπως είδαμε σε προηγούμενα κεφάλαια, για να εκτιμήσουμε τη μέση τιμή μιας μεταβλητής, πρέπει να κάνουμε πολλά επαναλαμβανόμενα πειράματα. Αντίστοιχα, για να εκτιμήσουμε ροπές διαδικασιών, πρέπει να κάνουμε πολλές επαναλήψεις δειγμάτων της διαδικασίας. Μερικές στάσιμες διαδικασίες, όμως, έχουν την εξής ενδιαφέρουσα και χρήσιμη ιδιότητα: η πραγματοποίηση ενός δείγματος της διαδικασίας (π.χ. το $X(t, \omega_0)$) είναι αρκετή για να περιγράψει ολόκληρη τη διαδικασία μέσω των στατιστικών της, δηλαδή οι μέσοι όροι της προκύπτουσας συνάρτησης $X(t, \omega_0)$ είναι ίσοι με τους στατιστικούς μέσους όρους. Θεωρούμε τους παρακάτω εκτιμητές μέσης τιμής και αυτοσυσχέτισης που είναι χρονικοί μέσοι όροι:

$$\hat{m}_X(T, \omega_0) = \frac{1}{2T} \int_{-T}^T X(t, \omega_0) dt$$

και

$$\hat{R}_X(\tau, T, \omega_0) = \frac{1}{2T} \int_{-T}^T X(t, \omega) X(t + \tau, \omega) dt.$$

Μια στάσιμη διαδικασία $X(t)$ λέγεται **εργοδική ως προς τη μέση τιμή** αν ισχύει

$$\hat{m}_X(T, \omega) \rightarrow m_X \quad \text{a.s.} \quad \text{όταν } T \rightarrow \infty.$$

Μια στάσιμη διαδικασία $X(t)$ λέγεται **εργοδική ως προς την αυτοσυσχέτιση** αν για κάθε $\tau \in \mathbb{R}$ ισχύει

$$\hat{R}_X(\tau, T, \omega) \rightarrow R_X(\tau) \quad \text{a.s.} \quad \text{όταν } T \rightarrow \infty.$$

Ο ισχυρός νόμος των μεγάλων αριθμών είναι ένα εργοδικό θεώρημα που λέει ότι ο μέσος όρος της αθροιστικής διαδικασίας συγκλίνει στη μέση τιμή με πιθανότητα 1 (ή ισοδύναμα almost surely).

Παράδειγμα: Έστω η στοχαστική διαδικασία $X(t) = A, \forall t$ όπου A είναι τ.μ. με μέση τιμή 0 και μεταβλητότητα 1. Συγκλίνει ο μέσος όρος; Είναι εργοδική;

Λύση: Αρχικά μπορούμε εύκολα να δούμε ότι $m_X(t) = E[X(t)] = E[A] = 0$ και να παρατηρήσουμε ότι η $X(t)$ είναι στάσιμη (έχει παντού την ίδια κατανομή και οι από κοινού κατανομές αναφέρονται στην ίδια τ.μ.).

Για τον αριθμητικό μέσο όρο έχουμε

$$\hat{m}_X(T, \omega_0) = \frac{1}{2T} \int_{-T}^T A(\omega_0) dt = A(\omega_0).$$

Άρα, ο $\hat{m}_X(T, \omega)$ συγκλίνει στην A βέβαια (sure convergence). Όμως, αυτό σημαίνει ότι ο $\hat{m}_X(T, \omega)$ δεν συγκλίνει στο $m_X(t)$ για κάθε ω , άρα η $X(t)$ δεν είναι εργοδική διαδικασία ως προς τη μέση τιμή. ■

4.4 Διαδικασίες Markov

Έστω οι χρονικές στιγμές t_1, t_2, \dots, t_n με $t_1 < t_2 < \dots < t_n$. Η ιδιότητα Markov περιγράφεται από την παρακάτω σχέση

$$P(X(t_n) \leq x_n | X(t_1) \leq x_1, X(t_2) \leq x_2, \dots, X(t_{n-1}) \leq x_{n-1}) = P(X(t_n) \leq x_n | X(t_{n-1}) \leq x_{n-1}). \quad (4.1)$$

Όταν μια διαδικασία ικανοποιεί αυτή τη σχέση για κάθε επιλογή των t_1, t_2, \dots, t_n την ονομάζουμε **διαδικασία Markov**. Η ιδιότητα λέει ότι η γνώση μιας παρελθοντικής κατάστασης είναι αρκετή με την έννοια ότι επιπλέον πληροφορία δεν μπορεί να δωθεί μέσω παλαιότερων στιγμών. Με άλλα λόγια, η πληροφορία του παρελθόντος του συστήματος εμπεριέχεται σε κάθε χρονική στιγμή. Επίσης, μπορούμε να σκεφτούμε την υπό συνθήκη ανεξαρτησία: η $X(t_n)$ είναι υπό συνθήκη του $X(t_{n-1})$ ανεξάρτητη από όλα τα παλαιότερα δείγματα της διαδικασίας.

Όταν οι τ.μ. $X(t_0, \omega)$ λαμβάνουν διακριτές τιμές, τότε αναφερόμαστε στις διαδικασίες Markov ως **αλυσίδες Markov**.

Η ιδιότητα Markov μπορεί να αντιστραφεί και να δειχθεί ότι αν οι τ.μ. $X_1 \rightarrow X_2 \rightarrow X_3$ σχηματίζουν αλυσίδα Markov, τότε το ίδιο ισχύει και αντίστροφα $X_3 \rightarrow X_2 \rightarrow X_1$. Επίσης, για τις αλυσίδες Markov ισχύει η εξής σημαντική ιδιότητα που είναι συνέπεια του ορισμού:

$$\begin{aligned} p_{X_1, \dots, X_n}(x_1, \dots, x_n) &= p_{X_1}(x_1) p_{X_2|X_1}(x_2|x_1) \dots p_{X_n|X_1, \dots, X_{n-1}}(x_n|x_1, \dots, x_{n-1}) \\ &\stackrel{\text{Markov}}{=} p_{X_1}(x_1) p_{X_2|X_1}(x_2|x_1) \dots p_{X_n|X_{n-1}}(x_n|x_{n-1}). \end{aligned}$$

Δηλαδή, η από κοινού ΣΜΠ για κάθε στιγμή εξαρτάται από την αρχική κατάσταση $p_{X_1}(x_1)$ και τις πιθανότητες μετάβασης από κατάσταση σε κατάσταση. Οι αλυσίδες Markov είναι σημαντικό εργαλείο μοντελοποίησης και θα αναφερθούμε σε αυτές εκτενώς σε επόμενο κεφάλαιο.

4.5 Διαδικασίες διακριτού χρόνου

Στις διαδικασίες αυτές, η παράμετρος της διαδικασίας t λαμβάνει τιμές από ένα διακριτό σύνολο, συνήθως στο \mathbb{N}^* , εκτός αν αναφέρεται αλλιώς. Επίσης, για λόγους παρουσίασης, θα χρησιμοποιήσουμε το συμβολισμό n αντί για t και θα αναφερόμαστε στη διαδικασία $\{X_n\}_{n=1,2,\dots}$.

4.5.1 i.i.d. διαδικασία

Μπορούμε εύκολα να παράγουμε μια στοχαστική διαδικασία λαμβάνοντας τ.μ. ανεξάρτητες και ομοίως κατανοημένες (i.i.d.) X_i , δημιουργώντας την ακολουθία $\{X_i\}_{i=1,2,\dots}$ που ονομάζεται **διαδικασία i.i.d.**. Η διαδικασία αυτή είναι η πιο εύκολη προς μελέτη, αφού μαθηματικώς είναι ο εκφυλισμός της διαδικασίας σε υποσύνολα ανεξάρτητων τυχαίων μεταβλητών. Η από κοινού κατανομή για οποιοδήποτε υποσύνολο δειγμάτων n_1, \dots, n_k θα είναι

$$F_{X_{n_1}, \dots, X_{n_k}}(x_1, x_2, \dots, x_k) = F_X(x_1) F_X(x_2) \dots F_X(x_k)$$

και για τις ροπές θα έχουμε

$$\begin{aligned} m_X(n) &= m_X \\ \text{Var}(X_n) &= \sigma_X^2 \\ C_X(n_1, n_2) &= \begin{cases} 0 & n_1 \neq n_2 \\ \sigma_X^2 & n_1 = n_2 \end{cases} \\ R_X(n_1, n_2) = C_X(n_1, n_2) + m_X(n_1)m_X(n_2) &= \begin{cases} m_X^2 & n_1 \neq n_2 \\ E[X^2] & n_1 = n_2 \end{cases}, \end{aligned}$$

όπου $E[X^2] = \sigma_X^2 + m_X^2$. Είναι σύνηθες σε τέτοιες περιπτώσεις να γράφουμε την αυτοσυσχέτιση ως $R_X(n_1, n_2) = \sigma_X^2 \delta_{n_1 n_2} + m_X^2$, όπου $\delta_{n_1 n_2}$ είναι η δέλτα συνάρτηση του Kronecker (δεν πρέπει να συγχέεται με την δέλτα Dirac) που ορίζεται ως

$$\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j. \end{cases}$$

Παρατηρήστε ότι η διαδικασία i.i.d. είναι στάσιμη. Λόγω της ανεξαρτησίας, θα έχουμε ότι η από κοινού κατανομή σε k δείγματα θα είναι ίση με $F_X(x_1)F_X(x_2) \dots F_X(x_k)$ ανεξάρτητα από την επιλογή των n_1, \dots, n_k .

Ερώτηση: Αποδείξτε ότι η διαδικασία i.i.d. είναι εργοδική ως προς το μέσο όρο.

Παράδειγμα: (Διαδικασία τυχαίου βήματος) Ένας μετρητής καθοδηγείται από ανεξάρτητους παλμούς $+1$ και -1 που είναι αποτελέσματα της τ.μ. $2X - 1$ όπου $X \sim \text{Bernoulli}(p)$. Δηλαδή

$$D_n = \begin{cases} +1 & X_n = 1 \\ -1 & X_n = 0. \end{cases}$$

όπου X_n είναι μια i.i.d. διαδικασία Bernoulli και η D_n είναι η διαδικασία τυχαίου βήματος. Βρείτε τη συνάρτηση μέσης τιμής, τη συνάρτηση μεταβλητότητας και την πιθανότητα $P(D_1 = +1, D_2 = -1, D_3 = +1, D_4 = -1)$.

Λύση: Για τη συνάρτηση μέσης τιμής έχουμε

$$m_D(n) = E[D] = E[2X - 1] = 2E[X] - 1 = 2p - 1,$$

για τη συνάρτηση μεταβλητότητας

$$\begin{aligned} \sigma_D^2(n) &= E[(D_n - m_D(n))^2] = E[(D - m_D)^2] = E[(2X - 1 - (2p - 1))^2] \\ &= 4E[(X - p)^2] = 4p(1 - p). \end{aligned}$$

Τέλος για την ζητούμενη πιθανότητα έχουμε

$$P(D_1 = +1, D_2 = -1, D_3 = +1, D_4 = -1) = p^2(1 - p)^2.$$

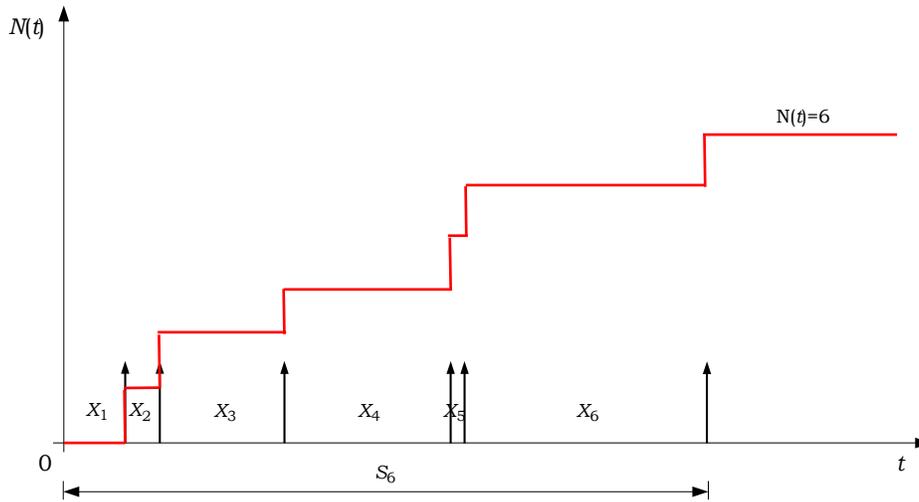
■

4.5.2 Διαδικασίες απαρίθμησης

Μια κατηγορία στοχαστικών διαδικασιών διακριτών τιμών (αποτελούνται από διακριτές τ.μ.) που ονομάζεται **διαδικασίες απαρίθμησης** (counting processes) ασχολείται με την απαρίθμηση συμβάντων από την αρχή του χρόνου. Τις διαδικασίες αυτές τις συμβολίζουμε με $N(t), t \geq 0$. Οι διαδικασίες απαρίθμησης είναι άμεσα συνδεδεμένες με διαδικασίες αθροισμάτων S_n καθώς ισχύει

$$N(t) = \sup\{n : S_n \leq t\}.$$

όπου η διαδικασία αθροισματωσ ορίζεται ως $S_n = \sum_{i=1}^n X_i$ και X_i είναι τ.μ. Η παραπάνω αντιστοίχιση μιας διαδικασίας απαρίθμησης σε μια διαδικασία αθροίσματος είναι φανερή στο σχήμα 4.4. Προσέξτε τη δυαδικότητα μεταξύ των δύο διαδικασιών: οι διαδικασίες απαρίθμησης λαμβάνουν διακριτές τιμές αλλά είναι συνεχούς χρόνου



Σχήμα 4.4: Η στοχαστική διαδικασία απαρίθμησης και η αντιστοίχισή της σε διαδικασία αθροίσματος.

διαδικασίες, ενώ οι διαδικασίες αθροίσματος λαμβάνουν συνεχείς τιμές αλλά είναι διακριτού χρόνου διαδικασίες. Και οι δύο διαδικασίες μοντελοποιούν την άφιξη διακριτών συμβάντων. Επίσης, ισχύει

$$P(N(t) \leq n) = P(S_n \leq t).$$

Παρατηρήστε ότι οι διαδικασίες απαρίθμησης και αθροίσματος είναι άυξουσες συναρτήσεις της χρονικής παραμέτρου και άρα δεν είναι στάσιμες.

Ανεξάρτητες αυξήσεις independent increments

Μια διαδικασία απαρίθμησης $N(t)$ έχει ανεξάρτητες αυξήσεις όταν οι διακριτές τυχαίες μεταβλητές

$$N(t_2) - N(t_1), N(t_3) - N(t_2), \dots, N(t_k) - N(t_{k-1})$$

είναι ανεξάρτητες για κάθε επιλογή t_1, t_2, \dots, t_k . Με άλλα λόγια, απαιτούμε οι αυξήσεις που αφορούν μη επικαλυπτόμενα διαστήματα να είναι ανεξάρτητες. Έστω οι διακριτές τ.μ. $X_i = N(t_{i+1}) - N(t_i)$. Ισοδύναμα, απαιτούμε η ακολουθία X_n να αποτελείται από ανεξάρτητες τ.μ.

Παράδειγμα: (Οι ανεξάρτητες αυξήσεις συνεπάγονται την ιδιότητα Markov) Έστω η διαδικασία $N(t)$ που έχει ανεξάρτητες αυξήσεις. Να δείξετε ότι έχει την ιδιότητα Markov.

Λύση: Παρατηρήστε ότι

$$\begin{aligned} P(N(t_k) = x_k | N(t_{k-1}) = x_{k-1}, \dots, N(t_1) = x_1) &= P(N(t_{k-1}) + X_k = x_k | N(t_{k-1}) = x_{k-1}, \dots, N(t_1) = x_1) \\ &= P(X_k = x_k - x_{k-1} | N(t_{k-1}) = x_{k-1}, \dots, N(t_1) = x_1) \\ &= P(X_k = x_k - x_{k-1} | N(t_{k-1}) = x_{k-1}) \\ &= P(N(t_{k-1}) + X_k = x_k | N(t_{k-1}) = x_{k-1}) \\ &= P(N(t_k) = x_k | N(t_{k-1}) = x_{k-1}), \end{aligned}$$

αφού το X_k δεν εξαρτάται από τα υπόλοιπα διαστήματα. ■

Διαδικασίες αθροίσματος ανεξάρτητων μεταβλητών

Όταν τα X_i είναι i.i.d. τ.μ., τότε η στοχαστική διαδικασία

$$S_n = \sum_{i=1}^n X_i$$

ονομάζεται διαδικασία αθροίσματος ανεξάρτητων μεταβλητών. Η από κοινού ΣΠΠ/ΣΜΠ της S_n μπορεί να υπολογιστεί χρησιμοποιώντας την συνέλιξη ΣΠΠ/ΣΜΠ ή τον πολλαπλασιασμό χαρακτηριστικών συναρτήσεων (όπως έχει περιγραφεί σε προηγούμενο κεφάλαιο). Η S_n είναι διαδικασία Markov καθώς ισχύει $S_n = S_{n-1} + X_n$. Επίσης, η διαδικασία αυτή έχει ανεξάρτητες αυξήσεις καθώς για δύο μη επικαλυπτόμενα διαστήματα $[n_0, n_1]$ και $[n_2, n_3]$ ισχύει

$$\begin{aligned} S_{n_1} - S_{n_0} &= X_{n_0+1} + \dots + X_{n_1} \\ S_{n_3} - S_{n_2} &= X_{n_2+1} + \dots + X_{n_3}, \end{aligned}$$

με τα δεξιά μέλη να είναι προφανώς ανεξάρτητα μεταξύ τους. Επίσης, για $m > n$ ισχύει

$$P(S_m - S_n = y) = P(S_{m-n} = y),$$

δηλαδή το άθροισμα των τ.μ. στο διάστημα $[n, m]$ έχει την ίδια ΣΜΠ με το άθροισμα των τ.μ. στο διάστημα $[0, m - n]$, οπότε λέμε ότι η S_n έχει **στάσιμες αυξήσεις** (stationary increments). Στην περίπτωση των ανεξάρτητων και στάσιμων αυξήσεων έχουμε:

$$\begin{aligned} P(S_{n_1} = y_1, S_{n_2} = y_2, S_{n_3} = y_3) &= \\ &= P(S_{n_1} = y_1, S_{n_2} - S_{n_1} = y_2 - y_1, S_{n_3} - S_{n_2} = y_3 - y_2) \\ &= P(S_{n_1} = y_1) P(S_{n_2} - S_{n_1} = y_2 - y_1) P(S_{n_3} - S_{n_2} = y_3 - y_2) \quad \text{ανεξάρτητες αυξήσεις} \\ &= P(S_{n_1} = y_1) P(S_{n_2-n_1} = y_2 - y_1) P(S_{n_3-n_2} = y_3 - y_2) \quad \text{στάσιμες αυξήσεις} \end{aligned}$$

Για συνεχείς τ.μ. X_i , η αντίστοιχη σχέση θα είναι:

$$f_{S_{n_1}, S_{n_2}, \dots, S_{n_k}}(y_1, y_2, \dots, y_k) = f_{S_{n_1}}(y_1) f_{S_{n_2-n_1}}(y_2 - y_1) \dots f_{S_{n_k-n_{k-1}}}(y_k - y_{k-1}).$$

Παράδειγμα: (1D Τυχαίος περίπατος) Έστω D_n η στοχαστική διαδικασία διακριτού χρόνου του προηγούμενου παραδείγματος (διαδικασία τυχαίου βήματος) και έστω S_n η διαδικασία αθροίσματος της D_n , δηλαδή

$$S_n = \sum_{i=1}^n D_i.$$

Η διαδικασία S_n ονομάζεται τυχαίος περίπατος (στο σχήμα 4.5 μπορείτε να δείτε δείγματα της διαδικασίας αυτής για $p = 0.5$). Βρείτε την ΣΜΠ του S_n , την συνάρτηση μέσης τιμής και την συνάρτηση αυτοσυμμεταβλητότητας.

Λύση: Για να βρισκόμαστε στη θέση j μετά από n βήματα θα πρέπει να έχουν έρθει $k + 1$ και $n - k - 1$, όπου

$2k - n = j$ δηλαδή $k = \frac{n+j}{2}$. Άρα, η ΣΜΠ θα μοιάζει με την κατανομή της Binomial

$$P(S_n = j) = \binom{n}{\frac{n+j}{2}} p^{\frac{n+j}{2}} (1-p)^{\frac{n-j}{2}}, \quad j \in \{-n, -n+1, \dots, 0, 1, \dots, n\}.$$

Παρατηρήστε ότι η $X = \frac{S_n+n}{2}$ θα είναι κατανομημένη κατά Binomial(n, p) και κατά συνέπεια η συνάρτηση μέσης τιμής θα είναι

$$E[S_n] = E[2X - n] = 2np - n = n(2p - 1) = nm_D.$$

Για να υπολογίσουμε την συνάρτηση αυτοσυμμεταβλητότητας θα χρησιμοποιήσουμε τις ιδιότητες των ανεξάρτητων στάσιμων αυξήσεων. Συγκεκριμένα, υποθέτουμε αρχικά $n < k$, θα έχουμε

$$\begin{aligned} C_S(n, k) &= E[(S_n - m_S(n))(S_k - m_S(k))] \\ &= E[(S_n - m_S(n))\{(S_k - m_S(k)) + (S_n - m_S(n)) - (S_n - m_S(n))\}] \\ &= E[(S_n - m_S(n))^2] + E[(S_n - m_S(n))(S_k - S_n + m_S(n) - m_S(k))] \\ &= \overset{\text{ανεξ. αυξ}}{E[(S_n - m_S(n))^2]} + E[(S_n - m_S(n))]E[(S_k - S_n + m_S(n) - m_S(k))] \\ &= E[(S_n - m_S(n))^2] + 0E[(S_k - S_n + m_S(n) - m_S(k))] \\ &= E[(S_n - m_S(n))^2] = \text{Var}(S_n) = n4p(1-p) = n\sigma_D^2. \end{aligned}$$

Επαναλαμβάνοντας το παραπάνω για $k < n$ βρίσκουμε ότι $C_S(n, k) = \text{Var}(S_k) = k\sigma_D^2$, που σημαίνει ότι γενικά για ανεξάρτητες και στάσιμες αυξήσεις ισχύει

$$C_S(n, k) = \min\{n, k\}\sigma_D^2,$$

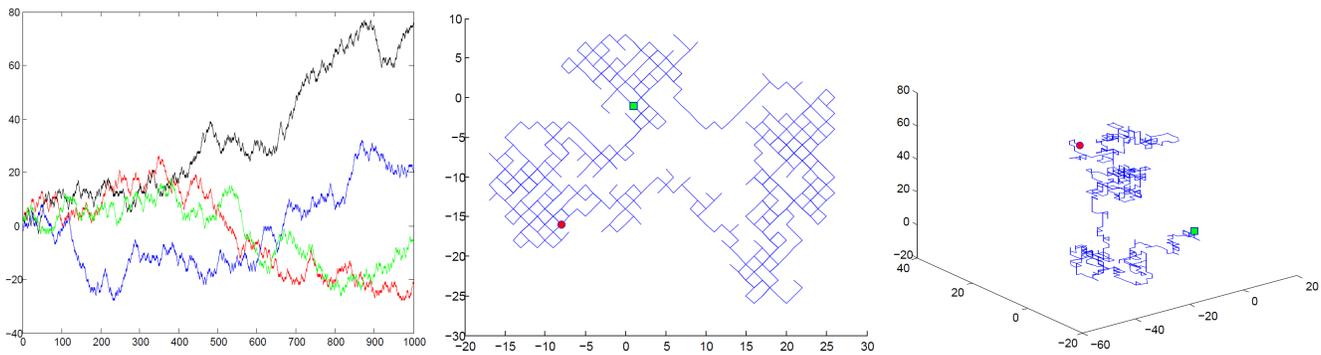
και ειδικά για τον τυχαίο περίπατο

$$C_S(n, k) = 4p(1-p) \min\{n, k\}.$$

Όταν $p > 0.5$, ο τυχαίος περίπατος 1D έχει την τάση να αυξάνει, ενώ για $p < 0.5$ έχει την τάση να μειώνεται. Όταν $p = 0.5$, παρουσιάζει ενδιαφέρουσα δυναμική συμπεριφορά. Όσο μεγαλώνει το n , αυξάνει η πιθανότητα το S_n να πάρει μεγάλες τιμές. Ασυμπτωτικά, η μέση τιμή της απόστασης από την αρχή μετά από N βήματα είναι $\sim \sqrt{\frac{2N}{\pi}}$. Στις δύο διαστάσεις, έχειδειχθεί ότι μετά από N βήματα, για $N \rightarrow \infty$, όλα τα σημεία του χώρου (συμπεριλαμβανομένου της αρχής) έχουν ίση πιθανότητα ως το σημείο που βρίσκεται ο τυχαίος περίπατος. Στις τρεις διαστάσεις, έχει αποδειχθεί ότι μετά από άπειρα βήματα, η πιθανότητα να ξαναφτάσει κανείς στην αρχή -αλλά και σε οποιοδήποτε άλλο σημείο- είναι μικρότερη της μονάδας! Συγκεκριμένα είναι 0.34053 (μια από τις σταθερές του Pólya).

Διαδικασία Poisson

Μια διαδικασία απαρίθμησης με ανεξάρτητες και στάσιμες αυξήσεις είναι διαδικασία Poisson. Η διαδικασία αυτή χρησιμοποιείται εκτενώς στη μοντελοποίηση τηλεπικοινωνιακών δικτύων και θα αναφερθούμε σε αυτή εκτενώς στο επόμενο κεφάλαιο.



Σχήμα 4.5: Τυχαίοι περίπατοι βασισμένοι στη διαδικασία τυχαίου βήματος για μια, δύο και τρεις διαστάσεις. Το πράσινο τετράγωνο δείχνει την αρχή και ο κόκκινος δίσκος το σημείο του τυχαίου περίπατου μετά από $n = 1000$ σημεία.

4.6 Διαδικασίες συνεχούς χρόνου

Σε αυτό το σημείο επανερχόμαστε στις διαδικασίες με συνεχή παράμετρο, που για λόγους ευκολίας της δίνουμε την έννοια του χρόνου. Όπως είδαμε, οι διαδικασίες απαρίθμησης είναι διαδικασίες συνεχούς χρόνου που λαμβάνουν διακριτές τιμές και μπορούν να αντιστοιχιστούν σε διαδικασίες διακριτού χρόνου που λαμβάνουν συνεχείς τιμές. Τεχνικά, οι διαδικασίες αυτές ανήκουν σε αυτή την ενότητα, κρίθηκε σκόπιμο όμως να παρουσιαστούν στην προηγούμενη ενότητα. Στην ενότητα αυτή θα μελετήσουμε διαδικασίες συνεχούς χρόνου που λαμβάνουν συνεχείς τιμές και συγκεκριμένα θα εστιάσουμε σε διαδικασίες Gauss.

4.6.1 i.i.d. Gauss

Αυτή είναι η πιο απλή μορφή διαδικασίας Gauss όπου τα δείγματα είναι ανεξάρτητες (ή ισοδύναμα ασυσχέτιστες) τ.μ. $\mathcal{N}(m, \sigma^2)$. Κατά τα γνωστά

$$m_X(t) = m \text{ και } \sigma_X^2(t) = \sigma^2.$$

Επίσης για την αυτοσυμμεταβλητότητα βρίσκουμε

$$C_X(t_1, t_2) = \sigma^2 \delta_{t_1, t_2}.$$

Κατά συνέπεια, η από κοινού ΣΠΠ k δειγμάτων θα είναι

$$f_{X(t_1), \dots, X(t_k)}(x_1, \dots, x_k) = \frac{1}{(2\pi\sigma^2)^{k/2}} e^{-\sum_{i=1}^k \frac{(x_i - m)^2}{2\sigma^2}}.$$

4.6.2 Γενική διαδικασία Gauss

Σε αντιστοιχία με τα τυχαία διανύσματα Gauss, μπορούμε να ορίσουμε την από κοινού ΣΠΠ ως

$$f_{\mathbf{X}}(\mathbf{x}) \doteq f_{X(t_1), X(t_2), \dots, X(t_k)}(x_1, x_2, \dots, x_k) = \frac{e^{-\frac{(\mathbf{x} - \mathbf{m})^T \mathbf{K}^{-1} (\mathbf{x} - \mathbf{m})}{2}}}{(2\pi)^{n/2} |\mathbf{K}|^{1/2}},$$

όπου $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ είναι το διάνυσμα στήλη με τις μεταβλητές της συνάρτησης, $\mathbf{m} = [m_X(t_1), m_X(t_2), \dots, m_X(t_k)]^T$ είναι το μέσο διάνυσμα και

$$\mathbf{K} = \begin{bmatrix} C_X(t_1, t_1) & C_X(t_1, t_2) & \dots & C_X(t_1, t_k) \\ C_X(t_2, t_1) & C_X(t_2, t_2) & \dots & C_X(t_2, t_k) \\ \vdots & \vdots & \ddots & \vdots \\ C_X(t_k, t_1) & C_X(t_k, t_2) & \dots & C_X(t_k, t_k) \end{bmatrix}$$

είναι ο πίνακας συμμεταβλητότητας που περιέχει τη συνάρτηση αυτοσυμμεταβλητότητας σε όλα τα πιθανά ζεύγη δειγμάτων. Η από κοινού ΣΠΠ καθορίζεται πλήρως από τα \mathbf{m}, \mathbf{K} .

Παράδειγμα: Έστω $X(t)$ διαδικασία Gauss με συνάρτηση μέσης τιμής $m_X(t) = 3t$ και συνάρτηση αυτοσυμμεταβλητότητας $C_X(t_1, t_2) = 9e^{-2|t_1 - t_2|}$. Βρείτε την πιθανότητα $P(X(1) + X(2) > 15)$. Είναι η διαδικασία αυτή WSS;

Λύση: Γνωρίζουμε ότι το άθροισμα μεταβλητών Gauss είναι κατανομημένο κατά Gauss με μέση τιμή το άθροισμα μέσων τιμών και μεταβλητότητα το άθροισμα όλων των στοιχείων του πίνακα συμμεταβλητότητας. Άρα $Y = X(1) + X(2) \sim \mathcal{N}(m_X(1) + m_X(2), \sigma_X^2(1) + \sigma_X^2(2) + 2E[(X(1) - m_X(1))(X(2) - m_X(2))])$. Έχουμε

$$E[X(1) + X(2)] = m_X(1) + m_X(2) = 3 + 6 = 9.$$

και

$$\begin{aligned} \text{Var}(X(1) + X(2)) &= \text{Var}(X(1)) + \text{Cov}(X(1), X(2)) + \text{Cov}(X(2), X(1)) + \text{Var}(X(2)) \\ &= C_X(1, 1) + C_X(1, 2) + C_X(2, 1) + C_X(2, 2) \\ &= 9(1 + e^{-2} + e^{-2} + 1) = 20.43. \end{aligned}$$

Δηλαδή $Y \sim \mathcal{N}(9, 20.43)$. Η ζητούμενη πιθανότητα θα είναι

$$\begin{aligned} P(X(1) + X(2) > 15) &= P(Y > 15) \\ &= P\left(\frac{Y - 9}{\sqrt{20.43}} > \frac{15 - 9}{\sqrt{20.43}}\right) \\ &= Q(1.327) = 0.0922, \end{aligned}$$

όπου $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt$ είναι η συμπληρωματική συνάρτηση σφάλματος $Q(x) = 1 - \text{erf}(x)$.

Η $X(t)$ εμφανώς δεν είναι WSS καθώς ενώ η συνάρτηση αυτοσυμμεταβλητότητας εξαρτάται μόνο από τη διαφορά $t_2 - t_1$, η συνάρτηση μέσης τιμής εξαρτάται από το χρόνο. ■

4.6.3 Διαδικασία Wiener

Ξεκινώντας από τον τυχαίο περίπατο της παραπάνω ενότητας, θα μελετήσουμε την διαδικασία συνεχούς χρόνου που προκύπτει σαν όριο της διαδικασίας αυτής. Συγκεκριμένα, θεωρούμε την συμμετρική διαδικασία βήματος h D_n (το D_n μπορεί να πάρει τις τιμές $\{-h, h\}$ με πιθανότητες $\{\frac{1}{2}, \frac{1}{2}\}$). Δημιουργούμε στη συνέχεια τη διαδικασία συνεχούς χρόνου $X_\delta(t)$ ως

$$X_\delta(t) = h(D_1 + D_2 + \dots + D_{t/\delta}) = hS_n$$

όπου είναι προφανές ότι σε χρονικό διάστημα δ η διαδικασία $X_\delta(t)$ κάνει διακριτά βήματα $\pm h$. Μέχρι στιγμής, η $X_\delta(t)$ είναι μια διαδικασία απαρίθμησης συμβάντων που συμβαίνουν περιοδικά στο χρόνο. Οι συναρτήσεις ροπών

του μοντέλου πρώτης τάξης θα είναι

$$\begin{aligned} E[X_\delta(t)] &= hE[S_n] = 0 \\ \text{Var}(X_\delta(t)) &= h^2 n \text{Var}(D_n) = h^2 n 4p(1-p) = h^2 n. \end{aligned}$$

Στη συνέχεια, μας ενδιαφέρει η στοχαστική διαδικασία που προκύπτει σαν όριο όταν ελαττώνουμε ταυτόχρονα το δ και το h (χάνουμε δηλαδή πολλά μικρά βήματα που στο όριο φαίνονται σαν συνεχή βήματα) έτσι ώστε να ισχύει πάντα $h = \sqrt{a\delta}$. η προκύπτουσα διαδικασία $X(t)$ είναι διαδικασία συνεχούς χρόνου, λαμβάνει τιμές στο \mathbb{R} και ονομάζεται **διαδικασία Wiener**. Παρατηρούμε αρχικά ότι

$$\begin{aligned} E[X(t)] &= 0 \\ \text{Var}(X(t)) &= (\sqrt{a\delta})^2 \frac{t}{\delta} = at. \end{aligned}$$

Στη συνέχεια παρατηρούμε ότι $X(t) = \lim_{\delta \rightarrow 0} hS_n$ είναι ένα άθροισμα άπειρων τ.μ. (καθώς $n = t/\delta$), οπότε από το κεντρικό οριακό θεώρημα γνωρίζουμε ότι το $X(t)$ θα είναι κατανομημένο κατά Gauss, δηλαδή θα ισχύει:

$$f_{X(t)}(x) = \frac{1}{\sqrt{2\pi at}} e^{-\frac{x^2}{2at}}.$$

Επίσης, λόγω του τυχαίου περιπάτου, γνωρίζουμε ότι η διαδικασία $X(t)$ έχει στάσιμες ανεξάρτητες αυξήσεις. Άρα, η από κοινού κατανομή μπορεί να υπολογιστεί ως εξής

$$\begin{aligned} f_{X(t_1), \dots, X(t_k)}(x_1, \dots, x_k) &= f_{X(t_1)}(x_1) f_{X(t_2-t_1)}(x_2 - x_1) \dots f_{X(t_k-t_{k-1})}(x_k - x_{k-1}) \\ &= \frac{\exp \left\{ -\frac{1}{2} \left[\frac{x_1^2}{at_1} + \frac{(x_2-x_1)^2}{a(t_2-t_1)} + \dots + \frac{(x_k-x_{k-1})^2}{a(t_k-t_{k-1})} \right] \right\}}{\sqrt{(2\pi a)^k t_1(t_2-t_1) \dots (t_k-t_{k-1})}}. \end{aligned}$$

Η διαδικασία αυτή χρησιμοποιείται για την μοντελοποίηση της κίνησης *Brown* (Brownian motion) στη Φυσική αλλά και στις Τηλεπικοινωνίες.

Ερώτηση: Να δείξετε ότι η διαδικασία Wiener είναι διαδικασία Gauss.

Κεφάλαιο 5

Στοχαστική διαδικασία Poisson

5.1 Ορισμός της διαδικασίας Poisson

Η διαδικασία Poisson μπορεί να θεωρηθεί ως διαδικασία απαρίθμησης συμβάντων ή ισοδύναμα ως διαδικασία σημείων (point processes). Χρησιμοποιείται για την μοντελοποίηση πολλών εφαρμογών, όπως οι αφίξεις τηλεφωνικών κλήσεων, αιτήματα σε ιστοσελίδες, θεωρία ουρών, τυχαίες θέσεις σημείων (δέντρων, σταγόνων βροχής), κρουστικός θόρυβος, έκπτωση ραδιενεργών ατόμων κ.α. Επιδέχεται πολλαπλούς ορισμούς που είναι τελικά ισοδύναμοι μεταξύ τους.

5.1.1 Πρώτος ορισμός ως διαδικασία σημείων

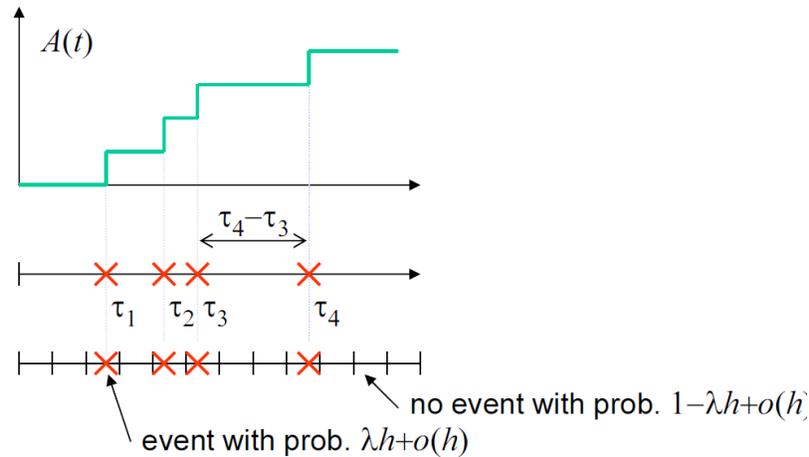
Θεωρούμε τη διαδικασία σημείων $X_n, n = 0, 1, \dots$. Η διαδικασία αυτή λέγεται διαδικασία Poisson με ένταση λ αν και μόνο αν εντός του μικρού διαστήματος $(t, t + h]$ έχουμε

$$P(1 \text{ γεγονός στο } (t, t + h]) = \lambda h - o(h)$$

και

$$P(\text{κανένα γεγονός στο } (t, t + h]) = 1 - \lambda h - o(h)$$

όπου το $o(h)$ συμβολίζει οποιαδήποτε υπογραμμική συνάρτηση (sublinear function) έτσι ώστε για $h \rightarrow 0$ έχουμε $\frac{o(h)}{h} \rightarrow 0$. Αυτό σημαίνει ότι τα γεγονότα συμβαίνουν ένα τη φορά με πιθανότητα 1, και ότι η έντασή τους είναι ίση με λ .



Σχήμα 5.1: Οι τρεις ορισμοί της διαδικασίας Poisson είναι ισοδύναμοι.

Από τον ορισμό μπορούμε να παρατηρήσουμε ότι ο αριθμός των συμβάντων σε δύο ξένα διαστήματα είναι ανεξάρτητος και ο χρόνος μεταξύ δύο συμβάντων έχει την ιδιότητα απώλειας μνήμης. Αυτό μας οδηγεί στο συμπέρασμα ότι ο χρόνος μεταξύ δύο συμβάντων είναι εκθετικά κατανομημένος.

5.1.2 Δεύτερος ορισμός ως διαδικασία σημείων

Μια διαδικασία σημείων $X_n, n = 1, 2, \dots$ είναι διαδικασία Poisson με ένταση λ αν και μόνο αν έχει στάσιμα, ανεξάρτητα και όμοια κατανομημένα διαστήματα μεταξύ των σημείων $X_{n+1} - X_n$ που είναι τ.μ. $\sim \exp(\lambda)$.

Ο αριθμός επιτυχιών μιας διαδικασίας i.i.d. Bernoulli στο διάστημα $[0, t]$ είναι κατανομημένος κατά Binomial. Αν πάρουμε πολύ μικρά βήματα σε αυτήν την διαδικασία έτσι ώστε $\frac{n}{t} \rightarrow 0$, η κατανομή Binomial τείνει οριακά στην κατανομή Poisson(λ) ενώ πρακτικά εξομοιώνουμε τη διαδικασία Poisson. Αυτό μας οδηγεί στον τρίτο ισοδύναμο ορισμό.

5.1.3 Τρίτος ορισμός ως διαδικασία απαρίθμησης

Μια διαδικασία απαρίθμησης $N(t), t \geq 0$ είναι διαδικασία Poisson με ένταση λ αν και μόνο αν έχει ανεξάρτητες αυξήσεις που ακολουθούν την κατανομή $\sim \text{Poisson}(\lambda\tau)$ για διαστήματα μήκους τ .

Από τους ορισμούς προκύπτει ότι η διαδικασία Poisson δεν είναι στάσιμη, έχει όμως στάσιμες αυξήσεις. Οι τρεις παραπάνω ορισμοί στέκουν ο καθένας ανεξάρτητος και συνεπάγονται ο ένας τους άλλους δύο.

5.2 Ιδιότητες της διαδικασίας Poisson

5.2.1 Βασικές ιδιότητες

Απόρροια των ορισμών, η διαδικασία Poisson με ένταση λ έχει τις παρακάτω βασικές ιδιότητες

- $N(0) = 0$.
- Αν λογιστεί σαν διαδικασία απαρίθμησης, έχει στάσιμες ανεξάρτητες αυξήσεις.
- Οι χρόνοι μεσοδιαστήματος (interarrival times) είναι κατά συνέπεια ανεξάρτητοι (όταν αναφέρονται σε μη επικαλυπτόμενα διαστήματα) και το μήκος τους κατανομημένο $\sim \exp(\lambda)$.



Σχήμα 5.2: Το παράδοξο του ωτοστόπ.

- Η πιθανότητα αναμονής μέχρι την επόμενη άφιξη είναι επίσης $\sim \exp(\lambda)$.
- Για ένα διάστημα τ , ισχύει πάντα

$$P(N(t + \tau) - N(t) = k) = \frac{e^{-\lambda\tau}(\lambda\tau)^k}{k!}.$$

- Η πιθανότητα σε ένα διάστημα $h \rightarrow 0$ να έχουμε παραπάνω από μια άφιξη τείνει στο μηδέν.

Παράδειγμα: (Το παράδοξο του ωτοστόπ - The hitchhiker's paradox): Σε έναν έρημο δρόμο, ένας ταξιδιώτης αποφασίζει να κάνει ωτοστόπ. Είναι γνωστό ότι από τον δρόμο αυτό, περνούν αυτοκίνητα σύμφωνα με την διαδικασία Poisson με ένταση λ (ας πούμε $\frac{1}{\lambda} = 10$ λεπτά). Πόσο χρόνο κατά μέσο όρο πρέπει να περιμένει ο ταξιδιώτης μέχρι να εμφανιστεί το πρώτο αυτοκίνητο; Πως εξηγείται αυτό;

Λύση: Γνωρίζουμε ότι τα αυτοκίνητα θα έχουν εκθετικούς χρόνους μεσοδιαστήματος με μέση τιμή 10 λεπτά.

Άρα μια πρώτη λογική εκτίμηση είναι να πούμε ότι ο ταξιδιώτης θα περιμένει κατά μέσο όρο 5 λεπτά, που είναι το μισό, υποθέτοντας ότι έρχεται στη μέση ενός μέσου μεσοδιαστήματος. Όμως, γνωρίζοντας ότι η διαδικασία Poisson έχει απώλεια μνήμης, μπορούμε να ισχυριστούμε ότι θα πρέπει να περιμένουμε έναν ολόκληρο χρόνο μεσοδιαστήματος που είναι 10 λεπτά. Τι από τα δύο συμβαίνει;

Συμβαίνει το δεύτερο, το οποίο υπονοεί ότι το μεσοδιάστημα που επέλεξε ο ταξιδιώτης να ξεκινήσει το ωτοστόπ είναι μεγαλύτερο από τα κανονικά μεσοδιαστήματα! Υποθέτοντας ότι τα μεσοδιαστήματα έχουν εκθετικά κατανομημένο μήκος, ο ταξιδιώτης καλείται να διαλέξει ένα στην τύχη, όμως πόσο είναι το μεγεθός του;

Όπως στο σχήμα 5.2.1, έστω X το μεσοδιάστημα που διαλέγει ο ταξιδιώτης και Y ένα οποιοδήποτε μεσοδιάστημα (γνωρίζουμε ότι $f_Y(x) = \lambda e^{-\lambda x}$, αλλά δεν γνωρίζουμε πως συμπεριφέρεται το X λόγω της υποκειμενικότητας του ταξιδιώτη). Η πιθανότητα να διαλέξει ο ταξιδιώτης ένα μεσοδιάστημα είναι ανάλογη του μήκους του (σαν να ρίχνουμε ένα σημείο ομοιόμορφα στο χρόνο), αλλά και της πιθανότητας εμφάνισης αυτού του μεσοδιαστήματος, δηλαδή

$$f_X(x)dx = Kx f_Y(x)dx,$$

όπου, $f_X(x)$ η ΣΠΠ του X και K μια σταθερά. Ολοκληρώνοντας, μπορούμε να υπολογίσουμε $K = \lambda$. Άρα βρίσκουμε ότι

$$f_X(x) = x\lambda^2 e^{-\lambda x}, x \geq 0$$

και υπολογίζουμε $E[X] = \int_0^\infty x f_X(x)dx = \frac{2}{\lambda} = 20$ λεπτά. Αυτό επιβεβαιώνει τόσο τον ορισμό (που λέει ότι λόγω της απώλειας μνήμης θα πρέπει να περιμένουμε για ένα ολόκληρο μεσοδιάστημα) όσο και τη λογική, αφού είναι πιθανότερο να διαλέξουμε στην τύχη ένα μεγάλο μεσοδιάστημα από ένα μικρότερο. ■

Η συνάρτηση αυτοσυμμεταβλητότητας μπορεί να υπολογιστεί χρησιμοποιώντας τους τύπους για διαδικασίες περιπάτου ή αναλυτικά, έστω $t_1 < t_2$

$$\begin{aligned} C_N(t_1, t_2) &= E[(N(t_1) - \lambda t_1)(N(t_2) - \lambda t_2)] \\ &= E[(N(t_1) - \lambda t_1)(N(t_2 - t_1) - \lambda(t_2 - t_1))] + E[(N(t_1) - \lambda t_1)^2] \\ &= \text{Var}(N(t_1)) = \lambda t_1. \end{aligned}$$

Όμοια, για $t_2 < t_1$ βρίσκουμε ότι $C_N(t_1, t_2) = \lambda t_2$, δηλαδή έχουμε ότι $C_N(t_1, t_2) = \lambda \min\{t_1, t_2\}$.

5.2.2 Δέσμευση ως προς k

Έστω μια διαδικασία Poisson με ένταση λ , και δεσμεύουμε ως προς το γεγονός στο διάστημα $[t, t + \tau]$ να έχουμε ένα ακριβώς συμβάν, δηλαδή $N(t + \tau) - N(t) = 1$. Στη συνέχεια, διερωτώμαστε ποιά είναι η πιθανότητα το συμβάν να έγινε στο πρώτο διάστημα $[t, t + x]$, $x \in [0, \tau]$, που είναι ουσιαστικά η κατανομή του υπό συνθήκη γεγονότος.

$$\begin{aligned} P(X \leq x) &= P(N(t+x) - N(t) = 1 | N(t+\tau) - N(t) = 1) \\ &= \frac{P(N(t+x) - N(t) = 1, N(t+\tau) - N(t) = 1)}{P(N(t+\tau) - N(t) = 1)} \\ &= \frac{P(N(t+x) - N(t) = 1, N(t+\tau) - N(t+x) = 0)}{P(N(t+\tau) - N(t) = 1)} \\ &= \frac{\frac{e^{-\lambda x} (\lambda x)^1}{1!} \frac{e^{-\lambda(\tau-x)} (\lambda(\tau-x))^0}{0!}}{\frac{e^{-\lambda\tau} (\lambda\tau)^1}{1!}} = \frac{x}{\tau}. \end{aligned}$$

Δηλαδή, το συμβάν είναι ομοιόμορφα κατανομημένο στο $[0, \tau]$. Το ίδιο συμβαίνει όταν δεσμεύουμε με το γεγονός να έχουμε k συμβάντα στο $[0, \tau]$, τότε η από κοινού κατανομή των k συμβάντων είναι ομοιόμορφη. Αν θέλουμε να εξομοιώσουμε μια διαδικασία Poisson σε ένα πεπερασμένο διάστημα, μπορούμε απλά να υπολογίσουμε τον αριθμό των σημείων εντός του διαστήματος – θα είναι τυχαίος και κατανομημένος $\sim \text{Poisson}(\lambda)$ – και στη συνέχεια να ρίξουμε τα σημεία εντός του διαστήματος με ομοιόμορφη κατανομή. Εναλλακτικά, μπορούμε να σημειώνουμε ένα συμβάν, και μετά να γεννάμε έναν εκθετικό χρόνο μέχρι το επόμενο συμβάν.

5.2.3 Εναπόθεση (superposition) δύο διαδικασιών Poisson

Έστω ότι σε ένα μπωλ με καυτή σούπα πέφτουν μύγες σύμφωνα με μια διαδικασία Poisson X_t με ένταση λ_x και μέλισσες σύμφωνα με μια διαδικασία Poisson Y_t με ένταση λ_y ανεξάρτητα μεταξύ τους. Τότε, μπορούμε ισοδύναμα να πούμε ότι έντομα πέφτουν στη σούπα σύμφωνα με μια διαδικασία Poisson Z_t με ένταση $\lambda_x + \lambda_y$.

Απόδειξη: Παρατηρούμε αρχικά ότι η Z_t είναι μια διαδικασία απαρίθμησης, δηλ. έχει τις τρεις βασικές ιδιότητες

- $Z_0 = 0$.
- Είναι άυξουσα.
- Είναι συνεχούς χρόνου και διακριτών καταστάσεων.

Επίσης, οι αυξήσεις είναι ανεξάρτητες λόγω ανεξαρτησίας μεταξύ των X_t, Y_t :

$$\begin{aligned}
& P(Z_{t_3-t_2} = k_2, Z_{t_1-t_0} = k_1) = P(X_{t_3-t_2} + Y_{t_3-t_2} = k_2, X_{t_1-t_0} + Y_{t_1-t_0} = k_1) \\
&= \sum_{y_2} \sum_{y_1} P(X_{t_3-t_2} = k_2 - y_2, X_{t_1-t_0} = k_1 - y_1 | Y_{t_3-t_2} = y_2, Y_{t_1-t_0} = y_1) P(Y_{t_3-t_2} = y_2, Y_{t_1-t_0} = y_1) \\
&= \sum_{y_2} \sum_{y_1} P(X_{t_3-t_2} = k_2 - y_2 | Y_{t_3-t_2} = y_2, Y_{t_1-t_0} = y_1) P(X_{t_1-t_0} = k_1 - y_1 | Y_{t_3-t_2} = y_2, Y_{t_1-t_0} = y_1) \\
&\times P(Y_{t_3-t_2} = y_2) P(Y_{t_1-t_0} = y_1) \\
&= \sum_{y_2} \sum_{y_1} P(X_{t_3-t_2} = k_2 - y_2 | Y_{t_3-t_2} = y_2) P(X_{t_1-t_0} = k_1 - y_1 | Y_{t_1-t_0} = y_1) P(Y_{t_3-t_2} = y_2) P(Y_{t_1-t_0} = y_1) \\
&= P(X_{t_3-t_2} + Y_{t_3-t_2} = k_2) P(X_{t_1-t_0} + Y_{t_1-t_0} = k_1) \\
&= P(Z_{t_3-t_2} = k_2) P(Z_{t_1-t_0} = k_1)
\end{aligned}$$

Απομένει μόνο να δείξουμε ότι οι αυξήσεις είναι στάσιμες ή ότι ακολουθούν κατανομή Poisson($(\lambda_x + \lambda_y)\tau$) σε ένα διάστημα $[0, \tau]$. Πράγματι,

$$\begin{aligned}
P(Z_{t+\tau} - Z_t = k) &= P(X_{t+\tau} + Y_{t+\tau} - X_t - Y_t = k) \\
&= \sum_{n=0}^k P(X_{t+\tau} - X_t = k - n) P(Y_{t+\tau} - Y_t = n) \\
&= \sum_{n=0}^k \frac{e^{-\lambda_x \tau} (\lambda_x \tau)^{k-n}}{(k-n)!} \frac{e^{-\lambda_y \tau} (\lambda_y \tau)^n}{n!} \\
&= e^{-(\lambda_x + \lambda_y)\tau} \tau^k \sum_{n=0}^k \frac{\lambda_x^{k-n} \lambda_y^n}{(k-n)! n!} \\
&= \frac{e^{-(\lambda_x + \lambda_y)\tau} \tau^k}{k!} \sum_{n=0}^k \frac{k!}{(k-n)! n!} \lambda_x^{k-n} \lambda_y^n \\
&= \frac{e^{-(\lambda_x + \lambda_y)\tau} ((\lambda_x + \lambda_y)\tau)^k}{k!},
\end{aligned}$$

αφού ισχύει $\sum_{n=0}^k \frac{k!}{(k-n)! n!} \lambda_x^{k-n} \lambda_y^n = (\lambda_x + \lambda_y)^k$. Την ίδια απόδειξη μπορούμε να επαναλάβουμε για οποιοδήποτε διάστημα $[t, t + \tau]$. ■

Συμπερασματικά, η εναπόθεση (superposition) δύο ανεξάρτητων διαδικασιών Poisson είναι μια διαδικασία Poisson με ένταση ίση με το άθροισμα των εντάσεων των δύο διαδικασιών.

5.2.4 Λέπτυνση (thinning) διαδικασίας Poisson

Σε ένα κλαμπ καταφθάνουν πελάτες σύμφωνα με μια διαδικασία $X_t \sim \text{Poisson}$ με ένταση λ . Ο πορτιέρης απομακρύνει (ρίχνει πόρτα) τους πελάτες με μια πιθανότητα p ίδια για όλους τους πελάτες. Το ερώτημα είναι αν μπορούμε να χαρακτηρίσουμε τη διαδικασία άφιξης των πελατών Y_t όπως παρατηρείται από τον μπαρμαν του καταστήματος. Παρακάτω θα αποδείξουμε ότι η διαδικασία Y_t είναι διαδικασία Poisson με ένταση λp και η επιλογή των πελατών ονομάζεται λέπτυνση της διαδικασίας (Poisson thinning).

Απόδειξη: Κατά τα γνωστά, η Y_t είναι προφανώς μια διαδικασία απαρίθμησης. Επίσης, η απόδειξη των ανεξάρτητων αυξήσεων είναι εύκολη, οπότε θα εστιάσουμε στην απόδειξη της κατανομής Poisson(λp) των

συμβάντων σε ένα διάστημα $[0, \tau]$:

$$\begin{aligned}
 P(Y_{t+\tau} - Y_t = k) &= \sum_{n=k}^{\infty} P(Y_{t+\tau} - Y_t = k | X_{t+\tau} - X_t = n) P(X_{t+\tau} - X_t = n) \\
 &= \sum_{n=k}^{\infty} P(k \text{ successes in } n \text{ trials}) P(X_{t+\tau} - X_t = n) \\
 &= \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \frac{e^{-\lambda\tau} (\lambda\tau)^n}{n!} \\
 &= \frac{e^{-\lambda\tau} (\lambda p\tau)^k}{k!} \sum_{n=k}^{\infty} \frac{(\lambda(1-p)\tau)^{n-k}}{(n-k)!} = \frac{e^{-\lambda\tau} (\lambda p\tau)^k}{k!} \sum_{i=0}^{\infty} \frac{(\lambda(1-p)\tau)^i}{i!} \\
 &= \frac{e^{-\lambda\tau} (\lambda p\tau)^k}{k!} e^{\lambda(1-p)\tau} = \frac{e^{-\lambda p\tau} (\lambda p\tau)^k}{k!},
 \end{aligned}$$

όπου χρησιμοποιήσαμε την ταυτότητα $e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$. ■

5.2.5 Ιδιότητα PASTA

Η διαδικασία Poisson χρησιμοποιείται ευρύτατα στην μοντελοποίηση αφίξεων, όπως για παράδειγμα σε συστήματα ουρών όπου έχουμε αφίξεις πακέτων, πελατών κ.ο.κ. Μια σημαντική ιδιότητα που απλοποιεί την μελέτη τέτοιων μοντέλων λέγεται ιδιότητα PASTA=Poisson arrivals see time averages. Η ιδιότητα αυτή λέει ότι οι αφίξεις (ή αλλιώς οι πελάτες) συναντούν το σύστημα σε στάσιμη κατάσταση. Αυτό πρακτικά σημαίνει ότι η αναλογία πελατών που συναντά το σύστημα σε μια κατάσταση $A(t)$ είναι ίδια με την αναλογία που θα συναντήσει ένας ουδέτερος παρατηρητής δειγματοληπτόντας το σύστημα σε αυθαίρετα σημεία. Με τον όρο time average εννοούμε τη συνάρτηση

$$\frac{1}{t} \int_0^t A(\tau) d\tau$$

και με τον όρο 'τι βλέπουν οι αφίξεις', εννοούμε τη συνάρτηση

$$\frac{1}{N(t)} \sum_{i=1}^{N(t)} A(t^-).$$

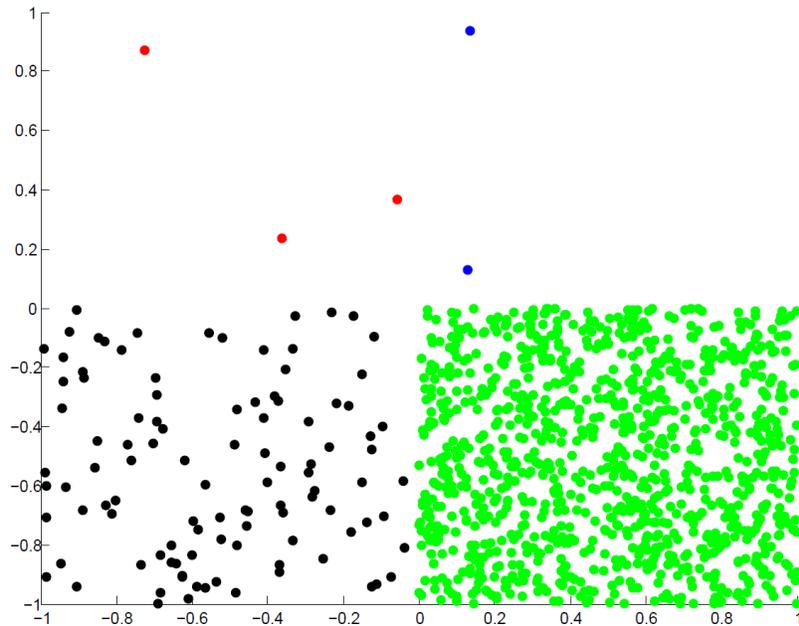
Η παραπάνω ιδιότητα λέει ότι αν ένα από τα παραπάνω όρια υπάρχει, τότε υπάρχει και το άλλο και είναι ίσα. Στην περίπτωση μιας διαδικασίας Poisson (δηλ. $A(t) = N(t)$) τα όρια δεν υπάρχουν καθώς οι συναρτήσεις αποκλίνουν. Όμως, σε συστήματα ουρών με αφίξεις Poisson, η σημαντική αυτή ιδιότητα σε συνδυασμό με το νόμο του Little επιτρέπει την μοντελοποίηση των συστημάτων ουρών χρησιμοποιώντας την μέθοδο της μέσης τιμής (mean value approach).

5.3 Άλλες μορφές της διαδικασίας Poisson

5.3.1 Μη ομογενής

Μια διαδικασία Poisson λέγεται **ομογενής** όταν η έντασή της είναι ανεξάρτητη της παραμέτρου του χρόνου. Μπορούμε να σκεφτούμε μη ομογενείς διαδικασίες Poisson όπου η ένταση είναι όντως συνάρτηση του χρόνου $\lambda(t)$. Σε αυτή την περίπτωση, ορίζουμε ως μέση χρονική ένταση την

$$m(t) = \int_0^t \lambda(x) dx$$



Σχήμα 5.3: Ένα δείγμα του τμήματος $[-1, 1] \times [-1, 1]$ μιας μη ομογενούς χωρικής διαδικασίας Poisson με $\lambda_1 = 1, \lambda_2 = 10, \lambda_3 = 100, \lambda_4 = 1000$ όπου ο δείκτης αναφέρεται στο τεταρτημόριο.

και ισχύει

$$P(N_\tau - N_0 = k) = \frac{e^{-m(\tau)} m(\tau)^k}{k!}.$$

Παρατηρήστε ότι για $\lambda(t) = c$ προκύπτει η πιθανότητα k αυξήσεων σε μια ομογενή διαδικασία Poisson.

5.3.2 Μείγμα Poisson

Έστω ότι ομάδες πελατών φτάνουν σε ένα κατάστημα σύμφωνα με μια διαδικασία Poisson X_t με ένταση λ . Όμως, κάθε ομάδα πελατών, μπορεί να είναι μια παρέα, μια οικογένεια ή και ένα μόνο άτομο. Η θετική τ.μ. Y_i μας δίνει τον αριθμό ατόμων σε κάθε ομάδα πελατών και είναι i.i.d. για κάθε άφιξη ομάδας πελατών. Ενδιαφερόμαστε για την σταχαστική διαδικασία αρίξεως ατόμων Z_t που ονομάζεται μείγμα Poisson (compound Poisson). Προφανώς θα ισχύει

$$Z_t = \sum_{i=1}^{X_t} Y_i$$

και η Z_t είναι μια διαδικασία απαρίθμησης που έχει όμως βήματα ≥ 1 (πριν είχαμε βήματα του 1). Μπορούμε να υπολογίσουμε την χαρακτηριστική συνάρτηση ενός δείγματος της Z_t :

$$\begin{aligned} \Phi_{Z_t}(\omega) &= E[e^{j\omega Z(t)}] = \sum_{n=0}^{\infty} E[e^{j\omega Z(t)} | X(t) = n] \frac{e^{-\lambda t} (\lambda t)^n}{n!} \\ &= e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} E[e^{j\omega(Y_1 + \dots + Y_n)}] = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \Phi_Y^n(\omega) \\ &= e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t \Phi_Y(\omega))^n}{n!} = e^{\lambda t (\Phi_Y(\omega) - 1)}. \end{aligned}$$

Το αποτέλεσμα αυτό μπορούμε να το χρησιμοποιήσουμε για να υπολογίσουμε τις ροπές:

$$\begin{aligned} m_Z(t) &= E[Z(t)] = \frac{\Phi_{Z_t}(0)'}{j} = \frac{e^{\lambda t(\Phi_Y(\omega)-1)}(\lambda t(\Phi_Y(\omega)-1))'}{j} \Big|_{\omega=0} \\ &= \lambda t \frac{e^{\lambda t(\Phi_Y(0)-1)}\Phi_Y'(0)}{j} = \lambda t E[Y], \end{aligned}$$

αφού $\Phi_Y(0) = 1$ και $\Phi_Y'(0) = jE[Y]$. Επίσης, για τη συνάρτηση μεταβλητότητας έχουμε

$$\begin{aligned} \sigma_Z^2(t) &= \text{Var}(Z(t)) = -\Phi_{Z_t}(0)'' - E[Z(t)]^2 = \\ &= -\lambda t e^{\lambda t(\Phi_Y(\omega)-1)}[\Phi_Y''(\omega) + \lambda t(\Phi_Y'(\omega))^2] \Big|_{\omega=0} - E[Z(t)]^2 \\ &= \lambda t E[Y^2] - (j\lambda t E[Y])^2 - E[Z(t)]^2 = \lambda t E[Y^2]. \end{aligned}$$

5.3.3 Χωρική διαδικασία Poisson

Μέχρι τώρα μελετήσαμε τη διαδικασία Poisson που είδαμε ότι ονομάζεται και διαδικασία σημείων (point process) όταν μελετάμε τη διακριτού χρόνου μορφή της. Αν τα σημεία βρίσκονται στο διδιάστατο χώρο, τότε αναφερόμαστε στην χωρική διαδικασία Poisson (spatial Poisson process ή 2d Poisson). Η διαδικασία αυτή έχει επίσης στάσιμες και ανεξάρτητες αυξήσεις και ο ρυθμός εμφάνισης των σημείων εξαρτάται από την ένταση που είναι τώρα η πυκνότητα των σημείων στο επίπεδο. Για σταθερή ένταση (δηλαδή για ομογενή χωρική διαδικασία Poisson) η πιθανότητα εμφάνισης k σημείων εντός μιας επιφάνειας A θα είναι

$$P(X_A = k) = \frac{e^{-\lambda A}(\lambda A)^k}{k!}.$$

Επίσης, ισχύουν όμοιες ιδιότητες με την περίπτωση μιας διάστασης. Μπορούμε δηλαδή να δεσμεύσουμε ως προς μια επιφάνεια και τότε τα σημεία θα είναι ομοιόμορφα κατανεμημένα στο χώρο. Επίσης, μπορούμε να εναποθέσουμε δύο χωρικές διαδικασίες, να λεπτύνουμε μια διαδικασία απορρίπτοντας σημεία με μια πιθανότητα ή να ορίσουμε μια μη ομογενή χωρική διαδικασία όπως στο σχήμα 5.3. Τα συμβάντα εντός δύο επιφανειών A_1 και A_2 που δεν έχουν κοινά σημεία είναι ανεξάρτητα (ανεξάρτητες αυξήσεις). Γενικεύοντας, μπορούμε να μιλάμε για διαδικασίες Poisson στον n -διάστατο χώρο όπου κάθε σημείο είναι πλέον ένα διάνυσμα με n στοιχεία.

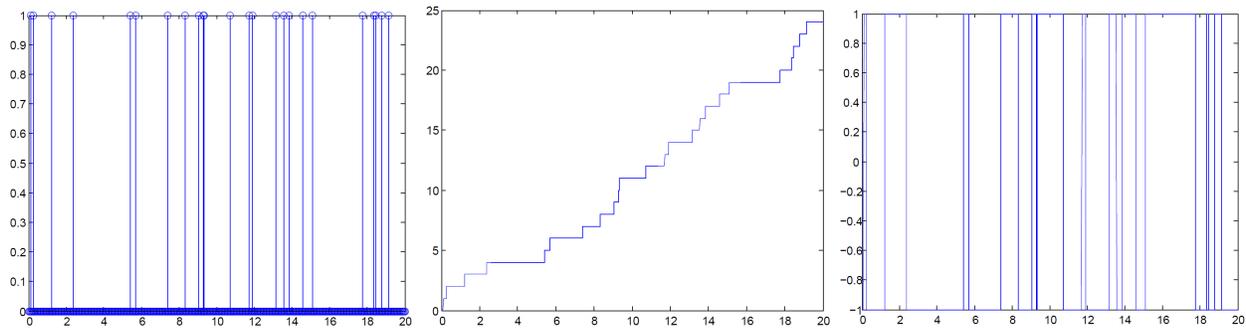
Δοκιμάστε στο matlab αυτό: `poiss.m`

5.3.4 Τυχαίο σήμα τηλέγραφου

Το τυχαίο σήμα τηλέγραφου (random telegraph signal) είναι μια διαδικασία βασισμένη στη διαδικασία Poisson. Έστω, λοιπόν μια διαδικασία $X(t)$ η οποία λαμβάνει τιμές στο $\{-1, 1\}$. Αρχικά έχουμε $X(0) = -1$ με πιθανότητα $\frac{1}{2}$ και $X(0) = 1$ πάλι με πιθανότητα $\frac{1}{2}$. Στη συνέχεια, η $X(t)$ αλλάζει πρόσημο κάθε φορά που η $N(t)$ αυξάνει κατά μια τιμή, όπου $N(t)$ είναι μια διαδικασία Poisson με ένταση λ .

Μπορούμε να υπολογίσουμε την ΣΜΠ ενός δείγματος της $X(t)$ χρησιμοποιώντας τις υπό συνθήκη ΣΜΠ

$$P(X(t) = i) = P(X(t) = i | X(0) = 1) P(X(0) = 1) + P(X(t) = i | X(0) = -1) P(X(0) = -1),$$



Σχήμα 5.4: Ένα δείγμα του τυχαίου σήματος τηλέγραφου, η διαδικασία σημείων Poisson (αριστερά), η διαδικασία απαρίθμησης (μέση) και το τυχαίο σήμα τηλέγραφου (δεξιά) σε έναν ορίζοντα 20 δευτερολέπτων. Η διαδικασία Poisson έχει μοναδιαία ένταση ($\lambda = 1$).

όπου $i \in \{-1, 1\}$ και

$$\begin{aligned} P(X(t) = i | X(0) = i) &= P(N(t) = \text{άρτιος}) = \sum_{j=0}^{\infty} \frac{e^{-\lambda t} (\lambda t)^{2j}}{(2j)!} \\ &= e^{-\lambda t} \frac{1}{2} (e^{\lambda t} + e^{-\lambda t}) = \frac{1}{2} (1 + e^{-2\lambda t}), \\ P(X(t) = -i | X(0) = i) &= P(N(t) = \text{περιττός}) = \sum_{j=0}^{\infty} \frac{e^{-\lambda t} (\lambda t)^{2j+1}}{(2j+1)!} \\ &= e^{-\lambda t} \frac{1}{2} (e^{\lambda t} - e^{-\lambda t}) = \frac{1}{2} (1 - e^{-2\lambda t}). \end{aligned}$$

Κάνοντας πράξεις βρίσκουμε

$$\begin{aligned} P(X(t) = 1) &= \frac{1}{2} \frac{1}{2} (1 + e^{-2\lambda t}) + \frac{1}{2} \frac{1}{2} (1 - e^{-2\lambda t}) = \frac{1}{2} \\ P(X(t) = -1) &= 1 - P(X(t) = 1) = \frac{1}{2}. \end{aligned}$$

Δηλαδή, οποιοδήποτε σημείο της διαδικασίας $X(t)$ έχει πάντα την ίδια πιθανότητα να είναι $+1$ ή -1 . Η συνάρτηση μέσης τιμής θα είναι

$$m_X(t) = 1P(X(t) = 1) + (-1)P(X(t) = -1) = 0,$$

για τη συνάρτηση μεταβλητότητας

$$\sigma_X^2 = E[X^2(t)] = 1^2P(X(t) = 1) + (-1)^2P(X(t) = -1) = 1$$

και για τη συνάρτηση αυτοσυμμεταβλητότητας

$$\begin{aligned} C_X(t_1, t_2) &= E[X(t_1)X(t_2)] = 1P(X(t_1) = X(t_2)) + (-1)P(X(t_1) \neq X(t_2)) \\ &= \frac{1}{2}(1 + e^{-2\lambda|t_2-t_1|}) - \frac{1}{2}(1 - e^{-2\lambda|t_2-t_1|}) = e^{-2\lambda|t_2-t_1|}, \end{aligned}$$

δηλαδή η συσχέτιση δύο σημείων φθίνει με το πέρασμα του χρόνου. Στο σχήμα 5.4 φαίνεται ένα δείγμα του τυχαίου σήματος τηλέγραφου για μια διαδικασία Poisson με μοναδιαία ένταση ($\lambda = 1$).

Δοκιμάστε στο matlab αυτό: `telegraph.m`



Σχήμα 5.5: Η διαδικασία Poisson λαμβάνεται ως έξοδος του συστήματος που έχει ως είσοδο μια σειρά από συναρτήσεις δέλτα και συνάρτηση απόκρισης το ολοκλήρωμα της εισόδου (αριστερά). Αν αντί για αυτή τη συγκεκριμένη συνάρτηση απόκρισης χρησιμοποιηθεί μια γενική αιτιατή συνάρτηση $h(t)$, το αποτέλεσμα είναι η διαδικασία Poisson shot noise (δεξιά).

5.3.5 Shot noise

Ένας άλλος τρόπος να σκεφτούμε τη διαδικασία Poisson $N(t)$ είναι σαν ένα άθροισμα από βηματικές συναρτήσεις $u(t - t_i)$. Η βηματική συνάρτηση ορίζεται ως

$$u(t - t_i) = \begin{cases} 1 & t \geq t_i \\ 0 & t < t_i \end{cases}$$

και η διαδικασία Poisson μπορεί να οριστεί ως

$$N(t) = \sum_{i=1}^{\infty} u(t - t_i),$$

με $t_i > 0$ να ορίζουν τις χρονικές στιγμές των συμβάντων. Επεκτείνοντας αυτό το μοντέλο, ορίζουμε τη σειρά συναρτήσεων δέλτα Dirac (train of delta functions)

$$Z(t) = \sum_{i=1}^{\infty} \delta(t - t_i), \quad t_i > 0.$$

Γνωρίζοντας τη σχέση $u(t - t_i) = \int_{-\infty}^t \delta(t - t_i) dt$, μπορούμε να καταλήξουμε στο συμπέρασμα ότι ένα δείγμα της διαδικασίας Poisson μπορεί να ληφθεί στην έξοδο του συστήματος του σχήματος 5.3.5 αν χρησιμοποιηθεί στην είσοδο το σήμα $Z(t)$ που περιέχει όλα τα σημεία συμβάντων του δείγματος. Ακολουθώντας τον ίδιο συνειρμό αλλά χρησιμοποιώντας σαν σύστημα ένα φίλτρο με συνάρτηση απόκρισης μια γενική αιτιατή $h(t)$ (δηλαδή $h(t) = 0, t < 0$), λαμβάνουμε στην έξοδο ένα στοχαστικό σήμα $Y(t)$ που ονομάζεται Poisson shot noise.

Η μέση τιμή του $Y(t)$ υπολογίζεται ως εξής

$$\begin{aligned} E[Y(t)] &= E_N[E[Y(t)|N(t)]] = E_N[E[\sum_{i=1}^N h(t - t_i)]] \\ &= E_N[\sum_{i=1}^N E[h(t - t_i)]] = E_N[N \int_0^t \frac{h(t-x)}{t} dx] \\ &= \lambda \int_0^t h(t-x) dx, \end{aligned}$$

όπου χρησιμοποιήσαμε το $E[h(t - t_i)] = \int_0^t \frac{h(t-x)}{t} dx$ που προκύπτει από το γεγονός ότι οι χρονικές στιγμές t_i είναι ομοιόμορφα κατανομημένες στο $[0, t]$ αφού έχουμε δεσμεύσει ως προς το πλήθος τους N .

Παράδειγμα: (Παρεμβολές σε ένα ασύρματο δίκτυο): Έστω ένα δίκτυο όπου ασύρματα τερματικά βρίσκονται στο χώρο σύμφωνα με μια δισδιάστατη διαδικασία Poisson. Κάθε τερματικό εκπέμπει με ισχύ P ενώ το ηλεκτρομαγνητικό σήμα υφίσταται απώλειες σύμφωνα με το νόμο του Friis, δηλαδή η συνάρτηση απωλειών είναι $l(d) = ad^{-2}$, όπου d είναι η απόσταση. Αν υποθέσουμε ότι όλα τα τερματικά εκπέμπουν ταυτόχρονα, βρείτε το ποσό παρεμβολών.

Λύση: Οι παρεμβολές σε ένα αυθαίρετο σημείο (θεωρούμε την αρχή των αξόνων), θα είναι το άθροισμα των σημάτων που φτάνουν στο σημείο αυτό:

$$I_O = \sum_{i \in X} Pl(|X_i - O|) = a \sum_{i \in X} |X_i - O|^{-2}.$$

Παρατηρούμε ότι το I_O είναι μια τ.μ. που εξαρτάται από τις θέσεις των σημείων X_i . Πιο συγκεκριμένα, το I_O είναι άλλο ένα παράδειγμα Shot noise. ■

Κεφάλαιο 6

Διαδικασίες Markov

Στο κεφάλαιο αυτό θα ασχοληθούμε με διαδικασίες διακριτού χώρου καταστάσεων που έχουν την ιδιότητα Markov και ονομάζονται **αλυσίδες Markov**. Οι διαδικασίες αυτές χρησιμοποιούνται για να περιγράψουν συστήματα που βρίσκονται σε μια από $|S|$ δυνατές καταστάσεις ενώ η συμπεριφορά του συστήματος εξαρτάται μόνο από τις πιθανότητες μετάβασης, θα μελετήσουμε δηλαδή αλυσίδες Markov πρώτης τάξης.

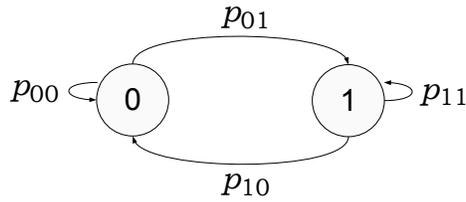
Μια αλυσίδα Markov X_t λέγεται **χρονο-αμετάβλητη** (time-homogeneous) όταν η υπό συνθήκη κατάσταση δεν επηρεάζεται από το χρόνο παρατήρησης, δηλαδή όταν

$$P(X_{t+\tau} = j | X_t = i) = P(X_\tau = j | X_0 = i),$$

και για αλυσίδες διακριτού χρόνου

$$P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i).$$

Στις χρονοαμετάβλητες αλυσίδες, οι πιθανότητες μετάβασης δεν μεταβάλλονται με την πάροδο του χρόνου. Παρακάτω θα ασχοληθούμε μόνο με χρονο-αμετάβλητες αλυσίδες Markov.



Σχήμα 6.1: Μια απλή αλυσίδα Markov διακριτού χρόνου.

6.1 Αλυσίδες Markov διακριτού χρόνου

6.1.1 Εισαγωγή

Οι αλυσίδες Markov διακριτού χρόνου (Discrete time Markov chains ή DTMC), περιγράφουν συστήματα που εξελίσσονται σε διακριτές χρονικές στιγμές $n = 0, 1, \dots$. Η κατάσταση του συστήματος στο χρόνο n (ανήκει στο σύνολο καταστάσεων $\mathcal{S} = \{0, 1, \dots, s\}$) εξαρτάται από την κατάσταση στο χρόνο $n - 1$ ενώ επιπλέον γνώση του παρελθόντος δεν μας βοηθάει. Κατά συνέπεια, είναι δυνατόν να προβλέψουμε σε ποιά κατάσταση θα βρεθούμε στην επόμενη χρονική στιγμή γνωρίζοντας μόνο την τωρινή κατάσταση και τις πιθανότητες μετάβασης που είναι ουσιαστικά οι υπό συνθήκη πιθανότητες:

$$p_{ij} \doteq \text{P}(X_{n+1} = j | X_n = i).$$

Επιπλέον, παρατηρούμε ότι οι πιθανότητες μετάβασης p_{ij} δεν εξαρτώνται από το χρόνο, κάτι που οφείλεται στο ότι η αλυσίδα είναι χρονο-αμετάβλητη.

Παράδειγμα: Έστω η αλυσίδα Markov με χώρο καταστάσεων το $\mathcal{S} = \{0, 1\}$ και πιθανότητες μετάβασης $p_{01} = p$, $p_{10} = q$, όπως στο σχήμα 6.1. Για τις πιθανότητες μετάβασης μπορούμε να γράψουμε:

$$\begin{aligned} p_{01} &\doteq \text{P}(X_{n+1} = 1 | X_n = 0) &= p \\ p_{00} &\doteq \text{P}(X_{n+1} = 0 | X_n = 0) &= 1 - p \\ p_{10} &\doteq \text{P}(X_{n+1} = 0 | X_n = 1) &= q \\ p_{11} &\doteq \text{P}(X_{n+1} = 1 | X_n = 1) &= 1 - q. \end{aligned}$$

Ας υποθέσουμε ότι τη χρονική στιγμή 0 το σύστημα βρίσκεται στην κατάσταση 0, δηλαδή $X_0 = 0$. Τότε ποιά είναι η πιθανότητα στην επόμενη χρονική στιγμή να είναι στην κατάσταση 1; Ποιά είναι η πιθανότητα τη στιγμή 2 να είναι στην κατάσταση 1;

$$\text{P}(X_1 = 1) = p_{01}\text{P}(X_0 = 0) + p_{11}\text{P}(X_0 = 1) = p_{01}1 + p_{11}0 = p$$

και

$$\text{P}(X_2 = 1) = p_{01}\text{P}(X_1 = 0) + p_{11}\text{P}(X_1 = 1) = p_{01}(1 - p) + p_{11}p = p(2 - p - q).$$

Κάθε χρονοαμετάβλητη αλυσίδα Markov διακριτού χρόνου καθορίζεται από έναν τετραγωνικό πίνακα που περιέχει τις πιθανότητες μετάβασης και ονομάζεται **στοχαστικός πίνακας μετάβασης** (stochastic transition matrix). Ο πίνακας λέγεται στοχαστικός διότι έχει την ιδιότητα ότι το άθροισμα κάθε σειράς είναι μονάδα

και τα στοιχεία του λαμβάνουν τιμές στο $[0, 1]$, με άλλα λόγια κάθε σειρά του πίνακα είναι μια ΣΜΠ. Για μια αλυσίδα με $|\mathcal{S}| = s + 1$ πιθανές καταστάσεις ορίζουμε τον $s + 1 \times s + 1$ τετραγωνικό πίνακα ως

$$\mathbf{P} = \mathbf{P}^{(1)} = [p_{ij}] = \begin{bmatrix} p_{00} & p_{01} & \dots & p_{0s} \\ p_{10} & p_{11} & \dots & p_{1s} \\ \cdot & \cdot & \dots & \cdot \\ p_{s0} & p_{s1} & \dots & p_{ss} \end{bmatrix},$$

όπου ο εκθέτης (1) θυμίζει ότι αναφερόμαστε σε πιθανότητες μετάβασης ενός βήματος. Με τον ίδιο τρόπο μπορούμε να ορίσουμε τους πίνακες $\mathbf{P}^{(2)}$ και $\mathbf{P}^{(n)}$ που περιέχουν τις πιθανότητες μετάβασης από την κατάσταση i στην κατάσταση j σε 2 και n βήματα αντίστοιχα, όπου κάθε βήμα αναφέρεται στο ότι ο χρόνος έχει αυξηθεί κατά ένα. Για την πιθανότητα n βημάτων γράφουμε

$$p_{ij}^{(n)} \doteq \mathbf{P}(X_{\ell+n} = j | X_{\ell} = i), i, j \in \mathcal{S}.$$

Από το θεώρημα ολικής πιθανότητας έχουμε ότι $\sum_j p_{ij}^{(n)} = 1$. Αν συνδυάσουμε αυτό με την ιδιότητα Markov, λαμβάνουμε τις εξισώσεις Chapman-Kolmogorov:

$$p_{ij}^{(n)} = \sum_{h \in \mathcal{S}} p_{ih}^{(m)} p_{hj}^{(n-m)}, \forall 0 < m < n, i, j \in \mathcal{S}.$$

Οι εξισώσεις Chapman-Kolmogorov αναλύουν μια πιθανότητα μετάβασης πολλαπλών βημάτων σε δύο μικρότερα βήματα στο χρόνο, όπου φυσικά πρέπει να ληφθούν υπόψη όλες οι ενδιάμεσες καταστάσεις. Για $m = 1$ λαμβάνουμε

$$p_{ij}^{(n)} = \sum_{h \in \mathcal{S}} p_{ih}^{(1)} p_{hj}^{(n-1)},$$

κάτι που σε μορφή πινάκων γράφεται και ως

$$\mathbf{P}^{(n)} = \mathbf{P}\mathbf{P}^{(n-1)} = \mathbf{P}\mathbf{P} \dots \mathbf{P} = \mathbf{P}^n.$$

Υποθέτουμε τώρα ότι τη χρονική στιγμή 0, η διαδικασία βρίσκεται σε μια από τις πιθανές καταστάσεις με πιθανότητα που περιγράφεται από μια ΣΜΠ. Θα ονομάσουμε αυτήν την ΣΜΠ **αρχική κατανομή** και θα τη συμβολίσουμε με το διάνυσμα γραμμή $\mathbf{v}(0)$ (όπου υπονοείται $v_i(0) = \mathbf{P}(X_0 = i), \forall i \in \mathcal{S}$). Με άλλα λόγια, η ΣΜΠ του X_0 είναι $\mathbf{v}(0)$. Στην γενική περίπτωση είναι προφανές ότι σε κάθε βήμα η κατανομή αλλάζει, δηλαδή η πιθανότητα το σύστημα να βρεθεί σε μια κατάσταση τροποποιείται. Στο χρόνο n το σύστημα βρίσκεται στην κατάσταση X_n που περιγράφεται από την κατανομή $\mathbf{v}(n)$. Ισχύει λόγω της ιδιότητας Markov ότι

$$\mathbf{v}(n) = \mathbf{v}(n-1)\mathbf{P} = \mathbf{v}(0)\mathbf{P}^{(n)} = \mathbf{v}(0)\mathbf{P}^n.$$

Δηλαδή, η κατανομή την χρονική στιγμή n εξαρτάται μόνο από την αρχική κατανομή, τον πίνακα μεταβάσεων και τον χρόνο n που έχει παρέλθει. Η ανάλυση με αυτόν τον τρόπο μιας αλυσίδας Markov, ονομάζεται μεταβατική ανάλυση (transient analysis) και είναι συνήθως δύσχερη.

6.1.2 Στάσιμη κατανομή - equilibrium

Ορίζουμε ως **στάσιμη κατανομή** \mathbf{v} κάθε κατανομή για την οποία ισχύει

$$\mathbf{v} = \mathbf{v}\mathbf{P}, \|\mathbf{v}\| = 1, \quad (6.1)$$

δηλαδή μια ΣΜΠ έτσι ώστε η αλυσίδα την επόμενη χρονική στιγμή να έχει την ίδια ΣΜΠ. Με άλλα λόγια, αν η αλυσίδα βρεθεί σε μια στάσιμη κατανομή την χρονική στιγμή n_0 , τότε η ΣΜΠ παραμένει σταθερή για όλες τις χρονικές στιγμές $n > n_0$. Σε μια αλυσίδα είναι δυνατόν να υπάρχουν περισσότερες από μια στάσιμες κατανομές¹.

Στη συνέχεια ορίζουμε την **οριακή κατανομή** ως το όριο

$$\tilde{\mathbf{v}} \doteq \lim_{n \rightarrow \infty} \mathbf{v}(n).$$

Φυσικά το όριο αυτό, άρα και η οριακή κατανομή, δεν ορίζονται πάντοτε. Παρατηρήστε ότι ισχύει

$$\tilde{\mathbf{v}} = \lim_{n \rightarrow \infty} \mathbf{v}(n) = \lim_{n \rightarrow \infty} \mathbf{v}(0)\mathbf{P}^{(n)} = \mathbf{v}(0) \lim_{n \rightarrow \infty} \mathbf{P}^{(n)} \doteq \mathbf{v}(0)\tilde{\mathbf{P}},$$

όπου $\tilde{\mathbf{P}}$ είναι ο οριακός πίνακας μετάβασης, δηλαδή ο πίνακας n βημάτων όπου n μεγαλώνει αυθαιρέτως πολύ ώστε ο πίνακας $\mathbf{P}^{(n)}$ να διαφέρει ελάχιστα από τον πίνακα $\mathbf{P}^{(n+1)}$. Όταν τα δύο όρια $\tilde{\mathbf{v}}$ και $\tilde{\mathbf{P}}$ υπάρχουν και δεν έχουν μηδενικά στοιχεία, τότε λέμε ότι το σύστημα επέρχεται σε ισορροπία (steady state ή equilibrium) μετά την πάροδο αρκετού χρόνου. Επιπλέον, η οριακή κατανομή $\tilde{\mathbf{v}}$ είναι στάσιμη και μοναδική. Το σύστημα εξισώσεων

$$\tilde{\mathbf{v}} = \tilde{\mathbf{v}}\tilde{\mathbf{P}}, \|\tilde{\mathbf{v}}\| = 1, .$$

έχει μοναδική λύση. Επιπρόσθετα, καθώς ο οριακός πίνακας μετάβασης δεν εξαρτάται από το χρόνο ή από την κατάσταση i , θα πρέπει να έχει τη μορφή

$$\tilde{\mathbf{P}} = \begin{bmatrix} \tilde{v}_0 & \tilde{v}_1 & \dots & \tilde{v}_s \\ \tilde{v}_0 & \tilde{v}_1 & \dots & \tilde{v}_s \\ \cdot & \cdot & \dots & \cdot \\ \tilde{v}_0 & \tilde{v}_1 & \dots & \tilde{v}_s \end{bmatrix}$$

καθώς μόνο αυτή η μορφή εγγυάται ότι για οποιαδήποτε αρχική κατάσταση, το σύστημα καταλήγει σε ισορροπία μετά από πολύ χρόνο, δηλαδή ότι $\tilde{\mathbf{v}} = \mathbf{v}(0)\tilde{\mathbf{P}}$ για κάθε $\mathbf{v}(0)$.

Παράδειγμα: Έστω η αλυσίδα του προηγούμενου παραδείγματος με πίνακα μετάβασης $\mathbf{P} = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$.

Υπάρχει οριακή κατανομή; Αν ναι, υπολογίστε την.

Λύση: Παρακάτω θα μάθουμε ιδιότητες των αλυσίδων Markov που βοηθούν στο να καθορίσουμε την ύπαρξη της οριακής κατανομής. Προς το παρόν, μπορούμε να δώσουμε μια κατασκευαστική απόδειξη. Αρχικά υπολογίζουμε μια στάσιμη κατανομή βρίσκοντας μια λύση του συστήματος

$$\begin{aligned} v_0 &= v_0 0.9 + v_1 0.2 \\ v_1 &= v_0 0.1 + v_1 0.8 \\ v_0 + v_1 &= 1 \end{aligned}$$

έστω η λύση $v_0 = \frac{2}{3}$ και $v_1 = \frac{1}{3}$. Στη συνέχεια φτιάχνουμε τον πίνακα $\mathbf{P}' = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$ και παρατηρούμε ότι $\mathbf{P}' = \lim_{n \rightarrow \infty} \mathbf{P}^n$. Δηλαδή ο πίνακας \mathbf{P}' είναι ένας οριακός πίνακας μετάβασης. Άρα, κατασκευάζοντας παραδείγματα, δείξαμε ότι υπάρχουν τα όρια κατανομής και πινάκων μετάβασης και άρα το σύστημα έχει κατανομή ισορροπίας η οποία είναι και η στάσιμη κατανομή που έχουμε βρεί $[\frac{2}{3} \ \frac{1}{3}]$. ■

Δοκιμάστε στο matlab αυτό: `stochastic_matrix.m`

¹Η πιο γενική μορφή αλυσίδας, όμως, που ονομάζεται εργοδική, έχει μοναδική στάσιμη κατανομή.

²Σε αυτήν την περίπτωση το σύστημα έχει μοναδική λύση.

6.1.3 Κατηγοριοποίηση καταστάσεων και αλυσίδων

Χαρακτηρισμός καταστάσεων

Έστω $i, j \in \mathcal{S}$ δύο καταστάσεις μιας αλυσίδας Markov. Η κατάσταση j λέγεται **προσβάσιμη από την i** αν για κάποιο n η πιθανότητα το σύστημα να μεταβεί από την i στην j σε n βήματα είναι θετική (μη μηδενική). Απαιτείται δηλαδή να ισχύει

$$P(X_{n+\ell} = j | X_\ell = i) > 0, \text{ έστω για μια επιλογή } \ell, n.$$

Σε αντίθετη περίπτωση, όταν δηλαδή οι παραπάνω πιθανότητες είναι μηδενικές για κάθε n (η επιλογή του ℓ δεν έχει σημασία ούτως ή άλλως αφού το σύστημα είναι χρονοαμετάβλητο), λέμε ότι η κατάσταση i είναι μη προσβάσιμη από την j . Έστω το σύνολο καταστάσεων $A(i)$ που περιέχει όλες τις καταστάσεις που είναι προσβάσιμες από την i , και προφανώς $A(i) \subseteq \mathcal{S}$. Τότε, αν και μόνο αν η κατάσταση i είναι προσβάσιμη από όλες τις καταστάσεις του συνόλου $A(i)$, λέμε ότι η κατάσταση i είναι **επαναληπτική** (recurrent). Η ονομασία αυτή εμπεριέχει το νόημα ότι η κατάσταση i θα είναι πάντα μια πιθανή κατάσταση που το σύστημα μπορεί να επισκεφτεί. Αν αντίθετα το σύστημα μπορεί να βρεθεί σε μια κατάσταση $k \in A(i)$ έτσι ώστε $i \notin A(k)$, τότε γνωρίζουμε ότι η κατάσταση i δεν θα είναι πλέον προσβάσιμη. Σε αυτήν την περίπτωση, η κατάσταση i ονομάζεται **μεταβατική** (transient). Το σύστημα είναι δυνατόν να βρεθεί σε μια μεταβατική κατάσταση· είναι όμως βέβαιο ότι μετά την πάροδο αρκετού χρόνου, το σύστημα θα βρεθεί σε μια από τις καταστάσεις που καθιστούν μη προσβάσιμη την μεταβατική κατάσταση και άρα το σύστημα δεν θα ξαναβρεθεί ποτέ σε αυτήν. Η μεταβατικότητα και η επαναληπτικότητα είναι ιδιότητες καταστάσεων που προσδίδονται από τον πίνακα μετάβασης. Για παράδειγμα, ένα ζευγάρι καταστάσεων (i, j) είναι επαναληπτικό αν $p_{ij} > 0$ και $p_{ji} > 0$. Όταν σκεφτόμαστε μεγαλύτερα σύνολα (ή αλλιώς κλάσεις) καταστάσεων, όμως, οι σχέσεις δεν είναι τόσο απλές. Σε πολλές αλυσίδες, πάντως, μπορούμε τελικά να χωρίσουμε το σύνολο καταστάσεων σε δύο ομάδες, μια που περιέχει όλες τις επαναληπτικές κλάσεις που επικοινωνούν μεταξύ τους και μια με τις υπόλοιπες καταστάσεις που είναι μεταβατικές.

Μια κατάσταση i λέγεται **κατάσταση απορρόφησης** (absorbing state) αν ισχύει $p_{ii} = 1$. Δηλαδή, αν το σύστημα βρεθεί σε μια τέτοια κατάσταση τότε παραμένει εκεί.

Μια κατάσταση i λέγεται **περιοδική** κατάσταση (periodic state) αν είναι προσβάσιμη μόνο σε ακέραια πολλαπλάσια χρόνου d . Δηλαδή απαιτούμε να ισχύει

$$d = \gcd\{n : P(X_n = i | X_0 = i) > 0\},$$

όπου γσδ είναι ο μέγιστος κοινός διαιρέτης. Λέμε τότε ότι η i είναι περιοδική με περίοδο d . Αν $d = 1$ τότε η κατάσταση λέγεται **απεριοδική**. Η περιοδικότητα μιας κατάστασης είναι ιδιότητα που χαρακτηρίζει μια αλυσίδα καθώς μπορεί να αποδειχθεί ότι οι καταστάσεις που επικοινωνούν έχουν όλες την ίδια περίοδο.

Χαρακτηρισμός αλυσίδων

Μια αλυσίδα λέγεται **μη μειώσιμη** (irreducible) αν όλες οι καταστάσεις της είναι προσβάσιμες από τις υπόλοιπες. Προφανώς, μια αλυσίδα που περιέχει έστω και μια κατάσταση απορρόφησης δεν είναι μη μειώσιμη.

Σε περίπτωση αλυσίδων με άπειρες αλλά αριθμήσιμες καταστάσεις (countably infinite) ο μέσος χρόνος επαναληπτικότητας πρέπει να ληφθεί υπόψη για να κατηγοριοποιηθεί μια κατάσταση ως προς την επαναληπτικότητα της. Έστω $f_i^{(n)}$ η πιθανότητα της πρώτης επανάληψης μετά από n βήματα. Τότε η συνολική πιθανότητα επανάληψης είναι

$$f_i = \sum_{n=1}^{\infty} f_i^{(n)}.$$

Μια κατάσταση i λέγεται επαναληπτική (επανάληψη του παραπάνω ορισμού) όταν $f_i = 1$. Αν $f_i < 1$ τότε λέγεται μεταβατική. Ο μέσος χρόνος επαναληπτικότητας μπορεί να οριστεί ως

$$m_i = \sum_{n=1}^{\infty} n f_i^{(n)}.$$

Όταν ο μέσος αυτός χρόνος είναι πεπερασμένος ($m_i < \infty$) τότε η κατάσταση ονομάζεται **θετικά επαναληπτική** (positive recurrent ή recurrent non-null). Αντίθετα, αν $m_i = \infty$, τότε ονομάζεται **μηδενικά επαναληπτική** (recurrent null) και αυτό σημαίνει ότι ενώ οι μεταβάσεις επιτρέπουν την επίσκεψη σε αυτή την κατάσταση, αυτό αναμένεται να γίνει πολύ αργά, τόσο αργά που πρακτικά το σύστημα δεν θα την επισκεφτεί ποτέ. Μια αλυσίδα ονομάζεται θετικά επαναληπτική αν όλες οι καταστάσεις της είναι θετικά επαναληπτικές.

Ορισμός Μια DTMC λέγεται **εργοδική** όταν είναι απεριοδική, μη μειώσιμη και θετικά επαναληπτική.

6.1.4 Βασικά θεωρήματα για τις αλυσίδες Markov διακριτού χρόνου

Θεώρημα 4 Οι καταστάσεις μιας μη μειώσιμης DTMC είναι όλες ίδιες (περιοδικές ή απεριοδικές, μεταβατικές, μηδενικά επαναληπτικές ή θετικά επαναληπτικές).

Θεώρημα 5 Όταν μια DTMC είναι απεριοδική τότε η οριακή κατανομή $\tilde{\mathbf{v}}$ ορίζεται και το σύστημα έχει κατανομή ισοροπίας.

Θεώρημα 6 Όταν μια DTMC είναι απεριοδική και μη μειώσιμη, τότε η οριακή κατανομή $\tilde{\mathbf{v}}$ ορίζεται και δεν εξαρτάται από την αρχική κατανομή $\mathbf{v}(0)$.

Θεώρημα 7 Όταν μια DTMC είναι εργοδική τότε η οριακή κατανομή $\tilde{\mathbf{v}}$ ορίζεται και είναι μοναδική. Η λύση δίνεται από το σύστημα

$$\begin{aligned} \mathbf{v} &= \mathbf{vP} \\ \sum_{i=0}^s v_i &= 1. \end{aligned}$$

Παράδειγμα: (Υπολογισμός στάσιμης κατανομής) Δίνεται ο παρακάτω πίνακας μετάβασης μιας Μαρκοβιανής αλυσίδας διακριτού χρόνου τριών καταστάσεων:

$$P = \begin{bmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}.$$

Ικανοποιούνται οι συνθήκες για την ύπαρξη οριακών πιθανοτήτων; Εξηγήστε. Μακροπρόθεσμα, τι ποσοστό του χρόνου βρίσκεται η αλυσίδα σε κάθε μία από τις τρεις καταστάσεις;

Λύση: Η Μαρκοβιανή αλυσίδα είναι μη μειώσιμη, απεριοδική και θετικά επαναληπτική, άρα και εργοδική. Τα ποσοστά του χρόνου δίνονται από τις οριακές πιθανότητες v_0, v_1, v_2 , που προκύπτουν λύνοντας το σύστημα εξισώσεων

$$v_j = \sum_{i=0}^2 v_i p_{ij}, \quad j = 0, 1, 2$$

μαζί με τη σχέση $\sum_{i=0}^2 v_i = 1$. Έχουμε:

$$\begin{cases} v_0 = 0.5v_0 + 0.3v_1 + 0.2v_2, \\ v_1 = 0.4v_0 + 0.4v_1 + 0.3v_2, \\ v_2 = 0.1v_0 + 0.3v_1 + 0.5v_2, \\ v_0 + v_1 + v_2 = 1. \end{cases}$$

Οποιοσδήποτε δύο από τις τρεις πρώτες και η τελευταία δίνουν τη λύση

$$v_0 = \frac{21}{62}, \quad v_1 = \frac{23}{62}, \quad v_2 = \frac{9}{31}.$$

■

6.1.6 Μεταβατική ανάλυση - Χρόνοι πρώτης άφιξης

Σε αυτή την ενότητα μας ενδιαφέρει η ανάλυση μονοπατιών σε μια DTMC. Αρχικά, ορίζουμε ως **χρόνο ταξιδιού** (sojourn time) T_i την τ.μ. που εκφράζει το χρόνο που παραμένει το σύστημα στην κατάσταση i κάθε φορά που την επισκέπτεται. Η τ.μ. είναι προφανώς γεωμετρική με πιθανότητα επιτυχίας $1 - p_{ii}$ και κατά συνέπεια για τις δύο πρώτες ροπές της γνωρίζουμε ότι

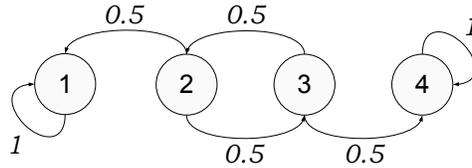
$$\begin{aligned} E[T_i] &= \frac{1}{1 - p_{ii}} \\ \text{Var}(T_i) &= \frac{p_{ii}}{(1 - p_{ii})^2}. \end{aligned}$$

Στη συνέχεια, ορίζουμε ως **χρόνο πρώτης άφιξης** (hitting time) μιας κατάστασης i ξεκινώντας από την κατάσταση j την τ.μ.

$$K_j(i) \doteq \inf\{n : X_n = i, X_0 = j\},$$

με την παραδοχή ότι το infimum ενός κενού συνόλου είναι το ∞ . Για αρμονία με την προηγούμενη παράγραφο αναφέρεται ότι $P(K_i(i) = n) = f_i^{(n)}$. Με όμοιο τρόπο μπορούμε να ορίσουμε τον χρόνο πρώτης άφιξης σε ένα σύνολο καταστάσεων $A \subseteq \mathcal{S}$ ξεκινώντας από την κατάσταση j .

$$K_j(A) \doteq \inf\{n : X_n \in A, X_0 = j\},$$



Σχήμα 6.3: Αλυσίδα Markov διακριτού χρόνου με καταστάσεις απορρόφησης.

Επίσης, μπορούμε να ορίσουμε την προσβασιμότητα μιας κατάστασης i (ή ενός συνόλου καταστάσεων A) από την κατάσταση j εκ νέου χρησιμοποιώντας την πιθανότητα

$$h_j(i) = P(K_j(i) < \infty)$$

και

$$h_j(A) = P(K_j(A) < \infty).$$

Συχνά μας ενδιαφέρει ο μέσος χρόνος πρώτης άφιξης $k_j(i) \doteq E[K_j(i)]$ και έχουμε

$$k_j(i) \doteq E[K_j(i)] = \sum_{n < \infty} nP(K_j(i) = n) + \infty(1 - h_j(i))$$

και

$$k_j(A) \doteq E[K_j(A)] = \sum_{n < \infty} nP(K_j(A) = n) + \infty(1 - h_j(A)).$$

Για να είναι πεπερασμένος ο μέσος χρόνος πρώτης άφιξης, προφανώς απαιτούμε ότι $h_j(i) = 1$ ή $h_j(A) = 1$. Το σημαντικό αποτέλεσμα είναι ότι μπορούμε πάντα να υπολογίζουμε τους μέσους χρόνους πρώτης άφιξης και την πιθανότητα απορρόφησης σε μια κατάσταση αν γνωρίζουμε το πίνακα μετάβασης \mathbf{P} λύνοντας ένα απλό σύστημα γραμμικών εξισώσεων.

Παράδειγμα: (Υπολογισμός μέσων χρόνων πρώτης άφιξης και πιθανοτήτων απορρόφησης) Στην αλυσίδα του σχήματος 6.3, και δεδομένου ότι $\mathbf{v}(0) = [0 \ 1 \ 0 \ 0]$, ζητείται η πιθανότητα να βρεθεί το σύστημα στην κατάσταση 4 και το μέσο χρόνο πρώτης άφιξης σε κατάσταση απορρόφησης.

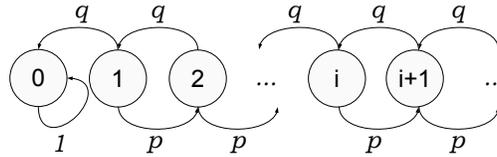
Λύση: Αρχικά, μπορούμε να παρατηρήσουμε ότι $h_1(4) = 0$, $h_4(4) = 1$, $k_4(\{1, 4\}) = k_1(\{1, 4\}) = 0$. Για τις πιθανότητες απορρόφησης, χρησιμοποιώντας τις εξισώσεις Chapman-Kolmogorov μπορούμε να γράψουμε το παρακάτω γραμμικό σύστημα εξισώσεων

$$\begin{cases} h_2(4) = 0.5h_3(4) + 0.5h_1(4) \\ h_3(4) = 0.5h_2(4) + 0.5h_4(4) \end{cases}$$

από το οποίο προκύπτει ότι $h_2(4) = \frac{1}{3}$. Στη συνέχεια, προσθέτοντας σε κάθε βήμα χρόνο 1, μπορούμε να κατασκευάσουμε ένα αντίστοιχο σύστημα εξισώσεων για τους μέσους χρόνους πρώτης άφιξης. Θέτουμε πρώτα $A = \{1, 4\}$ και έχουμε

$$\begin{cases} k_2(A) = 0.5k_3(A) + 0.5k_1(A) + 1 \\ k_3(A) = 0.5k_2(A) + 0.5k_4(A) + 1 \end{cases}$$

από το οποίο προκύπτει ότι $k_2(A) = 2$. ■



Σχήμα 6.4: DTMC για το πρόβλημα καταστροφής του τζογαδόρου.

Παράδειγμα: (Η καταστροφή του τζογαδόρου - Gambler's ruin) Στην αλυσίδα του σχήματος 6.4 έχουμε πιθανότητες μετάβασης $p_{0,0} = 1$, $p_{i,i-1} = q$ και $p_{i,i+1} = p$ για $i = 1, 2, \dots$. Ένας τζογαδόρος με περιουσία i ευρώ μπαίνει σε ένα καζίνο. Σε κάθε στοιχείο παίζει ένα ευρώ και είτε κερδίζει με πιθανότητα p , είτε χάνει με πιθανότητα $q = 1 - p$. Ποιά είναι η πιθανότητα να φύγει χάνοντας όλη την περιουσία του;

Λύση: Έστω $h_i(0)$ η ζητούμενη πιθανότητα. Η λύση του προβλήματος θα είναι η μικρότερη³ μη αρνητική λύση του συστήματος

$$\begin{cases} h_0(0) = 1 \\ h_i(0) = ph_{i+1}(0) + qh_{i-1}(0), \text{ για } i = 1, 2, \dots \end{cases}$$

Η λύση αυτού του συστήματος αναδρομής είναι απαιτητική. Δίνεται στις σελίδες 57–58 του [6] και είναι

$$h_i(0) = \begin{cases} A + B \left(\frac{q}{p}\right)^i & \text{αν } q \neq p \\ A + Bi & \text{αν } q = p, \end{cases}$$

όπου A, B σταθερές. Στη συνέχεια διακρίνουμε τις εξής περιπτώσεις:

- ($p < q$) που είναι συνήθως η περίπτωση ενός πραγματικού καζίνο. Η λύση της αναδρομής έχει την μορφή $h_i(0) = A + B \left(\frac{q}{p}\right)^i$. Λόγω του $0 \leq h_i(0) \leq 1$, θα πρέπει $B = 0$ και άρα $h_i(0) = 1$ για κάθε i .
- ($p > q$) η λύση της αναδρομής είναι της ίδιας μορφής. Συγκεκριμένα, θα πρέπει $B = 1 - A$, οπότε λαμβάνουμε

$$h_i(0) = \left(\frac{q}{p}\right)^i + A \left(1 - \left(\frac{q}{p}\right)^i\right).$$

Επιλέγοντας το $A = 0$ εγγυάται την μικρότερη μη αρνητική λύση και άρα $h_i(0) = \left(\frac{q}{p}\right)^i$.

- ($p = q$) η λύση της αναδρομής είναι της μορφής $h_i(0) = A + Bi$. Και πάλι λόγω των περιορισμών προκύπτει $h_i(0) = 1$.

Άρα, η μόνη ελπίδα του τζογαδόρου είναι να έχει μεγαλύτερη πιθανότητα να κερδίσει από το καζίνο, στην οποία περίπτωση μπορεί να αποφύγει τη χρεωκοπία. Ενδιαφέρον παρουσιάζουν άλλες μορφές του προβλήματος όπου ο στόχος του τζογαδόρου είναι να αποκτήσει έναν αριθμό N ευρώ. ■

³Μικρότερη σε αυτή την περίπτωση είναι μια λύση που κάθε στοιχείο της είναι μικρότερο από τα στοιχεία μιας άλλης λύσης. Το ότι πρέπει να διαλέξουμε την μικρότερη λύση προκύπτει από το θεώρημα 1.3.2 στο [6].

6.2 Αλυσίδες Markov συνεχούς χρόνου

6.2.1 Εισαγωγή

Μια στοχαστική διαδικασία συνεχούς χρόνου και διακριτού χώρου παραμέτρου X_t λέγεται αλυσίδα Markov συνεχούς χρόνου (Continuous time Markov chain ή CTMC) αν ικανοποιεί την σχέση 4.1. Μια CTMC έχει διακριτό χώρο καταστάσεων \mathcal{S} (όπως και η DTMC), όμως οι μεταβάσεις μεταξύ των καταστάσεων γίνονται σε ακαθόριστα χρονικά σημεία (όχι κάθε 1 μονάδα χρόνου). Όταν η X_t είναι επιπλέον χρονοαμετάβλητη, λόγω της έλλειψης μνήμης, η διάρκεια που το σύστημα παραμένει σε μια κατάσταση επιβάλλεται να είναι κατανομημένη εκθετικά, καθώς η εκθετική κατανομή είναι η μόνη συνεχής κατανομή με αυτήν την ιδιότητα. Έστω T_i ο χρόνος παραμονής στην κατάσταση i . Εφόσον η CTMC είναι χρονοαμετάβλητη, θα ισχύει

$$P(T_i > x + t | T_i > t) \stackrel{\text{Markov}}{=} P(T_i > x),$$

δηλαδή η τ.μ. T_i έχει την ιδιότητα απώλειας μνήμης, άρα είναι εκθετικά κατανομημένη.

Οι **πιθανότητες μετάβασης** εξαρτώνται πλέον από το χρονικό πλαίσιο που τις παρατηρούμε. Αρχικά ορίζουμε

$$p_{ij}(t, t + \tau) \doteq P(X_{t+\tau} = j | X_t = i),$$

που λόγω της ιδιότητας μη μεταβολής με το χρόνο τροποποιείται σε

$$p_{ij}(\tau) \doteq P(X_\tau = j | X_0 = i),$$

όπου τ είναι το χρονικό διάστημα στο οποίο ορίζουμε την πιθανότητα μετάβασης. Επίσης, ορίζουμε ότι $p_{ij}(0) = 1$, αν $i = j$ και $p_{ij}(0) = 0$, αν $i \neq j$. Αν συλλέξουμε τις πιθανότητες μετάβασης σε έναν πίνακα λαμβάνουμε τον πίνακα μετάβασης $\mathbf{P}(\tau) \doteq \{p_{ij}(\tau)\}$ που εξαρτάται από τα χρονικά διαστήματα τ και έχουμε ότι $\mathbf{P}(0) = \mathbf{I}$. Οι **πιθανότητες κατάστασης** ορίζονται όπως και στην DTMC, τις συμβολίζουμε, όμως, με διαφορετικό γράμμα προς αποφυγή σύγχυσης

$$\pi_i(\tau) \doteq P(X_\tau = i),$$

και τις συλλέγουμε στο διάνυσμα $\boldsymbol{\pi}$. Λόγω της ιδιότητας Markov μπορούμε να γράψουμε

$$\pi_j(\tau) = p_{ij}(\tau)\pi_i(0),$$

όπου $\pi_i(0)$ είναι η αρχική κατανομή κατάστασης.

6.2.2 Κατηγοριοποίηση CTMC

Όπως δείξαμε παραπάνω, ο χρόνος παραμονής της διαδικασίας σε μια κατάσταση είναι εκθετικά κατανομημένος. Όταν ο χρόνος αυτός παρέλθει, το σύστημα πρέπει να μεταβεί σε μια νέα κατάσταση (διάφορη της προηγούμενης) με κάποια πιθανότητα. Αυτή η παρατήρηση μας επιτρέπει να δούμε την CTMC από άλλη σκοπιά. Μπορούμε να υποθέσουμε μια υποκείμενη DTMC με πιθανότητες μετάβασης \tilde{p}_{ij} η οποία επιλέγει τις μεταβάσεις όταν οι χρόνοι παραμονής λήξουν με τον περιορισμό ότι $p_{ii} = 0$. Η DTMC αυτή ονομάζεται **ενσωματωμένη** (embedded Markov chain) και αποδεικνύεται ότι η CTMC κληρονομεί όλες τις ιδιότητες κατηγοριοποίησης από αυτήν.

6.2.3 Εξισώσεις Chapman-Kolmogorov για τις CTMC

Σε πλήρη αντιστοιχία με τις DTMC, χρησιμοποιώντας τον νόμο της ολικής πιθανότητας καταλήγουμε στις εξισώσεις Chapman-Kolmogorov για τις CTMC

$$p_{ij}(\tau) = \sum_{k \in \mathcal{S}} p_{ik}(w)p_{kj}(\tau - w), \text{ για κάθε } 0 < w < \tau.$$

Δυστυχώς, οι εξισώσεις αυτές λόγω της εξάρτησης του χρόνου δεν επιδέχονται άμεση λύση όπως πριν. Για το λόγο αυτό τις μετατρέπουμε σε διαφορικές εξισώσεις παραγωγίζοντας ως προς το χρόνο. Αρχικά ορίζουμε τους ρυθμούς

$$q_{ij} = \lim_{\tau \rightarrow 0} \frac{p_{ij}(\tau) - \delta_{ij}}{\tau}, \quad i \neq j,$$

$$q_{ii} = \lim_{\tau \rightarrow 0} \frac{p_{ii}(\tau) - 1}{\tau},$$

οι οποίοι ονομάζονται **ρυθμοί μετάβασης**. Τα q_{ij} περιγράφουν το ρυθμό μεταβάσεων (εξόδου από το i) προς το j ενώ τα q_{ii} περιγράφουν το ρυθμό εισόδου στο i . Οι ρυθμοί μετάβασης έχει αποδειχθεί ότι ορίζονται υπό πολύ γενικές συνθήκες (πέραν του σκοπού του συγγραμματος). Εφόσον ισχύει πάντα $\sum_{j \in \mathcal{S}} p_{ij}(t) = 1$ για κάθε t , προκύπτει άμεσα ότι

$$\sum_{j \in \mathcal{S}} q_{ij} = 0.$$

Επίσης, μπορούμε να συλλέξουμε τους ρυθμούς μετάβασης σε έναν πίνακα \mathbf{Q} που λέγεται **απειροστικός πίνακας γεννήτρια** (infinitesimal Generator matrix). Ο πίνακας αυτός έχει τις εξής ιδιότητες

- $0 \leq -q_{ii} < \infty, \forall i$
- $q_{ij} \geq 0, \forall i \neq j$
- $\sum_{j \in \mathcal{S}} q_{ij} = 0, \forall i$

και η μορφή του για $|\mathcal{S}| = s$ είναι η εξής

$$\mathbf{Q} = \begin{bmatrix} -\sum_{j \neq 1} q_{1j} & q_{12} & \cdots & q_{1s} \\ q_{21} & -\sum_{j \neq 2} q_{2j} & \cdots & q_{2s} \\ \vdots & \vdots & \cdots & \vdots \\ q_{sj} & q_{s2} & \cdots & -\sum_{j \neq s} q_{sj} \end{bmatrix}.$$

Ο πίνακας αυτός είναι ανεξάρτητος του χρονικού πλαισίου και περιγράφει πλήρως μια CTMC. Από τα παραπάνω καταλήγουμε ότι

$$\mathbf{Q} \doteq \lim_{\tau \rightarrow 0} \frac{\mathbf{P}(\tau) - \mathbf{I}}{\tau}.$$

Επίσης, μπορούμε να γράψουμε τις εξισώσεις Chapman-Kolmogorov σε μορφή πινάκων, οπότε θα έχουμε

$$\mathbf{P}(\tau + \Delta\tau) = \mathbf{P}(\tau)\mathbf{P}(\Delta\tau) = \mathbf{P}(\Delta\tau)\mathbf{P}(\tau)$$

και να τις χρησιμοποιήσουμε για να δείξουμε ότι

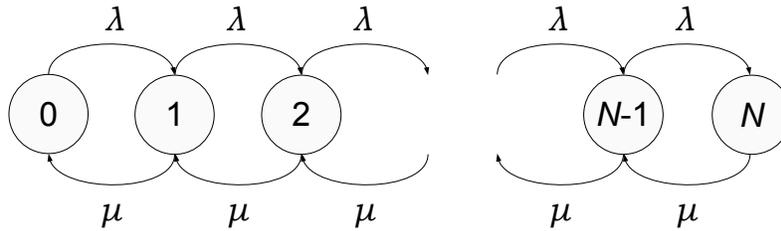
$$\frac{d\mathbf{P}(\tau)}{d\tau} = \lim_{\Delta\tau \rightarrow 0} \frac{\mathbf{P}(\tau + \Delta\tau) - \mathbf{P}(\tau)}{\Delta\tau} = \mathbf{P}(\tau) \lim_{\Delta\tau \rightarrow 0} \frac{\mathbf{P}(\Delta\tau) - \mathbf{I}}{\Delta\tau} = \lim_{\Delta\tau \rightarrow 0} \frac{\mathbf{P}(\Delta\tau) - \mathbf{I}}{\Delta\tau} \mathbf{P}(\tau)$$

και τελικά καταλήγουμε

$$\frac{d\mathbf{P}(\tau)}{d\tau} = \mathbf{P}(\tau)\mathbf{Q}$$

ή

$$\frac{d\mathbf{P}(\tau)}{d\tau} = \mathbf{Q}\mathbf{P}(\tau)$$



Σχήμα 6.5: Μια απλή αλυσίδα Markov συνεχούς χρόνου.

όπου η πρώτη εξίσωση ονομάζεται forward Chapman-Kolmogorov equation και η δεύτερη backward Chapman-Kolmogorov equation. Οι εξισώσεις αυτές γράφονται επίσης με τη μορφή

$$\frac{d\pi_j(t)}{dt} = \sum_i q_{ij}\pi_i(t). \quad (6.2)$$

Παράδειγμα: Για την CTMC του σχήματος 6.5 κατασκευάστε τον απειροστικό πίνακα γεννήτρια.

Λύση: Θα είναι

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 & \dots & 0 & 0 \\ \mu & -\lambda - \mu & \lambda & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ 0 & 0 & 0 & \dots & -\lambda - \mu & \lambda \\ 0 & 0 & 0 & \dots & \mu & -\mu \end{bmatrix}.$$

■

6.2.4 Στάσιμη ανάλυση

Ως στάσιμη κατανομή ορίζουμε την κατανομή $\tilde{\pi}$ για την οποία αν υπάρχει t έτσι ώστε $\pi(t) = \tilde{\pi}$, τότε για κάθε $t_1 > t$ θα έχουμε $\pi(t_1) = \tilde{\pi}$. Κατά συνέπεια η χρονική παράγωγος θα είναι ίση με μηδέν και θα ισχύει

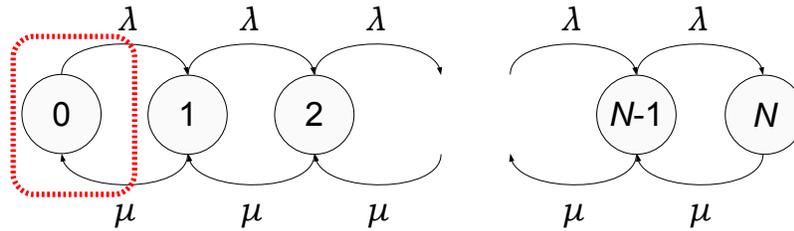
$$0 = \sum_i q_{ij}\tilde{\pi}_i, \forall j.$$

Οι εξισώσεις αυτές ονομάζονται **εξισώσεις ισορροπίας** (balance equations). Επίσης, θα πρέπει πάντα να ισχύει $\sum_i \tilde{\pi}_i = 1$.

Μια εναλλακτική μορφή των εξισώσεων ισορροπίας λαμβάνουμε αν ακολουθήσουμε τον παρακάτω συλλογισμό. Ισχύει $q_{jj} = -\sum_{i \neq j} q_{ji}$ για κάθε j . Άρα καταλήγουμε στο

$$\begin{aligned} \sum_i q_{ij}\tilde{\pi}_i &= 0 \Leftrightarrow \\ \sum_{i \neq j} q_{ij}\tilde{\pi}_i &= \tilde{\pi}_j \sum_{i \neq j} q_{ji}. \end{aligned}$$

Οι εξισώσεις ισορροπίας λένε ότι ο ζυγισμένος ρυθμός που μπαίνει και βγαίνει γύρω από μια κατάσταση πρέπει να είναι μηδέν. Ισοδύναμα μπορούμε να σκεφτούμε ότι χαράσσοντας μια καμπύλη που χωρίζει την αλυσίδα σε δύο μέρη (όπως στο σχήμα 6.6) θα πρέπει ο συνολικός ρυθμός μετάβασης (θετικών ροών που εισέρχονται



Σχήμα 6.6: Σε κάθε καμπύλη που χωρίζει την αλυσίδα σε δύο μέρη, στην κατάσταση ισορροπίας οι ζυγισμένοι ρυθμοί μετάβασης θα πρέπει να αθροίζονται σε μηδέν.

και αρνητικών που εξέρχονται) ζυγισμένος με τις στάσιμες πιθανότητες να είναι ίσος με μηδέν. Αυτή η αρχή ισορροπίας γενικεύεται και μας επιτρέπει να υπολογίζουμε την κατανομή στην κατάσταση ισορροπίας χαράσσοντας οποιεσδήποτε καμπύλες επιθυμούμε.

Παράδειγμα: Για την αλυσίδα του σχήματος 6.5 υποθέστε ότι υπάρχει στάσιμη κατάσταση και υπολογίστε την.

Λύση: Λαμβάνοντας την καμπύλη του σχήματος 6.6, εύκολα υπολογίζουμε ότι

$$\begin{aligned} \lambda \tilde{\pi}_0 &= \mu \tilde{\pi}_1 \Leftrightarrow \\ \tilde{\pi}_1 &= \frac{\lambda}{\mu} \tilde{\pi}_0. \end{aligned}$$

Στη συνέχεια μετακινούμε την καμπύλη έτσι ώστε να περιέχει τις καταστάσεις 0 και 1. Τότε λαμβάνουμε

$$\begin{aligned} \lambda \tilde{\pi}_1 &= \mu \tilde{\pi}_2 \Leftrightarrow \\ \tilde{\pi}_2 &= \frac{\lambda}{\mu} \tilde{\pi}_1 \Leftrightarrow \\ \tilde{\pi}_2 &= \left(\frac{\lambda}{\mu}\right)^2 \tilde{\pi}_0. \end{aligned}$$

Συνεχίζοντας με όμοιο τρόπο καταλήγουμε ότι

$$\tilde{\pi}_i = \left(\frac{\lambda}{\mu}\right)^i \tilde{\pi}_0, i = 1, \dots, N. \quad (6.3)$$

Τελικά θα έχουμε

$$\begin{aligned} \sum_i \tilde{\pi}_i &= 1 \Leftrightarrow \\ \tilde{\pi}_0 + \tilde{\pi}_1 + \dots + \tilde{\pi}_N &= 1 \Leftrightarrow \\ \tilde{\pi}_0 + \left(\frac{\lambda}{\mu}\right) \tilde{\pi}_0 + \dots + \left(\frac{\lambda}{\mu}\right)^N \tilde{\pi}_0 &= 1 \Leftrightarrow \\ \tilde{\pi}_0 &= \frac{1}{1 + \sum_{i=1}^N \left(\frac{\lambda}{\mu}\right)^i} \end{aligned}$$

το οποίο υπό τον όρο $\lambda \neq \mu$ και χρησιμοποιώντας το γεωμετρικό άθροισμα, μετατρέπεται σε

$$\tilde{\pi}_0 = \frac{1 - \frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{\mu}\right)^{N+1}}$$

το οποίο μαζί με την (6.3) δίνουν την πλήρη λύση στο πρόβλημα.

Για την περίπτωση $\lambda = \mu$, έχουμε ότι $\tilde{\pi}_0 = \tilde{\pi}_1 = \dots = \tilde{\pi}_N = \frac{1}{N+1}$, δηλαδή το σύστημα βρίσκεται σε όλες τις καταστάσεις με την ίδια πιθανότητα. Αυτή η περίπτωση ονομάζεται **κρίσιμη κατάσταση**. ■

6.2.5 Μεταβατική ανάλυση

Γράφοντας την εξίσωση (6.2) σε διανυσματική μορφή λαμβάνουμε

$$\frac{d\boldsymbol{\pi}(t)}{dt} = \boldsymbol{\pi}(t)\mathbf{Q}.$$

Η λύση αυτή της διαφορικής εξίσωσης είναι η γενίκευση της απλής μορφής σε διανυσματική

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)e^{\mathbf{Q}t}$$

όπου $e^{\mathbf{Q}t} \doteq \sum_{k=0}^{\infty} \frac{(\mathbf{Q}t)^k}{k!}$ ονομάζεται εκθετικός πίνακας (matrix exponential). Δεδομένης της αρχικής κατάστασης και του πίνακα γεννήτρια, μπορούμε λοιπόν να υπολογίσουμε την κατανομή σε κάθε χρονική στιγμή. Ο υπολογισμός αυτός είναι, όμως, ιδιαίτερα περίπλοκος ήδη για μικρό αριθμό καταστάσεων και συνήθως επιστρατεύονται αριθμητικές μέθοδοι.

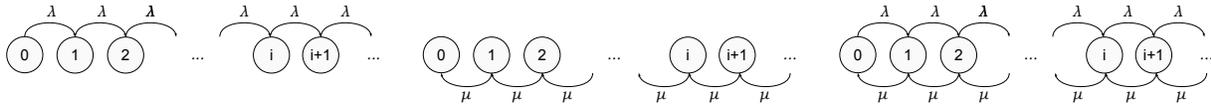
6.2.6 Διαδικασία Γεννήσεων-Θανάτων

Μια αλυσίδα CTMC, ονομάζεται **διαδικασία γεννήσεων** (birth process) αν έχει την ιδιότητα ότι κάθε μετάβαση αφορά μόνο δύο γειτονικές καταστάσεις έτσι ώστε $P(X(t + \Delta t) - X(t) > 1) = 0$ για αυθαίρετα μικρό Δt . Όμοια, ονομάζεται **διαδικασία θανάτων** (death process) αν έχει την ιδιότητα ότι κάθε μετάβαση αφορά μόνο δύο γειτονικές καταστάσεις έτσι ώστε $P(X(t + \Delta t) - X(t) < -1) = 0$ για αυθαίρετα μικρό Δt . Τέλος, διαδικασίες **γεννήσεων-θανάτων** (birth-death processes) ονομάζονται εκείνες για τις οποίες έχουμε $P(|X(t + \Delta t) - X(t)| > 1) = 0$ για αυθαίρετα μικρό Δt . Αυτό σημαίνει ότι σε μια τέτοια διαδικασία οι προσαυξήσεις γίνονται μια τη φορά. Δεν είναι δυνατόν, δηλαδή, να έχουμε μεταβάσεις που αυξάνουν/μειώνουν την κατάσταση (ή τον πληθυσμό) κατά περισσότερο από 1 τη φορά. Δηλαδή έχουμε

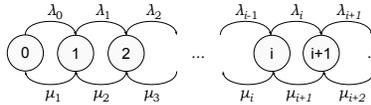
$$p_{i,j}(t) = \begin{cases} > 0 & j = i + 1 \\ = 0 & j \neq i + 1, \end{cases}$$

για κάθε t . Αν επιπλέον υποθέσουμε ότι οι ρυθμοί μετάβασης παραμένουν σταθεροί σε κάθε κατάσταση, τότε οι γεννήσεις θα γίνονται σύμφωνα με μια διαδικασία Poisson. Στο σχήμα 6.7 φαίνονται παραδείγματα τέτοιων διαδικασιών.

Οι διαδικασίες γεννήσεων-θανάτων είναι πολύτιμες για την μοντελοποίηση πληθυσμών. Για παράδειγμα, σε ένα δίκτυο επικοινωνιών, μπορούμε να θεωρήσουμε ότι ο πληθυσμός των χρηστών που ζητούν υπηρεσία είναι μια τέτοια διαδικασία.



Σχήμα 6.7: Διαδικασίες γεννήσεων, θανάτων και γεννήσεων-θανάτων.



Σχήμα 6.8: Διαδικασία γεννήσεων-θανάτων με μεταβλητούς ρυθμούς.

Θεωρήστε την διαδικασία γεννήσεων-θανάτων του σχήματος 6.8. Για να βρούμε τη στάσιμη κατανομή σε αυτήν την αλυσίδα χρησιμοποιούμε τις εξισώσεις ισορροπίας

$$\begin{aligned}\lambda_0 \pi_0 &= \mu_1 \pi_1 \\ \lambda_j \pi_j - \mu_{j+1} \pi_{j+1} &= \lambda_{j-1} \pi_{j-1} - \mu_j \pi_j, \forall j > 0.\end{aligned}$$

από το οποίο παίρνουμε ότι

$$\pi_j = \pi_0 \prod_{i=1}^j \frac{\lambda_i}{\mu_{i+1}}$$

και αθροίζοντας στη μονάδα την κατανομή, λαμβάνουμε τελικά

$$\pi_j = \frac{\prod_{i=1}^j \frac{\lambda_i}{\mu_{i+1}}}{1 + \sum_{k \in \mathcal{S} \setminus \{0\}} \prod_{i=1}^k \frac{\lambda_i}{\mu_{i+1}}}$$

Όταν η αλυσίδα έχει στάσιμη κατανομή, τότε το παραπάνω άθροισμα συγκλίνει.

Μερικά πολύ σημαντικά παραδείγματα είναι:

- Το μοντέλο ουράς $M/M/1$ (Αφίξεις Poisson και εκθετικοί χρόνοι εξυπηρέτησης) όταν $\lambda_j = \lambda$ και $\mu_j = \mu$ για κάθε j .
- Το μοντέλο ουράς $M/M/c$ (Όταν υπάρχουν c εξυπηρετητές) όταν $\lambda_j = \lambda$ για κάθε j και $\mu_j = \min\{j, c\}\mu$.
- Το μοντέλο πληθυσμιακής ανάπτυξης $\lambda_j = j\lambda + a$ και $\mu_j = j\mu$ για κάθε j , όπου το a εκφράζει ένα ποσό ενδημίας από ξένα κράτη.

Βιβλιογραφία

- [1] Robert G. Gallager, “Circularly-Symmetric Gaussian random vectors”, <http://www.rle.mit.edu/rgallager/documents/CircSymGauss.pdf>.
- [2] Jean Walrand, “Lecture Notes on Probability Theory and Random Processes”, Department of Electrical Engineering and Computer Sciences. University of Berkeley, 2004.
- [3] Dimitri P. Bertsekas and John N. Tsitsiklis, Introduction to Probability (2nd edition), Athena Scientific, July 2008.
- [4] Alberto Leon-Garcia, Probability, Statistics and Random Processes for Electrical Engineering (third edition), Pearson International Edition, Prentice Hall, 2009.
- [5] Gunter Bolch, Stefan Greiner, Hermann de Meer and Kishor S. Trivedi, Queueing Networks and Markov Chains, Wiley Interscience, 2006.
- [6] J. R. Norris, Markov Chains, Cambridge Series in Statistical and Probabilistic Mathematics, 1997.
- [7] Sheldon Ross, Applied Probability Models with Optimization Applications, Dover Publications, 1970.
- [8] Samuli Aalto, Lectures for Introduction to Teletraffic Theory, <http://www.netlab.tkk.fi/samuli/courses.shtml>
- [9] Bruce Hajek, Lectures for An Exploration of Random Processes for Engineers, <http://www.ifp.illinois.edu/hajek/Papers/randomprocesses.html>
- [10] M. Matsumoto and T. Nishimura, Mersenne Twister Home Page, <http://www.math.sci.hiroshima-u.ac.jp/m-mat/MT/emt.html>