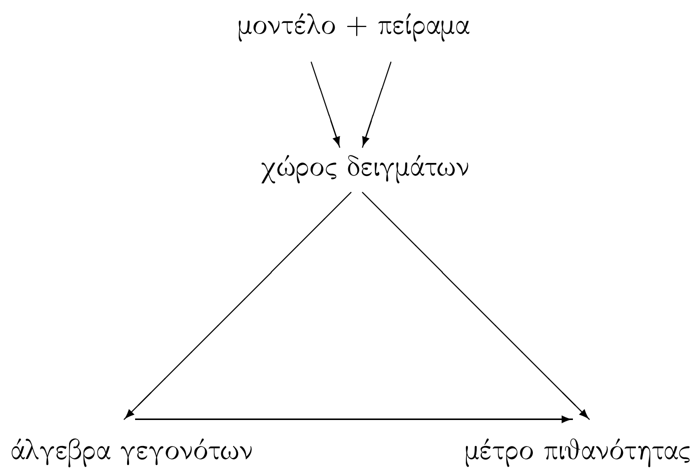
Κεφάλαιο 1

Πιθανότητες

Το διάγραμμα που ακολουθεί περιγράφει την σχέση των θεμελιωδών εννοιών των πιθανοτήτων:



- Χώρος δειγμάτων ή δειγματοχώρος (sample space): Όλα τα δυνατά αποτελέσματα ενός πειραματικού μοντέλου *σε πλήρη λεπτομέρεια* και *χωρίς καμία επικάλυψη* (στο χώρο).

- Άλγεβρα γεγονότων (algebra of events): Γεγονότα ή ενδεχόμενα (events) είναι ένα σύνολο από σημεία ή περιοχές σε ένα (συνήθως δισδιάστατο) χώρο. Η άλγεβρα γεγονότων είναι αξιωματικά ορισμένη (7 αξιώματα). Απεικόνιση χρησιμοποιώντας διαγράμματα Venn. Τα ενδεχόμενα περιέχουν (κανένα), ένα ή περισσότερα δείγματα του δειγματοχώρου.

- Μέτρο πιθανότητας (probability measure): Ένας (μοναδικός) πραγματικός αριθμός που αντιστοιχεί σε κάθε αποτέλεσμα του χώρου δειγμάτων. Αξιωματικά ορισμένο (3 αξιώματα).

1.1 'Αλγεβρα γεγονότων

Ορισμοί:

- Βέβαιο ενδεχόμενο (universal event) Ω είναι το σύνολο που περιέχει όλα τα σημεία στο δειγματοχώρο ενός πειράματος.

- Το συμπλήρωμα (complement) A' ή $\bar{A}$ ενός ενδεχομένου A είναι το σύνολο όλων των σημείων του Ω που δεν συμπεριλαμβάνονται στο A.

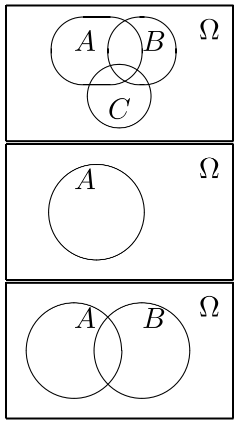
- Το κενό σύνολο ή κενό ενδεχόμενο (null set) $\emptyset$ ή ϕ δεν περιέχει κανένα σημείο του δειγματοχώρου.

- Η ένωση (union) $+$ ή $\cup$ δύο ενδεχομένων A και B είναι το ενδεχόμενο που περιέχει όλα τα σημεία που εμπεριέχονται στο A ή B.

- Η τομή (intersection) $\cdot$ ή $\cap$ δύο ενδεχομένων A και B είναι το ενδεχόμενο που περιέχει όλα τα σημεία που εμπεριέχονται στο A και στο B.

'Εχουμε ορίσει λοιπόν δύο πράξεις: την πρόσθεση (ή ένωση) και τον πολλαπλασιασμό (ή τομή) δύο ενδεχομένων. Οι πράξεις αυτές ορίζονται πάνω σε ενδεχόμενα ή αλλιώς σύνολα σημείων σε ένα δισδιάστατο χώρο (που αντιστοιχούν σε πειραματικά αποτελέσματα ενός μοντέλου). Τα παρακάτω 7 αξιώματα ορίζουν την άλγεβρα γεγονότων:

1. $A + B = B + A$
2. $A + (B + C) = (A + B) + C$
3. $A(B + C) = AB + AC$
4. $(A')' = A$
5. $(AB)' = A' + B'$
6. $AA' = \emptyset$
7. $A\Omega = A$



Οι ακόλουθες ιδιότητες μπορεί να αποδειχθούν με χρήση των αξιωμάτων (στα δεξιά η βασική ιδέα της κάθε απόδειξης):

- $AB = BA$ $\quad [(A' + B')' = (B' + A')']$
- $A(BC) = (AB)C$ $\quad [(A' + (B' + C'))' = ((A' + B') + C')']$
- $A + \emptyset = A$ $\quad [(A\Omega)' = A']$
- $A\emptyset = \emptyset$ $\quad [AA' = \emptyset \Rightarrow (A + \emptyset)A' = \emptyset]$
- $A + \Omega = \Omega$ $\quad [(A'\emptyset)' = \emptyset']$
- $A + A' = \Omega$ $\quad [A + A' = (A'(A')')' = (A'A)' = \emptyset' = \Omega]$
- $A + A = A$ $\quad [A + A = A\Omega + A\Omega = A(\Omega + \Omega) = A\Omega = A]$
- $AA = A$ $\quad [AA = (A' + A')' = (A')' = A]$
- $A + AB = A$ $\quad [A + AB = A\Omega + AB = A(\Omega + B) = A\Omega = A]$
- $A + A'B = A + B$ $\quad [A + A'B = A(\Omega + B) + A'B = ...]$
- $A + BC = (A + B)(A + C)$ $\quad [AA + BA + AC + BC = A + BC]$

Σημείωση 1: Θα μπορούσε κανείς να επιλέξει διαφορετικά τα 7 αξιώματα (π.χ., ιδιότητα $A + A' = \Omega$ αντί για αξίωμα 6) καταλήγοντας στο ίδιο αποτέλεσμα.

Σημείωση 2: Η άλγεβρα γεγονότων σχετίζεται με την άλγεβρα Boole η οποία όμως ορίζεται πάνω στους αριθμούς $\{0, 1\}$ και όχι πάνω σε σύνολα (ενδεχομένων). Οι δύο πράξεις $+ \equiv \cup$ και $. \equiv \cap$, τα αξιώματα και οι ιδιότητες είναι ίδιες τόσο στην άλγεβρα Boole όσο και στην άλγεβρα ενδεχομένων.

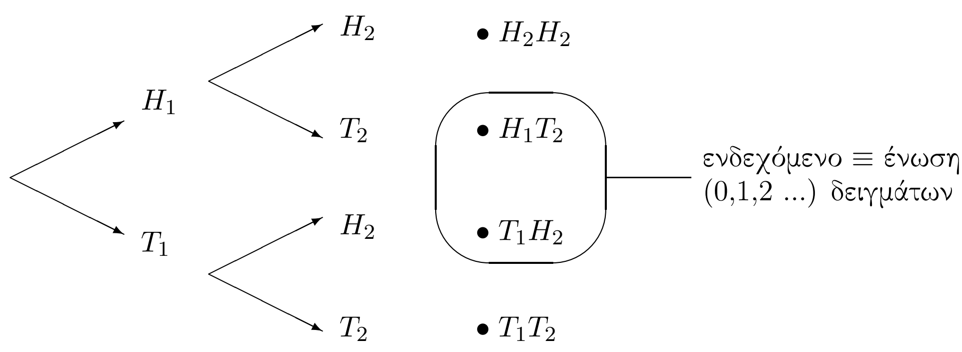
Σημείωση 3: Συγκρίνοντας την άλγεβρα ενδεχομένων με την άλγεβρα πραγματικών αριθμών με πράξεις την πρόσθεση και πολλαπλασιασμό η κύρια διαφορά είναι οι ιδιότητες $A + A = A$ και $A \cdot A = A$ που δεν ισχύουν για πραγματικούς αριθμούς (idempotency).

Δυο ακόμα σημαντικοί ορισμοί:

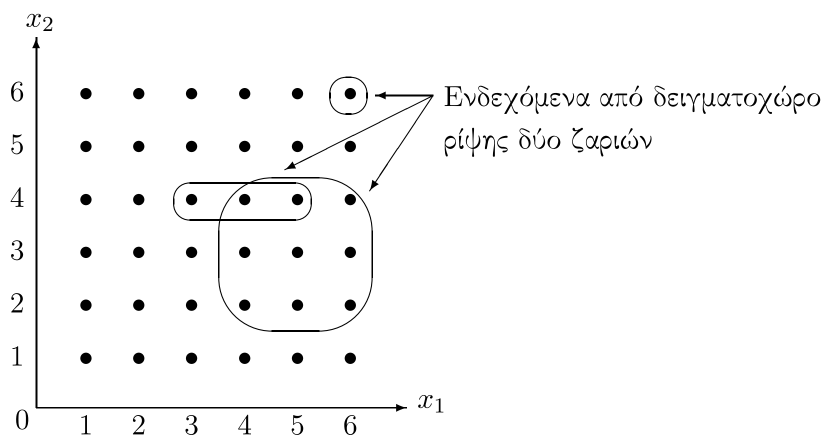
- Τα ενδεχόμενα A και B είναι ασυμβίβαστα ή ξένα (mutually exclusive) όταν η τομή τους είναι το κενό σύνολο, δηλαδή, $A \cap B = \emptyset$.
- Διαμερισμός (collectively exhaustive events) είναι μια συλλογή από ασυμβίβαστα γεγονότα A_i, $i = 1, 2, 3..., N$ των οποίων η ένωση είναι το βέβαιο ενδεχόμενο, δηλαδή, $\bigcup_{i=1}^{N} A_i = \Omega$.

1.2 Δειγματοχώρος

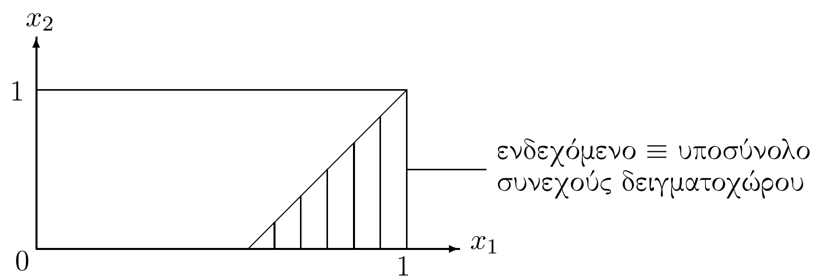
Παράδειγμα δειγματοχώρου από ρίψη δύο νομισμάτων (4 δείγματα). Τα ενδεχόμενα ορίζονται σαν ένωση δειγμάτων:



Τα δείγματα είναι πάντα ασυμβίβαστα μεταξύ τους αφού απαριθμούν όλα τα δυνατά αποτελέσματα ενός πειράματος. Η ένωση όλων των δειγμάτων είναι το βέβαιο ενδεχόμενο αφού απαριθμούν *όλα* τα δυνατά αποτελέσματα ενός πειράματος. Τα ενδεχόμενα ορίζονται σαν ένωση δειγμάτων και δεν είναι απαραίτητα ασυμβίβαστα, όπως φαίνεται στο παρακάτω δειγματοχώρο από την ρίψη δύο ζαριών:



Σε περίπτωση που τα αποτελέσματα του πειράματος παίρνουν συνεχείς τιμές (π.χ. μέτρηση αντίστασης κυκλώματος) πάλι τα δείγματα αποτελούν διαμερισμό, δηλαδή, είναι ξένα μεταξύ τους και η ένωση τους είναι το βέβαιο ενδεχόμενο. Ενδεχόμενα ορίζονται ως περιοχές ή ένωση περιοχών του (δισδιάστατου) δειγματοχώρου (αντίστοιχα σε μονοδιάστατο δειγματοχώρο είναι διαστήματα). Για παράδειγμα ο δειγματοχώρος δύο μετρήσεων x_1 και x_2 με τιμές στο [0,1] είναι:



Οι θεμελιώδεις ιδιότητες του δειγματοχώρου:

Τα δείγματα του δειγματοχώρου είναι ασυμβίβαστα ενδεχόμενα.
Τα δείγματα του δειγματοχώρου (όλα μαζί) αποτελούν διαμερισμό.
Ο δειγματοχώρος είναι ντετερμινιστικός (οι πιθανότητες είναι μια γενίκευση).
Όλα τα προβλήματα πιθανοτήτων ξεκινούν και επιστρέφουν στον δειγματοχώρο.

1.3 Μέτρο Πιθανότητας

Σε κάθε δείγμα του δειγματοχώρου και σε κάθε ενδεχόμενο αντιστοιχίζουμε ένα πραγματικό αριθμό στο διάστημα $[0, 1]$. Επιπρόσθετα ορίζουμε την πιθανότητα ενός ενδεχομένου ως το άθροισμα της πιθανότητας των δειγμάτων που το απαρτίζουν. Ο αριθμός αυτός ονομάζεται μέτρο πιθανότητας (probability measure) $P(.)$ και ορίζεται αξιωματικά ως εξής:

1. Για κάθε ενδεχόμενο A η πιθανότητά του είναι μη αρνητική: $P(A) \geq 0$

2. Η πιθανότητα του βέβαιου ενδεχομένου είναι ίση με ένα: $P(\Omega) = 1$

3. Η πιθανότητα της ένωσης ασυμβίβαστων γεγονότων είναι ίση με το άθροισμα της πιθανότητας των ενδεχομένων:
   εάν $AB = \emptyset$ τότε $P(A + B) = P(A) + P(B)$

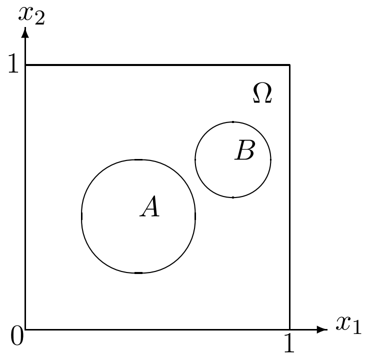
Η πιθανότητας $P(A)$ μπορεί να ερμηνευθεί ως ο λόγος του αριθμού n_A των πειραμάτων[1] με αποτέλεσμα το ενδεχόμενο A προς τον συνολικό αριθμό πειραμάτων n, όταν το $n \to \infty$. Δηλαδή,

$$P(A) = \lim_{n\to\infty} \frac{n_A}{n}$$

Παρατηρούμε ότι αυτή η ερμηνεία της πιθανότητας ικανοποιεί και τα τρία αξιώματα του μέτρου πιθανότητας επειδή $0 \leq n_A \leq n$ και άρα $P(A) \geq 0$. Επίσης $P(\Omega) = 1$ (επειδή $n/n = 1$). Τέλος για ξένα ενδεχόμενα A, B ισχύει ότι $(n_A + n_B)/n = (n_A/n) + (n_B/n)$ (ισχύει και στο όριο $n \to \infty$) και άρα $P(A + B) = P(A) + P(B)$. Η συχνότητα πειραματικής εμφάνισης ενός ενδεχομένου στο όριο όταν το πλήθος των πειραμάτων τείνει στο άπειρο είναι η πιθανότητα (ή μια ερμηνεία της πιθανότητας) αυτού του ενδεχομένου. Στην πράξη μπορούμε να εκτιμούμε πιθανότητες εμπειρικά χρησιμοποιώντας πειραματικά μοντέλα για πολύ μεγάλες τιμές του n (όπως εξηγούμε στο Κεφ. 1.10).

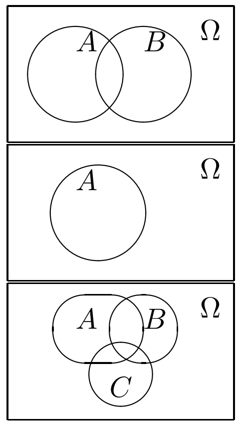
[1] Αυτή η φυσική ερμηνεία του μέτρου πιθανότητας προϋποθέτει την ύπαρξη ενός πειραματικού μοντέλου που παράγει διαφορετικά ενδεχόμενα.

Μια άλλη φυσική ερμηνεία της πιθανότητας είναι ως το εμβαδόν των ενδεχομένων σε ένα (δισδιάστατο) διάγραμμα Venn όπου το εμβαδόν του βέβαιου ενδεχομένου Ω είναι 1, π.χ., ο συνεχής δειγματοχώρος που απεικονίζεται με ένα τετράγωνο πλευράς 1. Πράγματι το εμβαδόν ενός ενδεχομένου παίρνει τιμές μεταξύ 0 και 1, ενώ το άθροισμα του εμβαδού των ξένων ενδεχομένων A και B είναι ίσο με το εμβαδόν του ενδεχομένου $A + B$.



Χρησιμοποιώντας την δεύτερη ερμηνεία του μέτρου πιθανότητας είναι εύκολο να δείξουμε τις παρακάτω ιδιότητες, π.χ., το εμβαδόν της ένωσης του A και B, ισούται με το εμβαδόν του A, συν το εμβαδόν του B, μείον το εμβαδόν της τομής του A και B :

- $P(A + B) = P(A) + P(B) - P(AB)$
- $P(A') = 1 - P(A)$
- $P(\emptyset) = 0$
- $P(A + B + C) = 1 - P(A'B'C')$
- $P(A + B + C) = P(A) + P(A'B) + P(A'B'C)$
- $P(A+B+C) = P(A)+P(B)+P(C)-P(AB)- P(BC) - P(AC) + P(ABC)$
- ...



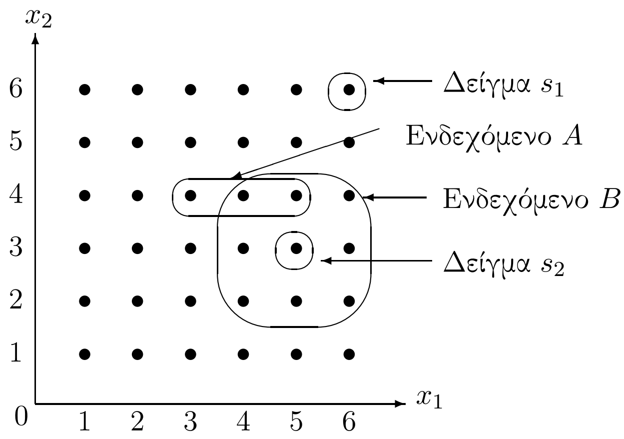
Σημαντικές ιδιότητες του μέτρου πιθανότητας:

- Στον δειγματοχώρο αθροίζω πιθανότητες δειγμάτων για να υπολογίσω πιθανότητες ενδεχομένων (διαμερισμός).
- Το άθροισμα των πιθανοτήτων ενδεχομένων που αποτελούν διαμερισμό είναι 1.
- Μέσα στο όρισμα του $P(\underbrace{A + CD + B(A + C'D)})$ χρησιμοποιούμε τα 7 αξιώματα (+ ιδιότητες) της άλγεβρας ενδεχομένων.

- Έξω στο όρισμα του $\underbrace{P(\;}A+CD+B(A+C'D))$ χρησιμοποιούμε τα 3 αξιώματα (+ ιδιότητες) του μέτρου πιθανότητας.

1.4 Πιθανότητα Υπό Συνθήκη

Έστω s_i δείγμα του δειγματοχώρου με μέτρο πιθανότητας $P(s_i)$. Η πιθανότητα υπό συνθήκη του s_i, δεδομένου ότι έχει συμβεί το ενδεχόμενο B με πιθανότητα $P(B)$, είναι $P(s_i)/P(B)$ εφόσον το s_i είναι μέλος του ενδεχομένου B και 0 αλλιώς. Για παράδειγμα στη διπλή ρίψη ζαριού όπου όλα τα δείγματα είναι ισοπίθανα με πιθανότητα 1/36, η πιθανότητα του s_1 (εξάρες) δεδομένου ότι έχει συμβεί το ενδεχόμενο B είναι 0 (γιατί s_1 και B ξένα μεταξύ τους). Η δεσμευμένη πιθανότητα $P(s_2|B) = (1/36)/(9/36) = 1/9$, γιατί το B 'περιέχει' το s_2, το οποίο έχει συνολικά 9 ισοπίθανα δείγματα.



Η φυσική ερμηνεία του ορισμού της υπό συνθήκης πιθανότητας είναι ότι το B είναι το βέβαιο ενδεχόμενο. Όλα τα δείγματα που ανήκουν στο B' είναι πλέον αδύνατον να συμβούν και άρα έχουν πιθανότητα 0. Αντίθετα, τα δείγματα που αποτελούν το B είναι πιθανά και το άθροισμα της δεσμευμένης πιθανότητάς τους πρέπει να είναι 1 (διαμερισμός). Άρα πρέπει να διαιρέσουμε τη δεσμευμένη πιθανότητα με το $P(B)$ ώστε να παίρνει τιμές $[0,1]$. Αντίστοιχα η δεσμευμένη πιθανότητα (conditional probability) $P(A|B)$ του ενδεχομένου A δεδομένου του B, όπως φαίνεται στο σχήμα είναι η πιθανότητα της τομής των A και B (προφανώς $P(AB'|B) = 0$), κανονικοποιημένη με την πιθανότητα $P(B)$. Για παράδειγμα στη διπλή ρίψη νομίσματος $P(A|B) = P(AB)/P(B) = (2/36)/(9/36) = 2/9$. Γενικά:

$$P(A|B) = P(AB + AB'|B) = P(AB|B) + P(AB'|B) = P(AB|B) \Rightarrow$$

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

1.5 Ανεξάρτητα Ενδεχόμενα

Ορισμός: τα ενδεχόμενα A, B είναι ανεξάρτητα (independent) αν και μόνο αν:

$$P(A|B) = P(A)$$

Από τον ορισμό προκύπτουν οι ιδιότητες (ισχύουν για A, B ανεξάρτητα):

$$P(AB) = P(A)P(B) \text{ και } P(B|A) = P(B)$$

Είναι εύκολο να δείξουμε ότι οι τρεις σχέσεις είναι ισοδύναμες χρησιμοποιώντας τον ορισμό της δεσμευμένης πιθανότητας (για ενδεχόμενα A, B με μη μηδενική πιθανότητα):

$$\underline{P(A|B) = P(A)} \Leftrightarrow \frac{P(AB)}{P(B)} = P(A) \Leftrightarrow \underline{P(AB) = P(A)P(B)} \Leftrightarrow$$

$$\frac{P(AB)}{P(A)} = P(B) \Leftrightarrow \underline{P(B|A) = P(B)}$$

Ορισμός: τα ενδεχόμενα A_i, $i = 1, 2...N$ είναι από κοινού ανεξάρτητα (mutually independent) αν και μόνο αν:

$$P(A_i|A_j A_k ... A_p) = P(A_i), \quad \forall i \neq j, k...p, \quad 1 \leq i, j...k, ..., p \leq N$$

Σημείωση: Ανεξαρτησία ανά ζεύγη των, π.χ., A_1, A_2, A_3, δεν συνεπάγεται από κοινού ανεξαρτησία.

Ορισμός: τα ενδεχόμενα A, B είναι ανεξάρτητα υπό συνθήκη C (conditionally independent) αν και μόνο αν:

$$P(A|BC) = P(A|C)$$

Προκύπτουν οι ιδιότητες (που είναι και ισοδύναμοι ορισμοί):

$$P(AB|C) = P(A|C)P(B|C) \text{ και } P(B|AC) = P(B|C)$$

Η ανεξαρτησία υπό συνθήκη είναι η πιο θεμελιώδης έννοια στην μοντελοποίηση ενδεχομένων. Είναι γνωστή και ως η Μαρκοβιανή ιδιότητα γιατί συναντάται σε Μαρκοβιανά μοντέλα (Markov models). Η έννοια της ανεξαρτησίας υπό συνθήκη συναντάται και στην καθημερινή μας ζωή σε αλυσίδες αιτίας αιτιατού (Markov chains). Για παράδειγμα τα τρία ενδεχόμενα A = 'άργησα να κοιμηθώ', B='άργησα να ξυπνήσω', και C = 'άργησα στο μάθημα' αποτελούν μια αλυσίδα (Markov) με σχέσεις άμεσης εξάρτησης $A \to B \to C$, δηλαδή, το A είναι μία πιθανή αιτία του B και το B του C. Το A και το C είναι επίσης εξαρτημένα, αλλά μόνο μέσω του B. Για παράδειγμα αν ξέρω ότι άργησα να κοιμηθώ η πιθανότητα του να αργήσω στο μάθημα αυξάνει και άρα $P(C|A) \neq P(C)$. Όμως άμα ξέρω ότι το B συνέβη τα A και C γίνονται ανεξάρτητα, δηλαδή άμα ξέρω ότι ξύπνησα αργά (B) δεν με ενδιαφέρει πλέον η αιτία του αργοπορημένου ξυπνήματος (A) για τον υπολογισμό της πιθανότητα του C (άργησα στο μάθημα) και άρα, $P(C|BA) = P(C|B)$, και τα A, C είναι ανεξάρτητα δεδομένου του B. Άλλα παραδείγματα ενδεχομένων είναι τα 'δεν είχα καλή ορατότητα', 'άργησα να φρενάρω', 'τράκαρα'. Άλλες συνηθισμένες σχέσεις εξάρτησης είναι δύο αιτίες και ένα αιτιατό, π.χ. $A \to C \leftarrow B$ όπου A='δεν διάβασα', B='ήταν δύσκολα τα θέματα' είναι δυο πιθανές αιτίες για το ενδεχόμενο C='δεν πέρασα το μάθημα', καθώς επίσης και μια αιτία και δύο αιτιατά, π.χ., $B \leftarrow A \to C$ όπου το ενδεχόμενο A= 'πρόγνωση καιρού' είναι η αιτία για τα ενδεχόμενα B='πάω εκδρομή' και C='πότισμα κήπου'.

Όταν δύο ενδεχόμενα με μη μηδενική πιθανότητα είναι ασυμβίβαστα τότε είναι πάντα εξαρτημένα (το αντίθετο δεν ισχύει). Η απόδειξη αυτής της ιδιότητας είναι απλή. Για ξένα ενδεχόμενα A, B έχουμε $P(AB) = P(\emptyset) = 0$ ενώ το γινόμενο $P(A)P(B)$ είναι μη μηδενικό, άρα $P(AB) \neq P(A)P(B)$. Εν γένει είναι πιο εύκολο να αναπαραστήσουμε γραφικά την ανεξαρτησία στον δειγματοχώρο παρά με διαγράμματα Venn.

1.6 Κανόνας αλυσίδας

Εφαρμόζοντας πολλές φορές τον ορισμό της υπό συνθήκη πιθανότητας προκύπτει ο κανόνας της αλυσίδας:

$$P(AB) = P(A|B)P(B) = P(B|A)P(A)$$

$$P(ABC) = P(A|BC)P(BC) = P(A|BC)P(B|C)P(C)$$

Όπου η πρώτη σχέση προκύπτει από τον ορισμό της υπό συνθήκη πιθανότητας για ενδεχόμενο A και συνθήκη BC.

$$P(ABC) = ... = P(B|AC)P(A|C)P(C) = P(B|AC)P(C|A)P(A)$$

$$P(ABC) = ... = P(C|AB)P(A|B)P(B) = P(C|AB)P(B|A)P(A)$$

Η από κοινού πιθανότητα ενδεχομένων εκφράζεται σαν συνάρτηση των πιθανοτήτων υπό συνθήκη. Αυτό είναι ιδιαίτερα χρήσιμο γιατί στα στατιστικά μοντέλα και τις πειραματικές μετρήσεις οι πιθανότητες συνήθως δίδονται με την μορφή δεσμευμένης πιθανότητας. Ποια από τις παραπάνω μορφές του κανόνα της αλυσίδας πρέπει να χρησιμοποιήσουμε για να βρούμε την από κοινού πιθανότητα; Η απάντηση βρίσκεται στις σχέσεις αιτίας αιτιατού (ή εν γένει χρονικής αλληλουχίας) μεταξύ των ενδεχομένων. Για παράδειγμα για μια χρονική (ή αιτιατή) αλυσίδα $A \to B \to C$ η μορφή $P(C|AB)P(B|A)P(A)$ είναι η πιο πρόσφορη γιατί αυτές είναι οι πιθανότητες που παρατηρούμε πειραματικά.

Η γενίκευση του κανόνα της αλυσίδας για N ενδεχόμενα (εφαρμόζουμε $N-1$ φορές τον ορισμό της δεσμευμένης πιθανότητας) είναι:

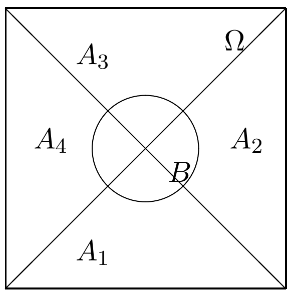
$$P(A_1A_2...A_N) = P(\bigcap_{i=1}^{N} A_i) = \prod_{i=1}^{N} P(A_i|A_{i-1}...A_1)$$

1.7 Οριακή Πιθανότητα ή Κανόνας Διαμερισμού

Η πιθανότητα του ενδεχομένου B μπορεί να εκφραστεί ως το άθροισμα των πιθανοτήτων της τομής του B με ένα διαμερισμό A_i, $i = 1, 2...N$ ως εξής:

$$P(B) = \sum_{i=1}^{N} P(A_i \cap B) \equiv \sum_{i=1}^{N} P(A_iB)$$

Η γραφική αναπαράσταση αυτής της ιδιότητας φαίνεται στο διπλανό σχήμα. Χρησιμοποιώντας την φυσική ερμηνεία του μέτρου πιθανότητας ως εμβαδόν, βλέπουμε ότι το εμβαδόν του ενδεχομένου B μπορεί να εκφραστεί σαν το άθροισμα του εμβαδού του B με κάθε ένα από τα ενδεχόμενα που αποτελούν το διαμερισμό, δηλαδή το άθροισμα του εμβαδού των A_1B, A_2B, A_3B και A_4B.



Ο κανόνας του διαμερισμού ισχύει (και είναι ιδιαίτερα χρήσιμος) και όταν τα A_i και B προέρχονται από διαφορετικούς δειγματοχώρους.

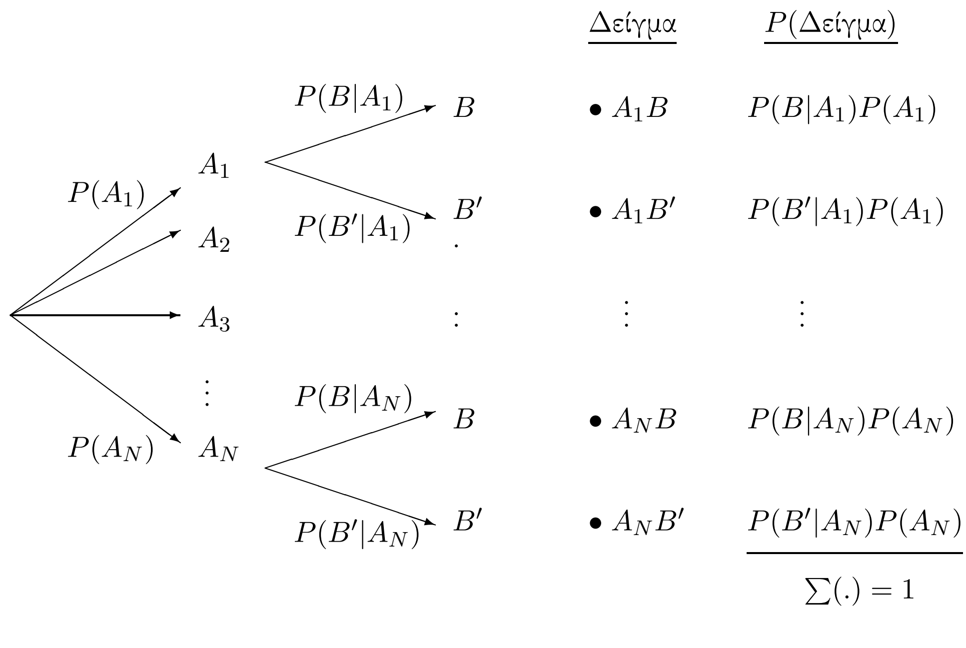
1.8 Κανόνας ή Θεώρημα Bayes

Ο κανόνας του Bayes (Bayes rule) για ένα διαμερισμό ενδεχομένων A_i, $i = 1, 2...N$ και ένα ενδεχόμενο B:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{N} P(A_i B)}$$

Παρά την απλή του μορφή (στην ουσία εφαρμόζουμε τον κανόνα της αλυσίδας στον αριθμητή και τον κανόνα του διαμερισμού στον παρανομαστή) είναι ιδιαίτερα σημαντικός στην στατιστική μοντελοποίηση. Σε μία αλυσίδα πειραμάτων $A \rightarrow B$ (χρονική ή αιτιατή αλυσίδα) συνήθως δίδονται οι απευθείας πιθανότητες (direct probabilities) της μορφής $P(B|A)$. Οι απευθείας πιθανότητες είναι χρήσιμες σε προβλήματα πρόβλεψης όπου προσπαθώ να βρω το πιο πιθανό αποτέλεσμα για μία αιτία και άρα μεγιστοποιώ πιθανότητες της μορφής $P(B|A)$ πάνω σε όλα τα δυνατά αποτελέσματα (αιτιατά B). Σε προβλήματα διάγνωσης όπου προσπαθούμε να βρούμε την πιο πιθανή αιτία που αντιστοιχεί σε ένα αποτέλεσμα (αιτιατό) θέλουμε να μεγιστοποιήσουμε την a posteriori πιθανότητα της μορφής $P(A|B)$, πάνω σε όλα τα A. Ο κανόνας του Bayes εκφράζει την a posteriori πιθανότητα $P(A|B)$, σαν συνάρτηση της απευθείας πιθανότητας $P(B|A)$ και της a priori πιθανότητας $P(A)$.

Ο κανόνας του Bayes 'αντιστρέφει' την (χρονική ή αιτιατή) αλυσίδα και είναι πολύτιμος σε προβλήματα διάγνωσης.



Το δέντρο πιθανότητας που βλέπουμε παραπάνω δείχνει γραφικά την διαδοχή δύο πειραμάτων το πρώτο με ενδεχόμενα $\{A_1, A_2, ...A_N\}$ και το δεύτερο με ενδεχόμενα $\{B, B'\}$. Ο από κοινού δειγματοχώρος του πειράματος αποτελείται από τα δείγματα $\{A_1B, A_1B', A_2B, A_2B', ...A_NB, A_NB'\}$. Η πιθανότητα του κάθε δείγματος υπολογίζεται με τον κανόνα της αλυσίδας χρησιμοποιώντας την a priori και απευθείας πιθανότητες που δίδονται, π.χ., $P(A_1B) = P(B|A_1)P(A_1)$. Το άθροισμα των πιθανοτήτων των δειγμάτων είναι ίσο με τη μονάδα. Για τον υπολογισμό της a posteriori πιθανότητας όμως πρέπει να χρησιμοποιήσουμε μια μορφή του κανόνα του Bayes, π.χ., $P(A_1|B) = P(A_1B)/\sum_{i=1}^{N} P(A_iB)$.

1.9 Δέντρα Δειγματοχώρου Πιθανοτήτων

Για την κατασκευή του δέντρου πιθανοτήτων (βλ. προηγούμενο διάγραμμα) ακολουθούνται τα ακόλουθα βήματα:

1. Δημιουργία δειγματοχώρου και ενδεχομένων, π.χ., στη ρίψη νομίσματος $A = \{H, T\}$, στη ρίψη ζαριού $B = \{1, 2, 3, 4, 5, 6\}$ κλπ

2. Δημιουργία χρονικής ή αιτιατής αλυσίδας, π.χ., αν η μαρτυρία του B βασίζεται στην μαρτυρία του A τότε $A \to B$.

3. Κατασκευή δέντρου ξεκινώντας από την αιτία A ή το χρονικά πρώτο πείραμα (αριστερά) και πηγαίνοντας προς το αιτιατό B (δεξιά).

4. Προσθέτουμε τις απευθείας και a priori πιθανότητες στο δέντρο, π.χ., $P(A_i)$, $P(B_j|A_i)$, $P(C_k|B_j, A_i)$.

5. Πολλαπλασιάζουμε τις πιθανότητες σε κάθε μονοπάτι/δείγμα για να υπολογίσουμε την πιθανότητα κάθε δείγματος στον από κοινού δειγματοχώρο (εφαρμόζουμε τον κανόνα της αλυσίδας), π.χ., $P(A_i, B_j, C_k) = P(C_k|B_j, A_i)P(B_j|A_i)P(A_i)$.

Είμαστε έτοιμοι να υπολογίσουμε όποιες άλλες πιθανότητες χρειάζεται. Για τον υπολογισμό από κοινού πιθανοτήτων αθροίζουμε την πιθανότητα των δειγμάτων που συμφωνούν με το ενδεχόμενο. Για τον υπολογισμό απευθείας πιθανοτήτων (εφόσον δεν δίνονται ήδη) χρησιμοποιούμε τον ορισμό της δεσμευμένης πιθανότητας. Για τον υπολογισμό a posteriori πιθανοτήτων χρησιμοποιούμε τον κανόνα του Bayes.

1.10 Αριθμητική Προσέγγιση Πιθανοτήτων

Είναι σχετικά εύκολο να υπολογίσουμε πιθανότητες αριθμητικά χρησιμοποιώντας μια γεννήτρια τυχαίων αριθμών. Πράγματι με την εντολή $rand(1,1)$ στο MATLAB μπορούμε να δημιουργήσουμε έναν 'τυχαίο' αριθμό (1×1) με τιμές ισοκατανεμημένες στο διάστημα $[0,1]$. Οπότε μπορούμε να υλοποιήσουμε το ενδεχόμενο A με πιθανότητα 0.3 ως $a = [rand(1,1) < 0.3]$; μια συνάρτηση που (κατά μέσο όρο) 3/10 φορές θα είναι 1 και 7/10 φορές θα είναι 0. Αν τρέξουμε αρκετές φορές αυτήν την εντολή μπορούμε να προσεγγίσουμε το $P(A)$ ως εξής:

```
no_trials = 1000;
a_counter = 0;
for i = 1:no_trials
  a = [rand(1,1) < 0.3];
  if (a == 1)
     a_counter = a_counter + 1;
   end
end
prob_a = a_counter / no_trials;
```

Όταν τρέξω τον παραπάνω κώδικα παίρνω 0.2880. Τον ξανατρέχω και παίρνω 0.3170. Βλέπω ότι κάθε φορά η πιθανότητα είναι διαφορετική αλλά κοντά στην πραγματική 0.3. Για να μειώσω το λάθος μου μπορώ να αυξήσω τον αριθμό των πειραμάτων, π.χ., για 10000 πειράματα παίρνω: 0.3024, 0.3033, 0.3003, 0.2988 που είναι πιο κοντά στο 0.3. Πράγματι σύμφωνα με την ερμηνεία του μέτρου πιθανότητας ως τον λόγο n_A/n, όπου n_A ο αριθμός των πειραμάτων που είχαν αποτέλεσμα A (αντιστοιχεί στην μεταβλητή a_counter) και n ο συνολικός αριθμός πειραμάτων (αντιστοιχεί στην σταθερά no_trials), το n πρέπει να τείνει στον άπειρο. Άρα η ακρίβεια υπολογισμού της πιθανότητας εξαρτάται από τον αριθμό των πειραμάτων n, μεγαλύτερα n συνήθως βελτιώνουν την ακρίβεια.

Χρησιμοποιώντας τις γεννήτριες τυχαίων αριθμών μπορούμε να λύνουμε αριθμητικά πολύπλοκα προβλήματα πιθανοτήτων. Για παράδειγμα, ποια είναι η πιθανότητα να φέρω 25 φορές γράμματα σε 50 ρίψεις ενός νομίσματος (θα δούμε στο Κεφ. 4 πως να υπολογίσουμε αναλυτικά αυτή την πιθανότητα). Αριθμητικά:

```
no_trials = 1000;
no_throws = 50;
p_counter = 0;
for i = 1:no_trials
  no_grammata = 0;
  for j = 1:no_throws
```

```
     a = [rand(1,1) < 0.5];
     if (a == 1)
        no_grammata = no_grammata + 1;
     end
   end
   if (no_grammata == 25)
     p_counter = p_counter + 1;
    end
end
prob_p = p_counter / no_trials;
```

Η πιθανότητα προκύπτει 0.11 που είναι μια καλή προσέγγιση. Ο ίδιος κώδικας χρησιμοποιώντας πίνακες:

```
no_trials = 1000;
no_throws = 50;
A = rand(no_throws, no_trials);
p_counter = sum([sum(A < 0.5) == 25]);
prob_p = p_counter / no_trials;
```

και σε μια γραμμή (παράδειγμα προς αποφυγή) :

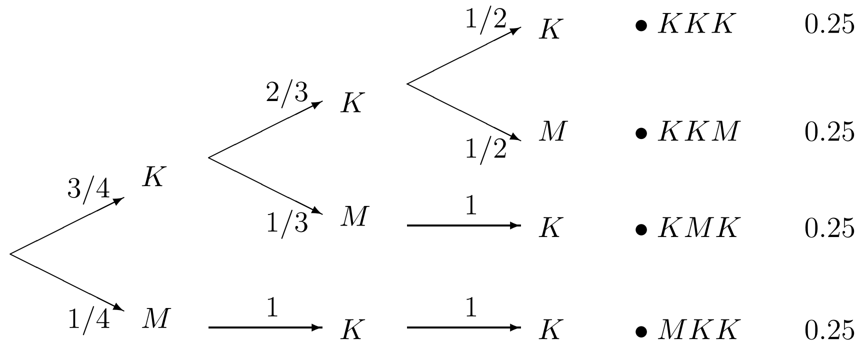
```
prob_p = sum([sum(rand(50,1000) < 0.5) == 25])/1000;
```

1.11 Παραδείγματα

Παράδειγμα 1: Ένα δοχείο περιέχει 3 κόκκινες και 1 μαύρη μπάλα. Τραβάω 3 μπάλες στη σειρά (χωρίς επανατοποθέτηση).

(α) Ποιο είναι το ενδεχόμενο να τραβήξω κόκκινη μπάλα στην πρώτη προσπάθεια; Στην δεύτερη; Στην τρίτη προσπάθεια;

Τα ενδεχόμενα για την 1η προσπάθεια είναι: $E_1 = K$, $E_1 = M$, για την 2η: $E_2 = K$, $E_2 = M$, και για την τρίτη: $E_3 = K$, $E_3 = M$. Οι σχέσεις εξάρτησης είναι $E_1 \rightarrow E_2 \rightarrow E_3$, δηλαδή κάθε προσπάθεια εξαρτάται από την προηγούμενη (χρονικά) προσπάθεια γιατί δεν επανατοποθετούμε τις μπάλες στο δοχείο. Στην πρώτη προσπάθεια οι πιθανότητες είναι $P(K) = 3/4$, $P(M) = 1/4$. Στην δεύτερη προσπάθεια $P(K|K) = 2/3$, $P(M|K) = 1/3$ ενώ $P(K|M) = 1$. Τέλος στην τρίτη προσπάθεια $P(K|KK) = 1/2$, $P(M|KK) = 1/2$, ενώ $P(K|KM) = 1$, $P(K|MK) = 1$ (προφανώς το ενδεχόμενο $P(MM) = 0$). Σχεδιάζουμε το δέντρο πιθανότητας και υπολογίζουμε την από κοινού πιθανότητα κάθε δείγματος $E_1E_2E_3$:



Για τον υπολογισμό ενδεχομένων απλά αθροίζω την πιθανότητα των δειγμάτων που το αποτελούν (γιατί ο από κοινού δειγματοχώρος του πειράματος είναι διαμερισμός), δηλαδή:

$$P(E_1 = K) = P(KKK) + P(KMK) + P(KKM) = 0.75$$

$$P(E_2 = K) = P(KKK) + P(MKK) + P(KKM) = 0.75$$

$$P(E_3 = K) = P(KKK) + P(MKK) + P(KMK) = 0.75$$

(β) Ποια είναι η πιθανότητα να τραβήξω τουλάχιστον δύο κόκκινες μπάλες στη σειρά;

Το μόνο δείγμα που δεν πληρεί την συνθήκη (β) είναι το KMK όποτε αθροίζω την πιθανότητα των υπολοίπων δειγμάτων:

$$P((b)) = P(KKK) + P(KKM) + P(MKK) = 0.75$$

(γ) Ποια είναι η πιθανότητα να τραβήξω κόκκινη μπάλα στην 3η προσπάθεια δεδομένου ότι τράβηξα κόκκινη στην 1η προσπάθεια;

$$P(E_3 = K|E_1 = K) = \frac{P((E_1 = K) \cap (E3 = K))}{P(E_1 = K)} = \frac{P(KKK + KMK)}{0.75} = \frac{2}{3}$$

(δ) Είναι τα ενδεχόμενα 'τραβάω κόκκινη μπάλα στην 1η προσπάθεια', 'τραβάω κόκκινη μπάλα στην 3η προσπάθεια' ανεξάρτητα;

Τα ενδεχόμενα είναι εξαρτημένα γιατί η από κοινού πιθανότητα δεν ισούται με το γινόμενο των πιθανοτήτων, δηλαδή:

$$P(E_1 = K, E_3 = K) = 0.5 \neq P(E_1 = K)P(E_3 = K) = 0.75 \times 0.75$$

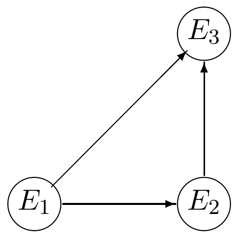
(ε) Είναι τα αποτελέσματα της 1ης και 3ης προσπάθειας ανεξάρτητα, δεδομένου του αποτελέσματος της 2ης προσπάθειας;

Αρκεί να αποδείξουμε ότι η σχέση $P(E_3|E_2, E_1) = P(E_3|E_2)$ δεν ισχύει για κάποιο από τα δείγματα[2] $E_1E_2E_3$ στο δειγματοχώρο μας, π.χ., για το δείγμα KKK:

$$P(E_3 = K|E_2 = K) = \frac{P(E_2 = K, E_3 = K)}{P(E_2 = K)} = \frac{P(KKK + MKK)}{0.75} = \frac{2}{3}$$

$$P(E_3 = K|E_2 = K, E_1 = K) = \frac{P(KKK)}{P(E_1 = K, E_2 = K)} = \frac{0.25}{P(KKK + MKK)} = \frac{1}{2}$$

και άρα $P(E_3|E_2, E_1) \neq P(E_3|E_2)$. Συμπεραίνουμε ότι οι σχέσεις μεταξύ των E_1, E_2 και E_3 είναι:



δηλαδή υπάρχει απευθείας εξάρτηση τόσο μεταξύ $E_1 \rightarrow E_2$ $E_2 \rightarrow E_3$, όσο και μεταξύ $E_1 \rightarrow E_3$. Άρα τα E_1, E_2 και E_3 δεν αποτελούν μια αλυσίδα αιτίας αιτιατού (Markov chain).

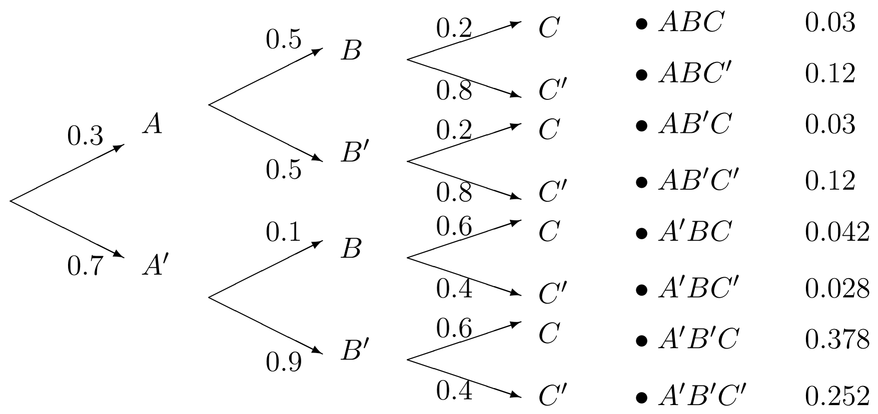
<u>Παράδειγμα 2:</u> Η πιθανότητα να διαβάσει ο Αλέξης για το μάθημα ΤΗΛ 311 είναι 0.3. Εάν διαβάσει έχει πιθανότητα 0.5 να περάσει το μάθημα, ενώ αν δεν διαβάσει έχει πιθανότητα 0.1 να περάσει. Εάν δεν διαβάσει θα έχει 0.6 πιθανότητα να βγει με την Φαίη το Σάββατο, ενώ άμα διαβάσει θα έχει πιθανότητα μόνο 0.2 να βγει με την Φαίη.

α) Ποια είναι η από κοινού πιθανότητα να βγει ο Αλέξης με την Φαίη το Σάββατο και να περάσει το μάθημα;

Τα ενδεχόμενα που δίνονται είναι τα A = ‘διαβάζει ο Αλέξης για το μάθημα ΤΗΛ 311’, B = ‘περνάει ο Αλέξης το μάθημα ΤΗΛ 311’, και C = ‘βγαίνει ο Αλέξης με την Φαίη το Σάββατο’ (καθώς επίσης και τα συμπληρώματά τους: A', B' και C'). Οι πιθανότητες που δίνονται είναι η a priori $P(A) = 0.3$, και οι απευθείας πιθανότητες $P(B|A) = 0.5$, $P(B|A') = 0.1$, $P(C|A) = 0.2$,

[2] Για να είναι δυο πειράματα ανεξάρτητα πρέπει όλα τα ενδεχόμενα του ενός πειράματος να είναι ανεξάρτητα με τα ενδεχόμενα του άλλου πειράματος (βλ. επίσης Κεφ. 2).

$P(C|A') = 0.6$. Παρατηρούμε ότι οι σχέσεις αιτίας αιτιατού μπορούν να κωδικοποιηθούν ως: $B \leftarrow A \rightarrow C$ δηλαδή το A επηρεάζει άμεσα το B και C, π.χ., αν διαβάσεις είναι πολύ πιθανό να περάσεις το μάθημα αλλά να μην βγεις με την Φαίη (και το ανάποδο). Άρα ξεκινώντας από την αιτία φτιάχνουμε το δέντρο του δειγματοχώρου και υπολογίζουμε την (από κοινού) πιθανότητα των δειγμάτων (Σημείωση: Δεν έχει σημασία αν ακολουθήσω την σειρά A, B, C ή A, C, B):



Για τον υπολογισμό ενδεχομένων απλά αθροίζω την πιθανότητα των δειγμάτων που το αποτελούν, δηλαδή:

$$P(BC) = P(ABC + A'BC) = P(ABC) + P(A'BC) = 0.03 + 0.042 = 0.072$$

β) Ποια είναι η πιθανότητα του ενδεχομένου να βγει ο Αλέξης με την Φαίη το Σάββατο ή να περάσει το μάθημα;

Ψάχνω για δείγματα που 'περιέχουν' το B ή το C. Άρα:

$$P(B + C) = P(ABC + A'BC + ABC' + A'BC' + AB'C + AB'C) =$$

$$= 0.03 + 0.12 + 0.03 + 0.042 + 0.028 + 0.378 = 0.628$$

Ισοδύναμα:

$$P(B + C) = P((B'C')') = 1 - P(B'C') = 1 - P(AB'C' + A'B'C') = 0.628$$

γ) Ποια είναι η πιθανότητα να έχει διαβάσει ο Αλέξης για το μάθημα δεδομένου ότι πέρασε το μάθημα;

Πρέπει να υπολογίσουμε το $P(A|B)$ και άρα χρησιμοποιούμε τον κανόνα του Bayes:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.5 \times 0.3}{P(ABC + ABC' + A'BC + A'BC')} =$$

$$= \frac{0.15}{0.03 + 0.12 + 0.042 + 0.028} = \frac{0.15}{0.22} \approx 0.682$$

δ) Είναι τα γεγονότα 'πέρασα το μάθημα ΤΗΛ 311' και 'βγήκα με την Φαίη το Σάββατο' ανεξάρτητα;

Υπολογίζω τις πιθανότητες $P(BC)$, $P(B)P(C)$, και τις συγκρίνω:

$$P(C) = P(ABC + AB'C + A'BC + A'B'C) = 0.48$$

$$P(B)P(C) = 0.22 \times 0.48 \approx 0.106$$

Από το ερώτημα (α) $P(BC) = 0.072$, άρα:

$$P(BC) \neq P(B)P(C) \Rightarrow B,\ C \quad \text{εξαρτημένα}$$

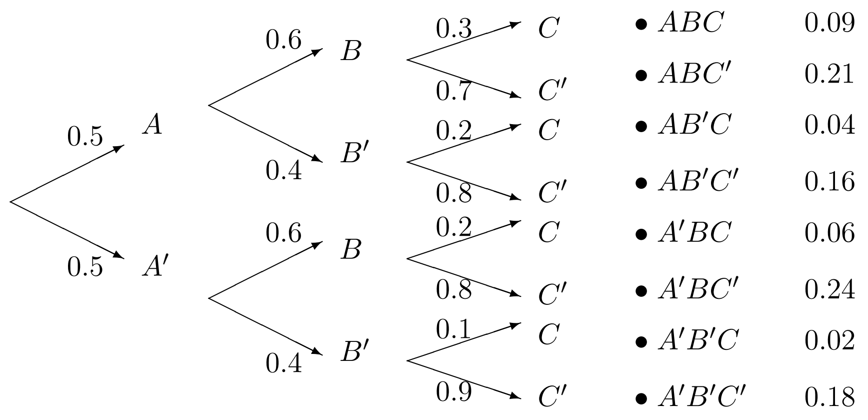
δ) Είναι τα γεγονότα 'πέρασα το μάθημα ΤΗΛ 311' και 'βγήκα με την Φαίη το Σάββατο' ανεξάρτητα δεδομένου ότι ο Αλέξης διάβασε για το μάθημα;

Από τον ορισμό του προβλήματος $P(C|BA) = P(C|A)$, οπότε B, C ανεξάρτητα υπό συνθήκη A.

Παράδειγμα 3: Ο Αστέρας Εξαρχείων έχει φτάσει στον τελικό του Champion League. Ο Ζούπος, κορυφαίος επιθετικός του Αστέρα είναι τραυματίας και η πιθανότητα να ξεκινήσει στον τελικό είναι 0.5. Η πιθανότητα να βρέξει στον τελικό είναι 0.6. Ο Θωμάς εκτιμά ότι άμα ξεκινήσει ο Ζούπος και βρέξει ο Αστέρας έχει πιθανότητα να σηκώσει το τρόπαιο 0.3. Άμα ξεκινήσει ο Ζούπος και δεν βρέξει η πιθανότητα αυτή είναι 0.2. Άμα δεν ξεκινήσει ο Ζούπος και βρέξει η πιθανότητα είναι επίσης 0.2. Τέλος άμα δεν παίξει ο Ζούπος και δεν βρέξει η πιθανότητα είναι 0.1.

α) Ποια είναι η πιθανότητα να σηκώσει ο Αστέρας το τρόπαιο σύμφωνα με το μοντέλο του Θωμά;

Τα ενδεχόμενα του προβλήματός μας είναι $A =$'ξεκινάει τον τελικό ο Ζούπος ', $B =$'βρέχει στον τελικό' και $C =$'σηκώνει ο Αστέρας Εξαρχείων το τρόπαιο' (καθώς και τα συμπληρώματά τους A', B', και C'). Οι πιθανότητες που δίνονται είναι οι $P(A) = 0.5$, $P(B) = 0.6$, $P(C|AB) = 0.3$, $P(C|A'B) = P(C|AB') = 0.2$, $P(C|A'B') = 0.1$. Σύμφωνα με το μοντέλο του Θωμά τα A και B είναι αιτίες του C και άρα η τοπολογία του δικτύου είναι $A \rightarrow C \leftarrow B$. Το δέντρο πιθανότητας λοιπόν ξεκινάει από τις δύο αιτίες (αριστερά) και πηγαίνει προς το αιτιατό (δεξιά). Άρα ο από κοινού δειγματοχώρος και οι πιθανότητες των δειγμάτων έχουν ως εξής (δεν έχει σημασία άμα η σειρά είναι A, B, C ή B, A, C):



Άρα αθροίζοντας την πιθανότητα των δειγμάτων η πιθανότητα να σηκώσει το τρόπαιο ο Αστέρας είναι:

$$P(C) = P(ABC + A'BC + AB'C + A'B'C) = 0.09 + 0.04 + 0.06 + 0.02 = 0.21$$

β) Δεδομένου ότι κέρδισε ο Αστέρας, ποια είναι η πιο πιθανή αιτία (βροχή ή Ζούμπος);

Σύμφωνα με το κανόνα του Bayes υπολογίζουμε την a posteriori πιθανότητα για αυτό το πρόβλημα διάγνωσης:

$$P(A|C) = \frac{P(AC)}{P(C)} = \frac{P(ABC + AB'C)}{P(C)} = \frac{0.13}{0.21} \approx 0.62$$

$$P(B|C) = \frac{P(BC)}{P(C)} = \frac{P(ABC + A'BC)}{P(C)} = \frac{0.15}{0.21} \approx 0.71$$

Άρα πιο πιθανή αιτία είναι η βροχή για τη νίκη του Αστέρα.

(γ) Σύμφωνα με το μοντέλο του Θωμά τα ενδεχόμενα A και B είναι ανεξάρτητα. Είναι όμως τα A και B ανεξάρτητα δεδομένου του C;

Πρέπει να συγκρίνουμε το $P(A|BC)$ με το $P(A|C)$ σύμφωνα με τον ορισμό της υπό συνθήκη ανεξαρτησίας

$$P(A|BC) = \frac{P(ABC)}{P(BC)} = \frac{P(ABC)}{P(ABC + A'BC)} = \frac{0.09}{0.15} = 0.6$$

Το $P(A|C) \approx 0.62$ από το ερώτημα (β), οπότε $P(A|BC) \neq P(A|C)$ και άρα A, B είναι εξαρτημένα υπό τη συνθήκη C. Η γνώση του αποτελέσματος λοιπόν κάνει τις αιτίες του εξαρτημένες!!! Η γενίκευση αυτού του αποτελέσματος στα δίκτυα Bayes είναι γνωστή ως explaining away.

(δ) Σύμφωνα με τις εκτιμήσεις του θα πρέπει ο Θωμάς να παίξει τον αγώνα στο Στοίχημα άμα η νίκη δίνει απόδοση 6:1 και η ήττα 1.2:1; [Σημείωση: Το αναμενόμενο κέρδος ορίζεται ως το γινόμενο της απόδοσης επί την πιθανότητα του ενδεχομένου μείον το αρχικό κεφάλαιο.]

Για να υπολογίσουμε την πιθανότητα p που το Στοίχημα δίνει για νίκη του Αστέρα (q για ήττα) θεωρούμε ότι το αναμενόμενο κέρδος μας (σύμφωνα με το μοντέλο του Στοιχήματος (p, q)) είτε παίξουμε νίκη του Αστέρα είτε όχι θα είναι 0 (έτσι ώστε το Στοίχημα κατά μέσο όρο να μην βγάζει ούτε να χάνει χρήματα). Έστω ότι παίζουμε 1 ευρώ για τον Αστέρα. Το αναμενόμενο κέρδος μας σύμφωνα με το μοντέλο πρόβλεψης του Στοιχήματος είναι $6p$ μείον το κεφάλαιό μας 1. Το συνολικό κέρδος πρέπει να είναι 0. Άρα:

$$E(\text{Κέρδος}) = [6p + 0q] - 1 = 0 \Rightarrow p = 1/6 \approx 0.167$$

Αντίστοιχα έστω ότι παίζουμε 1 ευρώ ότι θα χάσει ο Αστέρας:

$$E(\text{Κέρδος}) = [0p + 1.2q] - 1 = 0 \Rightarrow q = 1/1.2 \approx 0.833$$

[Σημείωση: Στην πράξη το άθροισμα p και q θα είναι συνήθως μεγαλύτερο του 1 (συνήθως μεταξύ 1.1 και 1.2) ώστε το Στοίχημα να κερδίζει 10% με 20% κατά μέσο όρο. Οι πιθανότητες p και q πρέπει τότε να κανονικοποιηθούν, π.χ., $p' = p/(p+q)$ και $q' = 1 - p' = q/(p+q)$, ώστε να αθροίζουν σε 1. Στο συγκεκριμένο πρόβλημα υποθέτουμε ότι το Στοίχημα είναι μη κερδοσκοπικός οργανισμός, οπότε $p + q = 1$].
Συγκρίνοντας τα αποτελέσματα των δύο μοντέλων βλέπουμε ότι η πιθανότητα νίκης σύμφωνα με τον Θωμά είναι 0.21 ενώ σύμφωνα με το Στοίχημα 0.167. Άρα ο Θωμάς πρέπει να ποντάρει στη νίκη του Αστέρα που του δίνει για κάθε ευρώ αναμενόμενο κέρδος σύμφωνα με το μοντέλο του:

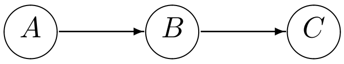
$$E(\text{Κέρδος}) = [6P(C) + 0P(C')] - 1 = 6 \times 0.21 - 1 = 1.26 - 1 = 0.26$$

Άρα το αναμενόμενο κέρδος του Θωμά είναι 26 λεπτά για κάθε ευρώ που ποντάρει στον Αστέρα.

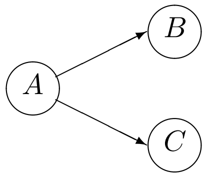
Ενθουσιασμένος ο Θωμάς με την ενδελεχή του ανάλυση έπαιξε 50 ευρώ στη νίκη του Αστέρα. Δυστυχώς όμως τα προγνωστικά επαληθεύτηκαν και ο Αστέρας έχασε. Το αναμενόμενο κέρδος του Θωμά ήταν 13 ευρώ. Το πραγματικό του κέρδος ήταν -50 ευρώ.

1.12 Θεμελιώδεις σχέσεις εξάρτησης

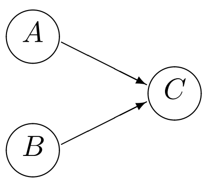
Ένας γράφος που αναπαριστά τις απευθείας εξαρτήσεις ενδεχομένων (ή τυχαίων μεταβλητών εν γένει όπως θα δούμε στο επόμενο κεφάλαιο) ονομάζεται δίκτυο Bayes (Bayesian network). Είδαμε ήδη στα παραπάνω παραδείγματα κάποια απλά δίκτυα. Τα τρία θεμελιώδη δομικά υλικά τέτοιων δικτύων είναι:



Όπου όλα τα ζεύγη (A, B), (A, C), (B, C) είναι εξαρτημένα, αλλά τα A και C είναι ανεξάρτητα δεδομένου του B, δηλαδή, $A \perp C | B$. [Σημείωση: Το παραπάνω δεν ισχύει για οποιαδήποτε A, B, C που αποτελούν χρονική αλυσίδα (π.χ., βλέπε παράδειγμα 1), ισχύει, π.χ., όταν υπάρχει η σχέση αιτίας αιτιατού μεταξύ των A, B και B, C (αλλά όχι μεταξύ απευθείας μεταξύ A, C)] Το δεύτερο παράδειγμα:



Όπου τα (A, B), (A, C), (B, C) είναι εξαρτημένα, αλλά τα B και C είναι ανεξάρτητα δεδομένου του A, δηλαδή, $B \perp C | A$. Αυτό είναι το δίκτυο με μια αιτία και δύο αιτιατά (παράδειγμα 2). Τέλος:



Όπου τα (A, C), (B, C) είναι εξαρτημένα, αλλά τα (A, B) είναι ανεξάρτητα. Δεδομένου όμως του C, τα A και B είναι εξαρτημένα! Αυτό είναι το δίκτυο με δυο αιτίες και ένα αιτιατό (παράδειγμα 3). Συνοψίζοντας:

ΤΟΠΟΛΟΓΙΑ	ΕΞΑΡΤΗΜΕΝΑ	ΑΝΕΞΑΡΤΗΤΑ
$A \rightarrow B \rightarrow C$	$(A, B), (B, C), (A, C)$	$(\mathbf{A}, \mathbf{C}) \vert \mathbf{B}$
$B \leftarrow A \rightarrow C$	$(A, B), (B, C), (A, C)$	$(\mathbf{B}, \mathbf{C}) \vert \mathbf{A}$
$A \rightarrow C \leftarrow B$	$(\mathbf{A}, \mathbf{B}) \vert \mathbf{C}, (B, C), (A, C)$	(A, B)

Με βάση τις παραπάνω σχέσεις μπορούμε να βρούμε τις ανεξαρτησίες για οποιαδήποτε δίκτυο Bayes (Bayesian network).

1.13 Σύνοψη

- Η πιθανότητα ορίζεται από τα 7 αξιώματα της άλγεβρας ενδεχομένων και τα 3 αξιώματα του μέτρου πιθανότητας.
- Η πράξη της πρόσθεσης αντιστοιχεί σε ένωση συνόλων ενδεχομένων (λογική πράξη ή). Η πράξη του πολλαπλασιασμού αντιστοιχεί σε τομή συνόλων ενδεχομένων (λογική πράξη και).
- Ο δειγματοχώρος περιέχει όλα τα ενδεχόμενα ενός πειραματικού μοντέλου με κάθε λεπτομέρεια. Τα δείγματα αποτελούν διαμερισμό (ξένα μεταξύ τους και η ένωση τους είναι το βέβαιο ενδεχόμενο).
- Το μέτρο πιθανότητας αναπαρίσταται γραφικά ως το εμβαδόν ενός δισδιάστατου Ευκλείδειου μοναδιαίου δειγματοχώρου.
- Πιθανότητα του A υπό συνθήκη B (όπου B είναι πλέον το βέβαιο ενδεχόμενο) αποδεικνύεται ίση με την πιθανότητα της τομής AB κανονικοποιημένη με την πιθανότητα του B.
- Η ανεξαρτησία A, B είναι μια στατιστική ιδιότητα των A, B που μας λέει ότι η γνώση του ενός δεν επηρεάζει τις στατιστικές ιδιότητες του άλλου, δηλαδή, $P(A|B) = P(A)$ οπότε και $P(B|A) = P(A)$. Η σχέση ανεξαρτησίας είναι στατιστική ιδιότητα που έχει να κάνει με το μέτρο πιθανότητας και δεν είναι εύκολο να αναπαρασταθεί με διαγράμματα Venn.
- Ανεξαρτησία υπό συνθήκη C (δηλαδή το C είναι το βέβαιο ενδεχόμενο). δεν συνεπάγεται ανεξαρτησία (ούτε το αντίθετο ισχύει).
- Ο κανόνας της αλυσίδας μας βοηθάει να υπολογίσουμε την από κοινού πιθανότητα ενδεχομένων.
- Ο κανόνας του διαμερισμού μας βοηθάει να υπολογίσουμε πιθανότητες ενδεχομένων ως άθροισμα της πιθανότητας των δειγμάτων που αποτελούν αυτό το ενδεχόμενο (για προβλήματα πρόβλεψης).
- Ο κανόνας του Bayes μας βοηθάει να υπολογίσουμε a posteriori πιθανότητες ως συνάρτηση των απευθείας και a priori πιθανοτήτων σε προβλήματα διάγνωσης
- Το δέντρο πιθανότητας είναι το κύριο εργαλείο για την λύση προβλημάτων πιθανοτήτων. Αναπαριστά τα ενδεχόμενα (στους κόμβους του δέντρου), τις σχέσεις εξάρτησης μεταξύ ενδεχομένων, τις απευθείας πιθανότητες

(πάνω στα κλαδιά του), τον από κοινού δειγματοχώρο και την πιθανότητα κάθε δείγματος (γινόμενο πιθανοτήτων μονοπατιού).

- Υπάρχουν θεμελιώδεις σχέσεις ανεξαρτησίας υπό συνθήκη που προκύπτουν από τις σχέσεις αιτίας αιτιατού μεταξύ ενδεχομένων.

Ασκήσεις

1. Αποδείξτε τις ιδιότητες της άλγεβρας ενδεχομένων χρησιμοποιώντας τα 7 αξιώματα.

2. Αποδείξτε τις ιδιότητες του μέτρου πιθανότητας χρησιμοποιώντας τα 3 αξιώματα.

3. Απλοποιήστε την πιθανότητα $P(A + A'B + AB'C' + ABC' + AB'C + ABC)$

4. Απλοποιήστε την πιθανότητα $P(ABC) = P(A|BC)P(B|C)P(C)$ δεδομένου ότι A, B ανεξάρτητα υπό συνθήκη C και B, C ανεξάρτητα.

5. Ένας τηλεπικοινωνιακός δίαυλος στέλνει 0 με πιθανότητα 0.6 και 1 με πιθανότητα 0.4. Η υπό συνθήκη πιθανότητα λάθους μετάδοσης (0 δεδομένου ότι στείλαμε 1 ή 1 δεδομένου ότι στείλαμε 0) είναι 0.1. Υπολογίστε την πιθανότητα να γίνει λάθος στη μετάδοση. Υπολογίστε την πιθανότητα να γίνει λάθος δεδομένου ότι στείλαμε 1. Υπολογίστε την πιθανότητα να στείλαμε 1, δεδομένου ότι λάβαμε 0.

6. Έχω δύο δοχεία. Το ένα περιέχει δυο κόκκινες και μια μαύρη μπάλα. Το δεύτερο μια κόκκινη και μια μαύρη. Η πιθανότητα να επιλέξω το πρώτο δοχείο είναι 0.3. Τραβάω μια μπάλα, άμα είναι μαύρη την επανατοποθετώ στο ίδιο δοχείο. Άμα είναι κόκκινη την τοποθετώ στο άλλο δοχείο. Επιλέγω ένα δοχείο (πάλι με πιθανότητα 0.3 το πρώτο) και τραβάω και μια δεύτερη μπάλα. Ποια είναι η πιθανότητα να τραβήξω δύο κόκκινες μπάλες. Ποια είναι η πιθανότητα να επιλέξω διαφορετικά δοχεία στις δυο προσπάθειες. Ποια είναι η πιθανότητα να έχω επιλέξει το πρώτο δοχείο στην πρώτη προσπάθεια δεδομένου ότι τράβηξα μια κόκκινη και μια μαύρη μπάλα. Σχεδιάστε τις εξαρτήσεις του δικτύου για τα 4 ενδεχόμενα (επιλογή δοχείου 1η, 2η, επιλογή μπάλας 1η, 2η).

7. Υπολογίστε την πιθανότητα να μπω με μια ζαριά σε πεντάπορτο στο τάβλι. Υπολογίστε την πιθανότητα αυτή αριθμητικά με τη χρήση υπολογιστή.

8. Δημιουργήστε ένα μοντέλο για τα ενδεχόμενα 'άργησα να κοιμηθώ', 'άργησα να ξυπνήσω', 'άργησα στο μάθημα'. Η επιλογή των πιθανοτήτων να γίνει ώστε να ταιριάζει στο προσωπικό σας μοντέλο. Αποδείξτε τις σχέσεις ανεξαρτησίας υπό συνθήκη για αυτό το δίκτυο.

9. Προσθέστε την πιθανή αιτία 'δεν μπορούσα να κοιμηθώ' στο παραπάνω δίκτυο και επαναλάβατε.

10. Εργάζεστε στο Στοίχημα. Υπολογίστε την πιθανότητα να καταλάβει μια Ελληνική ομάδα το Champion League τα επόμενα 10 χρόνια (φτιάξτε ένα αναλυτικό μοντέλο πρόβλεψης), καθώς και την απόδοση ώστε το Στοίχημα να κερδίζει 10% του τζίρου κατά μέσο όρο.

11. Τι αναπαριστά η παρακάτω πιθανότητα:

```
prob_p = sum([sum(rand(30,10000) < 0.2) <= 14])/10000;
```

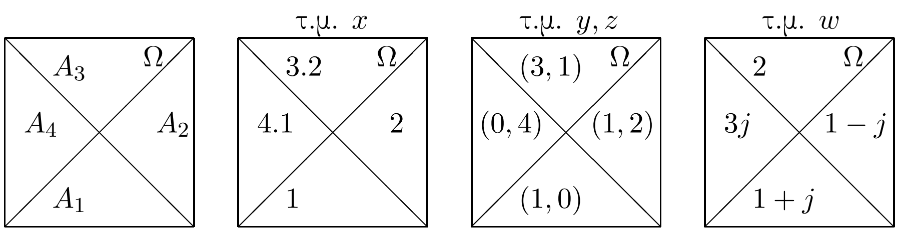
12. Έχω δυο κάλπικα νομίσματα με πιθανότητα 0.4 και 0.6 να φέρουν γράμματα αντίστοιχα. Επιλέγω το ένα από τα δύο και το αποτέλεσμα 30 ρίψεων είναι 16 γράμματα και 14 κορώνες. Υπολογίστε αριθμητικά την πιθανότητα να έχω επιλέξει το πρώτο ή δεύτερο νόμισμα δεδομένου του αποτελέσματος του πειράματος. Πόσο σίγουροι είστε σχετικά με το νόμισμα που έχω επιλέξει;

Κεφάλαιο 2

Διακριτές Τυχαίες Μεταβλητές

Οι τυχαίες μεταβλητές αποτελούν μια γενίκευση των κλασικών πιθανοτήτων που μάθαμε στο Κεφάλαιο 1. Συγκεκριμένα αντιστοιχίζουμε έναν (ή περισσοτέρους) πραγματικούς αριθμούς σε κάθε ένα από τα δείγματα του δειγματοχώρου (δηλαδή στα αποτελέσματα ενός πειράματος). Οι αριθμοί αυτοί αποτελούν απλά 'ταμπέλες', π.χ., αντί να ονομάζουμε ένα ενδεχόμενο A, το ονομάζουμε[1] ενδεχόμενο 3 ή ενδεχόμενο $(3, 4)$. Όπως και στις κλασικές πιθανότητες θέλουμε κάθε δείγμα να έχει μια και μόνο μία 'ταμπέλα', αλλά είναι σύνηθες οι ταμπέλες δύο δειγμάτων να είναι ίδιες (σ' αυτή την περίπτωση λέμε ότι τα δύο δείγματα ανήκουν στο ίδιο ενδεχόμενο). Άρα η αντιστοίχηση δειγμάτων σε αριθμούς είναι συνάρτηση (κάθε δείγμα παίρνει μοναδική τιμή). Επίσης η συνάρτηση αυτή ορίζει ένα διαμερισμό πάνω στο δειγματοχώρο μας.

Για παράδειγμα πάνω στον ίδιο διαμερισμό του δειγματοχώρου Ω, έχουμε ορίσει κλασικές πιθανότητες, τυχαία μεταβλητή, ζεύγος τυχαίων μεταβλητών και μιγαδική τυχαία μεταβλητή, ως εξής:



Για παράδειγμα το ενδεχόμενο A_1 των κλασικών πιθανοτήτων έχει μετονο-

[1] Προφανώς ο αριθμός αυτός συνήθως έχει φυσική ερμηνεία.

μαστεί σε ενδεχόμενο $x = 1$ όπου x τυχαία μεταβλητή στο δεύτερο σχήμα. Αντίστοιχα το A_1 έχει μετονομαστεί σε ενδεχόμενο $(y = 1, z = 0)$ όπου y, z τυχαίες μεταβλητές στο τρίτο σχήμα, ή ενδεχόμενο $w = 1 + j$ όπου w μιγαδική τυχαία μεταβλητή στο τέταρτο σχήμα. Η πιθανότητα αυτού του ενδεχομένου όμως παραμένει η ίδια (ανεξάρτητα από το πως το ονομάζω) και καθορίζεται από το πειραματικό μοντέλο που παράγει τα συγκεκριμένα δείγματα ή ενδεχόμενα, δηλαδή, $P(A_1) \equiv P(x = 1) \equiv P(y = 1 \cap z = 0) \equiv P(w = 1 + j)$.

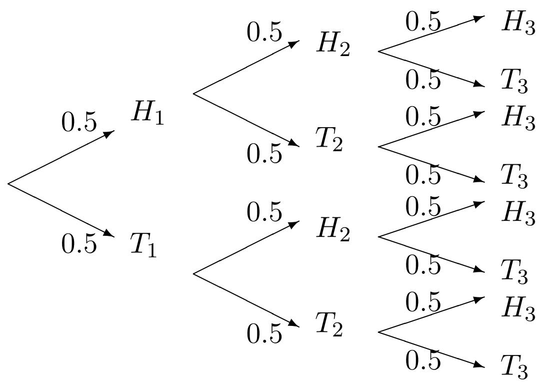
Ορισμός: Η διακριτή τυχαία μεταβλητή (discrete random variable) είναι μια συνάρτηση από κάθε δείγμα[2] του δειγματοχώρου σε ένα πραγματικό αριθμό.

Η τυχαία μεταβλητή δεν είναι τυχαία ούτε μεταβλητή.
Η τυχαία μεταβλητή είναι συνάρτηση από το δειγματοχώρο στο $\mathcal{R}$.
Η τυχαία μεταβλητή ορίζει ένα διαμερισμό του δειγματοχώρου.

Για παράδειγμα ορίζουμε δύο τυχαίες μεταβλητές πάνω στον δειγματοχώρο τριών ανεξάρτητων ρίψεων ενός νομίσματος[3]. Η τυχαία μεταβλητή h αντιστοιχίζει κάθε δείγμα με τον αριθμό από κορώνες (H = head) του δείγματος, π.χ., το δείγμα $H_1H_2H_3$ έχει τρεις κορώνες και άρα $h = 3$. Η τυχαία μεταβλητή r είναι ίση με την μεγαλύτερη αλληλουχία από κορώνες ή γράμματα για κάθε δείγμα, π.χ., το δείγμα $H_1H_2T_3$ έχει δύο κορώνες στη σειρά (αλλά όχι τρεις) και άρα $r = 2$, ενώ για το δείγμα $H_1T_2H_3$ η μεγαλύτερη αλυσίδα από κορώνες ή γράμματα είναι 1, οπότε $r = 1$. Ο δειγματοχώρος, η πιθανότητα κάθε δείγματος, και οι τιμές των τυχαίων μεταβλητών h, r για κάθε δείγμα είναι:

[2]Ισοδύναμα θα μπορούσαμε να ορίσουμε την τ.μ. σαν συνάρτηση από κάθε ενδεχόμενο σε ένα πραγματικό αριθμό. Γιατί;

[3]Η πιθανότητα κορώνα ή γράμματα για ένα 'δίκαιο' νόμισμα είναι 0.5. Άμα όμως γνωρίζουμε ποια πλευρά του νομίσματος είναι από πάνω (πριν ξεκινήσουμε την περιστροφή του νομίσματος) και πιάσουμε το νόμισμα στον αέρα τότε αυτή η πλευρά έχει πιθανότητα 0.51! Το παραπάνω απέδειξαν θεωρητικά και πειραματικά ο μαθηματικός Persi Diaconis και η ερευνητική του ομάδα (πηγή: http://en.wikipedia.org/wiki/Coin_flipping).



Δείγμα	$P(.)$	h	r
• $H_1H_2H_3$	$1/8$	3	3
• $H_1H_2T_3$	$1/8$	2	2
• $H_1T_2H_3$	$1/8$	2	1
• $H_1T_2T_3$	$1/8$	1	2
• $T_1H_2H_3$	$1/8$	2	2
• $T_1H_2T_3$	$1/8$	1	1
• $T_1T_2H_3$	$1/8$	1	2
• $T_1T_2T_3$	$1/8$	0	3

Ο υπολογισμός της πιθανότητας μιας τιμής τυχαίας μεταβλητής υπολογίζεται αθροίζοντας τα δείγματα που αντιστοιχούν σ' αυτή την τιμή, π.χ., το ενδεχόμενο $h = 2$ έχει πιθανότητα:

$$P(h = 2) = P(H_1H_2T_3 + H_1T_2H_3 + T_1H_2H_3) = 1/8 + 1/8 + 1/8 = 3/8$$

Αντίστοιχα υπολογίζουμε και την από κοινού πιθανότητα των τιμών δύο τυχαίων μεταβλητών, π.χ.,

$$P(h{=}2 \cap r{=}2) \equiv P(h{=}\,2, r{=}\,2) = P(H_1H_2T_3{+}T_1H_2H_3) = 1/8{+}1/8 = 1/4$$

Όπως είπαμε κάθε τυχαία μεταβλητή (ή ομάδα τυχαίων μεταβλητών) είναι ένας διαμερισμός του δειγματοχώρου και άρα ξέρουμε ότι το άθροισμα των πιθανοτήτων όλων των ενδεχομένων μια τυχαίας μεταβλητής είναι ίσο με τη μονάδα. Για παράδειγμα οι διαμερισμοί h, r, (h, r) καθώς και οι πιθανότητες κάθε ενδεχομένου είναι:

h	0	1	2	3	$\sum(.)$
$P(h)$	$1/8$	$3/8$	$3/8$	$1/8$	1

r	1	2	3	$\sum(.)$
$P(r)$	$1/4$	$1/2$	$1/4$	1

(h, r)	1	2	3	$\sum(.)$
0	0	0	$1/8$	$1/8$
1	$1/8$	$1/4$	0	$3/8$
2	$1/8$	$1/4$	0	$3/8$
3	0	0	$1/8$	$1/8$
$\sum(.)$	$1/4$	$1/2$	$1/4$	1

Πράγματι το άθροισμα των πιθανοτήτων όλων των ενδεχομένων είναι 1, π.χ.,

$$\sum_{h=0}^{3} \sum_{r=1}^{3} P(h, r) = 1$$

Επίσης είναι ενδιαφέρον να παρατηρήσουμε τον κανόνα του διαμερισμού, π.χ., για να βρούμε την τιμή του ενδεχομένου $h = 2$, μπορούμε να αθροίσουμε πάνω στον διαμερισμό r τις από κοινού πιθανότητες των ενδεχομένων $(h = 2, r)$

$$P(h{=}2) = \sum_{r=1}^{3} P(h{=}2, r) = 1/8 + 1/4 = 3/8$$

Εν γένει μπορούμε να γράψουμε ότι

$$\forall\, a \in \{0, 1, 2, 3\}, \quad P(h{=}a) = \sum_{r=1}^{3} P(h{=}a, r)$$

$$\forall\, b \in \{1, 2, 3\}, \quad P(r{=}b) = \sum_{h=0}^{3} P(h, r{=}b)$$

Ήδη βλέπουμε κάποια πλεονεκτήματα με την χρήση αριθμών αντί συμβόλων για την αναπαράσταση ενδεχομένων, όπως τη δυνατότητα χρήσης συναρτήσεων για την αναπαράσταση πιθανοτήτων. Για παράδειγμα η συνάρτηση $p_h(h_0)$ θα μπορούσε να αναπαριστά όλες τις πιθανότητες πάνω στον διαμερισμό της τ.μ. h: $\{h_0 = 0,\, h_0 = 1,\, h_0 = 2,\, h_0 = 3\}$.

2.1 Συνάρτηση Μάζας Πιθανότητας

Ορισμός: Η συνάρτηση μάζας πιθανότητας (probability mass function) $p_x(x_0)$ της διακριτής τυχαίας μεταβλητής x ισούται στο x_0 με το άθροισμα της πιθανότητας όλων των δειγμάτων που αντιστοιχούν στο ενδεχόμενο $x = x_0$.

Προφανώς $0 \leq p_x(x_0) \leq 1$ αξιωματικά από τον ορισμό του μέτρου πιθανότητας. Επίσης $\sum_{x_0} p_x(x_0) = 1$ γιατί τα ενδεχόμενα x αποτελούν διαμερισμό. Γενικεύοντας για δύο τυχαίες μεταβλητές:

Ορισμός: Η από κοινού συνάρτηση μάζας πιθανότητας (joint probability mass function) $p_{x,y}(x_0, y_0)$ των διακριτών τυχαίων μεταβλητών x, y ισούται στο (x_0, y_0) με το άθροισμα της πιθανότητας όλων των δειγμάτων που αντιστοιχούν στο ενδεχόμενο $x = x_0 \cap y = y_0$.

Από τον ορισμό του μέτρου πιθανότητας και επειδή τα ενδεχόμενα x, τα ενδεχόμενα y, και τα ενδεχόμενα (x, y) αποτελούν διαμερισμό μπορούμε να γράψουμε:

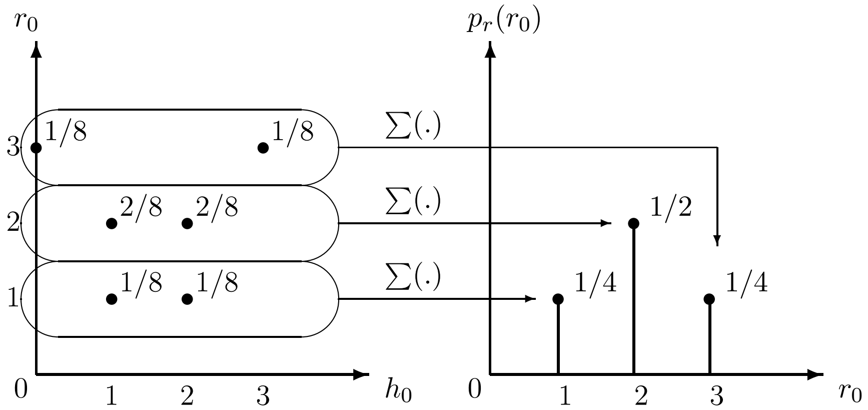
$$\forall\, x_0, y_0, \quad 0 \leq p_{x,y}(x_0, y_0) \leq 1$$

$$\sum_{x_0} \sum_{y_0} p_{x,y}(x_0, y_0) = 1$$

$$\forall\, x_0, \quad \sum_{y_0} p_{x,y}(x_0, y_0) = p_x(x_0)$$

$$\forall\, y_0, \quad \sum_{x_0} p_{x,y}(x_0, y_0) = p_y(y_0)$$

Οι δύο τελευταίες σχέσεις χρησιμοποιούν τη γνωστή ιδιότητα του διαμερισμού που στην περίπτωση των τυχαίων μεταβλητών (τ.μ.) ονομάζεται κανόνας οριακής πιθανότητας (marginal probability). Η από κοινού συνάρτηση μάζας πιθανότητας (Σ.Μ.Π.) των τ.μ. (h, r) από την τριπλή ρίψη νομίσματος, καθώς και η (οριακή) Σ.Μ.Π. της τ.μ. r (όπως υπολογίζεται από τον κανόνα του διαμερισμού) είναι:



Γενικότερα ό,τι έχουμε μάθει για κλασικές πιθανότητες στο Κεφάλαιο 1, ισχύει και για τις συναρτήσεις μάζας πιθανότητας, δηλαδή, ο ορισμός της υπό συνθήκη πιθανότητας, ο κανόνας της αλυσίδας, ο κανόνας του Bayes. Η κύρια διαφορά είναι ότι χρησιμοποιούμε μια συνάρτηση που αναπαριστά (μαζικά) όλες τις πιθανότητες των ενδεχομένων x και άρα οι σχέσεις πλέον ισχύουν για κάθε τιμή του x.

2.2 Συνάρτηση Κατανομής

Εκτός της συνάρτησης μάζας πιθανότητας $p_x(x_0) \equiv Prob(x = x_0)$, μπορώ να ορίσω και άλλες συναρτήσεις που περιγράφουν την πιθανότητα τ.μ., π.χ., $g_x(x_1, x_2) = Prob(x_1 < x \leq x_2)$, $h_x(x_0) = Prob(x \geq x_0)$, $w_x(x_0) = Prob(x \leq x_0^2)$. Κάθε μια από αυτές τις συναρτήσεις περιγράφουν (μαζικά) την πιθανότητα κάποιων ενδεχομένων και άρα και γι΄ αυτές ισχύουν οι βασικές ιδιότητες[4] που είδαμε στο Κεφάλαιο 1. Για παράδειγμα:

Ορισμός: Η συνάρτηση κατανομής πιθανότητας (cumulative probability mass function) ορίζεται ως: $F_x(x_0) \equiv Prob(x \leq x_0)$.

Ορισμός: Η από κοινού συνάρτηση κατανομής πιθανότητας (joint cumulative probability mass function) ορίζεται ως: $F_{x,y}(x_0, y_0) \equiv Prob(x \leq x_0, y \leq y_0)$.

Προφανώς $F_x(-\infty) = 0$ και $F_x(+\infty) = 1$. Επίσης εφόσον $x_1 < x_2 \Rightarrow F_x(x_1) \leq F_x(x_2)$. Τέλος: $Prob(x_1 < x \leq x_2) = F_x(x_2) - F_x(x_1)$. Αντίστοιχα $F_{x,y}(-\infty, -\infty) = 0$, $F_{x,y}(x_0, -\infty) = F_x(x_0)$, $F_{x,y}(+\infty, +\infty) = 1$. Η συνάρτηση κατανομής δεν χρησιμοποιείται πολύ για διακριτές τ.μ., πάντως είναι χρήσιμο να θυμόμαστε τη σχέση ανάμεσα στην Σ.Μ.Π. και την συνάρτηση κατανομής:

$$
\begin{array}{ccc}
\boxed{\begin{array}{c}\text{Συνάρτηση}\\ \text{Κατανομής}\\ F_x(x_0)\end{array}}
& \begin{array}{c} \xrightarrow{\quad p_x(x_0) = F_x(x_0 + \delta) - F(x_0 - \delta) \quad} \\[2ex] \xleftarrow{\quad F_x(x_0) = \sum_{x \leq x_0} p_x(x_0) \quad} \end{array}
& \boxed{\begin{array}{c}\text{Συνάρτηση}\\ \text{Μάζας}\\ \text{Πιθανότητας}\\ p_x(x_0)\end{array}}
\end{array}
$$

όπου δ είναι ένας πολύ μικρός αριθμός.

2.3 Σ.Μ.Π υπό συνθήκη και ανεξαρτησία Τ.Μ.

Ορίζουμε την υπό συνθήκη Σ.Μ.Π. $p_{x|y}(x_0|y_0)$ την συνάρτηση των τυχαίων μεταβλητών x και y, έτσι ώστε στο (x_0, y_0) η τιμή της να ισούται με την πιθανότητα να συμβεί το ενδεχόμενο x_0 δεδομένου ότι έχει συμβεί το y_0, δηλαδή, $p_{x|y}(x_0|y_0) \equiv Prob(x = x_0|y = y_0)$.

Η υπό συνθήκη Σ.Μ.Π. είναι μια συνάρτηση και του x και του y_0, ή αλλιώς για κάθε τιμή της συνθήκης $y = y_0$ έχουμε μια συνάρτηση του x. Βέβαια τα $x = x_0$ και $y = y_0$ είναι ενδεχόμενα, άρα όπως και για τις κλασικές πιθανότητες μπορώ να αποδείξω ότι:

$$\forall\, x_0, y_0, \quad p_{x|y}(x_0|y_0) = \frac{p_{x,y}(x_0, y_0)}{p_y(y_0)}$$

[4] Με την εξαίρεση της ιδιότητας του διαμερισμού που δεν ισχύει, γιατί τα ενδεχόμενα που αντιστοιχούν στις παραπάνω συναρτήσεις δεν αποτελούν διαμερισμό, π.χ. τα ενδεχόμενα $h \geq 1$ και $h \geq 2$ δεν είναι ξένα μεταξύ τους.

δηλαδή η υπό συνθήκη Σ.Μ.Π. είναι ίση με το λόγο της από κοινού Σ.Μ.Π. δια την οριακή Σ.Μ.Π. (της τ.μ. συνθήκης). Για παράδειγμα, η Σ.Μ.Π. του r υπό συνθήκη $h = 2$, είναι μια συνάρτηση του r που υπολογίζεται ως εξής:

$$p_{r|h}(r = 1|h = 2) = \frac{p_{h,r}(h = 2, r = 1)}{p_h(h = 2)} = \frac{1/8}{3/8} = \frac{1}{3}$$

$$p_{r|h}(2|2) = \frac{p_{h,r}(2,2)}{p_h(2)} = \frac{1/4}{3/8} = \frac{2}{3}$$

$$p_{r|h}(3|2) = \frac{p_{h,r}(2,3)}{p_h(2)} = \frac{0}{3/8} = 0$$

Βλέπω ότι ουσιαστικά έχω υπολογίσει τις πιθανότητες $Prob(r = 1|h = 2)$, $Prob(r = 2|h = 2)$, $Prob(r = 3|h = 2)$ όπως ακριβώς[5] και στο Κεφ. 1, απλά έχω χρησιμοποιήσει διαφορετικό φορμαλισμό και το αποτέλεσμα έχει τοποθετηθεί σε μια συνάρτηση:

$$p_{r|h}(r_0|h = 2) = \left\{ \begin{array}{ll} 1/3, & r_0 = 1 \\ 2/3, & r_0 = 2 \\ 0, & \text{αλλού} \end{array} \right\}$$

2.4 Ανεξαρτησία Τ.Μ.

Ορισμός: οι τυχαίες μεταβλητές x, y είναι ανεξάρτητες (independent) αν και μόνο αν:

$$\forall\ x_0, y_0, \quad p_{x|y}(x_0|y_0) = p_x(x_0)$$

δηλαδή, εάν κάθε ένα από τα ενδεχόμενα του διαμερισμού x είναι ανεξάρτητα με κάθε ένα από τα ενδεχόμενα του διαμερισμού y. Όπως και για τις κλασικές πιθανότητες προκύπτουν οι ιδιότητες (για τ.μ. x, y ανεξάρτητες):

$$\forall\ x_0, y_0, \quad p_{y|x}(y_0|x_0) = p_y(y_0)$$

$$\forall\ x_0, y_0, \quad p_{x,y}(x_0, y_0) = p_x(x_0)p_y(y_0)$$

Είναι εύκολο να δείξουμε ότι οι τρεις σχέσεις είναι ισοδύναμες χρησιμοποιώντας τον ορισμό της υπό συνθήκη Σ.Μ.Π. (βλ. Κεφ. 1), και άρα οι ιδιότητες αποτελούν και ισοδύναμους ορισμούς ανεξαρτησίας τ.μ. Η τελευταία ιδιότητα είναι ιδιαίτερα σημαντική: για ανεξάρτητες τ.μ. η από κοινού Σ.Μ.Π μπορεί να

[5] Εφόσον οι συναρτήσεις $p_{h,r}(h_0, r_0)$ και $p_h(h_0)$ δίνονται σε αναλυτική μορφή τότε μπορώ απλά να υπολογίσω το λόγο των συναρτήσεων και να εκφράσω το αποτέλεσμα επίσης σε αναλυτική μορφή. Κατανοούμε το πλεονέκτημα της χρήσης συναρτήσεων ενδεχομένων.

υπολογιστεί ως το γινόμενο των οριακών Σ.Μ.Π.[6].

Ορισμός: τα ενδεχόμενα x, y είναι ανεξάρτητα υπό συνθήκη $z = z_0$ (conditionally independent) αν και μόνο αν:

$$\forall\ x_0, y_0, \quad p_{x|y,z}(x_0|y_0, z\!=\!z_0) = p_{x|z}(x_0|z\!=\!z_0)$$

Προκύπτουν οι ιδιότητες (που είναι και ισοδύναμοι ορισμοί):

$$\forall\ x_0, y_0, \quad p_{y|x,z}(y_0|x_0, z\!=\!z_0) = p_{y|z}(y_0|z\!=\!z_0)$$

$$\forall\ x_0, y_0, \quad p_{x,y|z}(x_0, y_0|z\!=\!z_0) = p_{x|z}(x_0|z\!=\!z_0)\ p_{y|z}(y_0|z\!=\!z_0)$$

Η φυσική ερμηνεία της ανεξαρτησίας υπό συνθήκη είναι αντίστοιχη με αυτή των κλασικών πιθανοτήτων, απλά εδώ ζητάμε ανεξαρτησία για κάθε ζεύγος τιμών των τ.μ. πάνω στον διαμερισμό (x, y). Η έννοια της ανεξαρτησίας υπό συνθήκη είναι θεμελιώδης για προβλήματα μοντελοποίησης όπως είδαμε στο Κεφ. 1.

2.5 Κανόνας Αλυσίδας και Κανόνας Bayes

Οι κανόνες αλυσίδας και Bayes ισχύουν για κάθε ενδεχόμενο και άρα και για κάθε τιμή τυχαίας μεταβλητής. Συγκεκριμένα μπορούμε να γράψουμε:

$$\forall\ x_0, y_0, \quad p_{x,y}(x_0, y_0) = p_{x|y}(x_0|y_0)\ p_y(y_0) = p_{y|x}(y_0|x_0)\ p_x(x_0)$$

$$\forall\ x_0, y_0, z_0, \quad p_{x,y,z}(x_0, y_0, z_0) = p_{x|y,z}(x_0|y_0, z_0)\ p_{y|z}(y_0|z_0)\ p_x(x_0)$$

$$\ldots$$

και επίσης τον κανόνα του Bayes:

$$\forall\ x_0, y_0, \quad p_{x|y}(x_0|y_0) = \frac{p_{y|x}(y_0|x_0)p_x(x_0)}{p_y(y_0)} = \frac{p_{y|x}(y_0|x_0)p_x(x_0)}{\sum_{x_0} p_{y|x}(y_0|x_0)p_x(x_0)}$$

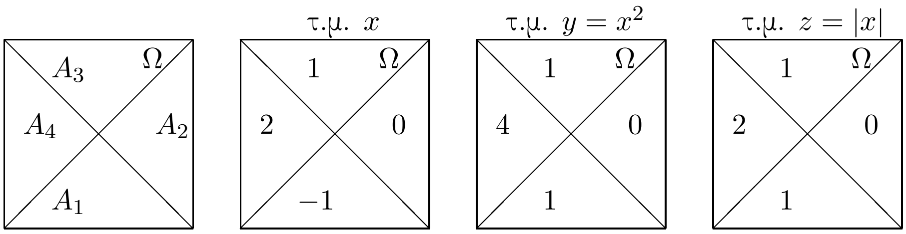
όπου στον παρανομαστή έχουμε χρησιμοποιήσει την ιδιότητα του διαμερισμού.

Μέχρι εδώ έχουμε απλά εισαγάγει έναν καινούργιο φορμαλισμό (συναρτήσεις ενδεχομένων) και έχουμε κάνει μια καλή επανάληψη των κλασικών πιθανοτήτων. Στη συνέχεια εισάγουμε δυο καινούργιες έννοιες: (α) συναρτήσεις τυχαίων μεταβλητών, και (β) αναμενόμενη τιμή τυχαίας μεταβλητής (και συναρτήσεων τυχαίων μεταβλητών).

[6] Εν γένει δεν υπάρχει τρόπος να υπολογιστεί η από κοινού Σ.Μ.Π. από τις οριακές Σ.Μ.Π. Το αντίθετο είναι πάντα εφικτό μέσω της ιδιότητας του διαμερισμού.

2.6 Συναρτήσεις Τυχαίων Μεταβλητών

Οι συνάρτηση μιας τυχαίας μεταβλητής x, π.χ., $y = x^2$ ή $z = |x|$, είναι επίσης τυχαία μεταβλητή[7]. Είναι σημαντικό να θυμόμαστε ότι τόσο η τ.μ. x όσο και οι συναρτήσεις της τ.μ. $y = x^2$, $z = |x|$ ορίζονται πάνω στον ίδιο δειγματοχώρο, και άρα η Σ.Μ.Π. των y, z προκύπτει όπως ακριβώς και για την x (δηλαδή αθροίζοντας την πιθανότητα των δειγμάτων που αντιστοιχούν στις y και z αντίστοιχα). Παρά το ότι οι x, y, z ορίζονται πάνω στον ίδιο δειγματοχώρο, ο τρόπος με τον οποίο διαμερίζουν τον δειγματοχώρο σε ενδεχόμενα μπορεί να είναι διαφορετικός (όπως βέβαια και ο αριθμός που αντιστοιχεί σε κάθε ενδεχόμενο). Για παράδειγμα βλέπουμε ένα δειγματοχώρο με τέσσερα ισοπίθανα δείγματα A_1, A_2, A_3, A_4, τον διαμερισμό και τιμές της τ.μ. x (όπως αντιστοιχούν στα τέσσερα δείγματα), καθώς και τους διαμερισμούς και τις τιμές των συναρτήσεων της x: τ.μ. $y = x^2$ και τ.μ. $z = |x|$.



Βλέπουμε ότι ο διαμερισμός του δειγματοχώρου είναι διαφορετικός[8] για την τ.μ. x ($\{A_1, A_2, A_3, A_4\}$) και τις τ.μ. y, z ($\{A_1 \cup A_3, A_2, A_4\}$). Επίσης παρά το ότι οι y, z ορίζουν τον ίδιο διαμερισμό ενδεχομένων οι τιμές των τ.μ. για το ενδεχόμενο A_4 είναι διαφορετικές. Τέλος, σημειώνουμε ότι η Σ.Μ.Π. των y, z μπορούν να υπολογιστούν (και) απευθείας από την Σ.Μ.Π. της x (χωρίς γνώση του πειραματικού δειγματοχώρου).

Παράδειγμα: Υπολογίστε την Σ.Μ.Π. μιας ζαριάς στο τάβλι. Υπολογίστε την Σ.Μ.Π. της τ.μ. που ισούται με το άθροισμα των δύο ζαριών (και ίση με το δύο φορές το άθροισμα των ζαριών όταν φέρουμε διπλές).

Η Σ.Μ.Π. της τ.μ. x που αντιστοιχεί στα αποτελέσματα της ρίψης ενός ζαριού

[7] Η απόδειξη: η τ.μ. x είναι μια συνάρτηση από τον χώρο ενδεχομένων στους πραγματικούς αριθμούς $\Omega \to \mathcal{R}$, η συνάρτηση τ.μ. $y = f(x)$ είναι μια συνάρτηση από πραγματικούς σε πραγματικούς αριθμούς $\mathcal{R} \to \mathcal{R}$ και άρα η σύνθεση των δύο συναρτήσεων είναι μια συνάρτηση από το χώρο ενδεχομένων στους πραγματικούς αριθμούς $\Omega \to \mathcal{R}$ (η νέα τ.μ. y).

[8] Ο διαμερισμός συναρτήσεων της διακριτής τ.μ. x μπορεί να έχει ίσο ή μικρότερο αριθμό ενδεχομένων από τον διαμερισμό της x (γιατί;).

είναι:

$$p_x(x_0) = \left\{ \begin{array}{ll} 1/6, & x_0 \in \{1,2,3,4,5,6\} \\ 0, & \text{αλλού} \end{array} \right\}$$

Η τ.μ. y (δεύτερο ζάρι) ακολουθεί την ίδια κατανομή και x, y είναι ανεξάρτητες, οπότε η από κοινού Σ.Μ.Π. είναι:

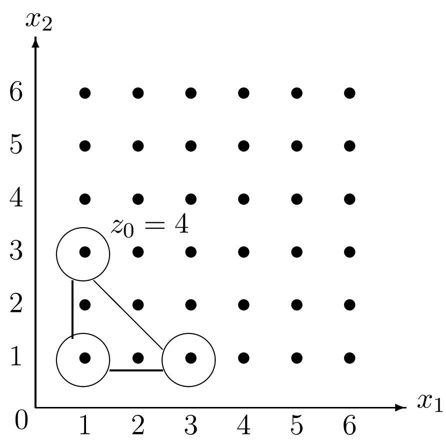
$$p_{x,y}(x_0, y_0) = p_x(x_0)p_y(y_0) = \left\{ \begin{array}{ll} 1/36, & x_0, y_0 \in \{1,2,3,4,5,6\} \\ 0, & \text{αλλού} \end{array} \right\}$$

Η τ.μ. z που αντιστοιχεί στο άθροισμα δυο ζαριών (λαμβάνοντας υπόψη και τις διπλές) μπορεί να εκφραστεί ως συνάρτηση των τ.μ. x, y ως εξής:

$$z = \left\{ \begin{array}{ll} x+y, & x \neq y \\ 4x, & x = y \end{array} \right\}$$

Αθροίζουμε την πιθανότητα όλων των δειγμάτων που αντιστοιχούν σε κάθε τιμή της z (π.χ., το ενδεχόμενο $z_0 = 4$ αντιστοιχεί στα δείγματα $\{(1,3),(3,1),(1,1)\}$):

$$p_z(z_0) = \left\{ \begin{array}{ll} 2/36, & z_0 = 3 \\ 3/36, & z_0 = 4 \\ 4/36, & z_0 = 5 \\ 4/36, & z_0 = 6 \\ 6/36, & z_0 = 7 \\ 5/36 & z_0 = 8 \\ 4/36 & z_0 = 9 \\ 2/36 & z_0 = 10 \\ 2/36 & z_0 = 11 \\ 1/36 & z_0 = 12 \\ 1/36 & z_0 = 16 \\ 1/36 & z_0 = 20 \\ 1/36 & z_0 = 24 \end{array} \right\}$$



Βλέπουμε ότι το πιο συνηθισμένο άθροισμα είναι το 7 και τα πιο σπάνια τα $12, 16, 20, 24$. Η πιθανότητα να φέρουμε ζαριά με άθροισμα μεγαλύτερο του 10 είναι μόλις $1/6$.

2.7 Αναμενόμενη Τιμή

Ορισμός: Η αναμενόμενη τιμή (expected value) μιας τυχαίας μεταβλητής x ορίζεται ως ο αριθμητικός μέσος όρος των τιμών που λαμβάνει με βάρος την πιθανότητα εμφάνισής τους:

$$E(x) = \sum_{x_0} x_0 \, p_x(x_0)$$

Η αναμενόμενη τιμή της τ.μ. x, $E(x)$ συχνά συμβολίζεται ως μ_x ή $\bar{x}$. Συχνά η μ_x δεν αντιστοιχεί σε ενδεχόμενο του πειράματος $p_x(\mu_x) = 0$, π.χ., για το παράδειγμα με το ζάρι η αναμενόμενη τιμή είναι:

$$E(x) = \sum_{x_0} x_0 \, p_x(x_0) = \sum_{x_0=1}^{6} x_0 \, \frac{1}{6} = \frac{1+2+3+4+5+6}{6} = 3.5$$

Παρά το ότι η τιμή $\mu_x = 3.5$ δεν έχει φυσικό νόημα, μας πληροφορεί ότι οι τιμές της τ.μ. x βρίσκονται γύρω από το 3.5. Δύο άλλοι τρόποι να εξηγήσουμε την έννοια της αναμενόμενης τιμής είναι:

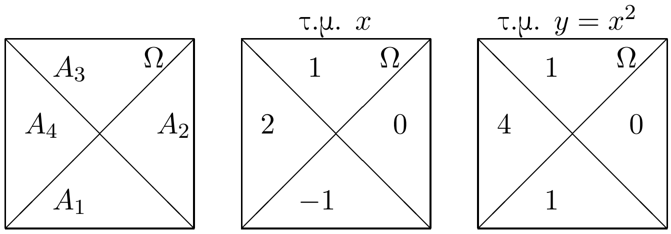
- Το 'κέντρο βάρους' της Σ.Μ.Π. $p_x()$. Πράγματι, για συμμετρικές κατανομές ο άξονας $x = \mu_x$ είναι ο άξονας συμμετρίας (δηλαδή 'κόβει' την Σ.Μ.Π. στη μέση).

- Η μέση τιμή μιας γεννήτρια n τυχαίων αριθμών που ακολουθούν την κατανομή $p_x()$ (όταν $n \to \infty$). Για παράδειγμα μια γεννήτρια που παράγει 0 και 1, με τριπλάσιο πλήθος 1 (από 0), π.χ., $\{0,1,1,0,1,0,1,1,0,0,1,1,1,...\}$, έχει μέση τιμή (στο όριο) $(1+1+1+0)/4 = 0.75$. Η Σ.Μ.Π. των τυχαίων αριθμών της γεννήτριας είναι $p_x(x_0 = 0) = 1/4$, $p_x(x_0 = 1) = 3/4$ και άρα η αναμενόμενη τιμή της κατανομής είναι $\sum_{x_0} x_0 p_x(x_0) = 0 \times 0.25 + 1 \times 0.75 = 0.75$.

Μπορούμε να γενικεύσουμε την έννοια την αναμενόμενης τιμής για συναρτήσεις τ.μ. ως εξής:

$$E(g(x)) \equiv \overline{g(x)} = \sum_{x_0} g(x_0) \, p_x(x_0)$$

Η πρώτη μας αντίδραση είναι γιατί χρησιμοποιούμε την $p_x(x_0)$ αντί για την $p_{g(x)}(g(x_0))$ για στον υπολογισμό του $E(g(x))$; Σύμφωνα με τον ορισμό της αναμενόμενης τιμής σίγουρα μπορούμε να υπολογίσουμε την αναμενόμενη τιμή της $g(x)$ και ως $\sum_{g(x_0)} g(x_0) p_{g(x)}(g(x_0))$. Προσέξτε ότι στην πρώτη περίπτωση αθροίζουμε πάνω στον διαμερισμό ενδεχομένων της τ.μ. x ενώ στην δεύτερη πάνω στον διαμερισμό της τ.μ. $g(x)$. Για επιβεβαίωση υπολογίστε την αναμενόμενη τιμή της τ.μ. $g(x) = x^2$, στο παρακάτω παράδειγμα:



Υποθέτουμε ότι $P(A_1)=P(A_2)=P(A_3)=P(A_4) = 1/4$, οπότε $p_x(x_0) = 1/4$, $x_0 \in \{-1, 0, 1, 2\}$ και $p_{g(x)}(1) = 1/2$, $p_{g(x)}(0) = p_{g(x)}(4) = 1/4$. Οπότε:

$$\sum_{g(x_0)} g(x_0) p_{g(x)}(g(x_0)) = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 4 \times \frac{1}{4} = 1.5$$

$$\sum_{x_0} g(x_0) p_x(x_0) = (-1)^2 \times \frac{1}{4} + 0^2 \times \frac{1}{4} + 1^2 \times \frac{1}{4} + 2^2 \times \frac{1}{4} = 1.5$$

Βλέπουμε ότι είναι ισοδύναμο να υπολογίσουμε το άθροισμα $g(x_0)p_x(x_0)$ πάνω στο διαμερισμό της τ.μ. x ή το άθροισμα $p_{g(x)}(g(x_0))$ πάνω στον διαμερισμό της τ.μ. $g(x)$. Είναι σημαντικό ότι μπορούμε να υπολογίσουμε την αναμενόμενη τιμή συναρτήσεων τ.μ. χωρίς να χρησιμοποιήσουμε την Σ.Μ.Π. τους!

Ορισμός: Η αναμενόμενη τιμή υπό συνθήκη[9] A μιας τυχαίας μεταβλητής x ορίζεται ως:

$$E(x|A) = \sum_{x_0} x_0 \; p_{x|A}(x_0|A)$$

Πάλι μπορούμε να γενικεύσουμε για συνάρτηση τυχαίων μεταβλητών $g(x)$:

$$E(g(x)|A) = \sum_{x_0} g(x_0) \; p_{x|A}(x_0|A)$$

Κλείνουμε με την γενίκευση για συναρτήσεις πολλών τυχαίων μεταβλητών:

$$E(g(x,y)) = \sum_{x_0} \sum_{y_0} g(x_0, y_0) \; p_{x,y}(x_0, y_0)$$

$$E(g(x,y)|A) = \sum_{x_0} \sum_{y_0} g(x_0, y_0) \; p_{x,y|A}(x_0, y_0|A)$$

Και πάλι αθροίζουμε πάνω σε όλα τα ενδεχόμενα του (από κοινού) διαμερισμού (x, y) την τιμή της συνάρτησης $g(x, y)$ επί την πιθανότητα εμφάνισης του ενδεχομένου. Οπότε η φυσική ερμηνεία της αναμενόμενης τιμής παραμένει η ίδια για μία ή περισσότερες τ.μ., αδέσμευτη ή υπό συνθήκη.

[9] Εάν A είναι το ενδεχόμενο που αντιστοιχεί στην τιμή μιας τ.μ., π.χ., $y = y_0$, μπορούμε να γράψουμε: $E(x|y = y_0) = \sum_{x_0} x_0 \; p_{x|y}(x_0|y = y_0)$. Παρά το ότι η $E(x)$ είναι σταθερά, η $E(x|y)$ είναι τ.μ. ορισμένη πάνω στον δειγματοχώρο της y (γιατί;).

2.8 Ροπές και Κεντρικές Ροπές

Ορισμός: Οι ροπές τάξης n της τ.μ. x ορίζονται ως οι αναμενόμενες τιμές των συναρτήσεων x^n. Άρα:

$$\text{Ροπή 1-ης τάξης:} \qquad E(x) = \sum_{x_0} x_0 \, p_x(x_0)$$

$$\text{Ροπή 2-ης τάξης:} \qquad E(x^2) = \sum_{x_0} x_0^2 \, p_x(x_0)$$

$$\text{Ροπή 3-ης τάξης:} \qquad E(x^3) = \sum_{x_0} x_0^3 \, p_x(x_0)$$

$$\ldots$$

Ορισμός: Οι κεντρικές ροπές τάξης n της τ.μ. x ορίζονται ως οι αναμενόμενες τιμές των συναρτήσεων $(x - E(x))^n$. Άρα:

$$\text{Κεντρική Ροπή 1-ης τάξης:} \qquad E(x - E(x)) = \sum_{x_0} (x_0 - E(x)) \, p_x(x_0) = 0$$

$$\text{Κεντρική Ροπή 2-ης τάξης:} \qquad E((x - E(x))^2) = \sum_{x_0} (x_0 - E(x))^2 \, p_x(x_0)$$

$$\text{Κεντρική Ροπή 3-ης τάξης:} \qquad E((x - E(x))^3) = \sum_{x_0} (x_0 - E(x))^3 \, p_x(x_0)$$

$$\ldots$$

Οι ροπές και οι κεντρικές ροπές περιγράφουν τη μορφή της Σ.Μ.Π.: το κέντρο βάρους της (ροπή 1-ης τάξης), τη διασπορά των τιμών γύρω από την αναμενόμενη τιμή (κεντρική ροπή 2-ης τάξης), το βαθμό συμμετρίας της Σ.Μ.Π. (κεντρική ροπή 3-ης τάξης) κ.λ.π. Συνήθως η στατιστική ανάλυση τ.μ. x περιορίζεται σε υπολογισμό των στατιστικών ιδιοτήτων μέχρι και δεύτερης τάξης δηλαδή στον υπολογισμό της ροπής 1-ης τάξης (αναμενόμενη τιμή μ_x) και την κεντρική ροπή 2-ης τάξης (διασπορά σ_x^2).

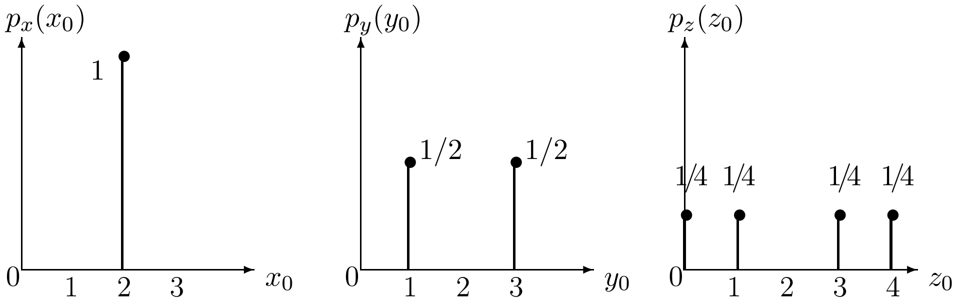
Ορισμός: Η κεντρική ροπή 2ης τάξης της τ.μ. x ονομάζεται διασπορά ή διακύμανση[10] (variance) και συμβολίζεται ως σ_x^2. Η τετραγωνική της ρίζα σ_x ονομάζεται τυπική απόκλιση σ_x (standard deviation). Άρα:

$$\sigma_x^2 \equiv E((x - E(x))^2) = \sum_{x_0} (x_0 - E(x))^2 \, p_x(x_0)$$

[10]Καμιά φορά η διακύμανση ορίζεται ως η ροπή 2ης τάξης (αντί της κεντρικής ροπής). Για αποφυγή σύγχυσης στη συνέχεια χρησιμοποιούμε μόνο τον όρο διασπορά για να περιγράφουμε την κεντρική ροπή 2ης τάξης.

Οι στατιστικές ιδιότητες υψηλότερης τάξης (higher order statistics) όπως η κεντρική ροπή 3-ης τάξης (skewness) και 4-ης τάξης (kurtosis) χρησιμοποιούνται λιγότερο στην πράξη[11].

Παράδειγμα: Υπολογίστε την αναμενόμενη τιμή και διασπορά για τις τ.μ. x, y, z των οποίων οι Σ.Μ.Π. δίδονται παρακάτω:



$$\mu_x \equiv E(x) = \sum_{x_0} x_0 \, p_x(x_0) = 2 \times 1 = 2$$

$$\mu_y \equiv E(y) = \sum_{y_0} y_0 \, p_y(y_0) = 1 \times 0.5 + 3 \times 0.5 = 2$$

$$\mu_z \equiv E(z) = \sum_{z_0} z_0 \, p_z(z_0) = 0 \times 0.25 + 1 \times 0.25 + 3 \times 0.25 + 4 \times 0.25 = 2$$

$$\sigma_x^2 \equiv E((x - E(x))^2) = \sum_{x_0} (x_0 - E(x))^2 \, p_x(x_0) = (2-2)^2 \times 1 = 0$$

$$\sigma_y^2 \equiv E((y - E(y))^2) = \sum_{y_0} (y_0 - E(y))^2 \, p_y(y_0) = (1-2)^2 \times 0.5 + (3-2)^2 \times 0.5 = 1$$

$$\sigma_z^2 = \sum_{z_0} (x_0 - E(z))^2 \, p_z(z_0) = ((-2)^2 + (-1)^2 + 1^2 + 2^2) \times 0.25 = 2.5$$

Βλέπουμε ότι και οι τρεις τ.μ. έχουν την ίδια αναμενόμενη τιμή 2, επειδή οι τιμές και των τριών βρίσκονται γύρω από το 2. Η διασπορά των τιμών των x, y, z όμως είναι πολύ διαφορετική. Για την τ.μ.[12] x η διασπορά είναι $\sigma_x^2 = 0$, για την y, $\sigma_y^2 = 1$ και για την z, $\sigma_z^2 = 2.5$. Όσο μεγαλύτερη ποσότητα μάζας πιθανότητας έχουμε μακριά από την αναμενόμενη τιμή τόσο μεγαλύτερη είναι η τυπική απόκλιση και διασπορά της τ.μ.

[11] Ο λόγος είναι ότι απαιτείται μεγάλος όγκος δεδομένων για την ακριβή εκτίμηση (αριθμητικά) ροπών υψηλών τάξεων.

[12] Στην πράξη η x δεν είναι τ.μ. είναι η σταθερά $x = 2$.

2.9 Θεμελιώδεις στατιστικές ιδιότητες

Ιδιότητα 1: Για πραγματικό αριθμό $x_0 \in \mathcal{R}$, η αναμενόμενη τιμή του είναι $E(x_0) = x_0$ και η διασπορά του είναι 0. Ο πραγματικός αριθμός x_0 αντιστοιχεί με μια τ.μ. με Σ.Μ.Π. $p_x(x{=}x_0) = 1$ και 0 αλλού. Άρα η αναμενόμενη τιμή είναι x_0 και η διασπορά 0 (βλ. προηγούμενο παράδειγμα). Άρα και

$$E(E(x)) = E(x)$$

Ιδιότητα 2: Η αναμενόμενη τιμή την συνάρτησης τ.μ. $y = ax$ όπου $a \in R$ είναι

$$E(ax) = \sum_{x_0} a x_0 \, p_x(x_0) = a \sum_{x_0} x_0 \, p_x(x_0) = aE(x)$$

Ιδιότητα 3: **Η αναμενόμενη τιμή αθροίσματος $x + y$ τ.μ. ισούται με το άθροισμα των αναμενόμενων τιμών των τ.μ. x, y**

$$E(x{+}y) {=} \sum_{x_0}\sum_{y_0}(x_0{+}y_0)p_{x,y}(x_0,y_0) {=} \sum_{x_0}\sum_{y_0} x_0 p_{x,y}(x_0,y_0) {+} \sum_{x_0}\sum_{y_0} y_0 p_{x,y}(x_0,y_0)$$

$$= \sum_{x_0} x_0 \sum_{y_0} p_{x,y}(x_0,y_0) {+} \sum_{y_0} y_0 \sum_{x_0} p_{x,y}(x_0,y_0) = \sum_{x_0} x_0 \, p_x(x_0) {+} \sum_{y_0} y_0 \, p_y(y_0) \Rightarrow$$

$$E(x + y) = E(x) + E(y) \quad \text{ή} \quad \mu_{x+y} = \mu_x + \mu_y$$

όπου $\sum_{y_0} p_{x,y}(x_0, y_0) = p_x(x_0)$ λόγω του διαμερισμού (οριακή Σ.Μ.Π.). Η ιδιότητα αυτή **ισχύει για κάθε τ.μ.** x, y όπως αποδείξαμε.

Ιδιότητα 4 [Γενίκευση της 2,3]: Η αναμενόμενη τιμή κάθε γραμμικού συνδυασμού τ.μ. (ή συνάρτησης αυτών των τ.μ.) ισούται με τον γραμμικό συνδυασμό των αναμενόμενων τιμών των τ.μ. (ή αναμενόμενης τιμής της συνάρτησης αυτών):

$$\forall \text{ τ.μ. } x_i, f_i(x_i),\ a_i \in \mathcal{R},\ i = 1, 2...n : \quad E\left(\sum_{i=1}^{n} a_i f_i(x_i)\right) = \sum_{i=1}^{n} a_i E(f_i(x_i))$$

Ιδιότητα 5: **Η διασπορά τυχαίας μεταβλητής ισούται με τη διαφορά ανάμεσα στη ροπή δεύτερης τάξης και το τετράγωνο της αναμενόμενης τιμής της:**

$$E((x{-}E(x))^2) = E(x^2{-}2xE(x){+}(E(x))^2) = E(x^2){-}2E(x)E(x){+}(E(x))^2 \Rightarrow$$

$$\sigma_x^2 \equiv E((x - E(x))^2) = E(x^2) - [E(x)]^2 \quad \text{ή} \quad E(x^2) = \sigma_x^2 + [E(x)]^2$$

Προσοχή: Οι παρακάτω ιδιότητες ισχύουν ΜΟΝΟ για ανεξάρτητες τ.μ.

Ιδιότητα 6: Εάν οι τ.μ. x, y είναι ανεξάρτητες τότε η αναμενόμενη τιμή του γινομένου τους xy ισούται με το γινόμενο των αναμενόμενων τιμών των x, y.

$$E(xy) = \sum_{x_0}\sum_{y_0}(x_0 y_0)\; p_{x,y}(x_0, y_0) = \sum_{x_0}\sum_{y_0} x_0 y_0\; p_x(x_0)\; p_y(y_0) =$$

$$= \left(\sum_{x_0} x_0 p_x(x_0)\right)\left(\sum_{y_0} y_0 p_y(y_0)\right) \Rightarrow \text{ για ανεξ. τ.μ. } x,\, y\text{: } E(xy) = E(x)E(y)$$

Ιδιότητα 7 [Γενίκευση της 6]: Εάν οι τ.μ. x, y είναι ανεξάρτητες τότε η αναμενόμενη τιμή του γινομένου συναρτήσεών τους $f(x)g(y)$ ισούται με το γινόμενο των αναμενόμενων τιμών των συναρτήσεων $f(x)$, $g(y)$. Άρα:

$$\text{για ανεξ. τ.μ. } x,\, y\text{:} \quad E(f(x)g(y)) = E(f(x))E(g(y))$$

$$\text{για ανεξ. τ.μ. } x_1,\, x_2,\, \ldots\, x_n\text{:} \quad E\left(\prod_{i=1}^{n} f_i(x_i)\right) = \prod_{i=1}^{n} E(f_i(x_i))$$

π.χ., εάν x, y είναι ανεξάρτητες[13] τότε $E(x^2 y) = E(x^2)E(y)$, $E(x^2 y^3) = E(x^2)E(y^3)$.

Ιδιότητα 8: Εάν οι τ.μ. x, y είναι ανεξάρτητες τότε η διασπορά του αθροίσματός τους ισούται με το άθροισμα των διασπορών των x, y. Εν γένει:

$$\text{για ανεξ. τ.μ. } x,\, y\text{:} \quad \sigma^2_{x+y} = \sigma^2_x + \sigma^2_y$$

$$\text{για } a_i \in \mathcal{R} \text{ και ανεξ. τ.μ. } x_1,\, x_2,\, \ldots\, x_n\text{:} \quad \sigma^2_{\sum_{i=1}^{n} a_i x_i} = \sum_{i=1}^{n} a_i^2 \sigma^2_{x_i}$$

Η απόδειξη βασίζεται στη χρήση των ιδιοτήτων 3 και 6:

$$\sigma^2_{x+y} = E((x+y-\mu_{x+y})^2) = E((x+y-\mu_x-\mu_y)^2) = E([(x-\mu_x)+(y-\mu_y)]^2) =$$

$$E((x-\mu_x)^2) + E((y-\mu_y)^2) + 2E((x-\mu_x)(y-\mu_y)) = \sigma^2_x + \sigma^2_y + 2E(x-\mu_x)E(y-\mu_y)$$

Ο τελευταίος όρος είναι ίσος με το 0 γιατί η 1η κεντρική ροπή $E(x-\mu_x) = E(x) - E(\mu_x) = E(x) - \mu_x = 0$ και άρα $\sigma^2_{x+y} = \sigma^2_x + \sigma^2_y$ όταν x, y ανεξάρτητες.

[13] Γι' αυτό η ανεξαρτησία ονομάζεται και στατιστική ανεξαρτησία τονίζοντας ότι όλες οι συναρτήσεις στατιστικών ιδιοτήτων είναι διαχωρίσιμες ως προς x, y.

2.10 Γραμμική Ανεξαρτησία

Παρατηρούμε ότι πληθώρα χρήσιμων στατιστικών ιδιοτήτων (6, 7, 8) ισχύουν μόνο για ανεξάρτητες τυχαίες μεταβλητές. Επιπρόσθετα οι σημαντικές ιδιότητες 6, 8 ισχύουν όταν οι τ.μ. ικανοποιούν την συνθήκη $E(xy) = E(x)E(y)$. Η συνθήκη αυτή μας πληροφορεί ότι οι x, y έχουν μια ανεξαρτησία 'ελαφράς' μορφής ή αλλιώς ανεξαρτησία 1-ης τάξης.

Ορισμός: Οι τυχαίες μεταβλητές x, y είναι γραμμικά ανεξάρτητες ή ασυσχέτιστες (linearly independent) όταν (και μόνο όταν) ισχύει η συνθήκη $E(xy) = E(x)E(y)$, δηλαδή:

$$\text{οι τ.μ. } x, y \text{ είναι ασυσχέτιστες} \iff E(xy) = E(x)E(y)$$

Προκύπτει ότι από την ιδιότητα 6 ότι:

$$\text{οι τ.μ. } x, y \text{ είναι ανεξάρτητες} \implies \text{οι τ.μ. } x, y \text{ είναι ασυσχέτιστες}$$

Το αντίθετο δεν ισχύει. Η γραμμική ανεξαρτησία είναι λοιπόν υποπερίπτωση της πλήρους ανεξαρτησίας. Υπενθυμίζουμε ότι η ιδιότητα 7 ΔΕΝ ισχύει για γραμμικά ανεξάρτητες τ.μ., π.χ., για τ.μ. x, y ασυσχέτιστες ΔΕΝ ισχύει απαραίτητα ότι: $E(x^2y) = E(x^2)E(y)$. Συχνά δυο τ.μ. είναι εξαρτημένες, αλλά γραμμικά ανεξάρτητες[14]. Θα δούμε ότι οι έννοιες της γραμμικής ανεξαρτησίας και γραμμικής εξάρτησης είναι θεμελιώδεις σε προβλήματα στατιστικής μοντελοποίησης.

2.11 Συνδιασπορά και Συντελεστής Συσχέτισης

Ορισμός: Η ποσότητα $\sigma_{xy} \equiv \sigma_{yx} \equiv E[(x - E(x))(y - E(y))]$ ορίζεται ως η συνδιασπορά ή συνδιακύμανση (covariance) των τ.μ. x, y. Ο πίνακας με διαγώνια στοιχεία σ_x^2, σ_y^2 και στοιχεία εκτός διαγωνίου τα σ_{xy}, σ_{yx} ορίζεται ως ο πίνακας συνδιασποράς των τ.μ. x, y.

[14] Ένας άλλος λόγος για την εισαγωγή της έννοιας της γραμμικής ανεξαρτησίας είναι ότι η απόδειξη της ανεξαρτησίας τ.μ. απαιτεί την από κοινού Σ.Μ.Π. $p_{x,y}(x_0, y_0)$ που σε πολλές περιπτώσεις δεν είναι γνωστή και πρέπει να εκτιμηθεί από πειραματικά δεδομένα. Όπως θα δούμε στο Κεφάλαιο 5, σε προβλήματα στατιστικής μοντελοποίησης η εκτίμηση της $E(xy)$ απαιτεί λιγότερα δεδομένα από ότι η εκτίμηση της Σ.Μ.Π. $p_{x,y}(x_0, y_0)$.

Προκύπτει εξ ορισμού ότι η συνδιασπορά μιας τ.μ. με τον εαυτό της είναι η διασπορά της, δηλαδή, $\sigma_{xx} \equiv \sigma_x^2$. Επίσης η συνδιασπορά τ.μ. x, y με αναμενόμενες τιμές μ_x, μ_y, αντίστοιχα, μπορεί να εκφραστεί ως:

$$E((x-\mu_x)(y-\mu_y)) = E(xy) - \mu_x E(y) - E(x)\mu_y + \mu_x\mu_y = E(xy) - \mu_x\mu_y$$

Άρα η συνδιασπορά ασυσχέτιστων τ.μ. x, y είναι 0. Επίσης όταν η συνδιασπορά είναι 0 τότε $E(xy) = \mu_x\mu_y$ άρα οι τ.μ. x, y είναι ασυσχέτιστες. Άρα (εφόσον η διασπορά των x, y είναι μη μηδενική):

$$\text{οι τ.μ. } x,\, y \text{ είναι ασυσχέτιστες} \quad \Longleftrightarrow \quad \sigma_{xy} \equiv E[(x-E(x))(y-E(y))] = 0$$

Προφανώς $\sigma_{xy} = \sigma_{yx} = 0$ και για (στατιστικά) ανεξάρτητες τ.μ. (το αντίθετο δεν ισχύει). Με βάση την συνδιασπορά μπορούμε να ορίσουμε μια ποσότητα που υπολογίζει τον βαθμό γραμμικής εξάρτησης δυο τ.μ.

Ορισμός: Ο συντελεστής συσχέτισης ρ_{xy} δύο τ.μ. x, y με αναμενόμενη τιμή μ_x, μ_y και διασπορά σ_x^2, σ_y^2, ορίζεται ως ο λόγος της συνδιασποράς προς το γινόμενο των τυπικών αποκλίσεων των x, y, δηλαδή:

$$\rho_{xy} \equiv \frac{\sigma_{xy}}{\sigma_x\sigma_y} = \frac{E\left((x-\mu_x)(y-\mu_y)\right)}{\sqrt{E((x-\mu_x)^2)\; E((y-\mu_y)^2)}} = \frac{E(xy)-\mu_x\mu_y}{\sigma_x\sigma_y}$$

Ο συντελεστής συσχέτισης εκφράζει το βαθμό γραμμική σχέσης μεταξύ των τ.μ. x, y και παίρνει τιμές στο διάστημα $[-1, 1]$. Χαρακτηριστικές τιμές είναι οι $\rho_{xy} = -1$ και $\rho_{xy} = 1$ που αντιστοιχούν σε πλήρη γραμμική εξάρτηση, καθώς και η τιμή $\rho_{xy} = 0$ που υποδηλώνει γραμμική ανεξαρτησία.

Παράδειγμα 1: Υπολογίστε τον συντελεστή συσχέτισης μεταξύ των τ.μ. x και $y = ax$ όπου $a \in \mathcal{R}$.

Υποθέτουμε ότι η αναμενόμενη τιμή και διασπορά της x είναι μ_x,σ_x^2 αντίστοιχα

$$\mu_y = E(y) = E(ax) = aE(x) = a\mu_x$$

$$\sigma_y^2 = E((y-\mu_y)^2) = E((ax-a\mu_x)^2) = aE((x-\mu_x)^2) = a^2\sigma_x^2$$

$$\sigma_{xy} = E((x-\mu_x)(y-\mu_y)) = E((x-\mu_x)(ax-a\mu_x)) = aE((x-\mu_x)^2) = a\sigma_x^2$$

$$\rho_{xy} \equiv \frac{\sigma_{xy}}{\sigma_x\sigma_y} = \frac{a\sigma_x^2}{\sigma_x\sqrt{a^2\sigma_x^2}} = \frac{a}{|a|} = \left\{ \begin{array}{cc} 1, & a>0 \\ 0, & a=0 \\ -1, & a<0 \end{array} \right\}$$

Βλέπουμε ότι για $a > 0$, π.χ., $a = 3$, η τ.μ. $y = 3x$ συνδέεται με γραμμική σχέση με το x, π.χ., για $x = \{1, 2, 3...\}$, $y = \{3, 6, 9...\}$. Για $a < 0$, πάλι έχουμε πλήρη γραμμική εξάρτηση αλλά αύξηση του x σημαίνει μείωση του y. Τέλος για $a = 0$, θεωρούμε ότι $\rho_{xy} = 0$ αφού δεν υπάρχει εξάρτηση ανάμεσα στην τ.μ. x και την σταθερά $y = 0$.

Παράδειγμα 2: Υπολογίστε τον συντελεστή συσχέτισης μεταξύ των τ.μ. x και $z = x+y$, όπου x, y ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την ίδια κατανομή με αναμενόμενη τιμή μ και διασπορά σ^2.

$$\mu_z = E(z) = E(x+y) = E(x) + E(y) = \mu + \mu = 2\mu$$

$$x,\, y \text{ ανεξ.} \Rightarrow \sigma_z^2 = \sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 = 2\sigma^2 \qquad [\text{ιδιότητα } 8]$$

$$E(xz) = E(x(x+y)) = E(x^2) + E(xy) = \sigma^2 + \mu^2 + E(x)E(y) = \sigma^2 + 2\mu^2$$

$$\rho_{xz} = \frac{E(xz) - \mu_x\mu_z}{\sigma_x\sigma_z} = \frac{\sigma^2 + 2\mu^2 - 2\mu^2}{\sigma\sqrt{2\sigma^2}} = \frac{\sigma^2}{\sqrt{2}\,\sigma^2} = \frac{1}{\sqrt{2}} \approx 0.707$$

Βλέπουμε ότι ο συντελεστής συσχέτισης των x και $z = x+y$ (για ανεξάρτητα x, y) είναι 0.707, δηλαδή μεταξύ 0 (γραμμική ανεξαρτησία) και 1 (πλήρη γραμμική εξάρτηση). Πράγματι η γνώση της x δεν προσδιορίζει πλήρως της z (με γραμμικό τρόπο), π.χ., άμα $x = x_0$ η τιμή τη z θα κατανέμεται γύρω από την τιμή $\mu + x_0$ με διασπορά σ^2. Η γνώση της x μειώνει την διασπορά της z στο μισό (από $\sigma_z^2 = 2\sigma^2$ σε $\sigma_{z|x}^2 = \sigma^2$) αλλά δεν προσδιορίζει πλήρως την z, άρα $|\rho_{xz}| < 1$.

Παράδειγμα 3: Υπολογίστε τον συντελεστή συσχέτισης μεταξύ των τ.μ. x και $y = x^2$. Η Σ.Μ.Π. της x δίδεται: $p_x(x_0) = 1/5$ για $x_0 \in \{-2, -1, 0, 1, 2\}$ και 0 αλλού.

$$\mu_x = E(x) = \sum_{x_0} x_0\, p_x(x_0) = (-2 - 1 + 0 + 1 + 2) \times (1/5) = 0$$

$$\mu_y = E(x^2) = \sum_{x_0} x_0^2\, p_x(x_0) = (4 + 1 + 0 + 1 + 4) \times (1/5) = 2$$

$$E(xy) = E(x^3) = \sum_{x_0} x_0^3\, p_x(x_0) = (-8 - 1 + 0 + 1 + 8) \times (1/5) = 0$$

$$E(y^2) = E(x^4) = \sum_{x_0} x_0^4\, p_x(x_0) = (16 + 1 + 0 + 1 + 16) \times (1/5) = 6.4$$

$$\sigma_x^2 = E(x^2) - (\mu_x)^2 = 2 \qquad \sigma_y^2 = E(x^4) - (\mu_y)^2 = 6.4 - 4 = 2.4$$

$$\rho_{xy} = \frac{E(xy) - \mu_x\mu_y}{\sigma_x\sigma_y} = \frac{0 - 0 \times 2}{\sqrt{2}\sqrt{2.4}} = 0$$

Είναι σημαντικό να θυμόμαστε ότι ο συντελεστής συσχέτισης υπολογίζει το βαθμό γραμμικής σχέσης. Στο παράδειγμα οι δύο τ.μ. είναι πλήρως εξαρτημένες (δηλαδή η γνώση των τιμών της τ.μ. x προβλέπει ντετερμινιστικά την τιμή της y) αλλά ο συντελεστής συσχέτισης είναι 0 (αντί για 1) γιατί x, y γραμμικά ανεξάρτητες!

2.12 Γεννήτρια Τυχαίων Αριθμών Σ.Μ.Π.

Όπως είδαμε στο Κεφ. 1.10 μπορούμε να αντιστοιχίσουμε πιθανότητες σε ενδεχόμενα χρησιμοποιώντας την γεννήτρια τυχαίων αριθμών. Με τον ίδιο τρόπο μπορούμε να αντιστοιχίζουμε πιθανότητες με αριθμητικά ενδεχόμενα, δηλαδή, με τις τιμές μιας τ.μ. x. Επιπρόσθετα επειδή οι τ.μ. παίρνουν πραγματικές τιμές μπορούμε να δημιουργήσουμε μια γεννήτρια τυχαίων αριθμών που ακολουθεί μια συγκεκριμένη κατανομή. Για παράδειγμα, για την τυχαία μεταβλητή x που παίρνει την τιμή 1 με πιθανότητα 0.3 και την τιμή 0 με πιθανότητα 0.7, η γεννήτρια n τυχαίων αριθμών της x είναι:

```
n = 10000;
rnd = rand(1,n);
x = [rnd < 0.3];
```

οπότε περίπου 30% απο τους 10000 αριθμούς που δημιουργούμε έχουν τιμή 1 και οι υπόλοιποι 0. Αντίστοιχα η γεννήτρια της τ.μ. y του αποτελέσματος ρίψης ενός ζαριού είναι:

```
n = 10000;
rnd = rand(1,n);
y = [rnd <= 1/6]*1 + [1/6 < rnd].*[rnd <= 2/6]*2 +  ...
    [2/6<rnd].*[rnd<=3/6]*3  +  [3/6<rnd].*[rnd<=4/6]*4 + ...
    [4/6 < rnd].*[rnd <= 5/6]*5  +  [5/6 < rnd]*6;
```

όπου για κάθε τιμή $\{1, 2, 3, 4, 5, 6\}$ αντιστοιχίζουμε την μάζα πιθανότητας που της αντιστοιχεί (1/6), χωρίζοντας το διάστημα $[0, 1]$ σε έξι ανεξάρτητα ίσα υποδιαστήματα. Το παραπάνω θα μπορούσα να το είχα πετύχει πιο απλά ως:

```
y = ceil(rand(1,n)*6);
```

όπου δημιουργώ n αριθμούς στο διάστημα $[0, 6]$ και στη συνέχεια αντιστοιχίζω τους αριθμούς στο διάστημα $[0, 1]$ με 1, στο διάστημα $[1, 2]$ με 2 κ.λ.π. Έτσι δημιουργώ n αριθμούς που ακολουθούν την Σ.Μ.Π. $p_y(y_0) = 1/6$, για $y \in \{1, 2, 3, 4, 5, 6\}$ και 0 αλλού. Εν γένει για να δημιουργήσω μια γεννήτρια τυχαίων αριθμών οποιασδήποτε τ.μ. x που ακολουθεί Σ.Μ.Π. $p_x(x_0)$: (1) υπολογίζω την συνάρτηση κατανομής $F_x(x_0) = \sum_{x \le x_0} p_x(x_0)$, (2) δημιουργώ n τυχαίους ανεξάρτητους τυχαίους αριθμού ισοκατανεμημένους στο διάστημα $[0, 1]$ (rand()), (3) χωρίζω το διάστημα $[0, 1]$ σε υποδιαστήματα. με όρια τις τιμές πιθανότητας της συνάρτησης κατανομής $F_x(x_0)$, και (4) δίνω την τιμή της τ.μ. x που αντιστοιχεί στο διάστημα πιθανότητας σε όλους τους τυχαίους αριθμούς που πέφτουν μέσα στο διάστημα αυτό.

Παράδειγμα: Δημιουργήστε γεννήτρια τ.μ. x και $y = x^2$ με Σ.Μ.Π.

$$p_x(x_0) = \left\{ \begin{array}{ll} 1/4, & x_0 = 1 \\ 1/2, & x_0 = 2 \\ 1/4, & x_0 = 4 \\ 0, & \text{αλλού} \end{array} \right\} \qquad F_x(x_0) = \left\{ \begin{array}{ll} 0, & x_0 < 1 \\ 1/4, & x_0 < 2 \\ 3/4, & x_0 < 4 \\ 1, & 4 \le x_0 \end{array} \right\}$$

Χωρίζουμε το διάστημα $[0, 1]$ σε 3 υποδιαστήματα χρησιμοποιώντας σαν όρια τις τιμές πιθανότητας της $F_x(x_0)$ δηλαδή $\{0, 1/4, 3/4, 1\}$. Τα διαστήματα είναι: $[0, 1/4]$, $(1/4, 3/4]$, $(3/4, 1]$. Αντιστοιχίζουμε τις τιμές της τ.μ. $x = 1, 2, 4$ για τυχαίους αριθμούς που βρίσκονται σε κάθε ένα από τα παραπάνω διαστήματα. Δηλαδή:

```
n = 10000;
rnd = rand(1,n);
x = [rnd<=1/4]*1 + [1/4<rnd].*[rnd<=3/4]*2 + [3/4<rnd]*4;
y = x.^2;
```

έτσι ώστε οι τιμές $\{1, 2, 4\}$ να εμφανίζονται με πιθανότητα $\{1/4, 1/2, 1/4\}$ αντίστοιχα. Τέλος η γεννήτρια τ.μ. y μπορεί να δημιουργηθεί είτε μέσω του υπολογισμού της Σ.Μ.Π. της y ή απευθείας ως συνάρτηση της x.

Εφόσον έχουμε μια γεννήτρια τυχαίων αριθμών που ακολουθούν μια συγκεκριμένη κατανομή μπορούμε να εκτιμήσουμε την Σ.Μ.Π. (γνωστή και ως την εμπειρική κατανομή της τ.μ.) υπολογίζοντας το ιστόγραμμα τυχαίων αριθμών που παράγει η γεννήτρια, π.χ., στο προηγούμενο παράδειγμα

```
hist(x);
```

Το ιστόγραμμα υπολογίζει το πλήθος των αριθμών n_A που βρίσκονται σε κάθε διάστημα A. Για να εκτιμήσουμε την πιθανότητα οι τιμές της x να βρίσκονται στο διάστημα A αρκεί να διαιρέσουμε (κανονικοποιήσουμε) με το πλήθος των πειραμάτων n. Άρα το ιστόγραμμα (με το άξονα των y) κανονικοποιημένο δια το πλήθος των τυχαίων αριθμών n, εκτιμά την Σ.Μ.Π. που ακολουθούν οι τυχαίοι αριθμοί, με την προϋπόθεση ότι σε κάθε διάστημα αντιστοιχεί το πολύ μία από τις δυνατές τιμές της τ.μ. Για διακριτές τ.μ. (είναι ασφαλές) να επιλέγουμε πολύ μεγάλο πλήθος διαστημάτων (bins), π.χ., hist(x,10000).

2.13 Υπολογισμός Στατιστικών Ιδιοτήτων

Έστω ότι έχουμε μια γεννήτρια τυχαίων αριθμών που ακολουθούν την Σ.Μ.Π. $p_x(x_0)$. Οι τυχαίοι αριθμοί δημιουργούνται με συχνότητα που περιγράφεται από την Σ.Μ.Π., και άρα από πλήθος n τυχαίων αριθμών περίπου $np_x(x_0)$ από αυτούς θα έχουν την τιμή $x = x_0$. Άρα η μέση τιμή των τυχαίων αριθμών:

$$\frac{1}{n}(a_1 + a_2 + ...a_n) \approx \frac{1}{n}\sum_{x_0} x_0(np_x(x_0)) = \sum_{x_0} x_0 p_x(x_0) \equiv E(x)$$

είναι εκτιμητής της αναμενόμενης τιμής της τ.μ. x. Το ίδιο μπορούμε να πούμε για οποιαδήποτε συνάρτηση της x, π.χ., $y = x^2$, δηλαδή:

$$\frac{1}{n}(a_1^2 + a_2^2 + ...a_n^2) \approx \frac{1}{n}\sum_{x_0} x_0^2(np_x(x_0)) = \sum_{x_0} x_0^2 p_x(x_0) \equiv E(x^2)$$

Εν γένει επειδή η γεννήτρια τ.μ. x δημιουργεί τον αριθμό x_0 με συχνότητα $p_x(x_0)$, οι στατιστικές ιδιότητες της x (ή συναρτήσεων της x) μπορούν να εκτιμηθούν ως ο αριθμητικός μέσος όρος των αριθμών (ή συναρτήσεων των αριθμών) που δημιουργεί η γεννήτρια[15]. Άρα για να εκτιμήσω την αναμενόμενη τιμή, διασπορά και κεντρική ροπή 3ης τάξης του αποτελέσματος της ρίψης ενός ζαριού:

```
n = 10000;
x = ceil(rand(1,n)*6);
expected_value_x = mean(x);
variance_x = mean((x-mean(x)).^2);
skewness_x = mean((x-mean(x)).^3);
```

Για την διασπορά (ή τυπική απόκλιση) μπορούμε να χρησιμοποιήσουμε και τον εκτιμητή var(x) (ή std(x)) που όπως θα δούμε και στο Κεφ. 5 είναι ίσος με ((n-1)/n)*mean((x-mean(x)).^2). Για μεγάλες τιμές του n (όπως στο παράδειγμά μας) οι δύο εκτιμητές var(x) και mean((x-mean(x)).^2) είναι πρακτικά ισοδύναμοι. Με τον ίδιο τρόπο μπορούμε να εκτιμήσουμε στατιστικές ιδιότητες συναρτήσεων πολλών τ.μ., όπως για παράδειγμα την μέση τιμή του αθροίσματος ή γινομένου δυο ανεξαρτήτων ζαριών:

[15] Θα δούμε στο Κεφ. 5 κάποια θεωρητικά αποτελέσματα για την ποιότητα αυτών των αριθμητικών εκτιμητών στατιστικών ιδιοτήτων τ.μ.

```
n = 10000;
x = ceil(rand(1,n)*6);
y = ceil(rand(1,n)*6);
mean(x+y);
mean(x.*y);
```

Για την εκτίμηση του συντελεστή συσχέτισης μπορώ να χρησιμοποιήσω τον αριθμητικό εκτιμητή ή την εντολή corrcoef() στο MATLAB

```
var_x = mean((x-mean(x)).^2);
var_y = mean((y-mean(y)).^2);
corc_xy = mean((x-mean(x)).*(y-mean(y)))./ sqrt(var_x * var_y);
corrcoef(x,y);
```

Παράδειγμα 1: Υπολογίστε αναλυτικά και αριθμητικά την αναμενόμενη τιμή του αθροίσματος δύο ζαριών στο τάβλι (δηλαδή $\times 2$ για διπλές ζαριές). Ποία είναι η αναμενόμενη τιμή της z δεδομένου ότι το πρώτο ζάρι έφερε 6.

Υπενθυμίζουμε ότι η τιμή κάθε ζαριού είναι τ.μ. x, και y αντίστοιχα. Με βάση την ανεξαρτησία των x, y είχαμε υπολογίσει την από κοινού Σ.Μ.Π. στο παράδειγμα του Κεφ. 2.6. Επίσης έχουμε υπολογίσει την Σ.Μ.Π. της τ.μ. z που συμβολίζει το άθροισμα των δύο ζαριών ($z = x + y$ για $x \neq y$ και $z = 4x$ για $x = y$ (διπλές)). Μπορούμε λοιπόν να υπολογίζουμε την αναμενόμενη τιμή της z από τον τύπο ως:

$$\mu_z = E(z) = \sum_{z_0} z_0 p_z(z_0) = 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + ... + 24 \times \frac{1}{36} \approx 8.17$$

Πιο απλά:

$$\mu_z = \sum_{x_0}\sum_{y_0} z_0\, p_{x,y}(x_0,y_0) = \sum_{x_0}\sum_{y_0 \neq x_0}(x_0+y_0)\, p_{x,y}(x_0,y_0) + \sum_{x_0}\sum_{y_0 = x_0} 4x_0\, p_{x,y}(x_0,y_0)$$

$$= \sum_{x_0}\sum_{y_0}(x_0 + y_0)p_{x,y}(x_0, y_0) + \sum_{x_0} 2x_0 p_{x,y}(x_0, x_0) = E(x + y) + \sum_{x_0} 2x_0 \frac{1}{36}$$

$$= E(x) + E(y) + \frac{1}{18}\sum_{x_0} x_0 = 2\sum_{x_0} x_0 p_x(x_0) + \frac{21}{18} \approx 2 \times 3.5 + 1.17 = 8.17$$

Βλέπουμε ότι οι διπλές αυξάνουν την αναμενόμενη τιμή της z μόνο κατά 1.17 (από 7 σε 8.17). Αριθμητικά μπορούμε να εκτιμήσουμε την αναμενόμενη τιμή της z:

```
n = 10000;
x = ceil(rand(1,n)*6);
y = ceil(rand(1,n)*6);
z = (x+y).*[x ~= y] + 4*x.*[x == y];
mean(z);
```

Δεδομένου ότι φέραμε $x = 6$ η υπό συνθήκη Σ.Μ.Π. $p_{y|x}(y_0|x_0) = p_y(y_0)$ γιατί x, y ανεξάρτητες. Άρα η $p_{z|x}(z_0|x_0 = 6) = 1/6$, για κάθε ένα από τις τιμές του $y \in \{1, 2, 3, 4, 5, 6\}$ και η τιμή της ζ είναι αντίστοιχα $\{7, 8, 9, 10, 11, 24\}$(η τελευταία τιμή αντιστοιχεί στις εξάρες). Άρα:

$$\mu_z = E(z) = \sum_{z_0} z_0 p_z(z_0) = (7 + 8 + 9 + 10 + 11 + 24) \times \frac{1}{6} = \frac{69}{6} = 11.5$$

Αριθμητικά:

```
n = 10000;
x = 6;
y = ceil(rand(1,n)*6);
z = (x+y).*[x ~= y] + 4*x.*[x == y];
mean(z);
```

<u>Παράδειγμα 2</u> (Θεώρημα Συνολικής Πιθανότητας): Υπολογίστε την αναμενόμενη τιμή της τ.μ. $E(x|y)$ όπου x, y τ.μ.

Παρά το ότι η αναμενόμενη τιμή μιας τ.μ. είναι σταθερά, η υπό συνθήκη αναμενόμενη τιμή $E(x|y)$ παίρνει μία τιμή για κάθε τιμή της τ.μ. y, και άρα είναι μια συνάρτηση της τ.μ. y (θυμόμαστε ότι οι συναρτήσεις τ.μ. είναι επίσης τ.μ. ορισμένες πάνω στον ίδιο δειγματοχώρο). Πράγματι:

$$E[E(x|y)] = \sum_{y_0} E(x|y = y_0)\ p_y(y_0) = \sum_{y_0} \sum_{x_0} x_0\ p_{x|y}(x_0|y_0)\ p_y(y_0)$$

$$= \sum_{y_0} \sum_{x_0} x_0\ p_{y|x}(y_0|x_0)\ p_x(x_0) = \sum_{y_0} \left(\sum_{x_0} x_0\ p_x(x_0) \right)\ p_{y|x}(y_0|x_0)$$

$$= \sum_{y_0} E(x)\ p_{y|x}(y_0|x_0) = E(x) \sum_{y_0} p_{y|x}(y_0|x_0) = E(x)$$

Όπου χρησιμοποιήσαμε τον κανόνα του Bayes και το γεγονός ότι $E(x)$ σταθερά. Άρα η αναμενόμενη τιμή της τ.μ. x μπορεί να υπολογιστεί ως ο γραμμικός συνδιασμός των αναμενόμενων τιμών της x υπό συνθήκη τ.μ. y με βάρη την πιθανότητα της τ.μ., δηλαδή:

$$E(x) = \sum_{y_0} E(x|y = y_0)\ p_y(y_0)$$

Σύμφωνα με την παραπάνω σχέση, μπορούμε να υπολογίσουμε την αναμενόμενη τιμή του αθροίσματος δυό ζαριών στο τάβλι (βλ. Παράδειγμα 1) ως:

$$E(z) = E(z|x = y)\ Prob(x \neq y) + E(z|x \neq y)\ Prob(x = y)$$

Αλλά όταν $x \neq y$ τότε $z = x + y$ και όταν $x = y$ τότε $z = 4x$, οπότε:

$$E(z) = E(x+y)Prob(x=y)+E(4x)Prob(x\neq y) = 2E(x)\frac{30}{36}+4E(x)\frac{6}{36} \approx 8.17$$

Το θεώρημα συνολικής πιθανότητας (total probability theorem) είναι ο 'κανόνας διαμερισμού' για τις αναμενόμενες τιμές, δηλαδή ένας τρόπος να υπολογίσουμε την αναμενόμενη τιμή μιας τ.μ. x πάνω στον διαμερισμό που ορίζει η τ.μ. y.

2.14 Σύνοψη

- Η τυχαία μεταβλητή είναι μια συνάρτηση από τον χώρο ενδεχομένων στους πραγματικούς αριθμούς.
- Η τυχαία μεταβλητή ορίζει ένα διαμερισμό του δειγματοχώρου.
- Η τιμή μιας τυχαίας μεταβλητής αντιστοιχεί σε ένα ενδεχόμενο και άρα ο υπολογισμός πιθανοτήτων (αξιώματα, ιδιότητες, κανόνες) ισχύουν για κάθε τιμή μιας τ.μ.
- Η συνάρτηση μάζας πιθανότητας αντιστοιχίζει σε κάθε τιμή x_0 της τ.μ. x την πιθανότητα εμφάνισης του ενδεχομένου $x = x_0$.
- Οι βασικές έννοιες της υπό συνθήκη πιθανότητας, ανεξαρτησίας, ανεξαρτησίας υπό συνθήκη παραμένουν και για τ.μ., με μόνη διαφορά ότι οι ορισμοί και οι σχέσεις ισχύουν για κάθε τιμή των τ.μ. Για παράδειγμα ανεξαρτησία μεταξύ τ.μ. x, y ισοδυναμεί με ανεξαρτησία μεταξύ όλων των ζευγών ενδεχομένων (ή τιμών) $(x = x_0, y = y_0)$.
- Οι βασικοί κανόνες διαμερισμού, αλυσίδας και Bayes ισχύουν επίσης για κάθε τιμή των τ.μ. (Σημείωση: ο κανόνας του διαμερισμού όταν εφαρμόζεται πάνω στον διαμερισμό που ορίζει μια τ.μ. μετονομάζεται σε οριακή Σ.Μ.Π.).
- Οι συναρτήσεις τ.μ. είναι επίσης τ.μ.
- Η αναμενόμενη τιμή $E()$ τ.μ. (ή συνάρτησης τ.μ.) ορίζεται ως ο αριθμητικός μέσος όρος των τιμών της τ.μ. (ή συνάρτησης τ.μ.) με βάρος την πιθανότητα εμφάνισης κάθε τιμής.

- Οι θεμελιώδεις στατιστικές ιδιότητες τ.μ. x είναι η αναμενόμενη τιμή της $\mu_x \equiv E(x)$ (ροπή 1ης τάξης) και η διασπορά $\sigma_x^2 \equiv E((x-\mu_x)^2)$ (κεντρική ροπή 2ης τάξης).

- Η αναμενόμενη τιμή αθροίσματος τ.μ. ισούται με το άθροισμα των αναμενόμενων τιμών των τ.μ.

- Η γραμμική ανεξαρτησία είναι υποπερίπτωση της πλήρους ανεξαρτησίας. Οι τ.μ. x, y είναι ασυσχέτιστες ή γραμμικά ανεξάρτητες όταν και μόνο όταν $E(xy) = E(x)E(y)$.

- Ο βαθμός γραμμικής ανεξαρτησίας δυο τ.μ. x, y υπολογίζεται από τον συντελεστή συσχέτισης που ισούται με το λόγο της συνδιασποράς των x, y, δια το γινόμενο των τυπικών αποκλίσεων x και y.

- Εφόσον έχω μια γεννήτρια τυχαίων αριθμών που ακολουθούν την Σ.Μ.Π. της τ.μ. x μπορώ να εκτιμήσω την αναμενόμενη τιμή του x (ή συναρτήσεων του x) ως τον αριθμητικό μέσο όρο των τυχαίων αριθμών (ή συναρτήσεων των τυχαίων αριθμών).

Ασκήσεις

1. Δίδεται η Σ.Μ.Π. της x: $p_x(x_0) = 1/5$ όταν $x_0 \in \{-2, -1, 0, 1, 2\}$ και 0 αλλού. Υπολογίστε την Σ.Μ.Π. της $y = x^2$, την υπό συνθήκη Σ.Μ.Π. $p_{x|y}(x_0|y_0)$ και την από κοινού Σ.Μ.Π. $p_{x,y}(x_0, y_0)$. Δείξτε ότι οι x, y είναι εξαρτημένες τ.μ.

2. Δίδεται η από κοινού Σ.Μ.Π. $p_{x,y}(x_0, y_0) = k$ όταν $x_0 + y_0 = 3$ και $0 \leq x_0 \leq y_0 \leq 3$ (0 αλλού). Υπολογίστε τη σταθερά k. Υπολογίστε τις οριακές Σ.Μ.Π. $p_x(x_0)$, $p_y(y_0)$. Είναι οι x, y ανεξάρτητες τ.μ.;

3. Δημιουργήστε γεννήτρια τ.μ. x με Σ.Μ.Π.

$$p_x(x_0) = \left\{ \begin{array}{ll} 1/8, & x_0 = 1, 8 \\ 1/4, & x_0 = 2, 7 \\ 1/8, & x_0 = 3, 6 \\ 0, & \text{αλλού} \end{array} \right\}$$

4. Υπολογίστε την Σ.Μ.Π. της τ.μ. $y = x_1 + x_2$ όπου οι x_1, x_2 είναι ανεξάρτητες τ.μ. που ακολουθούν την Σ.Μ.Π. που δίδεται στην προηγούμενη άσκηση. Δημιουργήστε την γεννήτρια τ.μ. y.

5. Υπολογίστε την αναμενόμενη τιμή και διασπορά των τ.μ. x_1, x_2, y της προηγούμενης άσκησης. Υπολογίστε την αναμενόμενη τιμή για κάθε μία από τις συναρτήσεις: $x_1 + y$, $x_1 x_2$, $x_1^2 y$. Ποια είναι η αναμενόμενη τιμή και διασπορά της y υπό συνθήκη $x_1 = 2$. Υπολογίστε τις παραπάνω ποσότητες αριθμητικά χρησιμοποιώντας τις γεννήτριες τ.μ.

6. Δείξτε ότι η αναμενόμενη τιμή κάθε γραμμικού συνδυασμού τ.μ. (ή συνάρτησης αυτών των τ.μ.) ισούται με τον γραμμικό συνδυασμό των αναμενόμενων τιμών των τ.μ. (ή αναμενόμενης τιμής της συνάρτησης αυτών), δηλαδή:

$$\forall \text{ τ.μ. } x_i, f_i(x_i),\ a_i \in \mathcal{R},\ i{=}1,2...n:\ E\left(\sum_{i=1}^{n} a_i f_i(x_i)\right) = \sum_{i=1}^{n} a_i E(f_i(x_i))$$

7. Δείξτε ότι για ανεξάρτητες τ.μ. x_1, x_2, ... x_n και συναρτήσεις αυτών $f_i(x_i)$: $E\left(\prod_{i=1}^{n} f_i(x_i)\right) = \prod_{i=1}^{n} E(f_i(x_i))$.

8. Δείξτε την επόμενη σχέση για την διασπορά γραμμικού συνδυασμού ανεξάρτητων τ.μ. x_i με βάρη $a_i \in \mathcal{R}$: $\sigma^2_{\sum_{i=1}^{n} a_i x_i} = \sum_{i=1}^{n} a_i^2 \sigma^2_{x_i}$

9. Όταν ένας παίκτης 'κολλάει' τα ζάρια στο τάβλι η πιθανότητα να φέρει διπλές ζαριές διπλασιάζεται, π.χ., $p(6,6) = 1/18$. Όλες οι διπλές ζαριές παραμένουν ισοπίθανες μεταξύ τους (το ίδιο και για τις απλές ζαριές). Υπολογίστε την από κοινού Σ.Μ.Π. των δυο ζαριών σ' αυτή την περίπτωση και τις οριακές Σ.Μ.Π. Είναι τα δυο ζάρια ανεξάρτητα;

10. Υπολογίστε την αναμενόμενη τιμή και διασπορά του αθροίσματος των δύο ζαριών στην παραπάνω άσκηση αναλυτικά και αριθμητικά (οι διπλές ζαριές μετράνε δυο φορές). Ποια είναι η αναμενόμενη τιμή του αθροίσματος δεδομένου ότι το ένα ζάρι έφερε 6; Συγκρίνετε τα αποτελέσματά σας με την περίπτωση που ο παίκτης δεν κολλάει τα ζάρια.

11. Υπολογίστε τον συντελεστή συσχέτισης μεταξύ των δύο ζαριών στο παραπάνω πρόβλημα. Το ίδιο για έναν παίκτη που τριπλασιάζει την πιθανότητα να φέρει διπλές ζαριές κολλώντας τα ζάρια.

12. Ποιος είναι ο μέσος αριθμός από ζαριές που χρειάζεται ο παραπάνω παίκτης για να μπει σε ένα τετράπορτο ή πεντάπορτο;

13. Υπολογίστε την αναμενόμενη τιμή και διασπορά αθροίσματος $y = x_1 + x_2 + ... + x_n$, n ανεξάρτητων τ.μ. x_1, x_2, ... x_n που ακολουθούν την ίδια κατανομή με μέση τιμή μ και διασπορά σ^2.

14. Υπολογίστε τον συντελεστή συσχέτισης μεταξύ του αθροίσματος y και κάθε μια από τις x_i για την προηγούμενη άσκηση.

15. Υπολογίστε την ποσότητα $(\sigma_y^2 - \sigma_{y|x_i}^2)/\sigma_y^2$ (κανονικοποιημένη μείωση της διασποράς λόγω γνώσης της x_i) και του συντελεστή συσχέτισης για την παραπάνω άσκηση.

Κεφάλαιο 3

Συνεχείς Τυχαίες Μεταβλητές

Οι συνεχείς τυχαίες μεταβλητές ορίζονται ως συναρτήσεις από ένα συνεχή χώρο ενδεχομένων (με άπειρα δείγματα) στους πραγματικούς αριθμούς. Οι συνεχείς τ.μ. ορίζουν και αυτές ένα διαμερισμό στο χώρο ενδεχομένων με άπειρα όμως ενδεχόμενα. Οι ορισμοί και ιδιότητες των τ.μ. ισχύουν τόσο για διακριτές όσο και για συνεχείς με δυο διαφοροποιήσεις: (1) η Σ.Μ.Π αντικαθίσταται από την συνάρτηση πυκνότητας πιθανότητας (Σ.Π.Π.) (2) όλοι οι ορισμοί και ιδιότητες που αθροίζουν πάνω στον διαμερισμό μιας διακριτής τ.μ. αντικαθίστανται από ολοκληρώματα για συνεχείς τ.μ. (γιατί λαμβάνουν άπειρο πλήθος τιμών).

3.1 Συνάρτηση Πυκνότητας Πιθανότητας

Ορισμός: Η συνάρτηση πυκνότητας πιθανότητας (Σ.Π.Π.) $f_x(x)$ (probability density function (pdf)) συνεχούς τυχαίας μεταβλητής x ορίζεται ως:

$$Prob(a \leq x < b) = \int_a^b f_x(x)dx$$

Βλέπουμε ότι η Σ.Π.Π. ορίζει την πιθανότητα ενδεχομένων που αντιστοιχούν σε διαστήματα τιμών και όχι για συγκεκριμένες τιμές[1] της τ.μ. Η φυσική ερμηνεία της Σ.Π.Π. είναι:

- Η τιμή $f_x(x_0)$ ΔΕΝ εκφράζει πιθανότητα.

[1] Το μέτρο πιθανότητας που αντιστοιχεί σε μια συγκεκριμένη τιμή για συνεχείς τ.μ. πρέπει να είναι ούτως ή άλλως 0, γιατί αλλιώς η πιθανότητα του βέβαιου ενδεχομένου θα ήταν το άθροισμα άπειρων μη μηδενικών τιμών (δηλαδή άπειρο και όχι 1).

- Το γινόμενο $f_x(x_0)\Delta x$ εκφράζει το μέτρο πιθανότητας που αντιστοιχεί σε απειροστό ενδεχόμενο $[x_0, x_0 + \Delta x]$.
- Η σύγκριση των τιμών της Σ.Π.Π. σε δύο σημεία x_1, x_2, π.χ., $f_x(x_1) > f_x(x_2)$ μας πληροφορεί ότι η πιθανότητα στην περιοχή γύρω από το x_1 είναι μεγαλύτερη από την πιθανότητα γύρω από το x_2 (στην ουσία αυτό που συγκρίνουμε είναι $f_x(x_1)\Delta x > f_x(x_2)\Delta x$ ή αλλιώς την πιθανότητα του ενδεχομένων $[x_1, x_1 + \Delta x]$, $[x_2, x_2 + \Delta x]$).

Με βάση την παραπάνω ερμηνεία της Σ.Π.Π. μπορούμε να επαναλάβουμε όλους τους ορισμούς και ιδιότητες του Κεφ. 2 αντικαθιστώντας της Σ.Μ.Π. $p_x(.)$ με το γινόμενο της Σ.Π.Π. με το απειροστό ενδεχόμενο $f_x()dx$ (και αντικαθιστώντας αθροίσματα με ολοκληρώματα όπου χρειάζεται).
Προφανώς $f_x(x_0) \geq 0$ αλλά ΔΕΝ ισχύει ότι $f_x(x_0) \leq 1$ (γιατί;). Εξ ορισμού ισχύει ότι $\int_{-\infty}^{+\infty} f_x(x)\, dx = 1$ γιατί η πιθανότητα $Prob(-\infty < x < +\infty) = 1$.
Γενικεύοντας για δύο τυχαίες μεταβλητές:

Ορισμός: Η από κοινού συνάρτηση πυκνότητας πιθανότητας (joint probability density function) $f_{x,y}(x, y)$ για τις συνεχείς τυχαίες μεταβλητές x, y ορίζεται ως:

$$Prob((a \leq x < b)\ \cap\ (c \leq y < d)) = \int_a^b \int_c^d f_{x,y}(x, y)\, dy\, dx$$

Προφανώς $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{x,y}(x, y)\, dy\, dx = 1$.

Από τον ορισμό του μέτρου πιθανότητας και επειδή τα ενδεχόμενα x, τα ενδεχόμενα y, και τα ενδεχόμενα (x, y) αποτελούν διαμερισμό μπορούμε να γράψουμε:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{x,y}(x, y)\, dx\, dy = 1$$

$$\forall\ x_0, \quad \int_{-\infty}^{+\infty} f_{x,y}(x_0, y)\, dy = f_x(x_0)$$

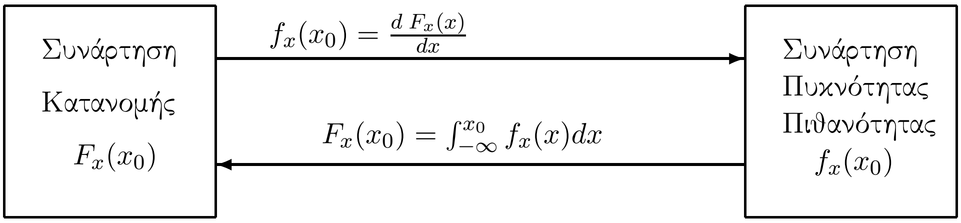
$$\forall\ y_0, \quad \int_{-\infty}^{+\infty} f_{x,y}(x, y_0)\, dx = f_y(y_0)$$

Οι δύο τελευταίες σχέσεις είναι παραδείγματα του κανόνα οριακής πιθανότητας (ιδιότητα διαμερισμού).

Ο ορισμός της συνάρτησης κατανομής είναι ο ίδιος για διακριτές ή συνεχείς τ.μ.:
Ορισμός: Η συνάρτηση κατανομής πιθανότητας (cumulative probability mass function) ορίζεται ως: $F_x(x_0) \equiv Prob(x \leq x_0)$.

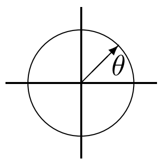
Ορισμός: Η από κοινού συνάρτηση κατανομής πιθανότητας (joint cumulative probability mass function) ορίζεται ως: $F_{x,y}(x_0, y_0) \equiv Prob(x \leq x_0, y \leq y_0)$.

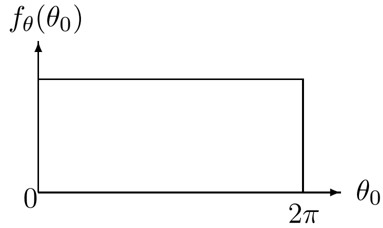
Η σχέση ανάμεσα στην Σ.Π.Π. και την συνάρτηση κατανομής για συνεχείς τ.μ.:



όπου πλέον αντί για άθροισμα και διαφορές υπολογίζω ολοκληρώματα και παραγώγους. Συχνά για συνεχείς τ.μ. η Σ.Π.Π. προκύπτει ως η παράγωγος της συνάρτησης κατανομής (βλ. Σ.Π.Π. συναρτήσεων συνεχών τ.μ.).

Παράδειγμα: Υπολογίστε την Σ.Π.Π. και συνάρτηση κατανομής της τ.μ. γωνίας θ δείκτη, που περιστρέφεται τυχαία (όλες οι δυνατές γωνίες είναι ισοπίθανες).





Η τ.μ. θ ακολουθεί ομογενή κατανομή στο διάστημα $[0, 2\pi)$ (δυνατές τιμές γωνίας) όπως φαίνεται στο παραπάνω σχήμα. Άρα $f_\theta(\theta_0) = a$ για $a \in [0, 2\pi)$ (και 0 αλλού). Η τιμή της σταθεράς a προκύπτει από:

$$\int_{-\infty}^{+\infty} f_\theta(\theta)\, d\theta = 1 \;\Rightarrow\; \int_0^{2\pi} a\, d\theta = 1 \;\Rightarrow\; a = \frac{1}{2\pi}$$

Η συνάρτηση κατανομής προκύπτει ως το ολοκλήρωμα της Σ.Π.Π., δηλαδή:

$$F_\theta(\theta_0) = \int_{-\infty}^{\theta_0} f_\theta(\theta)\, d\theta = \left\{ \begin{array}{ll} 0, & \theta_0 < 0 \\ \frac{\theta_0}{2\pi}, & 0 \le \theta_0 < 2\pi \\ 1, & \theta_0 \ge 2\pi \end{array} \right\}$$

3.2 Σ.Μ.Π υπό Συνθήκη και Ανεξαρτησία Τ.Μ.

Ορίζουμε την υπό συνθήκη Σ.Π.Π. $f_{x|y}(x|y)$ ως:

$$Prob(a \le x < b|y) = \int_a^b f_{x|y}(x|y)dx$$

Όπως και στην διακριτή περίπτωση προκύπτει ότι

$$\forall\ x_0, y_0 \quad f_{x|y}(x_0|y_0) = \frac{f_{x,y}(x_0, y_0)}{f_y y_0)}$$

Η απόδειξη έχει ως εξής. Η υπό συνθήκη πιθανότητα του ενδεχομένου $A = [x_0, x_0 + \Delta x]$ δεδομένου ότι έχει συμβεί το ενδεχόμενο $B = [y_0, y_0 + \Delta y]$ ισούται εξ ορισμού με $f_{x|y}(x_0|y_0)\Delta x$. Από κλασσικές πιθανότητες γνωρίζουμε ότι:

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{Prob((x_0 \le x < x_0 + \Delta x) \cap (y_0 \le y < y_0 + \Delta y))}{Prob(x_0 \le x < x_0 + \Delta x)}$$

$$= \frac{f_{x,y}(x_0, y_0)\Delta x \Delta y}{f_y(y_0)\Delta y} = \frac{f_{x,y}(x_0, y_0)\Delta x}{f_y(y_0)}$$

από τον ορισμό της από κοινού και της Σ.Π.Π. συνεχούς τ.μ. Άρα:

$$f_{x|y}(x_0|y_0)\Delta x = \frac{f_{x,y}(x_0, y_0)\Delta x}{f_y(y_0)} \Rightarrow f_{x|y}(x_0|y_0) = \frac{f_{x,y}(x_0, y_0)}{f_y(y_0)}$$

Δείξαμε την παραπάνω σχέση χρησιμοποιώντας ιδιότητες των κλασσικών πιθανοτήτων. Πράγματι εφόσον το (απειροστό) ενδεχόμενο συνεχούς τ.μ. $[x_0 \le x < x_0 + \Delta x]$ έχει μέτρο πιθανότητας $f_x(x_0)\Delta x$ μπορώ να επεκτείνω τους ορισμούς και ιδιότητες του Κεφ. 1,2 και για συνεχείς τ.μ.

Ορισμός: οι συνεχείς τυχαίες μεταβλητές x, y είναι ανεξάρτητες (independent) αν και μόνο αν:

$$\forall\ x_0, y_0, \quad f_{x|y}(x_0|y_0) = f_x(x_0)$$

δηλαδή, εάν κάθε ένα από τα (απειροστά) ενδεχόμενα του διαμερισμού x είναι ανεξάρτητα με κάθε ένα από τα (απειροστά) ενδεχόμενα του διαμερισμού y. Προκύπτουν οι ιδιότητες (για τ.μ. x, y ανεξάρτητες):

$$\forall\ x_0, y_0, \quad f_{y|x}(y_0|x_0) = f_y(y_0)$$

$$\forall\ x_0, y_0, \quad f_{x,y}(x_0, y_0) = f_x(x_0)p_y(y_0)$$

Είναι εύκολο να δείξουμε ότι οι τρεις σχέσεις είναι ισοδύναμες.

Ορισμός: τα ενδεχόμενα x, y είναι ανεξάρτητα υπό συνθήκη[2] $z = z_0$ (conditionally independent) αν και μόνο αν:

$$\forall\ x_0, y_0, \quad f_{x|y,z}(x_0|y_0, z\!=\!z_0) = f_{x|z}(x_0|z\!=\!z_0)$$

[2]Είναι σημαντικό να θυμόμαστε ότι η συνθήκη $z = z_0$ αντιστοιχεί στην πράξη στο απειροστό ενδεχόμενο $[z_0, z_0 + \Delta z]$.

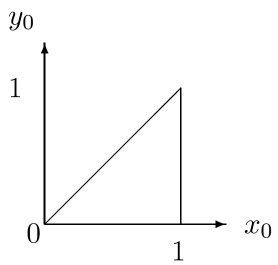
Προκύπτουν οι ιδιότητες (που είναι και ισοδύναμοι ορισμοί):

$$\forall\ x_0, y_0, \quad f_{y|x,z}(y_0|x_0, z=z_0) = f_{y|z}(y_0|z=z_0)$$
$$\forall\ x_0, y_0, \quad f_{x,y|z}(x_0, y_0|z=z_0) = f_{x|z}(x_0|z=z_0)\ f_{y|z}(y_0|z=z_0)$$

Παράδειγμα: Υπολογίστε τις οριακές $f_x()$, $f_y()$ και υπό συνθήκη Σ.Π.Π. $f_{x|y}$, $f_{y|x}$ για την παρακάτω από κοινού Σ.Π.Π. των τ.μ. x, y:

$$f_{x,y}(x_0, y_0) = \left\{ \begin{array}{ll} 3x_0, & 0 \le y_0 \le x_0 \le 1 \\ 0, & \text{αλλού} \end{array} \right\}$$

Εφόσον η από κοινού Σ.Π.Π. λαμβάνει μη μηδενικές τιμές μόνο για $0 \le y_0 \le x_0 \le 1$, ο (συνεχής) χώρος ενδεχομένων (x, y) είναι το τρίγωνο που φαίνεται στο διπλανό σχήμα.



Οι οριακές Σ.Π.Π. είναι (για $x \in [0,1]$, $y \in [0,1]$):

$$f_x(x_0) = \int_{-\infty}^{+\infty} f_{x,y}(x_0, y_0)\ dy_0 = \int_0^{x_0} 3x_0\ dy_0 = 3x_0^2$$

$$f_y(y_0) = \int_{-\infty}^{+\infty} f_{x,y}(x_0, y_0)\ dx_0 = \int_{y_0}^{1} 3x_0\ dx_0 = \frac{3x_0^2|_{y_0}^1}{2} = \frac{3}{2} - \frac{3y_0^2}{2}$$

Σημείωση: Τα όρια ολοκλήρωσης προκύπτουν από τη σχέση: $0 \le y_0 \le x_0 \le 1$, δηλαδή $0 \le y_0 \le x_0$ για το ολοκλήρωμα ως προς y και $y_0 \le x_0 \le 1$ για το ολοκλήρωμα ως προς x. Οι οριακές Σ.Π.Π. είναι λοιπόν:

$$f_x(x_0) = \left\{ \begin{array}{ll} 3x_0^2, & 0 \le x_0 \le 1 \\ 0, & \text{αλλού} \end{array} \right\} \qquad f_y(y_0) = \left\{ \begin{array}{ll} \frac{3}{2} - \frac{3y_0^2}{2}, & 0 \le y_0 \le 1 \\ 0, & \text{αλλού} \end{array} \right\}$$

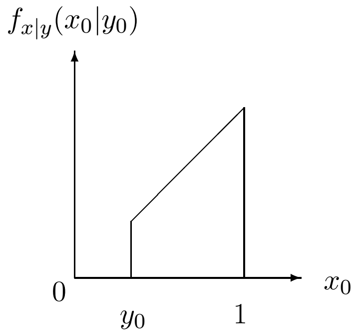
Η υπό συνθήκη Σ.Π.Π. προκύπτει ως το πηλίκο της από κοινού Σ.Π.Π. δια την οριακή Σ.Π.Π. της συνθήκης, δηλαδή (για $0 \le y_0 \le x_0 \le 1$):

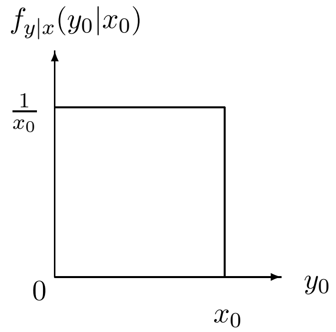
$$f_{x|y}(x_0|y_0) = \frac{f_{x,y}(x_0, y_0)}{f_y y_0)} = \frac{3x_0}{\frac{3}{2} - \frac{3y_0^2}{2}} = \frac{2x_0}{1 - y_0^2}, \qquad \text{όταν } 0 \le y_0 \le x_0 \le 1$$

$$f_{y|x}(y_0|x_0) = \frac{f_{x,y}(x_0, y_0)}{f_x x_0)} = \frac{3x_0}{3x_0^2} = \frac{1}{x_0}, \qquad \text{όταν } 0 \leq y_0 \leq x_0 \leq 1 \ \ (0 \text{ αλλού})$$

Επειδή $f_x() \neq f_{x|y}()$ προκύπτει ότι x, y εξαρτημένες τ.μ.

Οι γραφικές παραστάσεις των υπό συνθήκη Σ.Π.Π. έχουν ως εξής:





Δηλαδή η $f_{x|y}()$ είναι συνάρτηση τόσο του x όσο και του y. Για μια συγκεκριμένη τιμή του $y = y_0$ η Σ.Π.Π. είναι ανάλογη του x (το y_0 καθορίζει την κλίση της Σ.Π.Π.). Για συνθήκη $x = x_0$, η $f_{y|x}()$ είναι η ομογενής κατανομή στο $[0, x_0]$. Εν γένει για κάθε τιμή της συνθήκης ($x = x_0$ ή $y = y_0$) έχουμε μια διαφορετική Σ.Π.Π. για την αδέσμευτη τ.μ. (y ή x αντίστοιχα). Το σημαντικό είναι ότι μπορούμε να υπολογίσουμε την αναλυτική μορφή των υπό συνθήκη Σ.Π.Π. ως πηλίκο συναρτήσεων (από κοινού δια οριακή Σ.Π.Π.). Επίσης υπενθυμίζουμε ότι μπορούμε να υπολογίσουμε τις οριακές και υπό συνθήκη Σ.Π.Π. τ.μ. από την από κοινού Σ.Π.Π. (το αντίθετο δεν είναι εφικτό παρά μόνο εάν οι τ.μ. είναι ανεξάρτητες).

3.3 Κανόνας αλυσίδας και κανόνας Bayes

Οι κανόνες αλυσίδας και Bayes ισχύουν για κάθε απειροστό ενδεχόμενο και άρα και για κάθε τιμή συνεχούς τυχαίας μεταβλητής. Συγκεκριμένα μπορούμε να γράψουμε:

$$\forall\ x_0, y_0, \quad f_{x,y}(x_0, y_0) = f_{x|y}(x_0|y_0)\ f_y(y_0) = f_{y|x}(y_0|x_0)\ f_x(x_0)$$

$$\forall\ x_0, y_0, z_0, \quad f_{x,y,z}(x_0, y_0, z_0) = f_{x|y,z}(x_0|y_0, z_0)\ f_{y|z}(y_0|z_0)\ f_x(x_0)$$

$$\ldots$$

και επίσης τον κανόνα του Bayes:

$$\forall\ x_0, y_0, \quad f_{x|y}(x_0|y_0) = \frac{f_{y|x}(y_0|x_0) f_x(x_0)}{f_y(y_0)} = \frac{f_{y|x}(y_0|x_0) f_x(x_0)}{\int_{-\infty}^{+\infty} f_{y|x}(y|x_0) f_x(x_0)\ dy}$$

όπου στον παρανομαστή έχουμε χρησιμοποιήσει την ιδιότητα του διαμερισμού[3].

3.4 Μικτές Τυχαίες Μεταβλητές

Οι μικτές τ.μ. έχουν τμήμα της μάζας πιθανότητας κατανεμημένο σε συνεχή χώρο ενδεχομένων και την υπόλοιπη μάζα πιθανότητας κατανεμημένη σε διακριτά ενδεχόμενα, είναι δηλαδή ένα μείγμα διακριτής και συνεχούς τ.μ. Για να αναπαραστήσουμε τις πιθανότητες μικτών ενδεχομένων χρησιμοποιώντας μια συνάρτηση μπορούμε να επεκτείνουμε τον ορισμό την Σ.Π.Π. και για διακριτές τ.μ. Για παράδειγμα θεωρήστε την μικτή τ.μ. x, όπου η μισή μάζα πιθανότητας αντιστοιχεί στο ενδεχόμενο $x = 2$ και η υπόλοιπη είναι ισοκαναμενημένη στο διάστημα $[0, 1]$. Μπορούμε να ορίσουμε μια μικτή Σ.Π.Π. της x με τιμή $1/2$ στο διάστημα $[0, 1]$ (έτσι ώστε το ολοκλήρωμα της Σ.Π.Π. στο διάστημα αυτό να αντιστοιχεί στη μισή μάζα πιθανότητας). Η τιμή της Σ.Π.Π. στο σημείο $x = 2$ πρέπει να είναι $+\infty$ (impulse) έτσι ώστε το 'ολοκλήρωμα' της Σ.Π.Π. στο σημείο $x = 2$ να παίρνει πεπερασμένη τιμή την οποία ορίζουμε[4] να είναι $1/2$. Έτσι το ολοκλήρωμα $\int_{-\infty}^{+\infty} f_x(x)dx = 1/2 + 1/2 = 1$.

Η χρήση Σ.Π.Π. για μικτές τ.μ., αν και μαθηματικά ορθή, πολλές φορές δεν είναι πρακτική. Αντί για την χρήση μιας συνάρτησης για την περιγραφή μικτών τ.μ., συχνά χρησιμοποιούμε δύο συναρτήσεις, την Σ.Μ.Π. για το τμήμα της μάζας πιθανότητας που αντιστοιχεί σε διακριτά ενδεχόμενα και την Σ.Π.Π. για το τμήμα της μάζας πιθανότητας που αντιστοιχεί σε συνεχή ενδεχόμενα. Για το παραπάνω παράδειγμα, π.χ., $p_x(x_0) = 1/2$, για $x_0 = 2$ (0 αλλού) και $f_x(x_0) = 1/2$, για $x_0 \in [0, 1]$ (0 αλλού). Ο υπολογισμός τιμών πάνω στον διαμερισμό x πρέπει να γίνεται με προσοχή, π.χ.,

$$\sum_{x_0 \text{ διακρ.}} p_x(x_0) + \int_{x_0 \text{ συνεχ.}} f_x(x_0)dx_0 = 1$$

όπου το άθροισμα είναι πάνω στα διακριτά ενδεχόμενα και το ολοκλήρωμα πάνω στα συνεχή ενδεχόμενα του δειγματοχώρου. Αντίστοιχα μπορώ να υπολογίζω στατιστικές ιδιότητες (π.χ., αναμενόμενη τιμή μικτής τ.μ.).

[3] Σημειώνουμε ότι οι αποδείξεις των παραπάνω σχέσεων χρησιμοποιούν απειροστά ενδεχόμενα όπως και στην περίπτωση της υπό συνθήκη Σ.Π.Π.

[4] Το ολοκλήρωμα είναι γινόμενο διαστήματος μήκους 0 επί τιμή $+\infty$ και μπορεί να πάρει οποιαδήποτε πεπερασμένη τιμή επιθυμούμε.

3.5 Συναρτήσεις Τυχαίων Μεταβλητών

Όπως και για διακριτές τ.μ., οι συναρτήσεις συνεχών τ.μ. είναι επίσης τ.μ. (γιατί;). Υπενθυμίζουμε ότι οι τ.μ. και οι συναρτήσεις τους ορίζονται στον ίδιο πειραματικό δειγματοχώρο, απλά ορίζουν ένα διαφορετικό διαμερισμό πάνω σε αυτό τον δειγματοχώρο.

Για να υπολογίσω την Σ.Π.Π. τ.μ. $y = g(x)$ από την Σ.Π.Π. της τ.μ. x ακολουθώ τα βήματα:

- Υπολογίζω την συνάρτηση κατανομής της τ.μ. x, $F_x(x_0) = Prob(x \leq x_0)$.
- Υπολογίζω την συνάρτηση κατανομής της τ.μ. y:
$$F_y(y_0) = Prob(y \leq y_0) = Prob(g(x) \leq y_0) = ...$$
ως συνάρτηση την συνάρτηση κατανομής του x.
- Υπολογίζω την παράγωγο της συνάρτησης κατανομής: $f_y(y_0) = \frac{dF_y(y_0)}{dy}$.

Ένας άλλος πιο γενικός τρόπος υπολογισμού της Σ.Π.Π. συνάρτησης τ.μ. $y = g(x)$ είναι να υπολογίσουμε τις ρίζες της συνάρτησης $y = g(x)$ ως προς x. Έστω ότι οι ρίζες είναι x_1, x_2, ... x_n, n το πλήθος τότε:

$$f_y(y) = \frac{f_x(x_1)}{|g'(x_1)|} + \frac{f_x(x_2)}{|g'(x_2)|} + ... + \frac{f_x(x_n)}{|g'(x_n)|}$$

Παράδειγμα 1: Υπολογίστε την Σ.Π.Π. της τ.μ. $y = 1/x$ όταν η τ.μ. x ακολουθεί ομογενή κατανομή στο διάστημα [1,2].

Η συνάρτηση κατανομής της x είναι:

$$F_x(x_0) = \left\{ \begin{array}{ll} 0, & x_0 < 1 \\ x_0 - 1, & 1 \leq x_0 < 2 \\ 1, & x_0 > 2 \end{array} \right\}$$

Η συνάρτηση κατανομής της y είναι:

$$F_y(y_0) = Prob(y \leq y_0) = Prob\left(\frac{1}{x} \leq y_0\right) = Prob\left(x \geq \frac{1}{y_0}\right) = 1 - F_x(\frac{1}{y_0})$$

γιατί $x > 0$. Άρα:

$$f_y(y_0) = \frac{dF_y(y_0)}{dy} = \left(-\frac{1}{y_0}\right)' = \frac{1}{y_0^2}, \qquad \text{όταν} \quad 1 \leq \frac{1}{y_0} \leq 2$$

Άρα:

$$f_y(y_0) = \left\{ \begin{array}{ll} \frac{1}{y_0^2}, & \frac{1}{2} < y_0 \leq 1 \\ 0, & \text{αλλού} \end{array} \right\}$$

Χρησιμοποιώντας τις ρίζες της συνάρτησης $y = g(x)$: βλέπουμε ότι υπάρχει μόνο μια ρίζα $x_1 = 1/y$. Εφαρμόζουμε τον τύπο:

$$f_y(y_0) = \frac{f_x(x_1)}{|g'(x_1)|} = \frac{f_x(1/y_0)}{|(-1/x_1^2|} = \frac{f_x(1/y_0)}{|-1/(1/y_0)^2|} = \left\{ \begin{array}{ll} \frac{1}{y_0^2} & 1 < 1/y_0 \leq 2 \\ 0, & \text{αλλού} \end{array} \right\}$$

δηλαδή παίρνουμε το ίδιο αποτέλεσμα. Βλέπουμε ότι με την μέθοδο αυτή βρίσκουμε την γενική μορφή της Σ.Π.Π. τ.μ. $y = 1/x$ για οποιαδήποτε Σ.Π.Π. της τ.μ. x, δηλαδή:

$$f_y(y_0) = \frac{f_x(1/y_0)}{y_0^2}$$

π.χ., εάν η Σ.Π.Π. x είναι ανάλογη του $1/x^2$ στο διάστημα $[0.5, 1]$ ή y ακολουθεί ομογενή κατανομή στο $[1, 2]$.

<u>Παράδειγμα 2</u>: Υπολογίστε την Σ.Π.Π. της τ.μ. $y = |x|$.

Η συνάρτηση $y = |x|$ έχει δύο ρίζες $x_1 = y$ και $x_2 = -y$. Άρα:

$$f_y(y_0) = \frac{f_x(x_1)}{|g'(x_1)|} + \frac{f_x(x_2)}{|g'(x_2)|} = \frac{f_x(y_0)}{|1|} + \frac{f_x(-y_0)}{|-1|} = f_x(y_0) + f_x(-y_0)$$

για $y_0 \geq 0$ (και μηδέν για $y_0 < 0$. Παρατηρούμε ότι για συμμετρικές Σ.Π.Π. $f_x()$ γύρω από το 0, τότε η $f_y()$ είναι δύο φορές η τιμή της $f_x()$ για $y > 0$ και μηδέν παντού αλλού.

Μπορούμε να γενικεύσουμε την παραπάνω διαδικασία για τον υπολογισμό Σ.Π.Π. δύο τ.μ. ως εξής: Έστω η τ.μ. $z = g(x, y)$ από κοινού συνάρτηση των τ.μ. x, y. Η Σ.Π.Π. της z μπορεί να υπολογιστεί μέσω της από κοινού Σ.Π.Π. των x, y: $f_{x,y}()$. Όπως και για τη συνάρτηση μιας τ.μ.: (1) υπολογίζουμε την συνάρτηση κατανομής $F_{x,y}()$, (2) υπολογίζουμε τη συνάρτηση κατανομής της z, $F_z()$ ως $Prob(g(x, y) \leq z_0)$ και (3) υπολογίζουμε την $f_z()$ ως την παράγωγο της $F_z()$.

Μπορούμε επίσης να ορίσουμε μια δεύτερη τ.μ. $w = h(x, y)$ και να υπολογίσουμε την από κοινού Σ.Π.Π. των $z = g(x, y)$, $w = h(x, y)$ από την από κοινού Σ.Π.Π. των x,y. Έστω τα ζεύγη (x_1, y_1), (x_2, y_2) ... (x_n, y_n) είναι οι n το πλήθος ρίζες του συστήματος εξισώσεων $z = g(x, y)$, $w = h(x, y)$. Τότε η από κοινού Σ.Π.Π. $f_{z,w}()$ μπορεί να εκφραστεί ως:

$$f_{z,w}(z, w) = \frac{f_{x,y}(x_1, y_1)}{|J(x_1, y_1)|} + \frac{f_{x,y}(x_2, y_2)}{|J(x_2, y_2)|} + ... + \frac{f_{x,y}(x_n, y_n)}{|J(x_n, y_n)|}$$

όπου ο παρανομαστής είναι η ορίζουσα του πίνακα μερικών παραγώγων (Jacobian):

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

υπολογισμένη στις ρίζες του συστήματος εξισώσεων.

Παράδειγμα 3: Υπολογίστε την Σ.Π.Π. της τ.μ. $z = x/y$, δεδομένου ότι η από κοινού Σ.Π.Π. των x, y είναι ομογενής στην περιοχή $x \in [0,1]$, $y \in [0,1]$.

$$F_z(z_0) = Prob(z \leq z_0) = Prob\left(\frac{x}{y} \leq z_0\right)$$

Εφόσον $0 < z_0 \leq 1$ τότε $0 \leq x \leq yz_0$ και $0 \leq y \leq 1$ οπότε

$$Prob\left(\frac{x}{y} \leq z_0\right) = \int_0^1 \int_0^{yz_0} f_{x,y}(x,y)\, dx\, dy = \int_0^1 yz_0\, dy = \frac{z_0}{2}$$

Αντίστοιχα για $z_0 > 1$ τότε $0 \leq x \leq 1$ και $y > x/z_0$

$$Prob\left(\frac{x}{y} \leq z_0\right) = \int_0^1 \int_{x/z_0}^1 f_{x,y}(x,y)\, dy\, dx = \int_0^1 (1 - \frac{x}{z_0})\, dx = \frac{z_0}{2} = 1 - \frac{1}{2z_0}$$

Η Σ.Π.Π. είναι λοιπόν:

$$f_z(z_0) = \frac{dF_z(z_0)}{dz} = \left\{ \begin{array}{ll} \frac{1}{2}, & 0 \leq z_0 \leq 1 \\ \frac{1}{2z_0^2}, & z_0 > 1 \\ 0, & \text{αλλού} \end{array} \right\}$$

Αλλιώς μπορούμε να υπολογίσουμε τις ρίζες του συστήματος εξισώσεων $z = g(x,y) = x/y$ και $w = h(x,y) = x$. Υπάρχει μόνο μια ρίζα η $(x_1, y_1) = (w, z/w)$. Η Jacobian είναι:

$$J(x,y) = \begin{vmatrix} \frac{\partial g(x,y)}{\partial x} & \frac{\partial g(x,y)}{\partial y} \\ \frac{\partial h(x,y)}{\partial x} & \frac{\partial h(x,y)}{\partial y} \end{vmatrix} = \begin{vmatrix} \frac{1}{y} & -\frac{x}{y^2} \\ 1 & 0 \end{vmatrix} = \frac{x}{y^2}$$

Οπότε:

$$f_{z,w}(z_0, w_0) = \frac{f_{x,y}(x_1,y_1)}{|J(x_1,y_1)|} = \frac{f_{x,y}(w_0, w_0/z_0)}{|J(w_0, w_0/z_0)|} = \frac{f_{x,y}(w_0, w_0/z_0)}{|w_0/(w_0/z_0)^2|} = \frac{w_0}{z_0^2}$$

όταν $0 \leq w_0 \leq 1$ και $0 \leq w_0/z_0 \leq 1$. Η Σ.Π.Π. της z προκύπτει ως η οριακή της $f_{z,w}()$, δηλαδή[5]:

$$f_z(z_0) = \int_{-\infty}^{+\infty} f_{z,w}(z_0, w)\, dw = \left\{ \begin{array}{ll} \int_0^{z_0} [w/(z_0^2)]\, dw = \frac{1}{2}, & 0 \leq z_0 \leq 1 \\ \int_0^1 [w/(z_0^2)]\, dw = \frac{1}{2z_0^2}, & z_0 > 1 \\ 0, & \text{αλλού} \end{array} \right\}$$

[5] Τα όρια ολοκλήρωσης για την μεταβλητή w προκύπτουν από την συνθήκες, όταν $z_0 \leq 1$ τότε $0 \leq w_0 \leq z_0 \leq 1$, αλλιώς όταν $z_0 \geq 1$ τότε $0 \leq w_0 \leq 1 \leq z_0$.

και καταλήγουμε στο ίδιο αποτέλεσμα όπως και προηγουμένως.

3.6 Αναμενόμενη Τιμή και Ροπές

Ορισμός: Η αναμενόμενη τιμή (expected value) μιας συνεχούς τυχαίας μεταβλητής x ορίζεται ως:

$$E(x) = \int_{-\infty}^{+\infty} x \; f_x(x) \; dx$$

Η αναμενόμενη τιμή της τ.μ. x, $E(x)$ συχνά συμβολίζεται ως μ_x ή $\bar{x}$ όπως και για διακριτές τ.μ. Εν γένει η μόνη διαφορά με την διακριτή περίπτωση είναι η χρήση ολοκληρωμάτων αντί για αθροίσματα. Γενικεύοντας για συναρτήσεις συνεχών τ.μ.:

$$E(g(x)) \equiv \overline{g(x)} = \int_{-\infty}^{+\infty} g(x) \; f_x(x) \; dx$$

όπως και στην διακριτή περίπτωση.

Ορισμός: Η αναμενόμενη τιμή υπό συνθήκη A μιας τυχαίας μεταβλητής x ορίζεται ως:

$$E(x|A) = \int_{-\infty}^{+\infty} x \; f_{x|A}(x|A) \; dx$$

Πάλι μπορούμε να γενικεύσουμε για συνάρτηση τυχαίων μεταβλητών $g(x)$:

$$E(g(x)|A) = \int_{-\infty}^{+\infty} g(x) \; f_{x|A}(x|A) \; dx$$

Κλείνουμε με την γενίκευση για συναρτήσεις πολλών τυχαίων μεταβλητών:

$$E(g(x,y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) \; f_{x,y}(x,y) \; dy \; dx$$

$$E(g(x,y)|A) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) \; f_{x,y|A}(x,y|A) \; dy \; dx$$

Ορισμός: Οι ροπές τάξης n της συνεχούς τ.μ. x ορίζονται ως οι αναμενόμενες τιμές των συναρτήσεων x^n. Άρα:

Ροπή 1-ης τάξης: $$E(x) = \int_{-\infty}^{+\infty} x \; f_x(x) \; dx$$

Ροπή 2-ης τάξης: $$E(x^2) = \int_{-\infty}^{+\infty} x^2 \; f_x(x) \; dx$$

$$\text{Ροπή 3-ης τάξης:} \qquad E(x^3) = \int_{-\infty}^{+\infty} x^3 \, f_x(x) \, dx$$
$$\ldots$$

Ορισμός: Οι κεντρικές ροπές τάξης n της τ.μ. x ορίζονται ως οι αναμενόμενες τιμές των συναρτήσεων $(x - E(x))^n$. Άρα:

$$\text{Κεντρική Ροπή 1-ης τάξης:} \qquad E(x - E(x)) = \int_{-\infty}^{+\infty} (x - E(x)) \, f_x(x) \, dx = 0$$

$$\text{Κεντρική Ροπή 2-ης τάξης:} \qquad E((x - E(x))^2) = \int_{-\infty}^{+\infty} (x - E(x))^2 \, f_x(x) \, dx$$

$$\text{Κεντρική Ροπή 3-ης τάξης:} \qquad E((x - E(x))^3) = \int_{-\infty}^{+\infty} (x - E(x))^3 \, f_x(x) \, dx$$
$$\ldots$$

Όπως και για διακριτές τ.μ. η κεντρική ροπή 2ης τάξης της τ.μ. x ονομάζεται διασπορά ή διακύμανση (variance) και συμβολίζεται ως σ_x^2. Η τετραγωνική της ρίζα σ_x ονομάζεται τυπική απόκλιση σ_x (standard deviation).

Παράδειγμα: Υπολογίστε την $E(xy)$ και $E(y|x)$ για τις τ.μ. x, y, z των οποίων η από κοινού Σ.Μ.Π. είναι:

$$f_{x,y,z}(x_0, y_0, z_0) = \left\{ \begin{array}{ll} x_0 z_0 + 3y_0 z_0, & x_0, y_0, z_0 \in [0,1] \\ \\ 0, & \text{αλλού} \end{array} \right\}$$

Για να υπολογίσουμε την αναμενόμενη τιμή της xy και της $y|x$, χρειαζόμαστε τις Σ.Π.Π. $f_{x,y}(x,y)$ και $f_{y|x}(y|x)$, αντίστοιχα. Συγκεκριμένα:

$$f_{xy}(x_0, y_0) = \int_{-\infty}^{+\infty} f_{x,y,z}(x_0, y_0, z_0) \, dz_0 = \int_0^1 (x_0 z_0 + 3y_0 z_0) \, dz_0 = \frac{x_0 + 3y_0}{2}$$

για $x_0, y_0 \in [0,1]$. Η οριακή $f_x()$ και η $f_{y|x}()$ είναι:

$$f_x(x_0) = \int_{-\infty}^{+\infty} f_{x,y}(x_0, y_0) \, dy_0 = \int_0^1 \frac{x_0 + 3y_0}{2} \, dy_0 = \frac{x_0}{2} + \frac{3}{4}$$

$$f_{y|x}(y_0|x_0) = \frac{f_{xy}(x_0, y_0)}{f_x(x_0)} = \frac{\frac{x_0 + 3y_0}{2}}{\frac{x_0}{2} + \frac{3}{4}} = \frac{x_0 + 3y_0}{x_0 + 3/2}$$

για $x_0, y_0 \in [0,1]$. Οπότε:

$$E(xy) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy \, f_{x,y}(x,y) \, dy \, dx = \int_0^1 \int_0^1 xy \, \left(\frac{x}{2} + \frac{3y}{2} \right) \, dy \, dx =$$

$$\int_0^1 \frac{yx^3|_0^1}{6} + \frac{3y^2x^2|_0^1}{4}\, dy = \frac{y^2|_0^1}{12} + \frac{3y^3|_0^1}{12} = \frac{1}{3}$$

$$E(y|x) = \int_{-\infty}^{+\infty} y\ f_{y|x}(y|x)\ dy = \int_0^1 y\ \frac{x+3y}{x+3/2}\, dy = \frac{(x/2)+1}{x+3/2} = \frac{x+2}{2x+3}$$

Παρατηρούμε ότι η αναμενόμενη τιμή υπό συνθήκη x είναι συνάρτηση της τ.μ. x (και άρα η $E(y|x)$ είναι τ.μ.). Υπολογίστε για άσκηση την $E(E(y|x))$ και την $E(y)$.

3.7 Από Κοινού Στατιστικές Ιδιότητες

Ορισμός: Οι τυχαίες μεταβλητές x, y είναι γραμμικά ανεξάρτητες ή ασυσχέτιστες (linearly independent) όταν (και μόνο όταν) ισχύει η συνθήκη $E(xy) = E(x)E(y)$, δηλαδή:

$$\text{οι τ.μ. } x,\ y \text{ είναι ασυσχέτιστες} \iff E(xy) = E(x)E(y)$$

Προκύπτει ότι από την ιδιότητα 6 ότι:

$$\text{οι τ.μ. } x,\ y \text{ είναι ανεξάρτητες} \implies \text{οι τ.μ. } x,\ y \text{ είναι ασυσχέτιστες}$$

Το αντίθετο δεν ισχύει. Η γραμμική ανεξαρτησία είναι υποπερίπτωση της πλήρους ανεξαρτησίας.

Ορισμός: Η ποσότητα $\sigma_{xy} \equiv \sigma_{yx} \equiv E[(x - E(x))(y - E(y))]$ ορίζεται ως η συνδιασπορά ή συνδιακύμανση (covariance) των τ.μ. x, y. Ο πίνακας με διαγώνια στοιχεία σ_x^2, σ_y^2 και στοιχεία εκτός διαγωνίου τα σ_{xy}, σ_{yx} ορίζεται ως ο πίνακας συνδιασποράς των τ.μ. x, y.

Ορισμός: Ο συντελεστής συσχέτισης ρ_{xy} δύο τ.μ. x, y με αναμενόμενη τιμή μ_x, μ_y και διασπορά σ_x^2, σ_y^2, ορίζεται ως ο λόγος της συνδιασποράς προς το γινόμενο των τυπικών αποκλίσεων των x, y, δηλαδή:

$$\rho_{xy} \equiv \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{E\left((x-\mu_x)(y-\mu_y)\right)}{\sqrt{E((x-\mu_x)^2)\ E((y-\mu_y)^2)}} = \frac{E(xy) - \mu_x\mu_y}{\sigma_x\sigma_y}$$

Ο συντελεστής συσχέτισης εκφράζει το βαθμό γραμμική σχέσης μεταξύ των τ.μ. x, y και παίρνει τιμές στο διάστημα $[-1, 1]$ όπως και στην διακριτή περίπτωση.

Παράδειγμα: Υπολογίστε τον συντελεστή συσχέτισης μεταξύ των τ.μ. x, y όταν η από κοινού Σ.Π.Π. είναι η:

$$f_{x,y}(x_0, y_0) = \left\{ \begin{array}{ll} 3x_0, & 0 \leq y_0 \leq x_0 \leq 1 \\ 0, & \text{αλλού} \end{array} \right\}$$

Οι οριακές είναι οι $f_x(x_0) = 3x_0^2$, $x \in [0,1]$, και $f_y(y_0) = 3/2\ (1 - y_0^2)$, $y \in [0,1]$, οπότε:

$$E(x) = \int_0^1 x\ 3x^2 dx = \frac{3}{4} \qquad E(y) = \int_0^1 y\ \frac{3}{2}(1-y^2)dy = \frac{3}{4} - \frac{3}{8} = \frac{3}{8}$$

$$E(x^2) = \int_0^1 x^2\ 3x^2 dx = \frac{3}{5} \qquad E(y^2) = \int_0^1 y^2\ \frac{3}{2}(1-y^2)dy = \frac{1}{5}$$

$$E(xy) = \int_0^1 \int_0^x xy\ 3x\ dy\ dx = \int_0^1 \frac{3}{2} x^4 dx = 0.3$$

$$\sigma_x^2 = E(x^2) - [E(x)]^2 = 0.6 - 0.75^2 = 0.0375$$

$$\sigma_y^2 = E(y^2) - [E(y)]^2 = 0.2 - 0.375^2 \approx 0.0594$$

$$\sigma_{xy} = E(xy) - E(x)E(y) = 0.3 - \frac{3}{4} \times \frac{3}{8} = 0.0188$$

$$\rho_{xy} \equiv \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{0.0188}{\sqrt{0.0375}\sqrt{0.0594}} = \frac{0.0188}{0.1936 \times 0.2437} \approx 0.40$$

3.8 Θεμελιώδεις Στατιστικές Ιδιότητες

Συνοψίζουμε εδώ τις στατιστικές ιδιότητες που είδαμε στο Κεφ. 2 για διακριτές τ.μ. (για συνεχείς τ.μ. οι αποδείξεις ακολουθούν τα ίδια βήματα και αφήνονται ως άσκηση).

Ιδιότητα 1: Για τ.μ. x: $E(E(x)) = E(x)$

Ιδιότητα 2: Για $a \in R$ και τ.μ. x: $E(ax) = aE(x)$

Ιδιότητα 3: Για τ.μ x, y: $E(x+y) = E(x) + E(y)$

Ιδιότητα 4: Για τ.μ. x_i, $a_i \in \mathcal{R}$:

$$E\left(\sum_{i=1}^{n} a_i f_i(x_i)\right) = \sum_{i=1}^{n} a_i E(f_i(x_i))$$

Ιδιότητα 5: Για τ.μ x: $E((x - E(x))^2) = E(x^2) - [E(x)]^2$

Ιδιότητα 6:[6] Για **γραμμικά ανεξάρτητες** τ.μ. x, y: $E(xy) = E(x)E(y)$

Ιδιότητα 7: Για **ανεξάρτητες** τ.μ. x, y: $E(f(x)g(y)) = E(f(x))E(g(y))$

Ιδιότητα 7α: Για **ανεξάρτητες** τ.μ. x_i:

$$E\left(\prod_{i=1}^{n} f_i(x_i)\right) = \prod_{i=1}^{n} E(f_i(x_i))$$

Ιδιότητα 8: Για τη διασπορά **γραμμικά ανεξάρτητων** τ.μ. x, y ισχύει:

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2$$

Ιδιότητα 8α: Για τη διασπορά **γραμμικά ανεξάρτητων** τ.μ. x_i, $a_i \in \mathcal{R}$:

$$\sigma^2_{\sum_{i=1}^{n} a_i x_i} = \sum_{i=1}^{n} a_i^2 \sigma_{x_i}^2$$

Κανόνας Συνολικής Πιθανότητας: Για τ.μ x, y: $E(E(x|y)) = E(x)$

3.9 Γεννήτρια Τυχαίων Αριθμών Σ.Π.Π.

Όπως είδαμε στο Κεφ. 2 μπορούμε να δημιουργήσουμε μια γεννήτρια τυχαίων αριθμών που ακολουθούν οποιαδήποτε Σ.Μ.Π., ξεκινώντας από μια γεννήτρια που παράγει αριθμούς που ακολουθούν την ομογενή κατανομή. Η βασική ιδέα είναι να χωρίσουμε την μάζα πιθανότητας της ομογενούς κατανομής σε ενδεχόμενα που αντιστοιχούν στο μέτρο πιθανότητας της Σ.Μ.Π. που μας ενδιαφέρει. Τα ενδεχόμενα αυτά αντιστοιχούν σε διαστήματα της ομογενής κατανομής, π.χ., το διάστημα $[0.2, 0.5]$ αντιστοιχεί σε μάζα πιθανότητας 0.3, γιατί κατά μέσο όρο 30% από τους αριθμούς x που γεννούνται από την εντολή rand() ανήκουν σ' αυτό το διάστημα. Το τελευταίο βήμα της διαδικασίας είναι να δώσουμε σε όλους τους τυχαίους αριθμούς x του διαστήματος $[0.2, 0.5]$ την τιμή που αντιστοιχεί σε μάζα πιθανότητας 0.3 στην Σ.Μ.Π. που μας ενδιαφέρει.

Σε συνεχείς τ.μ. η διαδικασία ακολουθεί τα ίδια ακριβώς βήματα. Η κύρια διαφορά είναι ότι η Σ.Π.Π. δειγματοληπτείται και προσεγγίζεται από μια Σ.Μ.Π. με k τιμές. Για παράδειγμα θέλουμε να δημιουργήσουμε μια γεννήτρια τυχαίων αριθμών x που ακολουθούν την Σ.Π.Π. $f_x(x_0) = 3x_0^2$, για $x \in [0, 1]$. Η Σ.Π.Π. προσεγγίζεται για παράδειγμα σε $k = 100$ σημεία από την Σ.Μ.Π. $p_x(x_0) = 0.03\ x_0^2$, $x_0 \in \{0.005, 0.015, ..., 0.995\}$. Πράγματι η συνολική μάζα

[6] Προφανώς οι ιδιότητες 6, 8, 8α ισχύουν και για ανεξάρτητες τ.μ. (όχι μόνο για γραμμικά ανεξάρτητες).

πιθανότητας της Σ.Μ.Π. είναι 1 και προέρχεται από δειγματοληψία (και κανονικοποίηση με k) της Σ.Π.Π. Τα βήματα που ακολουθούμε στη συνέχεια είναι παρόμοια με αυτά που περιγράφονται στο Κεφ. 2: (1) Υπολογίζουμε την συνάρτηση κατανομής της διακριτοποιημένης τ.μ. x, π.χ., $F_x(x_0) = 0.01\ x_0^3$, $x_0 \in \{0.005, 0.015, ..., 0.995\}$. (2) Χωρίζουμε το διάστημα $[0,1]$ σε υποδιαστήματα με βάση τις τιμές της $F_x()$, δηλαδή $[0, 0.005^3]$, $[0.005^3, 0.015^3]$, ...,$[0.985^3, 0.995^3]$, έτσι ώστε η μάζα πιθανότητας της rand() που αντιστοιχεί σε κάθε διάστημα να είναι ίση με την $p_x()$. (3) Αντιστοιχίζουμε τις τιμές $x = 0.005, 0.015...$ στους τ.α που παράγει η rand() και πέφτουν σε κάθε ένα από τα αντίστοιχα διαστήματα $[0, 0.005^3]$, $[0.005^3, 0.015^3]$, ... Ο κώδικας που αντιστοιχεί στον παραπάνω παράδειγμα:

```
n = 10000;
k = 100;
sample_points = [0:1/k:1-1/k] + 1/(2*k);
rnd = rand(1,n);
x = zeros(size(rnd));
for i = sample_points
    x = x + [rnd<(i+1/(2*k)).^3].*[rnd>(i-1/(2*k)).^3] * i;
end
```

Παρατηρούμε ότι όταν το $k \to \infty$ τότε αντιστοιχίζουμε σε κάθε τ.α. rnd^3 τον αριθμό rnd. Άρα μπορούμε με μια εντολή να δημιουργήσουμε τ.α. που ακολουθούν την κατανομή $f_x()$, μετασχηματίζοντας τ.α. που ακολουθούν την ομογενή κατανομή μέσω της αντίστροφης συνάρτησης κατανομής $F^{-1}()$. Συγκεκριμένα:

```
rnd = rand(1,10000);
x = rnd.^(1/3);
```

Παράδειγμα: Δημιουργήστε μια γεννήτρια τ.α. y που ακολουθούν την κατανομή $f_y(y_0) = 4/y_0^2$, $y \in [2,4]$, με βάση την εντολή rand().

Η γεννήτρια τ.α. της $f_y()$ υπολογίζεται ως εξής: (1) Πρώτα υπολογίζουμε την συνάρτηση κατανομής $F_y(y_0) = 2 - 4/y_0$, (2) Μετά υπολογίζουμε την αντίστροφη συνάρτηση της συνάρτησης κατανομής:

$$2 - \frac{4}{y} = z \ \Rightarrow\ 2 - z = \frac{4}{y} \ \Rightarrow\ y = \frac{4}{2-z}$$

Άρα ο κώδικας που παράγει τ.α. y που ακολουθούν την $f_y()$ είναι:

```
rnd = rand(1,10000);
y = 4./(2-rnd);
```

Επειδή δεν είναι όλες οι συναρτήσεις αντιστρέψιμες ο μέθοδος υπολογισμού της συνάρτησης κατανομής μέσω διακριτοποίησης της Σ.Π.Π. είναι επίσης χρήσιμος. Για την δημιουργία γεννήτριας τ.μ. x, y, που ακολουθούν από κοινού Σ.Π.Π. $f_{x,y}(x_0, y_0)$ ακολουθούμε την ίδια διαδικασία. Σημείωση: Εφόσον οι x, y είναι ανεξάρτητες μπορούμε να τις δημιουργήσουμε ανεξάρτητα σύμφωνα με τις οριακές Σ.Π.Π. $f_x()$, $f_y()$.

Για να εκτιμήσουμε την Σ.Π.Π. από το αποτέλεσμα ενός πειράματος (ή από τ.α. που προέρχονται από γεννήτρια) υπολογίζουμε το ιστόγραμμα όπως και στην διακριτή περίπτωση.

```
hist(x,100);
```

Ο αριθμός των διαστημάτων που επιλέγουμε, π.χ., 100, έχει να κάνει με τον συνολικό αριθμό δειγμάτων που μας παρέχονται. Τέλος οι στατιστικές ιδιότητες εκτιμώνται όπως και στην διακριτή περίπτωση, π.χ., η αναμενόμενη τιμή ως ο αριθμητικός μέσος όρος των τ.α. (η ακρίβεια των εκτιμητών παρουσιάζεται στο Κεφ. 5).

3.10 Σύνοψη

- Οι συνεχείς τ.μ. ορίζονται πάνω σε συνεχή χώρο ενδεχομένων.
- Το μέτρο πιθανότητας του απειροστού ενδεχομένου $[x_0, x_0 + \Delta x]$ ισούται με $f_x(x_0)\Delta x$, όπου $f_x()$ η συνάρτηση πυκνότητας πιθανότητας της συνεχούς τ.μ. x.
- Η Σ.Π.Π. $f_x()$ μπορεί να υπολογιστεί ως η παράγωγος της συνάρτηση κατανομής $F_x()$ για συνεχείς τ.μ.
- Όλοι οι ορισμοί και ιδιότητες των διακριτών τ.μ. γενικεύονται για συνεχείς τ.μ., αντικαθιστώντας της Σ.Μ.Π. με την Σ.Π.Π. και τα αθροίσματα με ολοκληρώματα.
- Η Σ.Π.Π. συνάρτησης συνεχούς τ.μ. $y = g(x)$ υπολογίζεται ως η παράγωγος της συνάρτησης κατανομής της τ.μ. x (ή από τις ρίζες της $y = g(x)$ ως προς x).
- Μπορούμε να δημιουργήσουμε γεννήτριες τ.α. που ακολουθούν Σ.Π.Π. $f_x()$ μετασχηματίζοντας τ.α. αριθμούς που ακολουθούν την ομογενή κατανομή μέσω της αντίστροφης συνάρτησης κατανομής $F_x^{-1}()$.

Ασκήσεις

1. Αποδείξτε τον κανόνα της αλυσίδας και τον κανόνα του Bayes για συνεχείς τ.μ. (ξεκινώντας από τις κλασικές πιθανότητες).

2. Δείξτε ότι $E(E(x|y)) = E(x)$ για συνεχείς τ.μ.

3. Υπολογίστε τις οριακές και υπό συνθήκη Σ.Π.Π. των τ.μ. x, y που ακολουθούν από κοινού Σ.Π.Π. $f_{x,y}(x_0, y_0) = a x_0 y_0^2$, $x, y \in [0, 1]$, $a \in \mathcal{R}$. Είναι οι x, y ανεξάρτητες;

4. Υπολογίστε τις οριακές και υπό συνθήκη Σ.Π.Π. των τ.μ. x, y που ακολουθούν ομογενή από κοινού Σ.Π.Π. όταν $0 \leq \max\{x, y\} \leq 1$. Είναι οι x, y ανεξάρτητες;

5. Υπολογίστε την Σ.Π.Π. της συνάρτησης τ.μ. x, $y = x^2 - 3x + 2$ όταν η τ.μ. x ακολουθεί ομογενή κατανομή στο $[0, 1]$.

6. Υπολογίστε αναλυτικά την μορφή της Σ.Π.Π. της συνάρτησης τ.μ. x, $y = x^n$, $n \in \mathcal{N}$.

7. Υπολογίστε αναλυτικά την μορφή της Σ.Π.Π. της συνάρτησης ανεξάρτητων τ.μ. x, y, $z = x + y$.

8. Υπολογίστε αναλυτικά την μορφή της Σ.Π.Π. της συνάρτησης ανεξάρτητων τ.μ. x, y, $z = x^2 + y^2$.

9. Υπολογίστε τον συντελεστή συσχέτισης των x, z όπου $z = x^2 + y^2$, και x, y ανεξάρτητες τ.μ. με τις ίδιες στατιστικές ιδιότητες.

10. Το ίδιο όταν ο συντελεστής συσχέτισης των x, y είναι 0.5.

11. Υπολογίστε αριθμητικά (χρησιμοποιώντας γεννήτριες τ.α.) τον συντελεστή συσχέτισης μεταξύ των τ.μ. x, y όταν η από κοινού Σ.Π.Π. είναι:

$$f_{x,y}(x_0, y_0) = \left\{ \begin{array}{ll} 3x_0, & 0 \leq y_0 \leq x_0 \leq 1 \\ 0, & \text{αλλού} \end{array} \right\}$$

12. Στην παραπάνω άσκηση δημιουργήστε μια γεννήτρια ζευγών τ.μ. x, y που ακολουθούν Σ.Π.Π. $f_{x,y}(x_0, y_0) = a x_0$, $a \in \mathcal{R}$, $x, y \in [0, 1]$. Στη συνέχεια επιλέξτε μόνο τα ζεύγη δειγμάτων (x_0, y_0) για τα οποία ισχύει $y_0 \leq x_0$. Υπολογίστε τις στατιστικές ιδιότητες και τον συντελεστή συσχέτισης και συγκρίνετε με το αποτέλεσμα της προηγούμενης άσκησης.

13. Έστω x, y ανεξάρτητες τ.μ. με από κοινού Σ.Π.Π. $f_{x,y}()$. Περιγράψτε τα βήματα για την δημιουργία ζευγών τ.α. που ακολουθούν την παραπάνω κατανομή στην περίπτωση που: (α) οι οριακές Σ.Π.Π. είναι αντιστρέψιμες, και (β) οι οριακές Σ.Π.Π. δεν είναι αντιστρέψιμες.

14. Το ίδιο όταν οι x, y είναι εξαρτημένες τ.μ.

15. Υπολογίστε αριθμητικά (χρησιμοποιώντας γεννήτρια τ.α.) την $E(xy)$ και $E(y|x)$ για τις τ.μ. x, y, z των οποίων η από κοινού Σ.Μ.Π. είναι:

$$f_{x,y,z}(x_0, y_0, z_0) = \left\{ \begin{array}{ll} x_0 z_0 + 3y_0 z_0, & x_0, y_0, z_0 \in [0,1] \\ & \\ 0, & \text{αλλού} \end{array} \right\}$$

Κεφάλαιο 4

Συνήθεις Κατανομές και Θεωρήματα

Στο κεφάλαιο αυτό θα μελετήσουμε τις πιο σημαντικές κατανομές που χρησιμοποιούνται σε προβλήματα στατιστικής μοντελοποίησης.

4.1 Διωνυμική Κατανομή

Σε ένα πείραμα με δυο ενδεχόμενα με πιθανότητα p και $1-p$ (π.χ., ρίψη νομίσματος, μετάδοση ενός bit πληροφορίας σε τηλεπικοινωνιακό δίαυλο), ορίζουμε την τ.μ. Bernoulli x με τιμές $x=1$ για το ενδεχόμενο με πιθανότητα p και $x=0$ για το άλλο ενδεχόμενο. Έχει επικρατήσει το ενδεχόμενο $x=1$ με πιθανότητα p να αναφέρεται ως επιτυχές (π.χ., το ενδεχόμενο επιτυχούς μετάδοσης του bit). Η Σ.Μ.Π. Bernoulli είναι:

$$p_x(x_0) = \left\{ \begin{array}{ll} 1-p, & x_0 = 0 \\ p, & x_0 = 1 \\ 0, & \text{αλλού} \end{array} \right\}$$

Οι στατιστικές ιδιότητες είναι:

$$\mu_x = E(x) = \sum_{x_0} x\, p_x(x_0) = 1 \times p + 0 \times (1-p) = p$$

$$E(x^2) = \sum_{x_0} x^2\, p_x(x_0) = 1^2 \times p + 0^2 \times (1-p) = p$$

$$\sigma_x^2 = E((x-\mu_x)^2) = E(x^2) - [E(x)]^2 = p - p^2 = p\,(1-p)$$

Η διωνυμική (binomial) τ.μ. k περιγράφει τον συνολικό αριθμό επιτυχιών σε ένα πείραμα με δυο ενδεχόμενα που επαναλαμβάνεται n φορές. Τα ενδεχόμενα των πειραμάτων είναι ανεξάρτητα μεταξύ τους και η πιθανότητα επιτυχίας σε κάθε πείραμα είναι p. Άρα η διωνυμική τ.μ. k είναι το άθροισμα n ανεξάρτητων τ.μ. Bernoulli $x_1, x_2 ... x_n$ που ακολουθούν την ίδια κατανομή (independent identically distributed (i.i.d.)), δηλαδή, $k = \sum_{i=1}^{n} x_i$. Επειδή ο αριθμός των συνδυασμών k αριθμών σε n-άδες (ανεξαρτήτου σειράς) είναι $\left(\begin{array}{c} n \\ k \end{array}\right)$ η διωνυμική Σ.Μ.Π. είναι:

$$p_k(k_0) = \left(\begin{array}{c} n \\ k_0 \end{array}\right) p^{k_0}(1-p)^{n-k_0} = \frac{n!}{(n-k_0)!k_0!}\, p^{k_0}(1-p)^{n-k_0}$$

Οι στατιστικές ιδιότητες της διωνυμικής προκύπτουν άμεσα από την Bernoulli:

$$\mu_k = E(k) = E(\sum_{i=1}^{n} x_i) = \sum_{i=1}^{n} E(x_i) = \sum_{i=1}^{n} p = np$$

$$\sigma_k^2 = E((k-\mu_k)^2) = E((\sum_{i=1}^{n}(x_i-p))^2) = E(\sum_i (x_i-p)^2 + \sum_i \sum_{j\neq i}(x_i-p)(x_j-p)) =$$

$$= \sum_i E((x_i-p)^2) + \sum_i \sum_{j\neq i} E(x_i-p)E(x_j-p) = \sum_i \sigma_{x_i}^2 + 0 = np(1-p)$$

όπου $E((x_i-p)(x_j-p)) = E(x_i-p)E(x_j-p)$ λόγω ανεξαρτησίας x_i, x_j.

Παράδειγμα: Διαβάζουμε 1000 bit από ένα CD. Η πιθανότητα λάθους ανάγνωσης ενός bit είναι 0.01. Τα λάθη είναι ανεξάρτητα μεταξύ τους. Ποια είναι η αναμενόμενη τιμή και η διασπορά του συνολικού αριθμού λαθών στην ανάγνωση του CD.

Ο συνολικός αριθμός λαθών k ακολουθεί την διωνυμική κατανομή με $n = 1000$ και $p = 0.01$. Άρα $\mu_k = np = 10$ και $\sigma_k^2 = np(1-p) \approx 10$.

Μπορούμε να ορίσουμε την τυχαία μεταβλητή l ως τον αριθμό από ρίψεις που χρειάζεται μέχρι να έχουμε την πρώτη επιτυχία. Η τ.μ. l ακολουθεί την γεωμετρική (geometric) κατανομή

$$p_l(l_0) = p(1-p)^{l_0-1}, \quad l_0 = 1, 2, ...$$

δηλαδή, την πιθανότητα να έχουμε $l_0 - 1$ αποτυχίες ακολουθούμενες από μια επιτυχία. Οι στατιστικές ιδιότητες[1] της γεωμετρικής είναι: $\mu_l = 1/p$, $\sigma_l^2 =$

[1] Ένας τρόπος να υπολογίσουμε την αναμενόμενη τιμή της l είναι αναδρομικά δηλαδή $E(l) = 1 \times p + (E(l)+1) \times (1-p)$ (εξηγήστε πως προκύπτει αυτή η αναδρομική σχέση).

$(1-p)/p^2$. Αντίστοιχα μπορούμε να ορίσουμε την τυχαία μεταβλητή l_r ως τον αριθμό από ρίψεις που χρειάζεται μέχρι να έχουμε r επιτυχίες. Η τ.μ. l_r ακολουθεί την κατανομή Pascal.

Τέλος η γενίκευση της binomial τ.μ. σε ένα πείραμα με περισσότερα από δυο αποτελέσματα λέγεται multinomial τ.μ.

4.2 Κατανομή Poisson

Η κατανομή Poisson προκύπτει από την διωνυμική κατανομή για συνεχή χρόνο, όταν δηλαδή το n τείνει στο άπειρο. Η Poisson τ.μ. k αναπαριστά τον συνολικό αριθμό επιτυχιών (ή αφίξεων όπως ονομάζονται συνήθως) σε ένα χρονικό διάστημα t. Προφανώς για να έχουμε πεπερασμένο αριθμό αφίξεων η πιθανότητα επιτυχίας p τείνει στο 0, ώστε το γινόμενο np (αναμενόμενος αριθμός επιτυχιών) να είναι σταθερός. Σε συνεχή χρόνο θεωρούμε ότι οι επιτυχίες (ή αφίξεις) γίνονται με σταθερό ρυθμό λ και ο αναμενόμενος αριθμός επιτυχιών σε χρόνο t είναι λt (λόγω ανεξαρτησίας μεταξύ των αφίξεων). Άρα η Σ.Μ.Π. Poisson προκύπτει για $n \to \infty$, $p \to 0$ και $np = \lambda t$. Συγκεκριμένα:

$$\lim_{n\to\infty} \frac{n!}{(n-k)!k!}\, p^k (1-p)^{n-k} = \lim_{n\to\infty} \frac{n!}{n^k(n-k)!}\, \frac{(np)^k}{k!} \left(1 - \frac{np}{n}\right)^{n-k}$$

Το όριο του πρώτου όρου είναι 1 και ο τρίτος όρος είναι η προσέγγιση του εκθετικού $\exp\{-np\}$. Αντικαθιστώντας $\lambda t = np$ η Σ.Μ.Π. Poisson είναι:

$$p_k(k_0, t) = \frac{(\lambda t)^{k_0}}{k_0!} e^{-\lambda t}, \quad t \geq 0, \quad k = 0, 1, 2...$$

Οι στατιστικές ιδιότητες της Poisson προκύπτουν άμεσα από αυτές της διωνυμικής:

$$\mu_k = E(k) = np = \lambda t$$

$$\sigma_k^2 = E((k - \mu_k)^2) = np \lim_{p\to 0}(1-p) = np = \lambda t$$

Η κατανομή Poisson χρησιμοποιείται για μοντελοποίηση πανομοιότυπων ανεξάρτητων γεγονότων στο χρόνο, π.χ., αφίξεις πακέτων σε ένα δίκτυο υπολογιστών, αφίξεις πελατών σε ουρά πολυκαταστήματος, Εφόσον ισχύει η συνθήκη της ανεξαρτησίας τα φαινόμενα αυτά περιγράφονται στατιστικά με μια μόνο παράμετρο: τον ρυθμό άφιξης λ.

Η γενίκευση της γεωμετρικής κατανομής στο συνεχή χρόνο, είναι η συνεχής τ.μ. l που ορίζεται ως ο χρόνος που χρειάζεται μέχρι την πρώτη άφιξη (first-order interarrival time). Η Σ.Π.Π του χρόνου πρώτης άφιξης l είναι η εκθετική κατανομή (exponential):

$$p_l(l_0) = \lambda e^{-\lambda l_0}, \quad l_0 \geq 0$$

με αναμενόμενη τιμή και διασπορά $1/\lambda$ και $1/\lambda^2$, αντίστοιχα. Η συνεχής τ.μ. l_r που ορίζεται ως ο χρόνος που χρειάζεται για r αφίξεις, ακολουθεί την Σ.Π.Π. Erlang (που είναι η γενίκευση της Σ.Μ.Π. Pascal σε συνεχή χρόνο).

Παράδειγμα: Βρείτε την διαφορική εξίσωση που περιγράφει την κατανομή Poisson ως άθροισμα ανεξάρτητων τ.μ. Bernoulli με πιθανότητα επιτυχίας σε απειροστό χρονικό διάστημα Δt ίση με $\lambda\Delta t$.

Υπολογίζουμε την πιθανότητα $P(k, t + \Delta t)$ των k επιτυχιών στο διάστημα $t + \Delta t$ αναδρομικά ως συνάρτηση της πιθανότητας επιτυχιών στο διάστημα t. Δεδομένου ότι μέσα στο διάστημα Δt μπορούμε να έχουμε μία επιτυχία με πιθανότητα $\lambda\Delta t$ ή καμία με πιθανότητα $1 - \lambda\Delta t$ (και με βάση την ανεξαρτησία των επιτυχιών):

$$P(k, t + \Delta t) = \lambda\Delta t \, P(k - 1, t) + (1 - \lambda\Delta t) \, P(k, t)$$

δηλαδή ή συνέβησαν $k - 1$ επιτυχίες μέσα στο διάστημα t και άλλη μία μέσα στο Δt, ή είχα εξαρχής k επιτυχίες στο t (και καμία μετά). Από τα παραπάνω:

$$\frac{P(k, t + \Delta t) - P(k, t)}{\Delta t} + \lambda P(k, t) = \lambda P(k - 1, t)$$

Στο όριο, όταν $\Delta t \to 0$ παίρνω την διαφορική εξίσωση:

$$\frac{d}{dt}P(k, t) + \lambda P(k, t) = \lambda P(k - 1, t)$$

της οποίας η λύση είναι η Σ.Μ.Π. Poisson (μπορείτε να το επαληθεύσετε παίρνοντας την παράγωγό της).

4.3 Ομογενής Κατανομή και Γεννήτρια Κατανομών

Η ομογενής τ.μ. x (uniform) είναι ομοιόμορφα κατανεμημένη στο διάστημα $[a, b]$. Η ομογενής Σ.Π.Π.:

$$f_x(x_0) = \left\{ \begin{array}{ll} \frac{1}{b-a}, & x_0 \in [a, b] \\ 0, & \text{αλλού} \end{array} \right\}$$

Οι στατιστικές ιδιότητές της ομογενούς τ.μ.:

$$\mu_x = E(x) = \int_{-\infty}^{\infty} x f_x(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{x^2|_a^b}{2(b-a)} = \frac{a+b}{2}$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f_x(x) dx = \frac{x^3|_a^b}{3(b-a)} = \frac{a^2+ab+b^2}{3}$$

$$\sigma_x^2 = E(x^2) - [E(x)]^2] = \frac{a^2+b^2-2ab}{12} = \frac{(a-b)^2}{12}$$

Η ομογενής κατανομή είναι ιδιαίτερα σημαντική για υπολογιστικές προσομοιώσεις πιθανοτήτων. Σχεδόν σε όλες τις γλώσσες προγραμματισμού υπάρχει μια γεννήτρια τυχαίων αριθμών που προσομοιώνει την ομογενή κατανομή στο διάστημα $[0, 1]$. Καλούμαστε λοιπόν να δημιουργήσουμε τις κατανομές που χρειαζόμαστε μέσω της ομογενής κατανομής. Για παράδειγμα μια γεννήτρια (δίκαιου) ζαριού που παράγει 10000 παραδείγματα της τ.μ. y καθώς και το ιστόγραμμα (εμπειρική κατανομή) της y, είναι:

```
y = ceil(rand(1,10000)*6);
hist(y,100);
```

Η μέση τιμή και διασπορά της y μπορεί στην συνέχεια να υπολογιστεί

```
>> disp(mean(y))
   3.4841
>> disp(var(y))
   2.9098
```

ή να φτιάξουμε συναρτήσεις της y, π.χ.,

```
y = ceil(rand(1,10000)*6);
z = ceil(rand(1,10000)*6);
w = (y+z).*[y ~= z] + 4*y.*[y == z];
hist(w,100);
```

όπου η τ.μ. w είναι το άθροισμα μιας ζαριάς στο τάβλι.

Παράδειγμα: Ξεκινώντας από την διωνυμική κατανομή φτιάξτε μια γεννήτρια τ.μ. Poisson με $\lambda t = 1$ στο MATLAB.

Υλοποιούμε την τ.μ. Poisson ως την διωνυμική με (μεγάλο) $n = 1000$ και $p = 1/n = 0.001$, ώστε $np = \lambda t = 1$. Δημιουργούμε 10000 δείγματα της Poisson.

```
x = rand(1000,10000);
y = sum([x<0.001]);
hist(y,100);
```

Συγκρίνουμε το αποτέλεσμα με την εντολή

```
plot(pdf('pois',[0:10],1);
```

που υπολογίζει την πιθανότητα της Poisson για $k = 0, 1, 2...10$ και $\lambda t = 1$. Είναι η μέση τιμή και η διασπορά 1;

4.4 Κανονική Κατανομή

Η κανονική (normal) τ.μ. x έχει Σ.Π.Π.:

$$f_x(x_0) = \frac{1}{\sqrt{2\pi}\,\sigma}\; e^{-\frac{(x_0-\mu)^2}{2\sigma^2}}$$

όπου η αναμενόμενη τιμή[2] είναι: $\mu_x = E(x) = \mu$ και η διασπορά είναι: $\sigma_x^2 = E((x-m)^2) = \sigma^2$. Συνήθως το ότι η τ.μ. x που ακολουθεί κανονική κατανομή με αναμενόμενη τιμή μ και διασπορά σ^2 συμβολίζεται ως: $x \sim \mathcal{N}(\mu, \sigma^2)$. Το ορισμένο ολοκλήρωμα της $f_x(.)$ δεν μπορεί να υπολογιστεί αναλυτικά οπότε η πιθανότητες ενδεχομένων κανονικών τ.μ. υπολογίζονται αριθμητικά, π.χ., η πιθανότητα να είμαστε 1, 2 ή 3 τυπικές αποκλίσεις από την μέση τιμή[3]:

$$Prob\ (\mu - \sigma \leq x \leq \mu + \sigma) = \int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sqrt{2\pi}\,\sigma}\; e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \approx 0.68$$

$$Prob\ (\mu - 2\sigma \leq x \leq \mu + 2\sigma) \approx 0.95$$

$$Prob\ (\mu - 3\sigma \leq x \leq \mu + 3\sigma) \approx 0.997$$

[2]Είναι εύκολο να δούμε ότι το ολοκλήρωμα $\int_{-\infty}^{\infty}(x-\mu)f_x(x)dx = 0$ γιατί $(x-\mu)$ είναι αντισυμμετρική συνάρτηση και $f_x(x)$ συμμετρική ως προς τον άξονα $x = \mu$. Άρα το γινόμενο τους είναι αντισυμμετρική συνάρτηση με ολοκλήρωμα ίσο με το 0 και άρα $E(x) = \mu$.

[3]Οι αριθμοί αυτοί είναι ιδιαίτερα σημαντικοί στο πεδίο της στατιστικής ανάλυσης δεδομένων. Πολύ συχνά θεωρούμε ότι οι μετρήσεις μας ακολουθούν την κανονική κατανομή (βλ. κεντρικό οριακό θεώρημα) οπότε όταν δίνουμε διαστήματα εμπιστοσύνης για μία πρόβλεψη, π.χ., το κόμμα Χ θα πάρει 21.2% ± 1.3% στις εκλογές, το 1.3 είναι δύο φορές η τυπική απόκλιση των δειγμάτων μας και η πιθανότητα λάθους (η πρόβλεψη μας να είναι εκτός του διαστήματος) είναι περίπου 5%. Αντίστοιχα όταν θέλουμε να εξαιρέσουμε μετρήσεις που πιστεύουμε ότι είναι λανθασμένες (outliers) συνήθως εξαιρούμε μετρήσεις που είναι τρεις τυπικές αποκλίσεις μακριά από την μέση τιμή. Αν οι μετρήσεις μας ακολουθούν την κανονική κατανομή, τότε μαζί με outliers πετάμε και 0.3% από τις σωστές μετρήσεις μας (υπολογίστε πως επηρεάζει αυτό την εκτιμώμενη διασπορά).

Για τον αριθμητικό υπολογισμό των παραπάνω ολοκληρωμάτων χρησιμοποιούμε συνήθως την συνάρτηση κατανομής της κανονικής τ.μ. y με αναμενόμενη τιμή 0 και διασπορά 1, δηλαδή, $y \sim \mathcal{N}(0,1)$. Είναι εύκολο να δείξουμε ότι ο μετασχηματισμός $y = (x-\mu)/\sigma$ δίνει το επιθυμητό αποτέλεσμα. Η συνάρτηση κατανομής της $y \sim \mathcal{N}(0,1)$ ονομάζεται συνήθως $\Phi(.)$. Η Σ.Π.Π. και η συνάρτηση κατανομής της $y \sim \mathcal{N}(0,1)$ είναι

$$f_y(y_0) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y_0^2}{2}}$$

$$F_y(y_0) \equiv \Phi(y_0) = Prob(y \leq y_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_0} e^{-\frac{y^2}{2}} dy$$

Συνήθως δίδονται πίνακες της συνάρτησης $\Phi(.)$ ή συναρτήσεις που υπολογίζουν αριθμητικά το ολοκλήρωμα (σε υπολογιστές). Τα περισσότερα υπολογιστικά πακέτα χρησιμοποιούν την συνάρτηση $erf(y_0) = 1/\sqrt{2\pi} \int_0^{y_0} e^{-y^2} dy$ που σχετίζεται με την $\Phi(.)$ ως:

$$\Phi(y_0) = \frac{1}{2} + \frac{1}{2} erf\left(\frac{y_0}{\sqrt{2}}\right)$$

Εν γένει για $x \sim \mathcal{N}(\mu, \sigma^2)$:

$$Prob(x \leq x_0) = Prob\left(\frac{x-\mu}{\sigma} \leq \frac{x_0-\mu}{\sigma}\right) = \Phi\left(\frac{x_0-\mu}{\sigma}\right) = \frac{1}{2}\left[1 + erf\left(\frac{x_0-\mu}{\sigma\sqrt{2}}\right)\right]$$

Τέλος οι κεντρικές ροπές της κανονική κατανομής έχουν ως εξής:

$$E\left((x-\mu)^{2k}\right) = \frac{(2k)!}{2^k k!}\sigma^{2k}, \qquad E\left((x-\mu)^{2k-1}\right) = 0 \qquad k = 1,2,3...$$

δηλαδή οι άρτιες κεντρικές ροπές n τάξης είναι ανάλογες του σ^n και όλες οι περιττές κεντρικές ροπές[4] είναι 0.

Η κανονική ή Gaussian κατανομή είναι η σημαντικότερη κατανομή σε προβλήματα μοντελοποίησης γιατί (όπως θα δούμε αμέσως μετά) το αποτέλεσμα των περισσότερων αυξητικών φυσικών διαδικασιών, π.χ., το μήκος μια ποικιλίας ψαριού, ακολουθεί κανονική κατανομή. Χρησιμοποιείται επίσης σχεδόν αποκλειστικά για την μοντελοποίηση λαθών, π.χ., μετρήσεων. Επιπρόσθετα οι κανονικές τ.μ. έχουν μια ομάδα χρήσιμων ιδιοτήτων, μεταξύ αυτών:

- Η κανονική κατανομή μεγιστοποιεί την εντροπία πληροφορίας (από όλες τις κατανομές με γνωστή αναμενόμενη τιμή και διασπορά).

[4] Γιατί το ολοκλήρωμα από το $-\infty$ μέχρι το $+\infty$ αντισυμμετρικών συναρτήσεων είναι 0 (γινόμενο της αντισυμμετρικής συνάρτησης $(x-\mu)^{2k-1}$ και της συμμετρικής $f_x()$, με τον ίδιο άξονα συμμετρίας $x = \mu$).

- Το άθροισμα ανεξάρτητων κανονικών τ.μ. ακολουθεί κανονική κατανομή (επίσης η κανονική κατανομή είναι strictly stable).
- Εάν το άθροισμα ανεξάρτητων τ.μ. ακολουθεί την κανονική κατανομή τότε και κάθε μια από τις τυχαίες μεταβλητές που αποτελεί το άθροισμα ακολουθεί κανονική κατανομή (infinite divisibility).

Λόγω της χρησιμότητας της κανονικής κατανομής, οι ακόλουθες κατανομές συναρτήσεων κανονικών τ.μ. εμφανίζονται επίσης στην πράξη:

- Το τετράγωνο (ή άθροισμα τετραγώνων) κανονικών τ.μ. ακολουθεί chi-square κατανομή.
- Ο λόγος δυο ανεξάρτητων κανονικών τ.μ. ακολουθεί Cauchy κατανομή.
- Το εκθετικό e^x μια κανονικής τ.μ. x ακολουθεί log-normal κατανομή.
- Η τετραγωνική ρίζα $\sqrt{x^2+y^2}$ του αθροίσματος των τετραγώνων δυο ανεξάρτητων τ.μ. x, y, ακολουθεί Rayleigh κατανομή.

<u>Παράδειγμα 1:</u> Υπολογίστε την πιθανότητα η τ.μ. $x \sim \mathcal{N}(\mu,\sigma^2)$ να παίρνει τιμές μεταξύ a και b.

$$Prob(a \le x \le b) = Prob\left(\frac{a-\mu}{\sigma} \le \frac{x-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right)$$

Ξέρουμε ότι η τ.μ. $y=(x-\mu)/\sigma$ ακολουθεί $\mathcal{N}(0,1)$ οπότε:

$$Prob(a \le x \le b) = Prob\left(\frac{a-\mu}{\sigma} \le y \le \frac{b-\mu}{\sigma}\right) =$$

$$= Prob\left(\frac{b-\mu}{\sigma} \le y\right) - Prob\left(\frac{a-\mu}{\sigma} \le y\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Χρησιμοποιώντας ολοκληρώματα μπορούμε να δείξουμε την ίδια σχέση:

$$Prob(a \le x \le b) = \int_a^b \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}} dy =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(b-\mu)/\sigma} e^{-\frac{y^2}{2}} dy - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(a-\mu)/\sigma} e^{-\frac{y^2}{2}} dy = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

όπου έχουμε κάνει την αλλαγή μεταβλητής $y=(x-\mu)/\sigma$ (άρα $dy = dx/\sigma$).

<u>Παράδειγμα 2:</u> Υπολογίστε τον συντελεστή συσχέτισης μεταξύ των τ.μ. x και $y=x^2$ όταν η τ.μ. $x \sim \mathcal{N}(0,\sigma^2)$.

$$E(y) = E(x^2) = \sigma^2 + [E(x)]^2 = \sigma^2$$

Από τις τιμές των κεντρικών ροπών κανονικής τ.μ. (κι αφού $\mu = 0$):

$$E(xy) = E(xx^2) = E(x^3) = 0 \qquad E(y^2) = E(x^4) = 3\sigma^4$$

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x\sigma_y} = \frac{E(xy) - E(x)E(y)}{\sigma_x\sigma_y} = \frac{0}{\sqrt{\sigma^2 3\sigma^4}} = 0$$

Άρα οι (πλήρως εξαρτημένες) τ.μ. x και x^2 προκύπτουν γραμμικά ανεξάρτητες!

Παράδειγμα 3: Υπολογίστε τον συντελεστή συσχέτισης μεταξύ των τ.μ. x και $y = ax+b+n$ όταν $x \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $n \sim \mathcal{N}(\mu_n, \sigma_n^2)$ είναι ανεξάρτητες κανονικές τ.μ. και a, b είναι σταθερές.

$$\mu_y = E(y) = E(ax + b + n) = aE(x) + b + E(n) = a\mu_x + b + \mu_n$$

$$\sigma_y^2 = E\left((y - \mu_y)^2\right) = E\left((ax + b + n - (a\mu_x + b + \mu_n))^2\right) =$$

$$= E\left(a^2(x - \mu_x)^2\right) + E\left((n - mu_n)^2\right) = a^2\sigma_x^2 + \sigma_n^2$$

$$\sigma_{xy} = E\left((x - \mu_x)(y - \mu_y)\right) = E\left((x - \mu_x)[a(x - \mu_x) + (n - \mu_n)]\right) =$$

$$= aE\left((x - \mu_x)^2\right) + E\left((x - \mu_x)(n - \mu_n)\right) = a\sigma_x^2$$

όπου οι όροι που περιέχουν τις ανεξάρτητες τ.μ. x και n είναι 0. Άρα:

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x\sigma_y} = \frac{a\sigma_x^2}{\sigma_x\sqrt{a^2\sigma_x^2 + \sigma_n^2}} = \frac{1}{\sqrt{1 + (\sigma_n/(a\sigma_x))^2}}$$

Θεωρώντας ότι x είναι το σήμα, n είναι ο θόρυβος, και y είναι το θορυβώδες σήμα, ο συντελεστής συσχέτισης μεταξύ του σήματος και του θορυβώδους σήματος εξαρτάται από το κέρδος a και από το λόγο της διασποράς σήματος θορύβου σ_x/σ_n. Όταν το κέρδος $a \to \infty$ ή ο σηματοθορυβικός λόγος $(\sigma_x/\sigma_n) \to \infty$ (ή εν γένει παίρνουν μεγάλες τιμές), τότε $\rho_{xy} \to 1$ και τα δυο σήματα είναι πλήρως γραμμικά εξαρτημένα (επιθυμητό). Αντίθετα για μικρό κέρδος και μικρό σηματοθορυβικό λόγο τα x, y είναι γραμμικά ανεξάρτητα: $a \to 0$, $(\sigma_x/\sigma_n) \to 0 \Rightarrow \rho_{xy} \to 0$.

4.5 Κεντρικό Οριακό Θεώρημα

Η πιο σημαντική ιδιότητα της κανονικής κατανομής προκύπτει από το κεντρικό οριακό θεώρημα:

Κεντρικό Οριακό Θεώρημα (Central Limit Theorem): Το άθροισμα $r = x_1 + x_2 + ...x_n$, n ανεξάρτητων τ.μ. $x_1, x_2, ...x_n$ που ακολουθούν την ίδια κατανομή (i.i.d.) με μέση τιμή μ_x και διασπορά σ_x^2 προσεγγίζει την κανονική κατανομή για μεγάλες τιμές του n. Δηλαδή όταν $n \to \infty$, τότε $r \sim \mathcal{N}(n\mu, n\sigma^2)$. Πιο σωστά[5] όταν $n \to \infty$ τότε η τ.μ. $(r/\sqrt{n}) \sim \mathcal{N}(\mu, \sigma^2)$.

Η συνήθης απόδειξη του κεντρικού οριακού θεωρήματος (αν και απλή) χρησιμοποιεί χαρακτηριστικές συναρτήσεις Σ.Π.Π. που δεν τις έχουμε διδαχθεί. Αντ' αυτού (σε αδρές γραμμές) προτείνουμε μια απόδειξη που βασίζεται στον υπολογισμό κεντρικών ροπών της τ.μ. r. Οι κεντρικές ροπές άρτιας τάξης της $r/\sqrt{n}$ είναι

$$E\left(\left(\frac{r}{\sqrt{n}} - \mu_x\right)^{2k}\right) = \frac{1}{n^k}\, E\left(\left(\sum_i (x_i - \mu_x)\right)^{2k}\right) =$$

$$= \frac{1}{n^k}\, E(\sum_{i_1}\sum_{i_2}...\sum_{i_{2k}} [(x_{i_1} - \mu_x)(x_{i_2} - \mu_x)...(x_{i_{2k}} - \mu_x)])$$

Επειδή $n \to \infty$ με ενδιαφέρει να κρατήσω μόνο τους όρους του τύπου $i_1 = i_2 \neq i_3 = i_4 \neq ... \neq i_{2k-1} = i_{2k}$ (δηλαδή να ζευγαρώσω τους όρους) ώστε να μεγιστοποιήσω την τάξη[6] ως προς n. Πράγματι το πλήθος αυτών των όρων είναι της τάξης $O(n^k)$ και η τιμή κάθε όρου είναι σ_x^{2k}, οπότε

$$\lim_{n\to\infty} E\left(\left(\frac{r}{\sqrt{n}} - \mu_x\right)^{2k}\right) \sim \lim_{n\to\infty} \frac{1}{n^k}(n^k \sigma_x^{2k}) = \sigma_x^{2k}$$

Οι υπόλοιποι όροι μπορούν να αγνοηθούν όταν $n \to \infty$ γιατί είναι το πολύ $O(n^{k-1})$ το πλήθος. Δείξαμε λοιπόν ότι οι κεντρικές ροπές άρτιας τάξης είναι ανάλογες του σ_x^{2k} (μπορείτε ως άσκηση να υπολογίσετε το ακριβές πλήθος των ζευγαριών και από εκεί την ακριβή τιμή των άρτιων ροπών).

Για τους όρους περιττής τάξης $2k+1$ δεν μπορώ να ζευγαρώσω όλους τους όρους. Το καλύτερο που μπορώ να κάνω είναι να δημιουργήσω μια τριάδα και να ζευγαρώσω τους υπόλοιπους όρους (σε $k-1$ ζευγάρια), π.χ., $i_1 = i_2 \neq$

[5] Αποφεύγουμε έτσι στο όριο $n \to \infty$ άπειρη αναμενόμενη τιμή και διασπορά.

[6] Πράγματι αν, π.χ., πάρω τους όρους $i_1 = i_2 = ... = i_{2k}$ τότε το πλήθος τους είναι n.

$i_3 = i_4 \neq ... \neq i_{2k-1} = i_{2k} = i_{2k+1}$. Το πλήθος όρων τέτοιου τύπου είναι της τάξης $O(n^{(k-1)+1}) = O(n^k)$ και άρα οι ροπές περιττής τάξης στο όριο είναι:

$$\lim_{n\to\infty} E\left(\left(\frac{r}{\sqrt{n}} - \mu_x\right)^{2k+1}\right) = c \lim_{n\to\infty} \frac{1}{n^{k+0.5}} n^k = c \lim_{n\to\infty} \frac{1}{\sqrt{n}} = 0$$

όπου c (είναι όρος ανάλογος του γινομένου $\sigma_x^{2(k-1)} E((x-\mu_x)^3)$ και) δεν εξαρτάται από το n. Άρα οι κεντρικές ροπές περιττής τάξης της τ.μ. $(r/\sqrt{n})$ είναι στο όριο 0. Δείξαμε ότι όλες οι στατιστικές ιδιότητες της $(r/\sqrt{n})$ είναι ίδιες με κανονικής τ.μ. $\mathcal{N}(\mu, \sigma^2)$ όταν $n \to \infty$ και άρα $(r/\sqrt{n}) \sim \mathcal{N}(\mu, \sigma^2)$ (γιατί ισχύει το παραπάνω;).

Η απόδειξη αυτή είναι αποκαλυπτική ως προς της ταχύτητα σύγκλισης (σε κανονική κατανομή) αθροίσματος τυχαίων μεταβλητών. Ο ρυθμός σύγκλισης των περιττών ροπών είναι $1/\sqrt{n}$, δηλαδή πολύ αργός. Για τ.μ. x που έχουν Σ.Π.Π. με υψηλό βαθμό ασσυμετρίας (δηλαδή με υψηλές τιμές $E((x-\mu_x)^3)$ οπότε c μεγάλο) μπορεί να χρειαστεί ακόμα και $n > 1000$ (π.χ., για διωνυμικής κατανομές με p πολύ μικρό ή πολύ μεγάλο). Αντίθετα αν η Σ.Π.Π. της x είναι συμμετρική τότε $E((x-\mu_x)^{2k+1}) = 0$ και οι άρτιες ροπές συγκλίνουν σχετικά γρήγορα (ποιος είναι ο ρυθμός σύγκλισης των άρτιων ροπών;). Για παράδειγμα το άθροισμα 12 ανεξάρτητων τ.μ. που ακολουθούν ομογενή κατανομή προσεγγίζει πολύ καλά την κανονική κατανομή.

Επίσης είναι εύκολο να δούμε (με βάση την παραπάνω απόδειξη) ότι το κεντρικό οριακό θεώρημα ισχύει και όταν οι τ.μ. $x_1, x_2, ...x_n$ είναι 'περίπου' (i.i.d.), δηλαδή όταν έχουνε, π.χ., μικρό βαθμό γραμμικής εξάρτησης ή οι τ.μ. προέρχονται από κατανομές με παρόμοιες στατιστικές ιδιότητες 1ης, 2ης και 3ης τάξης (αναμενόμενη τιμή, διασπορά, κεντρική ροπή 3-ης τάξης).

Σημασία του Κεντρικού Οριακού Θεωρήματος: Οι φυσικές αυξητικές διαδικασίες συνήθως συντελούνται ως άθροισμα από ανεξάρτητες μικρές μεταβολές. Για παράδειγμα το μέγεθος οργανισμών (ύψος, βάρος, μήκος μαλλιών) μπορεί να προσεγγιστεί συνήθως ως άθροισμα τέτοιων ανεξάρτητων μεταβολών και άρα ακολουθεί (περίπου) κανονική κατανομή. Πολλές άλλες φυσικές μετρήσεις ακολουθούν επίσης (περίπου) κανονική κατανομή αν και η φυσική διαδικασία που τις δημιούργησε μπορεί να μην είναι πλήρως κατανοητή, π.χ., ανθρώπινες δεξιότητες. Για παράδειγμα οι βαθμοί στο τελικό διαγώνισμα ενός μαθήματος συνήθως ακολουθούν κανονική κατανομή γιατί η μαθησιακή διαδικασία μπορεί να προσομοιωθεί ως το άθροισμα των (i.i.d) ψηγμάτων γνώσης που απέκτησε ο φοιτητής. Τέλος υπάρχουν και πολλαπλασιαστικές αυξητικές διαδικασίες όπου $r = \prod_{i=1}^{n} x_i$, όπου μπορούμε να υποθέσουμε ότι $x_1, x_2 ... x_n$ i.i.d τ.μ. Για να εφαρμόσουμε το κεντρικό οριακό θεώρημα, ξαναγράφουμε αυτές τις διαδικασίες ως $\log(r) = \sum_{i=1}^{n} \log(x_i)$ και άρα στο όριο $n \to \infty$ η $\log r$ ακολουθεί κανονική κατανομή (και άρα η $r = e^{\log(r)}$ ακολουθεί log-normal κατανομή).

Ο συνδυασμός του κεντρικού οριακού θεωρήματος και των ιδιοτήτων της κανονικής κατανομής (μεγιστοποίηση εντροπίας, strict stability) καθιστούν την κανονική τ.μ. την πρώτη επιλογή σε προβλήματα μοντελοποίησης και στατιστικής ανάλυσης. Σε συστήματα τηλεπικοινωνιών ο θόρυβος μοντελοποιείται σχεδόν αποκλειστικά με κανονική τ.μ. (όπως και τα λάθη μετρήσεων).

Παράδειγμα: Υπολογίστε την πιθανότητα να έχουμε μεταξύ 2100 και 2200 επιτυχιών για μια διωνυμική κατανομή με $n = 3000$ και πιθανότητα επιτυχίας $p = 0.7$.

$$Prob(a \leq k \leq b) = Prob\left(\frac{a-np}{\sqrt{np(1-p)}} \leq \frac{k-\mu_k}{\sigma_k} \leq \frac{b-np}{\sqrt{np(1-p)}}\right)$$

όπου η διωνυμική τ.μ. $r = (k-\mu_k)/\sigma_k$ ακολουθεί (περίπου) κανονική κατανομή $\mathcal{N}(0,1)$ λόγω του κεντρικού οριακού θεωρήματος. Άρα:

$$Prob(2100 < k \leq 2200) = Prob(0 < r \leq 3.98) \approx \Phi(4) - \Phi(0) = 1 - 0.5 = 0.5$$

4.6 Αριθμητική Γεννήτρια Κατανομών

Μπορούμε να δημιουργήσουμε μια γεννήτρια αριθμών που ακολουθούν την κανονική κατανομή $\mathcal{N}(0,1)$ με βάση το κεντρικό οριακό θεώρημα αθροίζοντας 20 ομογενείς κατανομές αφαιρώντας τη αναμενόμενη τιμή και διαιρώντας με την τυπική απόκλιση ως εξής (υπενθυμίζουμε ότι η αναμενόμενη τιμή και διασπορά ομογενούς στο [0,1] είναι 0.5 και 1/12 αντίστοιχα):

```
N = 10000;
n = 20;
x = rand(n,N);        % generate n uniform pdfs in [0,1]
y = sum(x);
standard_normal = (y-n*0.5)/(sqrt(n*1/12)); % so that N(0,1)
hist(standard_normal,100);
```

Πράγματι οι ροπές 1ης, 2ης και 3ης τάξης είναι περίπου 0,1 και 0 αντίστοιχα

```
>> disp(mean(standard_normal))
   2.7730e-05
>> disp(var(standard_normal))
  1.0213
>> disp(mean(standard_normal.^3))
  0.0237
```

Συγκρίνετε τα αποτελέσματα με την γεννήτρια randn() του MATLAB. Για επαλήθευση, μπορούμε επίσης να υπολογίσουμε το συντελεστή συσχέτισης μεταξύ τ.μ. $x \sim \mathcal{N}(0,1)$ και $y = x^2$ ως εξής:

```
>> disp(corrcoef(standard_normal, standard_normal.^2))
    1.0000    0.0162
    0.0162    1.0000
```

που είναι κοντά στο 0 (Σημείωση: το αποτέλεσμα δίδεται ως $\begin{bmatrix} \rho_{xx} & \rho_{xy} \\ \rho_{yx} & \rho_{yy}) \end{bmatrix}$).

Σύμφωνα με το κεντρικό οριακό θεώρημα η διωνυμική κατανομή για μεγάλες τιμές του n προσεγγίζει την κανονική κατανομή (αφού η διωνυμική είναι άθροισμα από n i.i.d Bernoulli). Η ταχύτητα σύγκλισης εξαρτάται από το p επιτυχίας που καθορίζει τον βαθμό συμμετρίας της Σ.Μ.Π. Bernoulli. Για παράδειγμα συγκρίνουμε την ταχύτητα σύγκλισης για άθροισμα $n = 100$ Bernoulli με διαφορετικό βαθμό ασσυμετρίας:

```
N = 10000; n = 100; p1 = 0.3;  p2 = 0.1;
binomial1 = sum([rand(n,N) < p1]); % binomial n=100, p=0.3
binomial2 = sum([rand(n,N) < p2]); % binomial n=100, p=0.1
norm1 = (binomial1-n*p1)/(sqrt(n*p1*(1-p1)));
norm2 = (binomial2-n*p2)/(sqrt(n*p2*(1-p2)));
hist(norm1,n+1);
disp(mean(norm1.^3));
disp(mean(norm2.^3));
```

Το αποτέλεσμα (εκτίμησης) της ροπής 3ης τάξης είναι περίπου 0.07 για τη διωνυμική με $p = 0.3$ και 0.25 για την διωνυμική με $p = 0.1$. Βλέπουμε λοιπόν ότι για μικρές (ή μεγάλες) τιμές του p η ταχύτητα σύγκλισης είναι πολύ αργή. Εν γένει χρειαζόμαστε μερικές 100-άδες Bernoulli με τιμές του p κοντά στο 0.5 για να προσεγγίσουμε την κανονική κατανομή (Σημείωση: ένα άλλο πρόβλημα εδώ είναι ότι προσεγγίζουμε μια Σ.Π.Π. χρησιμοποιώντας μια Σ.Μ.Π., π.χ., για $n = 100$ έχουμε 101 δυνατά αποτελέσματα και άρα δειγματοληπτούμε την συνεχή Σ.Π.Π. σε 101 σημεία το πολύ).

Εκτός από την χρήση του κεντρικού οριακού θεωρήματος υπάρχουν αναλυτικοί τρόποι δημιουργίας δειγμάτων της κανονικής Σ.Π.Π. Ένας συνήθης αναλυτικός τρόπος για την δημιουργία κανονικών τ.μ. από ομογενείς τ.μ. είναι η μέθοδος των Box-Muller:

```
N = 10000;
x1 = rand(1,N);
x2 = rand(1,N);
```

```
y1 = sqrt(-2*log(x1)).*cos(2*pi*x2);
y2 = sqrt(-2*log(x1)).*sin(2*pi*x2);
hist(y1,100);
corrcoef(y1,y2);
hist(y1+y2,100);
```

Πράγματι βλέπουμε ότι τα $y1$, $y2$ ακολουθούν κανονική κατανομή και είναι ασυσχέτιστες (πώς μπορώ να δείξω ανεξαρτησία αριθμητικά;). Τέλος βλέπουμε ότι το άθροισμα $y1 + y2$ επίσης ακολουθεί κανονική κατανομή.

4.7 Ανισότητα Chebyshev

Για κάθε τ.μ. x με μέση τιμή μ_x και διασπορά σ_x^2, η ανισότητα Chebyshev προτείνει ένα άνω φράγμα της πιθανότητας οι τιμές της τ.μ. x να βρίσκονται εκτός του διαστήματος $[\mu_x - t, \mu_x + t]$

$$Prob(|x - \mu_x| \geq t) \leq \left(\frac{\sigma_x}{t}\right)^2$$

Η ανισότητα αυτή είναι σημαντική γιατί ισχύει για κάθε τ.μ. (ανεξαρτήτως κατανομής) και το φράγμα εξαρτάται μόνο από τις στατιστικές ιδιότητες 2ης τάξης της τ.μ. Η απόδειξη της ανισότητας είναι σχετικά απλή:

$$\sigma_x^2 = \int_{-\infty}^{+\infty} (x-\mu_x)^2 f_x(x)dx \geq \int_{-\infty}^{\mu_x - t} (x-\mu_x)^2 f_x(x)dx + \int_{\mu_x+t}^{+\infty} (x-\mu_x)^2 f_x(x)dx$$

$$\geq \int_{-\infty}^{\mu_x - t} t^2 f_x(x)dx + \int_{\mu_x+t}^{+\infty} t^2 f_x(x)dx \;=\; t^2 Prob(|x - \mu_x| \geq t)$$

όπου η αλλαγή του $(x - \mu_x)^2$ με t^2 μπορεί να γίνει γιατί $(x - \mu_x)^2 \geq t^2$ στη περιοχή υπολογισμού του ολοκληρώματος (στο τελευταίο βήμα η t είναι σταθερά και βγαίνει εκτός ολοκληρώματος).

Παράδειγμα: Υπολογίστε το πάνω φράγμα της πιθανότητας μια τ.μ. x να βρίσκεται μια, δύο ή τρεις τυπικές αποκλίσεις μακριά από την μέση τιμή της. Συγκρίνεται με την αντίστοιχη πιθανότητα για την κανονική τ.μ. $x \sim \mathcal{N}(0,1)$ (δηλαδή με την πιθανότητα να βρίσκεται εκτός του διαστήματος $[-1,1]$, $[-2,2]$ ή $[-3,3]$).

Σύμφωνα με την ανισότητα Chebyshev

$$Prob(|x - \mu_x| \geq \sigma_x) \leq \left(\frac{\sigma_x}{\sigma_x}\right)^2 \Rightarrow Prob(|x - \mu_x| \geq \sigma_x) \leq 1$$

$$Prob(|x - \mu_x| \geq 2\sigma_x) \leq \left(\frac{\sigma_x}{2\sigma_x}\right)^2 \Rightarrow Prob(|x - \mu_x| \geq \sigma_x) \leq 0.25$$

$$Prob(|x - \mu_x| \geq 3\sigma_x) \leq \left(\frac{\sigma_x}{3\sigma_x}\right)^2 \Rightarrow Prob(|x - \mu_x| \geq \sigma_x) < 0.12$$

Βλέπουμε πως στην περίπτωση της κανονικής κατανομής το όριο που προκύπτει από την ανισότητα είναι αρκετά μακριά από την αλήθεια, π.χ., $Prob(|x| \geq 2) \approx 0.05$ αντί για 0.25 (αντίστοιχα 0.003 και 0.12 για τρεις τυπικές αποκλίσεις μακριά). Επίσης το όριο είναι μεγαλύτερο ή ίσο του 1 για $t \leq \sigma_x$ (πρακτικά άχρηστο αφού η πιθανότητα είναι αξιωματικά πάντα μικρότερη του ή ίση με 1). Συνολικά η ανισότητα Chebyshev είναι πολύ σημαντική γιατί μας λέει, π.χ., ότι τουλάχιστον 3 στις 4 τιμές της x (κατά μέσο όρο) βρίσκονται στο διάστημα που ορίζουν δύο τυπικές αποκλίσεις από την μέση τιμή, ανεξαρτήτως της κατανομή που ακολουθεί η x!

4.8 Σύνοψη

- Το θεμελιώδες μοντέλο πιθανότητας που παράγει τις πιο χρήσιμες κατανομές είναι αυτό των ανεξάρτητων πανομοιότυπων πειραμάτων με δύο επιλογές.

- Η Bernoulli τ.μ. x μοντελοποιεί το αποτέλεσμα ενός πειράματος με δύο ενδεχόμενα, όπου $x = 1$ είναι η 'επιτυχής' έκβαση του πειράματος με πιθανότητα p ($x = 0$ αλλιώς).

- Η διωνυμική τ.μ. k είναι το άθροισμα n ανεξάρτητων Bernoulli τ.μ., και μοντελοποιεί τον αριθμό των επιτυχιών n ανεξάρτητων πανομοιότυπων πειραμάτων με πιθανότητα επιτυχίας p.

- Η Poisson τ.μ. k είναι αντίστοιχη της διωνυμικής για συνεχή χρόνο, δηλαδή όταν ο αριθμός των πειραμάτων τείνει στο άπειρο (ένα για κάθε χρονική στιγμή) και η πιθανότητα επιτυχίας p τείνει στο 0. Το γινόμενο $np = \lambda t$, όπου λ είναι ο ρυθμός επιτυχιών ή αφίξεων, και t είναι ο χρόνος.

- Η Poisson τ.μ. k μοντελοποιεί τον αριθμό από *ανεξάρτητες* αφίξεις σε ένα χρονικό διάστημα διάρκειας t.

- Η ομογενής κατανομή είναι ιδιαίτερα σημαντική γιατί (σχεδόν) όλες οι γεννήτριες τυχαίων αριθμών rand() παράγουν αριθμούς που την ακολουθούν.

- Η κανονική κατανομή είναι συνάρτηση (μόνο) της αναμενόμενης τιμής μ και διασποράς της σ.
- Η πιθανότητα οι τιμές μιας κανονικής τ.μ. να βρίσκονται 1, 2, ή 3 τυπικές αποκλίσεις από την αναμενόμενη τιμή είναι 0.68, 0.95, 0.997 αντίστοιχα.
- Σύμφωνα με το κεντρικό οριακό θεώρημα (Κ.Ο.Θ.) το άθροισμα n ανεξάρτητων τ.μ. που ακολουθούν την ίδια κατανομή συγκλίνει στην κανονική κατανομή (ταχύτητα σύγκλισης περιττών ροπών $1/\sqrt{n}$).
- Λόγω του Κ.Ο.Θ. κανονικές (ή log-normal) τ.μ. μπορούν να μοντελοποιήσουν αθροιστικά (ή πολλαπλασιαστικά) αυξητικά φαινόμενα, που απαντώνται συχνά στη φύση.
- Λόγω του Κ.Ο.Θ. και των χρήσιμων ιδιοτήτων της κανονικής κατανομής (μεγιστοποίηση εντροπίας, strict stability, infinite divisibility) η κανονική Σ.Π.Π. χρησιμοποιείται σε πολλά προβλήματα μοντελοποίησης, και σχεδόν αποκλειστικά για την μοντελοποίηση θορύβου ή λάθους μετρήσεων.

Ασκήσεις

1. Υπολογίστε την Σ.Μ.Π της κατανομής Pascal καθώς και τις στατιστικές της ιδιότητες.

2. Υπολογίστε την εκθετική κατανομή και την κατανομή Erlang χρησιμοποιώντας απειροστικό λογισμό (βλ. παράδειγμα για Poisson).

3. Ξεκινώντας από την διωνυμική κατανομή φτιάξτε μια γεννήτρια τ.μ. Poisson με $\lambda t = 1$ στο MATLAB. Συγκρίνετε την ακρίβεια της γεννήτριας για $n = 1000$, $n = 100$, $n = 10$ (και αντίστοιχα $p = 0.001$, $p = 0.01$, $p = 0.1$). Πόσο είναι το μέσο τετραγωνικό λάθος της Σ.Π.Π. ως συνάρτηση του n.

4. Υλοποιήστε την αλγόριθμο του Knuth για την δημιουργία Poisson τυχαίων αριθμών

   init: Let $L \leftarrow e^{-\lambda}$, $k \leftarrow 0$ and $p \leftarrow 1$
   do:
   $k \leftarrow k + 1$
   Generate uniform random number u in $[0, 1]$ and let $p \leftarrow p \times u$
   while $p \geq L$
   return $k - 1$

5. Δείξτε ότι το άθροισμα δυο ανεξάρτητων Poisson τ.μ., είναι επίσης Poisson τ.μ. Υπολογίστε τη μέση τιμή και διασπορά.

6. Δείξτε ότι εφόσον οι αφίξεις ακολουθούν κατανομή Poisson οι αφίξεις στο χρονικό διάστημα $[0, t]$ είναι ανεξάρτητες των αφίξεων στο διάστημα $(t, 2t]$.

7. Δείξτε πειραματικά ότι ισχύει το παραπάνω χρησιμοποιώντας γεννήτρια αριθμών που ακολουθούν την κατανομή Poisson.

8. Δείξτε ότι το άθροισμα δυο ανεξάρτητων κανονικών τ.μ., είναι επίσης κανονική τ.μ. Υπολογίστε τη μέση τιμή και διασπορά.

9. Δείξτε πειραματικά ότι ισχύει το παραπάνω χρησιμοποιώντας γεννήτρια αριθμών που ακολουθούν την κανονική κατανομή.

10. Αποδείξτε πειραματικά το θεώρημα του Cramer που λέει ότι εφόσον το άθροισμα δυο ανεξαρτήτων τυχαίων μεταβλητών x_1, x_2 είναι κανονική τ.μ., τότε και οι x_1, x_2 είναι κανονικές κατανομές! Ξεκινήστε από τη γεννήτρια randn(.) και χωρίστε κάθε αριθμό τυχαία σε δυο μέρη για να

δημιουργήσετε τις x_1, x_2 (είναι τα x_1, x_2 ανεξάρτητα που δημιουργήσατε ανεξάρτητα;).

11. Βρείτε το ακριβή αριθμό των όρων με τάξη πλήθους n^k στις κεντρικές ροπές $2k$ τάξης (απόδειξη κεντρικού οριακού θεωρήματος).

12. Συγκρίνεται αναλυτικά την ταχύτητα σύγκλισης άρτιων και περιττών ροπών στο κεντρικό οριακό θεώρημα. Εκτός του n ποια άλλη παράμετρος επηρεάζει την σύγκλιση;

13. Δείξτε (σε αδρές γραμμές) γιατί το κεντρικό οριακό θεώρημα ισχύει και για άθροισμα 'περίπου' (i.i.d.) τ.μ. Ορίστε πιο προσεκτικά τι σημαίνει 'περίπου' (i.i.d.) και εξηγήστε πως επηρεάζεται η ταχύτητα σύγκλισης.

14. Για διωνυμικές τ.μ. υπολογίστε την επίδραση του p στην ταχύτητα σύγκλισης των ροπών στο κεντρικό οριακό θεώρημα.

15. Υπολογίστε αριθμητικά το μέσο τετραγωνικό λάθος ανάμεσα στην κανονική Σ.Π.Π. και στη Σ.Π.Π. του αθροίσματος 100 ανεξάρτητων διωνυμικών τ.μ. (που ακολουθούν την ίδια κατανομή) ως συνάρτηση της πιθανότητας επιτυχίας p.

16. Υπολογίστε τον συντελεστή συσχέτισης της κανονικής τ.μ. x με αναμενόμενη τιμή 0, με κάθε μία από τις τυχαίες μεταβλητές x^{2k} και x^{2k+1}, όπου $k = 1, 2, 3...$

17. Συμφωνούν τα αποτελέσματά σας με την αριθμητική προσομοίωση όπου η κανονική τ.μ. δημιουργείται ως άθροισμα 20 ομογενών τ.μ., με την εντολή randn() ή με την μέθοδο Box-Muller.

18. Υπολογίστε τον συντελεστή συσχέτισης ανάμεσα σε τ.μ. x με αναμενόμενη τιμή 0 που ακολουθεί συμμετρική Σ.Π.Π. και την τ.μ. $y = x^2$.

19. Πόσο κοντά είναι το φράγμα της ανισότητας Chebyshev στη πιθανότητα για κανονική τ.μ. $x \sim \mathcal{N}(0, 1)$ για τιμές $t \in (1, 4]$.

20. Πόσο κοντά είναι το φράγμα της ανισότητας Chebyshev στη πιθανότητα για τ.μ. x με Σ.Μ.Π. $p_x(x_0) = 1/4$ όταν $x \in \{-3, -2, 2, 3\}$ (και 0 αλλού).

Κεφάλαιο 5

Στατιστική Μοντελοποίηση

Στα προηγούμενα κεφάλαια προτείναμε αριθμητικές μεθόδους υπολογισμού πιθανοτήτων, κατανομών και στατιστικών ιδιοτήτων τ.μ. Προϋπόθεση για τη χρήση τέτοιων αριθμητικών μεθόδων είναι η ύπαρξη μιας πειραματικής διαδικασίας που παράγει αριθμούς (ή ενδεχόμενα εν γένει) σύμφωνα με κάποιο μοντέλο. Για παράδειγμα, η εντολή [rand(1,1000) < 0.2] εκτελεί 1000 πειράματα με το εξής μοντέλο: τα πειράματα είναι ανεξάρτητα μεταξύ τους και κάθε πείραμα έχει δύο πιθανά ενδεχόμενα, 1 (με πιθανότητα $p = 0.2$) και 0 (με πιθανότητα $1 - p = 0.8$). Ο υπολογισμός των παραμέτρων του μοντέλου, δηλαδή της μεταβλητής p στο παράδειγμά μας[1], είναι ένα πρόβλημα εκτίμησης παραμέτρων με το οποίο ασχολείται ο κλάδος της στατιστικής (statistics).

Συχνά στην πράξη το μοντέλο είναι άγνωστο, δηλαδή το μόνο που έχουμε στη διάθεσή μας είναι ένα σύνολο αριθμών που παράγονται, π.χ., από μια φυσική διαδικασία όπως το σήμα φωνής. Ο σκοπός μας τότε είναι, όχι μόνο να υπολογίσουμε τις παραμέτρους του μοντέλου, αλλά και το μοντέλο αυτό καθ' αυτό. Με βάση τα δεδομένα μπορούμε να προχωρήσουμε στην εκτίμηση της Σ.Π.Π. (π.χ., χρησιμοποιώντας το ιστόγραμμα ως εμπειρικό εκτιμητή της κατανομής) και στη συνέχεια στην επιλογή ενός παραμετρικού μοντέλου, π.χ., άθροισμα από δύο κανονικές κατανομές. Στην συνέχεια μπορούμε να προχωρήσουμε στην εκτίμηση των παραμέτρων του μοντέλου, όπως συζητήσαμε στην προηγούμενη παράγραφο. Ο συνδυασμός μοντελοποίησης δεδομένων και στατιστικής εκτίμησης παραμέτρων ονομάζεται στατιστική μοντελοποίηση (modeling). Από τη στιγμή που έχω επιλέξει ένα μοντέλο και έχω εκτιμήσει παραμέτρους μπορώ να απαντήσω μια σειρά από ερωτήματα χρησιμοποιώντας πιθανότητες:

- Πόσο καλή είναι η επιλογή του μοντέλου μου ή αλλιώς πόσο καλά ταιριάζει

[1] Η Σ.Μ.Π., αναμενόμενη τιμή και διασπορά προκύπτουν ως συνάρτηση της p.

το μοντέλο στα δεδομένα μου; Για παράδειγμα είναι η κανονική κατανομή μια καλή επιλογή για τα δεδομένα μου (τεστ κανονικότητας);

- Πόσο καλή είναι η εκτίμηση των παραμέτρων του μοντέλου μου; Ποια είναι η αναμενόμενη τιμή του λάθους εκτίμησης;

- Είναι σωστή ή λάθος η υπόθεση σχετικά με τις τιμές των παραμέτρων ή στατιστικών ιδιοτήτων του μοντέλου; Ποιο είναι το διάστημα εμπιστοσύνης για τους εκτιμητές μου;

- Υπάρχει σχέση εξάρτησης ανάμεσα στα δείγματα που παρατηρώ; Για παράδειγμα αν οι μετρήσεις μου αφορούν το εισόδημα και το βαθμό ευτυχίας τυχαία επιλεγμένων ατόμων από μια δημοσκόπηση, υπάρχει γραμμική εξάρτηση μεταξύ των δύο;

Τέλος από τη στιγμή που έχω επιλέξει και εκπαιδεύσει μοντέλα μπορώ να τα χρησιμοποιήσω για την ταξινόμηση νέων δειγμάτων σε κατηγορίες. Για παράδειγμα, από δείγματα μετρήσεων ύψους εκτιμώ τις παραμέτρους δυο μοντέλων (κανονικές κατανομές) ένα για ενήλικες και ένα για παιδιά. Μπορώ τότε να ταξινομήσω μια καινούργια μέτρηση στην κατηγορία 'ενήλικας' ή 'παιδί', δηλαδή στο μοντέλο που είναι πιο 'κοντά' στη μέτρηση. Προβλήματα ταξινόμησης συναντάμε στην περιοχή της μηχανικής μάθησης (machine learning) και αναγνώρισης προτύπων (pattern recognition).

5.1 Εκτιμητές Μέσης Τιμής και Διασποράς

Ήδη έχουμε γνωρίσει εκτιμητές στατιστικών ιδιοτήτων τ.μ. από δείγματα. Για παράδειγμα έστω μια γεννήτρια που παράγει n ανεξάρτητους τ.α.[2] που ακολουθούν την Σ.Π.Π. $f_x()$ με μέση τιμή μ_x και διασπορά σ_x^2. Είδαμε ότι η αριθμητική μέση τιμή M_n και η αριθμητική διασπορά S_n είναι συνήθεις εκτιμητές:

$$M_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$S_n = \frac{1}{n}\sum_{i=1}^{n} (x_i - M_n)^2$$

[2] Ένας τ.α. (με άγνωστη τιμή) είναι εξ' ορισμού μια τ.μ. που ακολουθεί την Σ.Π.Π. της γεννήτριας που τον δημιούργησε. Όταν ξέρουμε την τιμή του, ο τ.α. δεν είναι πλέον 'τυχαίος', είναι απλά ένα ενδεχόμενο της παραπάνω τ.μ.

Επειδή τα δείγματα είναι ανεξάρτητοι τ.α. που ακολουθούν την ίδια Σ.Π.Π. $f_x()$ ξέρουμε ότι:

$$\mu_{x_i} = E(x_i) = \mu_x \qquad \sigma_{x_i}^2 = E((x_i - \mu_x)^2) = \sigma_x^2$$

Μπορούμε λοιπόν να υπολογίσουμε τις στατιστικές ιδιότητες των M_n και S_n που με τη σειρά τους είναι τ.μ. (ως συναρτήσεις τ.μ.). Συγκεκριμένα:

$$E(M_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\sum_{i=1}^{n}\mu_x = \frac{1}{n}n\mu_x = \mu_x$$

$$E((M_n - E(M_n))^2) = \sigma_{M_n}^2 = \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n}\sigma_{x_i}^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_x^2 = \frac{\sigma_x^2}{n}$$

όπου χρησιμοποιήσαμε την στατιστική ιδιότητα (8α) για άθροισμα ανεξάρτητων τ.μ. (βλ. Κεφ. 3). Βλέπουμε ότι η αναμενόμενη τιμή του εκτιμητή είναι η τιμή που θέλουμε να εκτιμήσουμε $E(M_n) = \mu_x$. Επίσης, η διασπορά της τ.μ. M_n (που είναι εξ ορισμού η αναμενόμενη τιμή του τετραγωνικού λάθους του εκτιμητή, δηλαδή, $E((M_n - \mu)^2)$) είναι αντιστρόφως ανάλογη του αριθμού των δειγμάτων n. Άρα, όσο περισσότερα δείγματα έχουμε, τόσο μειώνεται το λάθος εκτίμησης. Είναι επίσης σημαντικό να θυμόμαστε ότι η τ.μ. M_n ακολουθεί κανονική κατανομή για μεγάλο αριθμό δειγμάτων ($n \to \infty$).

Πόσα δείγματα χρειαζόμαστε για να έχουμε 'αρκετή' εμπιστοσύνη στον εκτιμητή μέσης τιμής; Η απάντηση εξαρτάται τόσο από το σ_x^2 όσο και από το n. Πάντως στην πράξη μπορούμε να θεωρήσουμε ότι λάθος μικρότερου του 5% της διασποράς σ_x^2 είναι αποδεκτό[3].

Αντίστοιχα για τον εκτιμητή της διασποράς τ.μ. S_n έχουμε:

$$E(S_n) = E\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - M_n)^2\right) = \frac{1}{n}\sum_{i=1}^{n} E((x_i - M_n)^2) = \frac{1}{n}\sum_{i=1}^{n} E(x_i^2 - 2x_iM_n + M_n^2)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(E(x_i^2) - \frac{2}{n}E(x_i^2) - 2E(x_i)E\left(M_n - \frac{x_i}{n}\right) + E(M_n^2)\right) =$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{n-2}{n}(\sigma_x^2 + \mu_x^2) - 2\,\frac{n-1}{n}\mu_x^2 + \frac{\sigma_x^2}{n} + \mu_x^2\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{n-1}{n}\sigma_x^2 = \frac{n-1}{n}\sigma_x^2$$

[3] Έτσι προκύπτει ο μαγικός αριθμός $n = 20$ που χρησιμοποιείται στην στατιστική μοντελοποίηση, δηλαδή, ότι χρειάζομαι τουλάχιστον 20 δείγματα για να εκπαιδεύσω μια παράμετρο του μοντέλου μου.

Βλέπουμε λοιπόν ότι η αναμενόμενη τιμή του εκτιμητή S_n δεν ισούται με την ποσότητα που θέλουμε να εκτιμήσουμε! Για τον λόγο αυτό συνήθως χρησιμοποιούμε τον εκτιμητή της διασποράς:

$$S_n^{unbiased} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - M_n)^2$$

έτσι ώστε $E(S_n^{unbiased}) = \sigma_x^2$. Για μεγάλες τιμές του n πρακτικά οι εκτιμητές S_n και $S_n^{unbiased}$ είναι ισοδύναμοι[4]. Μπορούμε να υπολογίσουμε και την διασπορά του εκτιμητή $S_n^{unbiased}$ (αναμενόμενη τιμή του τετραγώνου του λάθους):

$$E((S_n^{unbiased} - E(S_n^{unbiased}))^2) = \frac{1}{(n-1)^2}\, E\left(\left(\sum_{i=1}^{n} (x_i - M_n)^2\right)^2\right) - \sigma_x^4 = ...$$

Ποιες είναι οι ιδιότητες ενός καλού εκτιμητή; Για παράδειγμα γιατί να χρησιμοποιήσουμε τον αριθμητικό μέσο όρο για την εκτίμηση της αναμενόμενης τιμής και όχι τον παρακάτω:

$$M'_n = \frac{\max_i\{x_i\} - \min_i\{x_i\}}{2}$$

Γενικά, θέλουμε οι εκτιμητές μας να συγκλίνουν (στατιστικά) όσο το δυνατόν πιο γρήγορα στην πραγματική τιμή που θέλουμε να εκτιμήσουμε. Δηλαδή θέλουμε:

- Η αναμενόμενη τιμή του εκτιμητή να ισούται με την ποσότητα που θέλουμε να εκτιμήσουμε (ή τουλάχιστον το παραπάνω να ισχύει στο όριο όταν ο αριθμός των δειγμάτων εκτίμησης πηγαίνει στο άπειρο), π.χ.,

  $$E(M_n) = \mu_x \qquad lim_{n\to\infty} E(S_n) = \sigma_x^2$$

  Εφόσον ισχύει η παραπάνω ιδιότητα οι εκτιμητές λέγονται αμερόληπτοι (unbiased), π.χ., M_n. Όταν το παραπάνω ισχύει στο όριο $n \to \infty$ τότε οι εκτιμητές λέγονται ασυμπτωτικά αμερόληπτοι, π.χ., S_n (asymptotically unbiased).

- Η διασπορά του εκτιμητή να είναι όσο μικρότερη γίνεται και να μειώνεται όσο το δυνατόν γρηγορότερα ως συνάρτηση του αριθμού δειγμάτων n. Η διασπορά του εκτιμητή είναι ο στατιστικός μέσος όρος του τετραγώνου του λάθους, άρα (εφόσον έχουν το ίδιο συστηματικό λάθος) καλύτεροι εκτιμητές είναι αυτοί με τη μικρότερη διασπορά (more efficient). Για παράδειγμα είδαμε ότι η διασπορά του $M_n \sim 1/n$.

[4]Πράγματι για $n \to \infty$ τόσο η αναμενόμενη τιμή του S_n όσο και του $S_n^{unbiased}$ τείνουν στο σ_x^2 (asympotically unbiased).

- Τέλος, θέλουμε εκτιμητές που είναι ανθεκτικοί σε θόρυβο ή λανθασμένες μετρήσεις (robust). Παρατηρούμε, π.χ., ότι ο το λάθος του εκτιμητή M_n μπορεί να είναι μεγάλο αν μια από τις μετρήσεις μας είναι λανθασμένη (outlier).

Από όλους τους αμερόληπτους εκτιμητές, συνήθως προτιμούμε αυτόν με την μικρότερη διασπορά (efficient) γιατί ελαχιστοποιεί το μέσο τετραγωνικό λάθος[5]. Πράγματι, έστω P_n εκτιμητής της ποσότητας p (ο P_n είναι συνάρτηση των i.i.d. τ.α. x_i, $i = 1, 2, ..., n$). Εφόσον ο P_n είναι αμερόληπτος: $E(P_n) = p$, και το μέσο τετραγωνικό λάθος είναι ίσο με την διασπορά του εκτιμητή $\sigma^2_{P_n}$:

$$E((P_n - p)^2) = E((P_n - E(P_n))^2) = \sigma^2_{P_n}$$

Εφόσον ο εκτιμητής P_n δεν είναι αμερόληπτος, υπάρχει συστηματικό λάθος εκτίμησης b (bias), δηλαδή: $E(P_n) = p + b$. Το μέσο τετραγωνικό λάθος του εκτιμητή είναι:

$$E((P_n-p)^2) = E((P_n-(E(P_n)+b))^2) = E((P_n-E(P_n))^2+b^2-2b(P_n-E(P_n)))$$

$$= E((P_n - E(P_n))^2) + b^2 - 2bE(P_n - E(P_n)) = \sigma^2_{P_n} + b^2$$

δηλαδή, εν γένει, το μέσο τετραγωνικό λάθος ενός εκτιμητή είναι το άθροισμα της διασποράς του εκτιμητή και του τετραγώνου του συστηματικού λάθους του (bias). Μπορούμε να σχεδιάσουμε εκτιμητές με μικρή διασπορά, αλλά συχνά οι εκτιμητές αυτοί έχουν αυξημένο συστηματικό λάθος.

5.2 Συντελεστής Pearson

Ο συντελεστής Pearson είναι ο πιο γνωστός εκτιμητής του συντελεστή συσχέτισης δυο τ.μ. Έστω δείγματα x_i και y_i, $i = 1, 2, ..., n$ που παράγονται από γεννήτριες τ.α. ή από κάποιο πειραματικό μοντέλο ή φυσικό φαινόμενο. Ο συντελεστής Pearson (Pearson coefficient) ορίζεται ως:

$$P_n = \frac{\sum_{i=1}^{n}(x_i - M_x)(y_i - M_y)}{\sqrt{\sum_{i=1}^{n}(x_i - M_x)^2}\sqrt{\sum_{i=1}^{n}(y_i - M_y)^2}}$$

όπου $M_x = (1/n)\sum_{i=1}^{n} x_i$ και $M_y = (1/n)\sum_{i=1}^{n} y_i$ οι αριθμητικοί μέσοι όροι των x_i και y_i, αντίστοιχα (εκτιμητές της αναμενόμενης τιμής των τ.μ.). Όπως βλέπουμε ο συντελεστής Pearson ισούται με το πηλίκο του εκτιμητή της συνδιακύμανσης (x, y): $(1/n)\sum_{i=1}^{n}(x_i-M_x)(y_i-M_y)$ δια το γινόμενο των εκτιμητών[6]

[5] Για την ακρίβεια την αναμενόμενη τιμή του τετραγώνου του λάθους

[6] Παρατηρούμε ότι χρησιμοποιούμε τον ασυμπτωτικά αμερόληπτο εκτιμητή S_n εδώ.

τυπικής απόκλισης x, y: $\sqrt{(1/n)\sum_{i=1}^{n}(x_i - M_x)^2}\sqrt{(1/n)\sum_{i=1}^{n}(y_i - M_y)^2}$ (όπου τα $1/n$ στον αριθμητή και παρανομαστή απλοποιούνται). Άρα, πράγματι ο συντελεστής Pearson είναι εκτιμητής του συντελεστή συσχέτισης και οι τιμές του κυμαίνονται στο διάστημα $[-1, 1]$. Μπορούμε να υπολογίσουμε την αναμενόμενη τιμή και διασπορά του εκτιμητή Pearson (όπως για τους M_n, S_n).

Ο συντελεστής Pearson χαίρει ειδικής μνείας γιατί είναι το κύριο πειραματικό εργαλείο για να δείξουμε αν δυο ποσότητες x, y έχουν σχέση στατιστικής εξάρτησης[7]. Για παράδειγμα άμα θέλουμε να απλοποιήσουμε την τοπολογία ενός δικτύου Bayes μπορούμε να ψάξουμε για σχέσεις ανεξαρτησίας (ή ανεξαρτησίας υπό συνθήκη) μεταξύ των τ.μ. του μοντέλου μας χρησιμοποιώντας μετρήσεις. Επίσης στις κοινωνικές, οικονομικές, πολιτικές επιστήμες και στην ιατρική ο συντελεστής Pearson χρησιμοποιείται συχνά για να δείξει σχέση εξάρτησης μεταξύ δυο ποσοτήτων, π.χ., οικονομική κρίση και εγκληματικότητα. Είναι σημαντικό να θυμόμαστε ότι:

- Οι τιμές του συντελεστή Pearson υποδηλώνουν σχέση γραμμικής εξάρτησης ή ανεξαρτησίας και ΟΧΙ σχέση στατιστικής ανεξαρτησίας Δηλαδή, μπορεί ο συντελεστής Pearson να είναι κοντά στο 0 και οι ποσότητες να είναι εξαρτημένες (γιατί στατιστική ανεξαρτησία σημαίνει γραμμική ανεξαρτησία, αλλά όχι το αντίθετο!).

- Η εξάρτηση δυο ποσοτήτων δεν υποδηλώνει απαραίτητα σχέση αιτίας αιτιατού. Για παράδειγμα, εφόσον υπάρχει σχέση γραμμικής εξάρτησης μεταξύ οικονομικής κρίσης και εγκληματικότητας δεν σημαίνει απαραίτητα ότι η οικονομική κρίση είναι αιτία της εγκληματικότητας. Θυμόμαστε ότι σε αλυσίδες αιτίας αιτιατού ή γενικότερα σε δίκτυα Bayes δυο ποσότητες μπορεί να είναι εξαρτημένες αλλά να μην σχετίζονται άμεσα (στην περίπτωση αυτή μπορούμε να βρούμε μια συνθήκη κάτω από την οποία οι δυο ποσότητες είναι ανεξάρτητες). Η γενίκευση από εξάρτηση σε αιτιατότητα δεν μπορεί να αποδειχθεί εύκολα (πρέπει να μετρηθούν όλα τα πιθανά αίτια που αντιστοιχούν στις ποσότητες που μας ενδιαφέρουν). Εν γένει η στατιστική εξάρτηση είναι ένδειξη αιτιατότητας και όχι απόδειξη.

5.3 Διαστήματα Εμπιστοσύνης

Συνήθως δεν αρκεί ο υπολογισμός της τιμής ενός εκτιμητή, π.χ., του M_n, αλλά θέλουμε να υπολογίσουμε και ένα διάστημα εμπιστοσύνης (confidence interval)

[7] Προφανώς άμα έχουμε μεγάλο αριθμό δεδομένων μπορούμε να εκτιμήσουμε την από κοινού και οριακές Σ.Π.Π. και να συγκρίνουμε το γινόμενο των οριακών με την από κοινού για κάθε τιμή του ζεύγους (x, y). Πόσα δείγματα χρειαζόμαστε για να κάνουμε κάτι τέτοιο;

μέσα στο οποίο είμαστε 'σχεδόν σίγουροι' ότι βρίσκεται η πραγματική τιμή της παραμέτρου που έχουμε εκτιμήσει. Για παράδειγμα, τα δείγματά μας είναι ψήφοι (0 ή 1) σε κάλπη (exit polls) για τον υποψήφιο δήμαρχο Q και ο εκτιμητής μας είναι ο αριθμητικός μέσος όρος M_n που δίνει στον υποψήφιο Q το ποσοστό $M_n = 0.5$ (δηλαδή 50%). Έστω ότι η αριθμητικά υπολογισμένη διασπορά είναι $S_n = 0.25$ και $n = 2500$. Ορίζω ως διάστημα εμπιστοσύνης το διάστημα:

$$\left[M_n - 2\sqrt{\frac{S_n}{n}}, M_n + 2\sqrt{\frac{S_n}{n}}\right]$$

Παρατηρώ ότι το διάστημα εμπιστοσύνης αντιστοιχεί περίπου[8] στο διάστημα με όρια δύο τυπικές αποκλίσεις αριστερά ή δεξιά της μέσης τιμής του εκτιμητή, δηλαδή:

$$[\mu_{M_n} - 2\sigma_{M_n}, \mu_{M_n} + 2\sigma_{M_n}]$$

Υπενθυμίζουμε ότι $\mu_{M_n} = E(M_n) = \mu_x \approx M_n$ και $\sigma^2_{M_n} = \sigma^2_x/n \approx S_n/n$. Για το παράδειγμά μας ($\sqrt{S_n} = 0.5$) το διάστημα εμπιστοσύνης είναι το:

$$\left[0.5 - \frac{1}{\sqrt{2500}}, 0.5 + \frac{1}{\sqrt{2500}}\right] = [0.5-0.02, 0.5+0.02] = [50\%-2\%, 50\%+2\%]$$

Πόσο σίγουροι είμαστε για το ότι το αποτέλεσμα της κάλπης θα είναι το πολύ 2 ποσοστιαίες μονάδες πάνω ή κάτω από την τιμή 50%; Επειδή M_n είναι το άθροισμα $n = 2500$ τυχαίων μεταβλητών είναι ασφαλές (σύμφωνα με το Κ.Ο.Θ.) να υποθέσουμε ότι το M_n ακολουθεί κανονική κατανομή και άρα η πιθανότητα να βρισκόμαστε το πολύ δύο τυπικές αποκλίσεις μακριά από τη μέση τιμή είναι 95%. Αντίστοιχα η πιθανότητα λάθους (να βρισκόμαστε εκτός διαστήματος δηλαδή) είναι 5%. Η επιλογή των δυο τυπικών αποκλίσεων γύρω από την αναμενόμενη τιμή είναι συνήθης στην στατιστική και η πιθανότητα λάθους[9] 5% μεταφράζεται στο ότι είμαστε 'σχεδόν σίγουροι' για την ορθότητα του αποτελέσματος (δηλαδή είμαστε σωστοί 19 στις 20 φορές). Οι ερευνητές στη σωματιδιακή φυσική συνήθως θεωρούν ότι τρεις τυπικές αποκλίσεις είναι αρκετές (πιθανότητα λάθους 0.3%) για την πειραματική ανακάλυψη ενός καινούργιου σωματιδίου, π.χ., top quark (1995, Fermilab).

Συχνά καλούμαστε να απαντήσουμε για την στατιστική σημαντικότητα (statistical significance) των αποτελεσμάτων μας. Στην ουσία αυτό που είναι το ζητούμενο είναι να δείξουμε ότι η διαφορά ανάμεσα σε δυο τιμές (π.χ., αποτέλεσμα πειραματικού μοντέλου 1 και 2) είναι σημαντική, είναι δηλαδή πολύ

[8] Περίπου γιατί οι M_n και S_n είναι εκτιμητές των μ_x και σ^2_x.

[9] Προφανώς η πιθανότητα λάθους θα είναι 5% μόνο εάν τα δείγματα x_i είναι ανεξάρτητα μεταξύ τους και έχουν επιλεχθεί τυχαία και αντιπροσωπευτικά από τον πληθυσμό.

πιθανό η τιμή του μοντέλου 1 να είναι 'σχεδόν πάντα' μεγαλύτερη από την τιμή του μοντέλου 2 (εάν π.χ., επαναλάβουμε το πείραμά μας με καινούργια δεδομένα). Χοντρικά μπορούμε να πούμε κάτι τέτοιο, εάν η τιμή 2 είναι εκτός του διαστήματος εμπιστοσύνης της τιμής 1. Εν γένει υπάρχουν πιο λεπτομερή τεστ στατιστικής σημαντικότητας που λαμβάνουν υπόψη τους πολλές παραμέτρους ενός πειράματος καθώς και πιθανές λάθος μετρήσεις (outliers), π.χ., t-test, ANOVA.

5.4 Εκτίμηση Μεγιστοποίησης Πιθανότητας

Μπορούμε να σχεδιάσουμε εκτιμητές που βελτιστοποιούν κάποιο συγκεκριμένο κριτήριο. Έστω θ η ποσότητα που θέλουμε να εκτιμήσουμε, π.χ., η μέση τιμή της κανονικής κατανομής. Με βάση n ανεξάρτητες παρατηρήσεις x_1, x_2, ..., x_n μπορούμε να υπολογίσουμε τον βέλτιστο εκτιμητή $\hat{\theta}$ που μεγιστοποιεί την a posteriori πιθανότητα της ποσότητας που θέλουμε να εκτιμήσουμε δεδομένου των παρατηρήσεών μας, δηλαδή,

$$\hat{\theta} = \arg\max_{\theta} P(\theta|x_1, x_2, ..., x_n)$$

όπου $\arg\max$ συμβολίζει το όρισμα (εκτιμητή στην περίπτωσή μας) που μεγιστοποιεί την ποσότητα που ακολουθεί. Η παραπάνω σχέση μπορεί να ξαναγραφτεί ως:

$$\hat{\theta} = \arg\max_{\theta} \frac{P(x_1, x_2, ..., x_n|\theta)P(\theta)}{P(x_1, x_2, ..., x_n)} = \arg\max_{\theta} P(x_1, x_2, ..., x_n|\theta)\ P(\theta)$$

όπου χρησιμοποιήσαμε τον κανόνα του Bayes και το γεγονός ότι ο παρανομαστής $P(x_1, x_2, ..., x_n)$ δεν εξαρτάται[10] από την παράμετρο θ. Εφόσον έχουμε μια πρώτη a priori εκτίμηση της θ, π.χ., γνωρίζουμε ότι ο εκτιμητής μας ακολουθεί κανονική κατανομή με αναμενόμενη τιμή θ_0 και διασπορά σ_θ^2, μπορούμε να χρησιμοποιήσουμε αυτήν την πληροφορία στην παραπάνω εξίσωση και να μεγιστοποιήσουμε το γινόμενο $P(x_1, x_2..., x_n|\theta)\ P(\theta)$. Ο εκτιμητής που προκύπτει ονομάζεται εκτιμητής μεγιστοποίησης της a posteriori πιθανότητας (maximum a posteriori (MAP) estimator). Συχνά δεν έχουμε a priori γνώση της θ οπότε υποθέτουμε ότι η ποσότητα που θέλουμε να εκτιμήσουμε είναι μια σταθερά

[10] Άρα το όρισμα που $\hat{\theta}$ που μεγιστοποιεί τις πιθανότητες είναι το ίδιο με ή χωρίς την πιθανότητα των παρατηρήσεων $P(x_1, x_2, ..., x_n)$. Η τιμή της ποσότητας που μεγιστοποιούμε όμως είναι διαφορετική στις δύο περιπτώσεις και η παραπάνω ισότητα δεν ισχύει όταν χρησιμοποιήσουμε $\max$ αντί για $\arg\max$.

(και όχι μια τ.μ. όπως παραπάνω). Τότε $p(\theta = \hat{\theta}) = 1$ και ο εκτιμητής μας μεγιστοποιεί την απευθείας πιθανότητα:

$$\hat{\theta} = \arg\max_{\theta} P(x_1, x_2, ..., x_n|\theta) = \arg\max_{\theta} \prod_{i=1}^{n} P(x_i|\theta)$$

επειδή x_1, x_2, ..., x_n ανεξάρτητα. Ο εκτιμητής που προκύπτει ως το όρισμα που μεγιστοποιεί την απευθείας πιθανότητα ονομάζεται εκτιμητής μεγιστοποίησης πιθανότητας ή εκτιμητής μέγιστης πιθανοφάνειας (maximum likelihood (ML) estimator). Είναι ίσως ο πιο συνηθισμένος εκτιμητής που χρησιμοποιείται στην πράξη. Αυτό βέβαια δεν τον κάνει και τον 'καλύτερο' (π.χ., είναι ο εκτιμητής αυτός ανθεκτικός σε θόρυβο;).

Παράδειγμα 1: Έστω ότι η τ.μ. x ακολουθεί την κατανομή Bernoulli με πιθανότητα επιτυχίας p (ενδεχόμενο $x = 1$) και αποτυχίας $1 - p$ (ενδεχόμενο $x = 0$). Παρατηρούμε τα ακόλουθα ανεξάρτητα δείγματα: $\{1, 1, 1, 0, 0, 1, 0, 1, 1, 0\}$. Ποιος είναι ο εκτιμητής μεγιστοποίησης πιθανότητας για την ποσότητα p;

Στο παράδειγμα αυτό έχουμε $\theta \equiv p = P(x = 1)$ και $n = 10$ παρατηρήσεις $x_1 = 1$, $x_2 = 1$, ... $x_{10} = 0$. Έστω εν γένει ότι k από τις n παρατηρήσεις ήταν $x = 1$ (και άρα $n - k$ παρατηρήσεις ήταν $x = 0$). Τότε

$$P(x_1, x_2, ..., x_n|\theta) = \prod_{i=1}^{n} P(x_i|\theta) = \theta^k \; (1 - \theta)^{n-k}$$

επειδή οι παρατηρήσεις είναι ανεξάρτητες και $\theta \equiv p = P(x = 1)$, $1 - \theta \equiv 1 - p = P(x = 0)$. Για να βρούμε τον εκτιμητή που μεγιστοποιεί την παραπάνω ποσότητα θέλουμε η παράγωγος ως προς θ να ισούται με το 0:

$$\frac{d}{d\theta}\theta^k \; (1 - \theta)^{n-k} = 0 \;\Rightarrow\; k\theta^{k-1}(1 - \theta)^{n-k} - (n - k)\theta^k(1 - \theta)^{n-k-1} = 0$$

$$\Rightarrow\; \theta^{k-1}(1 - \theta)^{n-k-1}\,[k(1 - \theta) - (n - k)\theta] = 0 \;\Rightarrow\; k - n\theta = 0 \;\Rightarrow\; \hat{\theta} = \frac{k}{n}$$

Στο παράδειγμά μας ο εκτιμητής μεγιστοποίησης πιθανότητας του p έχει τιμή 6/10 δηλαδή 0.6. Εν γένει ο maximum likelihood (ML) εκτιμητής ενός ενδεχομένου A είναι:

$$\widehat{P(A)}_{ML} = \frac{\sum_{i=1}^{n} \delta(x_i = A)}{n}$$

όπου $\delta(x_i = A) = 1$ όταν $x_i = A$ (0 αλλού) και n ο συνολικός αριθμός παρατηρήσεων. Δηλαδή απλά μετράμε πόσες φορές είδαμε το ενδεχόμενο A και διαιρούμε με το πλήθος των πειραμάτων. Προφανώς στο όριο $n \to \infty$, ο εκτιμητής συγκλίνει στην πραγματική τιμή $P(A)$ (βλ. Κεφ. 1, ερμηνεία μέτρου πιθανότητας).

Παράδειγμα 2: Έστω ότι η τ.μ. x ακολουθεί την κανονική κατανομή με αναμενόμενη τιμή μ και διασποράς σ^2. Ποιος είναι ο εκτιμητής μεγιστοποίησης πιθανότητας για τις ποσότητες μ, σ^2;

Εδώ πρέπει να εκτιμήσουμε δύο ποσότητες την $\theta_1 \equiv \mu$ και την $\theta_2 \equiv \sigma^2$. Υποθέτουμε ότι έχουμε n μετρήσεις της τιμής x και θέλουμε να μεγιστοποιήσουμε

$$P(x_1, x_2, ..., x_n|\theta_1, \theta_2) = \prod_{i=1}^{n} P(x_i|\theta_1, \theta_2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta_2}} \exp\left\{-\frac{1}{2}\frac{(x_i-\theta_1)^2}{\theta_2}\right\}$$

Αντί να μεγιστοποιήσουμε την παραπάνω ποσότητα μπορούμε να μεγιστοποιήσουμε τον λογάριθμό της (επειδή ο λογάριθμος είναι μονοτονικά αύξουσα συνάρτηση). Το ζευγάρι ορισμάτων $(\hat{\theta}_1, \hat{\theta}_2)$ που μεγιστοποιεί τον λογάριθμο της πιθανότητας προκύπτει από τα σημεία μηδενισμού των μερικών παραγώγων ως προς θ_1 και θ_2, δηλαδή:

$$\left\{ \begin{array}{c} \frac{\partial}{\partial\theta_1} \sum_{i=1}^{n} \left(\ln\frac{1}{\sqrt{2\pi\theta_2}} - \frac{1}{2}\frac{(x_i-\theta_1)^2}{\theta_2}\right) = 0 \\ \frac{\partial}{\partial\theta_2} \sum_{i=1}^{n} \left(\ln\frac{1}{\sqrt{2\pi\theta_2}} - \frac{1}{2}\frac{(x_i-\theta_1)^2}{\theta_2}\right) = 0 \end{array} \right\} \Rightarrow \left\{ \begin{array}{c} \sum_{i=1}^{n} \frac{(x_i-\theta_1)}{\theta_2} = 0 \\ \sum_{i=1}^{n} \left(-\frac{1}{\theta_2} + \frac{(x_i-\theta_1)^2}{\theta_2^2}\right) = 0 \end{array} \right\}$$

$$\Rightarrow \left\{ \begin{array}{c} \sum_{i=1}^{n}(x_i - \theta_1) = 0 \\ \sum_{i=1}^{n}(\theta_2 - (x_i-\theta_1)^2) = 0 \end{array} \right\} \Rightarrow \left\{ \begin{array}{c} n\theta_1 = \sum_{i=1}^{n} x_i \\ n\theta_2 = \sum_{i=1}^{n}(x_i-\theta_1)^2 \end{array} \right\}$$

Οι ML εκτιμητές αναμενόμενης τιμής και διασποράς κανονικής κατανομής είναι

$$\widehat{\mu}_{ML} = \hat{\theta}_1 = M_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\widehat{\sigma^2}_{ML} = \hat{\theta}_2 = S_n = \frac{1}{n}\sum_{i=1}^{n}(x_i - M_n)^2$$

δηλαδή ο αριθμητικός μέσος όρος και η αριθμητική διασπορά. Παρατηρούμε ότι οι εκτιμητές ML μπορεί να μην είναι αμερόληπτοι (π.χ., S_n).

Ένας αρκετά διαφορετικός τρόπος να σχεδιάσουμε εκτιμητές είναι να εκτιμήσουμε απευθείας την υπό συνθήκη Σ.Π.Π. της τ.μ. θ δεδομένων των παρατηρήσεων, δηλαδή

$$f_{\theta|x_1,x_2,...,x_n}(\theta|x_1, x_2, ..., x_n) = \frac{f_{x_1,x_2,...,x_n|\theta}(x_1, x_2, ..., x_n|\theta) f_\theta(\theta)}{\int f_{x_1,x_2,...,x_n|\theta}(x_1, x_2, ..., x_n|\theta) f_\theta(\theta) d\theta}$$

όπου χρησιμοποιήσαμε τον κανόνα του Bayes για να εκφράσουμε την a posteriori Σ.Π.Π. ως συνάρτηση της Σ.Π.Π. της x και της a priori Σ.Π.Π. της θ. Για τον λόγο αυτό ο παραπάνω τρόπος εκτίμησης ονομάζεται Bayesian estimation. Όπως και η μέθοδος MAP προϋποθέτει την γνώση της a priori Σ.Π.Π. της θ. Συχνά επιλέγουμε την a priori Σ.Π.Π έτσι ώστε (1) να είναι εφικτός ο υπολογισμός αναλυτικά του ολοκληρώματος στον παρανομαστή, και (2) η a posteriori Σ.Π.Π. του θ να ανήκει στην ίδια οικογένεια κατανομών με την a priori Σ.Π.Π. της θ (conjugate prior).

5.5 Στατιστική Μοντελοποίηση

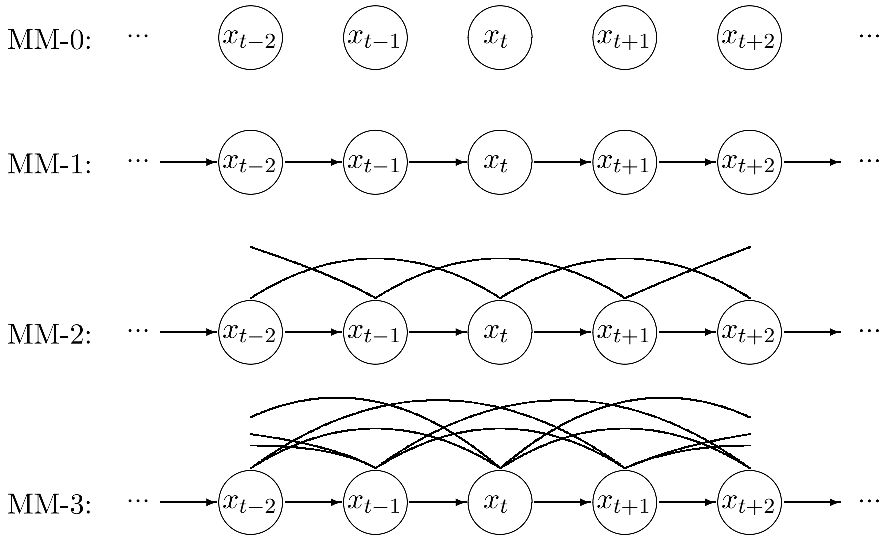
Στις προηγούμενες ενότητες είδαμε διάφορους τρόπους εκτίμησης παραμέτρων χρησιμοποιώντας πειραματικά δεδομένα. Μέχρι τώρα θεωρούμε ότι τα πειραματικά δεδομένα έχουν παραχθεί από ένα γνωστό μοντέλο, π.χ., η Σ.Π.Π. που ακολουθούν τα δείγματα στο Παράδειγμα 2 είναι η κανονική. Στην πράξη τόσο το μοντέλο όσο και η παράμετροι του μοντέλου μπορεί να είναι άγνωστες, δηλαδή, έχουμε να επιλύσουμε ένα πρόβλημα στατιστικής μοντελοποίησης. Πρέπει λοιπόν: (1) να επιλεγεί το στατιστικό μοντέλο που 'ταιριάζει' στα πειραματικά δεδομένα και (2) να υπολογιστούν οι παράμετροι του μοντέλου με βάση τα πειραματικά δεδομένα.

Τα στατιστικά μοντέλα χωρίζονται σε δύο κατηγορίες: παραμετρικά (parametric) και μη παραμετρικά (non-parametric). Τα παραμετρικά μοντέλα επιλέγουν μια Σ.Π.Π. (σε αναλυτική μορφή) και υπολογίζουν τις παραμέτρους που ορίζουν την κατανομή, π.χ., η κανονική Σ.Π.Π. ορίζεται πλήρως από δύο παραμέτρους: αναμενόμενη τιμή και διασπορά. Τα μη παραμετρικά στατιστικά μοντέλα εκτιμούν την εμπειρική Σ.Π.Π., δηλαδή, για κάθε περιοχή τιμών (bin) της τ.μ. εκτιμούν την αντίστοιχη μάζα πιθανότητας. Εν γένει τα παραμετρικά στατιστικά μοντέλα είναι πιο δημοφιλή από τα μη παραμετρικά. Η επιλογή της Σ.Π.Π. γίνεται κατόπιν στατιστικής ανάλυσης των πειραματικών δεδομένων, αν και συχνά χρησιμοποιούμε την κανονική κατανομή[11] ή τον γραμμικό συνδυασμό πολλών κανονικών κατανομών (mixture of Gaussians). Για παράδειγμα άμα θέλω να μοντελοποιήσω το ύψος των φοιτητών της τάξης μπορώ να χρησιμοποιήσω άθροισμα από δύο κανονικές κατανομές (για αγόρια και κορίτσια, αντίστοιχα). Άμα θέλω να μοντελοποιήσω το ύψος των παιδιών σε παιδικό σταθμό μπορώ να χρησιμοποιήσω άθροισμα από τέσσερις κανονικές κατανομές (για τις ηλικίες 2, 3, 4, 5), κ.λ.π.

Έως τώρα έχουμε υποθέσει ότι τα δεδομένα που εξετάζουμε είναι ανεξάρτη-

[11] Μπορώ να χρησιμοποιήσω τεστ κανονικότητας για να αποφασίσω αν τα δεδομένα μου 'ταιριάζουν' με την κανονική κατανομή.

τα μεταξύ τους δηλαδή η χρονική αλληλουχία δεν επηρεάζει την μοντελοποίηση και εκτίμηση παραμέτρων. Συχνά στην πράξη υπάρχει εξάρτηση μεταξύ παρατηρήσεων. Η εξάρτηση αυτή εκτείνεται προφανώς σε όλη την χρονική αλυσίδα, αλλά μπορώ να υποθέσω σχέσεις ανεξαρτησίας υπό συνθήκη για να απλοποιήσω το μοντέλο μου (χρησιμοποιώντας την ιδιότητα Markov). Για παράδειγμα σχεδιάζουμε τις σχέσεις άμεσης εξάρτησης[12] (δίκτυο Bayes) σε μια χρονική αλληλουχία μετρήσεων $\{..., x_{t-2}, x_{t-1}, x_t, x_{t+1}, x_{t+2}, ...\}$ υποθέτοντας ότι οι μετρήσεις ακολουθούν μοντέλο Markov) μηδενικής (MM-0), πρώτης (MM-1), δεύτερης (MM-2) και τρίτης (MM-3) τάξης:



Το μοντέλο μηδενικής (MM-0) τάξης υποθέτει πλήρη ανεξαρτησία μεταξύ μετρήσεων (δηλαδή είναι το μοντέλο που χρησιμοποιούσαμε μέχρι τώρα). Το μοντέλο πρώτης τάξης (MM-1) υποθέτει σχέση άμεσης εξάρτησης μεταξύ διαδοχικών μετρήσεων και σχέση ανεξαρτησίας υπό συνθήκη μεταξύ μετρήσεων που ΔΕΝ είναι διαδοχικές (εφόσον μια από τις μετρήσεις που μεσολαβούν είναι γνωστές), π.χ., x_{t+1}, x_{t-1} ανεξάρτητες υπό συνθήκη x_t. Χρησιμοποιώντας τον κανόνα της αλυσίδας, η από κοινού πιθανότητα όλων των μετρήσεων του μοντέλου (MM-1) ισούται με:

$$P(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} P(x_i|x_{i-1}, x_{i-2}, ..., x_1) = \prod_{i=1}^{n} P(x_i|x_{i-1})$$

[12] Οι θεμελιώδεις σχέσεις ανεξαρτησίας (υπό συνθήκη) για δίκτυα Bayes αναφέρονται στο Κεφ. 1.

γιατί δεδομένης της x_{i-1}, η x_i είναι ανεξάρτητη με κάθε μια από τις $x_{i-2}, ..., x_i$. Άρα οι παράμετροι του μοντέλου[13] (MM-1) είναι οι υπό συνθήκη πιθανότητες $P(x_i|x_{i-1})$. Υποθέτοντας ότι οι μετρήσεις x_i παίρνουν διακριτές τιμές (k το πλήθος) από το σύνολο $\{s_1, s_2, ..., s_k\}$, τότε το σύνολο των πιθανοτήτων που πρέπει να υπολογίσω είναι k^2 τον αριθμό[14]. Η εκτίμηση των παραμέτρων $a_{ij} \equiv P(x_{t+1} = s_j|x_t = s_i)$ μπορεί να γίνει με μεγιστοποίηση της πιθανότητας ως:

$$(\hat{a}_{ij})_{ML} = \frac{\text{count}(s_i, s_j)}{\text{count}(s_i)}$$

όπου η $\text{count}(s_i, s_j)$ υπολογίζει πόσες φορές έχω δει την μέτρηση s_i να ακολουθείται από την μέτρηση s_j. Οι τιμές s_i συχνά ονομάζονται καταστάσεις (states) και οι πιθανότητες a_{ij} πιθανότητες μετάβασης μεταξύ καταστάσεων (state transition probabilities). Επίσης όταν λέμε ότι οι $x_1, x_2, ..., x_n$ ακολουθούν μοντέλο Markov υπονοούμε το μοντέλο MM-1.

Το μοντέλο Markov δεύτερης τάξης (MM-2) υποθέτει σχέση άμεσης εξάρτησης μεταξύ τριών διαδοχικών μετρήσεων, άρα υπάρχει σχέση ανεξαρτησίας υπό συνθήκη μεταξύ μετρήσεων που τις χωρίζουν τουλάχιστον δύο μετρήσεις (δεδομένου ότι οι μετρήσεις που μεσολαβούν είναι γνωστές), π.χ., x_{t+2}, x_{t-1} ανεξάρτητες υπό συνθήκη x_t, x_{t+1}. Χρησιμοποιώντας τον κανόνα της αλυσίδας, η από κοινού πιθανότητα όλων των μετρήσεων του μοντέλου (MM-2) ισούται με:

$$P(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} P(x_i|x_{i-1}, x_{i-2}, ..., x_1) = \prod_{i=1}^{n} P(x_i|x_{i-1}, x_{i-2})$$

γιατί δεδομένων των x_{i-1}, x_{i-2}, η x_i είναι ανεξάρτητη με κάθε μια από τις x_{i-3}, ..., x_i. Άρα ο αριθμός των παραμέτρων του μοντέλου (MM-2) είναι είναι k^3 τον αριθμό[15] εφόσον οι μετρήσεις x_i παίρνουν διακριτές τιμές (k το πλήθος). Μπορούμε να υλοποιήσουμε ένα μοντέλο Markov δεύτερης τάξης χρησιμοποιώντας το μοντέλο MM-1 ορισμένο πάνω στο καρτεσιανό γινόμενο των καταστάσεων $\{s_1, s_2, ..., s_k\} \times \{s_1, s_2, ..., s_k\}$ (γιατί;). Τέλος για το μοντέλο MM-3 έχουμε:

$$P(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} P(x_i|x_{i-1}, x_{i-2}, ..., x_1) = \prod_{i=1}^{n} P(x_i|x_{i-1}, x_{i-2}, x_{i-3})$$

[13]Εφόσον ξέρουμε την από κοινού πιθανότητα $P(x_1, x_2, ..., x_n)$ μπορούμε να υπολογίσουμε οποιαδήποτε οριακή ή υπό συνθήκη πιθανότητα υποσυνόλου των $x_1, x_2, ..., x_n$. Άρα ξέρω τα 'πάντα' για το δίκτυο Markov.

[14]Στην πράξη πρέπει να γνωρίζουμε και τις αρχικές πιθανότητες $\pi_i \equiv P(x_1 = s_i)$, οπότε ο αριθμός παραμέτρων είναι $k^2 + k$.

[15]Πιο σωστά: $k^3 + k^2 + k$.

Τα μοντέλα Markov χρησιμοποιούνται λόγω της απλότητάς τους σε πολλές εφαρμογές τηλεπικοινωνιών και αναγνώρισης προτύπων.

Εν γένει οι μετρήσεις μου θα μπορούσαν να έχουν πιο πολύπλοκες σχέσεις άμεσης εξάρτησης. Σε κάθε περίπτωση μπορώ: (1) να σχεδιάσω το δίκτυο Bayes, (2) να βρω τις σχέσεις ανεξαρτησίας υπό συνθήκη, (3) να απλοποιήσω την από κοινού πιθανότητα του δικτύου χρησιμοποιώντας τις σχέσεις ανεξαρτησίας, (4) να εκτιμήσω τις τιμές των παραμέτρων της απλοποιημένης από κοινού πιθανότητας του δικτύου. Έχοντας την παραπάνω πληροφορία μπορώ πλέον να υπολογίσω της πιθανότητα (απευθείας ή υπό συνθήκη) μελών του δικτύου μου.

<u>Παράδειγμα 1:</u> Εκτιμήστε το MM-0 και MM-1 για την παρακάτω γραμματοσειρά $\{A, B, B, A, A, B, B, A, B, A, A, B, A, B, B, A, A, B, A, B\}$. Πιο μοντέλο ταιριάζει καλύτερα στα δεδομένα σας;

Το μοντέλο MM-0 υποθέτει ανεξαρτησία ανάμεσα στις παρατηρήσεις και άρα έχει μόνο μια παράμετρο την $P(x_t = A) \equiv P(A)$. Προφανώς $P(x_t = B) \equiv P(B) = 1 - P(A)$. Χρησιμοποιούμε εκτιμητές μεγιστοποίησης πιθανότητας:

$$\widehat{P(A)}_{ML} = \frac{\sum_{i=1}^{n} \delta(x_t = A)}{n}$$

Δεδομένου ότι έχουμε $n = 20$ μετρήσεις εκ των οποίων οι 10 είναι A έχουμε:

$$\widehat{P(A)}_{ML} = \widehat{P(B)}_{ML} = 0.5$$

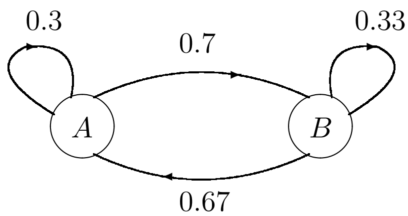
Το μοντέλο MM-1 υποθέτει άμεση εξάρτηση μόνο μεταξύ γειτονικών παρατηρήσεων και άρα έχει δύο παραμέτρους την $P(x_t = A|x_t = A) \equiv P(A|A)$ και την $P(x_t = A|x_t = B) \equiv P(A|B)$. Προφανώς ισχύει ότι $P(x_t = B|x_t = A) \equiv P(B|A) = 1 - P(A|A)$ και $P(x_t = B|x_t = B) \equiv P(B|B) = 1 - P(A|B)$. Χρησιμοποιούμε εκτιμητές μεγιστοποίησης πιθανότητας:

$$\widehat{P}(A|A)_{ML} = \frac{\text{count}(A, A)}{\text{count}(A)} = \frac{3}{10} = 0.3 \qquad \widehat{P}(B|A)_{ML} = 0.7$$

$$\widehat{P}(A|B)_{ML} = \frac{\text{count}(B, A)}{\text{count}(B)} = \frac{6}{9} \approx 0.67 \qquad \widehat{P}(B|B)_{ML} \approx 0.33$$

Σημείωση: Επειδή δεν ξέρουμε τι ακολουθεί το τελευταίο B δεν χρησιμοποιούμε αυτή την παρατήρηση στον εκτίμηση του $P(A|B)$. Συχνά το μοντέλο MM-1 αναπαρίσταται με το διάγραμμα καταστάσεων ως εξής:

MM-1:



όπου οι καταστάσεις αντιστοιχούν σε ενδεχόμενα και οι πιθανότητες μετάβασης αναγράφονται στις μεταβάσεις του γράφου. Συγκρίνοντας το μοντέλο MM-0 και MM-1 βλέπουμε ότι οι πιθανότητες μετάβασης είναι αρκετά κοντά, αν και στο MM-1 είναι πιο πιθανό να αλλάξουμε κατάσταση παρά να μείνουμε στην ίδια. Πάντως, δεν έχουμε αρκετά δεδομένα εκπαίδευσης για να μπορούμε να πούμε ότι οι διαφορές ανάμεσα στο μοντέλο MM-0 και MM-1 είναι στατιστικά σημαντικές. Το πιο πιθανό μοντέλο που εξηγεί τα δεδομένα μου είναι λοιπόν το πιο απλό, δηλαδή το MM-0. Δεδομένου ότι $P(A) \approx P(B)$ μπορώ να υποθέσω ότι A, B είναι οι δύο όψεις ενός (δίκαιου) νομίσματος.

Παράδειγμα 2: Ο καιρός το καλοκαίρι στην Κρήτη μοντελοποιείται με ένα μοντέλο Markov MM-1 με δύο καταστάσεις Α=αίθριος και Σ=συννεφιά (η πιθανότητα βροχής είναι αμελητέα και μπορεί να αγνοηθεί). Η πιθανότητα να είναι ο καιρός αίθριος και την επόμενη μέρα δεδομένου ότι είναι αίθριος σήμερα είναι $P(A|A) = 0.9$. Αντίστοιχα $P(\Sigma|\Sigma) = 0.5$. Υπολογίστε την πιθανότητα να έχουμε δέκα μέρες αίθριου καιρού στη σειρά δεδομένου ότι ο καιρός σήμερα είναι αίθριος. Ποιος είναι ο μέσος αριθμός αίθριων ημερών τον μήνα Αύγουστο στην Κρήτη σύμφωνα με το παραπάνω μοντέλο MM-1;

Έχουμε για την πιθανότητα να έχουμε δέκα μέρες αίθριου καιρού στη σειρά:

$$P(\underbrace{A, A, A, ..., A}_{10 \text{ φορές}}|A) = \prod_{i=1}^{10} P(A|A) = 0.9^{10} \approx 0.35$$

Για να υπολογίσω την αναμενόμενη τιμή της τ.μ. k που αντιστοιχεί στον αριθμό των αίθριων ημερών (από n συνολικά ημέρες) δημιουργώ την τ.μ. x_t με τιμή $x_t = 1$ αν ο καιρός είναι αίθριος την μέρα t και $x_t = 0$ αλλιώς. Ισχύει η ακόλουθη αναδρομική σχέση για την πιθανότητα αίθριου καιρού την $t+1$ μέρα ως συνάρτηση του καιρού την t μέρα (χρησιμοποιώντας το δέντρο πιθανότητας):

$$P(x_{t+1} = A) = P(x_t = A)P(A|A) + P(x_t = \Sigma)P(A|\Sigma)$$

Στο όριο $t \to \infty$ έχουμε $P(x_{t+1} = A) = P(x_t = A) = 1 - P(x_t = \Sigma)$, άρα

$$t \to \infty : \; P(A) = 0.9\; P(A) + 0.5\; (1 - P(A)) \Rightarrow 0.6P(A) = 0.5 \Rightarrow P(A) \approx 0.83$$

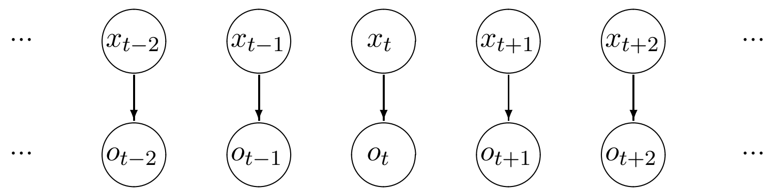
Άρα η αναμενόμενη τιμή της Bernoulli τ.μ. x είναι $E(x) \approx 1\times 0.83+0\times 0.17 = 0.83$. Η τ.μ. k είναι το άθροισμα $n = 31$ τ.μ. x_t και άρα, ο αναμενόμενος αριθμός από μέρες με αίθριο καιρό στην Κρήτη τον Αύγουστο είναι σύμφωνα με το μοντέλο μας[16]:

$$E(k) = E(x_1 + x_2 + ... + x_{31}) = 31 \times E(x_t) \approx 31 \times 0.83 = 25.7$$

5.6 Αναγνώριση Προτύπων

Συχνά δεν μπορώ να παρατηρήσω απευθείας τις τιμές x_i (δηλαδή τις καταστάσεις). Σ' αυτήν την περίπτωση λέμε ότι οι καταστάσεις είναι κρυφές (hidden). Αντ' αυτού παρατηρώ τις τιμές της τ.μ. o_i που έχει άμεση σχέση εξάρτησης με την κατάσταση x_i. Οι τ.μ. o_i μπορεί να αντιστοιχούν για παράδειγμα στο ύψος ενός παιδιού και οι τ.μ. x_i στην ηλικία του παιδιού. Υπάρχει (στατιστική) σχέση εξάρτησης μεταξύ του ύψους και της ηλικίας. Ο σκοπός μου είναι να μετρήσω το ύψος του παιδιού για να 'μαντέψω' την ηλικία. Θέλω δηλαδή να υπολογίσω την πιο πιθανή τιμή του ύψους x_i δεδομένου ότι γνωρίζω την ηλικία o_i. Το παραπάνω δίκτυο Bayes μπορεί να αναπαρασταθεί ως:

HMM-0:



όπου σε κάθε χρονική στιγμή t κάνω μια μέτρηση o_t και υπολογίζω την πιο πιθανή τιμή για το $x_t \in \{s_1, s_2, ..., s_k\}$. Το παραπάνω μοντέλο υποθέτει ανεξαρτησία μεταξύ καταστάσεων και μεταξύ μετρήσεων στο χρόνο. Το κριτήριο που χρησιμοποιώ είναι το κριτήριο μεγιστοποίησης της a posteriori πιθανότητας:

$$\hat{s} = \arg\max_{s_i} P(x_t = s_i|o_t) = \arg\max_{s_i} P(o_t|x_t = s_i)P(x_t = s_i)$$

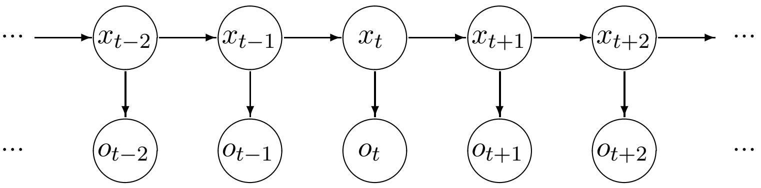
όπου έχω χρησιμοποιήσει πάλι το γεγονός ότι η πιθανότητα παρατήρησης $P(o_i)$ (στον παρανομαστή) δεν αλλάζει το αποτέλεσμα. Στο παράδειγμά μου, από όλες τις δυνατές ηλικίες $s_1, s_2, ..., s_k$ επιλέγω αυτή που μεγιστοποιεί την a posteriori

[16] Δεδομένου ότι μας ενδιαφέρει ο καιρός τον Αύγουστο μπορούμε να θεωρήσουμε ότι ανεξάρτητα από την αρχικές πιθανότητες π_A, π_Σ μετά από 61 ημέρες το μοντέλο Markov έχει συγκλίνει στην σταθερή κατάσταση.

πιθανότητα της ηλικίας δεδομένου του ύψους (παρατήρησης). Στην πράξη αυτό που κάνω είναι ότι ταξινομώ παιδιά σε ηλικιακές ομάδες με βάση το χαρακτηριστικό του ύψους. Το παραπάνω πρόβλημα ονομάζεται πρόβλημα ταξινόμησης (classification), και ο ταξινομητής που μεγιστοποιεί την a posteriori πιθανότητα ονομάζεται ταξινομητής Bayes (Bayes classifier). Μπορούμε να δείξουμε ότι ο ταξινομητής Bayes ελαχιστοποιεί την πιθανότητα λάθους ταξινόμησης εφόσον γνωρίζουμε την πραγματική a posteriori κατανομή.

Η γενίκευση του παραπάνω μοντέλου, όταν υπάρχει εξάρτηση μεταξύ των καταστάσεων, ονομάζεται κρυφό[17] μοντέλο Markov (hidden Markov model) και απεικονίζεται από το δίκτυο που ακολουθεί:

HMM-1:



Στην πράξη το παραπάνω μοντέλο αποτελείται από μια αλυσίδα Markov πρώτης τάξης (καταστάσεις x_t) και από παρατηρήσεις o_t που εξαρτώνται άμεσα μόνο από τις καταστάσεις που τις 'δημιουργούν'. Με βάση τις σχέσεις ανεξαρτησίας υπό συνθήκη μπορούμε να δείξουμε ότι η από κοινού πιθανότητα του δικτύου είναι:

$$P(x_1, x_2, ..., x_n, o_1, o_2, ..., o_n) = \prod_{i=1}^{n} P(x_i|x_{i-1})P(o_i|x_i)$$

και άρα το μοντέλο μου αποτελείται από τις πιθανότητες μετάβασης $a_{ij} \equiv P(x_{t+1} = s_j|x_t = s_i)$ (βλ. μοντέλο MM-1) και από τις πιθανότητες παρατήρησης $b_j(o_t) = P(o_t|x_t = s_j)$. Τόσο στο μοντέλο HMM-0 όσο και στο μοντέλο HMM-1 συνήθως προσπαθούμε να 'μαντέψουμε' την πιο πιθανή αλληλουχία καταστάσεων $X = (x_1, x_2, ..., x_n)$ δεδομένων των αντίστοιχων παρατηρήσεων $O = (o_1, o_2, ..., o_n)$, δηλαδή:

$$\hat{S} = \arg\max_S P(X = S|O) = \arg\max_S P(X = S, O)$$

όπου S είναι όλες οι πιθανές $n-$άδες (αλληλουχίες) καταστάσεων. Το παραπάνω πρόβλημα είναι ένα δυναμικό πρόβλημα ταξινόμησης[18] και ονομάζεται συνήθως πρόβλημα αποκωδικοποίησης (decoding) ή αναγνώρισης (recognition).

[17] Ο όρος κρυφό υπονοεί ότι οι καταστάσεις είναι συνήθως μη παρατηρήσιμες.

[18] Γιατί η πιο πιθανή κατάσταση σε μια χρονική στιγμή εξαρτάται από τις υπόλοιπες καταστάσεις, άρα ψάχνω την 'καλύτερη' αλληλουχία καταστάσεων.

Παράδειγμα 1: Έχω τα ακόλουθα αποτελέσματα από την ρίψη δύο νομισμάτων: $\{(1,H),(2,H),(2,T),(1,H),(2,H),(2,T),(1,H),(1,T),(1,T),(2,H),(1,H),$ $(2,H),(2,T),(1,H),(1,H),(2,T)\}$, όπου 1 ή 2 είναι το νόμισμα που χρησιμοποίησα και H ή T είναι το αποτέλεσμα της ρίψης. Οι ρίψεις είναι ανεξάρτητες μεταξύ τους. Υπολογίστε τις παραμέτρους του μοντέλου HMM-0. Δεδομένου ότι παρατηρώ H ποιο νόμισμα χρησιμοποίησα;

Έχουμε δυο καταστάσεις $x_t = 1$ ή $x_t = 2$ (νόμισμα 1 ή 2) και δύο τιμές παρατηρήσεων $o_t = H$ ή $o_t = T$. Συνολικά έχουμε $n = 16$ μετρήσεις. Εκτιμούμε πρώτα την a priori πιθανότητα χρησιμοποιώντας τα δείγματα εκπαίδευσης (παρατηρήσεις):

$$\widehat{p}(x_t = 1)_{ML} = \frac{\sum_{i=1}^{n} \delta(x_t = 1)}{n} = \frac{8}{16} = 0.5 \qquad \widehat{p}(x_t = 2)_{ML} = 0.5$$

Στη συνέχεια εκτιμούμε τις πιθανότητες παρατήρησης:

$$\widehat{p}(o_t = H|x_t = 1)_{ML} = \frac{\sum_{i=1}^{n} \delta(x_t = 1, o_t = H)}{\sum_{i=1}^{n} \delta(x_t = 1)} = \frac{6}{8} = 0.75$$

Αντίστοιχα εκτιμούμε ότι $p(o_t = T|x_t = 1) = 0.25$ και οι εκτιμητές της πιθανότητας $p(o_t = H|x_t = 2) = p(o_t = T|x_t = 2) = 0.5$. Αυτές είναι οι παράμετροι του μοντέλου HMM-0. Καλούμαστε να ταξινομήσουμε ένα καινούργιο δείγμα $(?, H)$ στην κατηγορία 1 ή 2 χρησιμοποιώντας τον ταξινομητή Bayes, δηλαδή:

$$\hat{s} = \arg \max_{s_i \in \{1,2\}} P(o_t = H|x_t = s_i)P(x_t = s_i) = \arg\max\{0.75 \times 0.5, 0.25 \times 0.5\}$$

Άρα ταξινομούμε το δείγμα $(?, H)$ στην κατηγορία 1, την πιο πιθανή.

Παράδειγμα 2: Για το παραπάνω παράδειγμα εκτιμήστε το μοντέλο HMM-1 και βρείτε την πιο πιθανή αλληλουχία νομισμάτων που αντιστοιχεί στις ρίψεις H, T, H.

Το μοντέλο HMM-1 έχει τις παραμέτρους του μοντέλου HMM-0 και επιπρόσθετες παραμέτρους τις πιθανότητες μεταβάσεις μεταξύ καταστάσεων $a_{ij} = p(x_t = s_j|x_{t-1} = s_i)$ όπου $s_i, s_j \in \{1, 2\}$. Εκτιμούμε τις πίνακα πιθανοτήτων μετάβασης $\mathbf{A}$ όπως στην προηγούμενη ενότητα χρησιμοποιώντας μεγιστοποίηση πιθανότητας[19], οπότε οι παράμετροι του μοντέλου μας είναι:

$$\mathbf{A} = \begin{bmatrix} 3/8 & 5/8 \\ 4/7 & 3/7 \end{bmatrix} \quad \mathbf{b}(H) = \begin{bmatrix} 0.75 \\ 0.5 \end{bmatrix} \quad \mathbf{b}(T) = \begin{bmatrix} 0.25 \\ 0.5 \end{bmatrix}$$

όπου $\mathbf{b}$ οι πιθανότητες παρατήρησης από το παράδειγμα 1. Η πιο πιθανή αλληλουχία νομισμάτων $S = (s_1, s_2, s_3)$ που αντιστοιχεί στις ρίψεις $o_1 = H, o_2 = T, o_3 = H$ υπολογίζεται ως:

$$\hat{S} = \arg\max_S P(X = S|O) = \arg\max_S P(X = S, O) =$$

$$= \arg\max_{s_1,s_2,s_3} \prod_{i=1}^{3} P(x_i = s_i|x_{i-1} = s_{i-1})P(o_i|x_i = s_i) =$$

$$= \arg\max_{s_1,s_2,s_3} \pi_{s_1} a_{s_1 s_2} a_{s_2 s_3} b_{s_1}(H) b_{s_2}(T) b_{s_3}(H)$$

όπου π_{s_i} η a priori πιθανότητα για την κατάσταση s_i, $b_{s_i}(o_i)$ η πιθανότητα παρατήρησης o_i στην κατάσταση s_i και a_{s_i,s_j} η πιθανότητα μετάβασης από την κατάσταση s_i στην κατάσταση s_j. Παρατηρούμε ότι μπορώ να 'σπάσω' την μεγιστοποίηση σε τοπικές μεγιστοποιήσεις ως προς s_1, s_2, s_3 (γιατί δεν είναι όλοι οι όροι συνάρτηση και των τριών καταστάσεων) ως εξής:

$$\hat{S} = \widehat{(s_1, s_2, s_3)} = \arg\max_{s_3}\{b_{s_3}(H) \ \max_{s_2}\{a_{s_2 s_3} b_{s_2}(T) \ \max_{s_1}\{\pi_{s_1} a_{s_1 s_2} b_{s_1}(H)\}\}\}$$

Η παρατήρηση αυτή (γνωστός και ως αλγόριθμος Viterbi) απλοποιεί σημαντικά το πρόβλημα αποκωδικοποίησης γιατί τώρα χρειάζεται να εξετάσω μόνο 3×2 δυνατούς συνδυασμούς καταστάσεων (μονοπάτια) αντί για 3^2 προηγουμένως. Συνήθως αναπαριστούμε τον αλγόριθμο Viterbi πάνω σε ένα δίκτυο καταστάσεων-παρατηρήσεων, π.χ.,

	$o_1 = H$		$o_2 = T$		$o_3 = H$
$s = 1$	$\pi_1 b_1(H)$	$\times a_{s_1 s_2}$	$b_1(T)$	$\times a_{s_2 s_3}$	$b_1(H)$
$s = 2$	$\pi_2 b_2(H)$		$b_2(T)$		$b_2(H)$

Επειδή η ποσότητα $\max_{s_1}\{\pi_{s_1} a_{s_1 s_2} b_{s_1}(H)\}$ εμπεριέχει την κατάσταση s_2 πρέπει να μεγιστοποιήσω την ποσότητα αυτή για $s_2 = 1$ **και** για $s_2 = 2$. Στην πράξη

[19] Συχνά οι καταστάσεις δεν είναι παρατηρήσιμες (ούτε) για τα δείγματα εκπαίδευσης, δηλαδή έχουμε ελλιπή δεδομένα: $\{(?, H), (?, H), (?, T), ...\}$. Σ' αυτήν την περίπτωση μπορούμε να χρησιμοποιήσουμε μια γενίκευση του αλγορίθμου μεγιστοποίησης της πιθανότητας σε ελλιπή δεδομένα όπως, π.χ., τον αλγόριθμο αναμενόμενης τιμής - μεγιστοποίησης της πιθανότητας (expectation-maximization).

μεγιστοποιώ ανάμεσα σε όλα τα μονοπάτια που καταλήγουν στην $s_2 = 1$ (και άρα κρατάω το καλύτερο) **και** ανάμεσα σε όλα τα μονοπάτια που καταλήγουν στην $s_2 = 2$. Δηλαδή:

$$s_2 = 1: \ \max_{s_1}\{\pi_{s_1} a_{s_1 s_2} b_{s_1}(H)\} = \max_{s_1}\{\pi_1 a_{11} b_1(H), \pi_2 a_{21} b_2(H)\} = M_1$$

$$s_2 = 2: \ \max_{s_1}\{\pi_{s_1} a_{s_1 s_2} b_{s_1}(H)\} = \max_{s_1}\{\pi_1 a_{12} b_1(H), \pi_2 a_{22} b_2(H)\} = M_2$$

Και στις δυο περιπτώσεις η πρώτη ποσότητα που αντιστοιχεί στην $s_1 = 1$ είναι μεγαλύτερη, δηλαδή, $M_1 = 0.5 \times (3/8) \times 0.75 \approx 0.14$ και $M_2 \approx 0.23$ που αντιστοιχούν στα μονοπάτια (1,1) και (1,2). Συνεχίζοντας με την ποσότητα $\max_{s_2}\{a_{s_2 s_3} b_{s_2}(T) M_{s_2}\}$, έχουμε:

$$s_3 = 1: \ \max_{s_2}\{a_{s_2 s_3} b_{s_2}(T) M_{s_2}\} = \max_{s_2}\{a_{11} b_1(T) M_1, a_{21} b_2(T) M_2\} = W_1$$

$$s_3 = 2: \ \max_{s_2}\{a_{s_2 s_3} b_{s_2}(T) M_{s_2}\} = \max_{s_2}\{a_{12} b_1(T) M_1, a_{22} b_2(T) M_2\} = W_2$$

Και στις δυο περιπτώσεις η πρώτη ποσότητα που αντιστοιχεί στην $s_2 = 2$ είναι μεγαλύτερη, δηλαδή οι $W_1 = (4/7) \times 0.5 \times 0.23 \approx 0.07$ και $W_2 \approx 0.05$ αντιστοιχούν στα μονοπάτια (1,2,1) και (1,2,2). Τέλος, βρίσκουμε ότι η ποσότητα $\max_{s_3}\{b_{s_3}(H) W_{s_3}\}$ μεγιστοποιείται για $s_3 = 1$ (με τιμή 0.05) και άρα το πιο πιθανό μονοπάτι είναι το $(1, 2, 1)$ ή αλλιώς $s_1 = 1$, $s_2 = 2$, $s_3 = 1$. Παρατηρούμε ότι το αποτέλεσμα είναι το ίδιο είτε χρησιμοποιήσουμε το μοντέλο HMM-0 είτε το HMM-1 (γιατί;).

5.6.1 Detection

Κλείνουμε αυτή τη σύντομη εισαγωγή στα προβλήματα αναγνώρισης προτύπων με μια υποπερίπτωση του γενικού προβλήματος ταξινόμησης. Τα προβλήματα ταξινόμησης (ή αναγνώρισης) με δύο μόνο καταστάσεις (ή κατηγορίες) ονομάζονται προβλήματα ανίχνευσης (detection). Λόγω της πρακτικής σημασίας αυτών των προβλημάτων αξίζουν ειδική μνεία. Παραδείγματα προβλημάτων ανίχνευσης είναι η ανάγνωση δυαδικών ψηφίων (bits) από CD, ιατρικές εξετάσεις για την ανίχνευση ασθενειών, συστήματα συναγερμού (ανίχνευσης διάρρηξης), ψηφιακά συστήματα τηλεπικοινωνιών κ.λ.π. Συνήθως η ανίχνευση του γεγονότος που ψάχνουμε (τιμή ψηφίου 1, διάρρηξη, ύπαρξη ασθένειας) έχει την τιμή $x = 1$ (γνωστή και ως υπόθεση H_1), ενώ η μη ανίχνευση έχει την τιμή $x = 0$ (συχνά ονομάζεται και μηδενική υπόθεση H_0).

Η διαδικασία επιλογής μοντέλου, εκτίμησης παραμέτρων μοντέλου και ταξινόμησης (ανίχνευσης) παραμένει η ίδια. Μπορώ πάντως με βάση την απόφαση του ταξινομητή μου $\hat{x}$ να ξεχωρίσω 4 περιπτώσεις:

$x = 1$	$\hat{x} = 1$	σωστή ανίχνευση (hit)
$x = 1$	$\hat{x} = 0$	λάθος απόρριψη (false reject, miss)
$x = 0$	$\hat{x} = 1$	λάθος ανίχνευση (false alarm)
$x = 0$	$\hat{x} = 0$	σωστή απόρριψη (correct reject)

Τα δύο είδη λάθους miss, false alarm συχνά έχουν διαφορετικό οικονομικό κόστος, π.χ., είναι προτιμότερη η λανθασμένη διάγνωση καρκίνου από την μη έγκαιρη διάγνωση της ασθένειας. Αντί λοιπόν να χρησιμοποιήσουμε τον ταξινομητή Bayes που δίνει ίση πιθανότητα στα δύο είδη λαθών, μπορούμε να επιλέξουμε ένα άλλο σημείο απόφασης (operation point) ώστε να ελαχιστοποιήσουμε το οικονομικό κόστος. Συχνά το αποτέλεσμα ενός προβλήματος detection αναπαρίσταται για ένα εύρος τιμών των σημείων απόφασης σε ένα διάγραμμα όπου στον άξονα των x βάζουμε το ποσοστό των λαθών ανίχνευσης και στον άξονα των y το ποσοστό των σωστών ανιχνεύσεων. Το διάγραμμα αυτό (false alarms vs. hits) ονομάζεται receiver operator curve (ROC).

5.7 Σύνοψη

- Ο αριθμητικός μέσος όρος M_n και η αριθμητική διασπορά $S_n^{unbiased}$ χρησιμοποιούνται συχνά ως εκτιμητές της αναμενόμενης τιμής και διασποράς τ.μ., γιατί είναι αμερόληπτοι και το μέσο τετραγωνικό λάθος μειώνεται ως $1/n$, όπου n το πλήθος των μετρήσεων (efficient).

- Μπορούμε να ορίσουμε εκτιμητές που ικανοποιούν συγκεκριμένα κριτήρια όπως μεγιστοποίηση της πιθανότητας (maximum likelihood).

- Συχνά χρησιμοποιούμε μοντέλα Markov για να μοντελοποιήσουμε μετρήσεις εξαρτημένες στο χρόνο. Η ιδιότητα της ανεξαρτησίας υπό συνθήκη των αλυσίδων Markov μειώνει σημαντικά τον αριθμό των παραμέτρων του μοντέλου.

- Η γενίκευση των αλυσίδων Markov σε προβλήματα ταξινόμησης (όπου χρησιμοποιούμε παρατηρήσεις για να εκτιμήσουμε τις καταστάσεις) ονομάζονται κρυφά μοντέλα Markov.

Ασκήσεις

1. Υπολογίστε αριθμητικά το μέσο τετραγωνικό λάθος του εκτιμητή της αναμενόμενης τιμής M_n και της διασποράς S_n ως συνάρτηση του αριθμού μετρήσεων n (Σημείωση: πρέπει να τρέξετε πολλές φορές το ίδιο πείραμα!).

2. Υπολογίστε αναλυτικά αν ο εκτιμητής $W_n = \sum_{i=1}^{n} x_i^3$ είναι αμερόληπτος εκτιμητής της ροπής 3ης τάξης (όπου x_i i.i.d.).

3. Υπολογίστε αριθμητικά το μέσο τετραγωνικό λάθος του εκτιμητή $M'_n = 0.5(\max\{x_i\} - \min\{x_i\})$ (όπου x_i i.i.d.). Είναι αμερόληπτος; Είναι ασυμπτωτικά αμερόληπτος;

4. Υπολογίστε αριθμητικά το μέσο τετραγωνικό λάθος του εκτιμητή γραμμικής συσχέτισης Pearson ως συνάρτηση του αριθμού των μετρήσεων n.

5. Υπολογίστε το διάστημα εμπιστοσύνης (τουλάχιστον) 95% για οποιαδήποτε τ.μ. ως συνάρτηση της αναμενόμενης τιμής και τυπικής απόκλισης χρησιμοποιώντας την ανισότητα Chebyshev.

6. Υπολογίστε τον εκτιμητή μεγιστοποίησης της πιθανότητας του p για την κατανομή Bernoulli με πιθανότητα επιτυχίας p.

7. Υπολογίστε αριθμητικά τις παραμέτρους των μοντέλων MM-0, MM-1 και MM-2 για 100 παρατηρήσεις που προέρχονται από γεννήτρια τ.α. Bernoulli με πιθανότητα επιτυχίας $p = 0.2$. Είναι οι παρατηρήσεις ανεξάρτητες σύμφωνα με το μοντέλο MM-2 που εκτιμήσατε αριθμητικά;

8. Σε ένα μοντέλο HMM-0 έχουμε δύο καταστάσεις s_1 και s_2 με κατανομές παρατηρήσεων $p(o_t|x_t = s_1) \sim \mathcal{N}(0, 1)$ και $p(o_t|x_t = s_2) \sim \mathcal{N}(2, 1)$. Δεδομένου ότι οι a priori πιθανότητες είναι ίσες $p(s_1) = p(s_2) = 0.5$ υπολογίστε το σημείο απόφασης σύμφωνα με το ταξινομητή του Bayes.

9. Για τον παραπάνω ταξινομητή υπολογίστε αριθμητικά το σημείο απόφασης και την πιθανότητα λάθος ταξινόμησης (χρησιμοποιήστε γεννήτριες κανονικών τ.α. που ακολουθούν ισοπίθανα μια από τις παραπάνω κατανομές).

10. Επαναλάβετε τα παραπάνω όταν οι a priori πιθανότητες είναι ίσες με $p(s_1) = 0.1$, $p(s_2) = 0.9$.

11. Βρείτε τις σχέσεις ανεξαρτησίας υπό συνθήκη (βλ. Κεφ. 1) σε ένα μοντέλο HMM-1 και δείξτε με βάση αυτές τις σχέσεις ότι:

$$P(x_1, x_2, ..., x_n, o_1, o_2, ..., o_n) = \prod_{i=1}^{n} P(x_i|x_{i-1})P(o_i|x_i)$$

12. Έστω τρία νομίσματα με πιθανότητα να φέρουν γράμματα 0.3, 0.5 και 0.7 αντίστοιχα. Έστω ότι χρησιμοποιώ ένα μοντέλο HMM-1 όπου ο πίνακας μετάβασης μεταξύ νομισμάτων είναι $a_{ij} = 1/3$, για $i, j = 1, 2, 3$ (η αρχική πιθανότητα να επιλέξω ένα από τα νομίσματα είναι επίσης $\pi_i = 1/3$, για $i = 1, 2, 3$). Είναι οι ρίψεις ανεξάρτητες; Ποια είναι η πιο πιθανή αλληλουχία από νομίσματα άμα παρατηρώ HTH.

13. Έστω ότι ο πίνακας πιθανοτήτων μετάβασης έχει τιμές $a_{ij} = 1/2$, για $i \neq j$ (και 0 στα στοιχεία της διαγωνίου). Επαναλάβατε το παραπάνω πρόβλημα.

14. Υλοποιήστε τον αλγόριθμο Viterbi και επιβεβαιώστε αριθμητικά το αποτέλεσμα της παραπάνω άσκησης.

15. Σε ένα πρόβλημα διάγνωσης ασθένειας τα αποτελέσματα του τεστ για τον υγιή πληθυσμό δίνεται από την κατανομή $p(o_t|x_t = 0) \sim \mathcal{N}(0, 1)$ και για αρρώστους $p(o_t|x_t = 1) \sim \mathcal{N}(4, 1)$. Δεδομένου ότι η a priori πιθανότητα κάποιος να είναι υγιής είναι $p(0) = 0.99$ υπολογίστε το σημείο απόφασης σύμφωνα με το ταξινομητή του Bayes.

16. Έστω ότι το οικονομικό κόστος στην παραπάνω άσκηση είναι δεκαπλάσιο όταν αποτύχουμε να διαγνώσουμε την ασθένεια (miss) από την λάθος διάγνωση της ασθένειας (false alarm). Πώς πρέπει να μετακινήσουμε το σημείο απόφασης για να ελαχιστοποιήσουμε την αναμενόμενη τιμή του οικονομικού κόστους;

Κεφάλαιο 6

Τυχαία Διανύσματα

Ορισμός: Το τυχαίο διάνυσμα $\mathbf{x} = [x_1, x_2, ..., x_n]^T$ είναι μια συνάρτηση από τον χώρο ενδεχομένων στο $n-$διδιάστατο χώρο πραγματικών αριθμών, δηλαδή, αντιστοιχίζει ενδεχόμενα σε $n-$άδες πραγματικών αριθμών.

Το τυχαίο διάνυσμα (τ.δ.) είναι στην πράξη n τ.μ., ομαδοποιημένες σε ένα διάνυσμα. Υπενθυμίζουμε ότι ανεξάρτητα από το πλήθος των τ.μ. το μέτρο πιθανότητας ορίζεται πάντα σε ένα (και μοναδικό) χώρο ενδεχομένων. Δηλαδή, το ενδεχόμενο A από τον χώρο ενδεχομένων Ω και η απεικόνιση του $\mathbf{x_0}$ σε μια $n-$άδα πραγματικών αριθμών έχουν εξ΄ ορισμού το ίδιο μέτρο πιθανότητας:

$$\begin{array}{lccc} \text{ενδεχόμενα:} & A \in \Omega & \rightarrow & \mathbf{x_0} \in \mathcal{R}^n \\ & \downarrow & & \downarrow \\ \text{μέτρο πιθανότητας:} & P(A) & \equiv & P(\mathbf{x_0}) \end{array}$$

Πρόκειται λοιπόν για καινούργιο φορμαλισμό ποσοτήτων που έχουμε εισαγάγει ήδη από τα Κεφ. 2 και 3. Πράγματι η Σ.Π.Π. (ή Σ.Μ.Π.) ενός τ.δ. $\mathbf{x} = [x_1, x_2, ..., x_n]^T$ δεν είναι τίποτα άλλο από την από κοινού Σ.Π.Π. (ή Σ.Μ.Π.) των $(x_1, x_2, ..., x_n)$, δηλαδή:

$$f_{\mathbf{x}}(\mathbf{x}) \equiv f_{x_1,x_2,...,x_n}(x_1, x_2, ..., x_n)$$

Προφανώς μπορούμε να γράψουμε όλες τις γνωστές ιδιότητες όπως την οριακή (διαμερισμός), την υπό συνθήκη Σ.Π.Π., τον κανόνα της αλυσίδας και τον κανόνα του Bayes, π.χ.,

$$\int_{\mathcal{R}^n} f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} f_{x_1,x_2,...,x_n}(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n = 1$$

$$f_{x_2,...,x_n}(x_2, ..., x_n) = \int_{-\infty}^{+\infty} f_{x_1,x_2,...,x_n}(x_1, x_2, ..., x_n) dx_1 = \int_{-\infty}^{+\infty} f_{\mathbf{x}}(\mathbf{x}) dx_1$$

$$f_{x_3,...,x_n}(x_3,...,x_n) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{x_1,x_2,...,x_n}(x_1,x_2,...,x_n)dx_1dx_2$$

$$f_{\mathbf{y}}(\mathbf{y}) = \int_{\mathcal{R}^n} f_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y})d\mathbf{x}$$

$$f_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}) = \frac{f_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y})}{f_{\mathbf{y}}(\mathbf{y})}$$

$$f_{\mathbf{x},\mathbf{y},\mathbf{z}}(\mathbf{x},\mathbf{y},\mathbf{z}) = f_{\mathbf{x}|\mathbf{y},\mathbf{z}}(\mathbf{x}|\mathbf{y},\mathbf{z})f_{\mathbf{y}|\mathbf{z}}(\mathbf{y}|\mathbf{z})f_{\mathbf{z}}(\mathbf{z})$$

$$f_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}) = \frac{f_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})f_{\mathbf{x}}(\mathbf{x})}{f_{\mathbf{y}}(\mathbf{y})} = \frac{f_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})f_{\mathbf{x}}(\mathbf{x})}{\int_{\mathcal{R}^n} f_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y})d\mathbf{x}}$$

όπου $\mathbf{x}$, $\mathbf{y} = [y_1, y_2, ..., y_m]^T$, $\mathbf{z} = [z_1, z_2, ..., z_l]^T$ τ.δ. Βλέπουμε ότι η χρήση διανυσμάτων είναι απλά ένας βολικός μαθηματικός φορμαλισμός.

6.1 Στατιστικές Ιδιότητες Τ.Δ.

Η αναμενόμενη τιμή ενός τ.δ. $\bar{x} = [x_1, x_2, ..., x_n]^T$ είναι το διάνυσμα των αναμενόμενων τιμών $[\mu_{x_1}, \mu_{x_2}, ..., \mu_{x_n}]^T$ των τ.μ. που το αποτελούν, δηλαδή,

$$\mu_{\mathbf{x}} \equiv E(\mathbf{x}) = E\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} E(x_1) \\ E(x_2) \\ \vdots \\ E(x_n) \end{bmatrix} \equiv \begin{bmatrix} \mu_{x_1} \\ \mu_{x_2} \\ \vdots \\ \mu_{x_n} \end{bmatrix}$$

Για τ.δ. χρησιμοποιούμε τον πίνακα συνδιασποράς (ή συνδιακύμανσης) για να περιγράψει τις στατιστικές ιδιότητες δεύτερης τάξης. Περιέχει στη διαγώνιο τις διασπορές των τ.μ. που αποτελούν το τ.δ. και στα στοιχεία εκτός διαγωνίου την συνδιασπορά όλων των ζευγαριών των τ.μ. Ο πίνακας συνδιασποράς $\mathbf{\Sigma_x}$ μπορεί να αναπαρασταθεί χρησιμοποιώντας γραμμική άλγεβρα ως[1]:

$$\mathbf{\Sigma_x} \equiv E([\mathbf{x} - \mu_{\mathbf{x}}][\mathbf{x} - \mu_{\mathbf{x}}]^T) =$$

$$= E\left(\begin{bmatrix} x_1 - \mu_{x_1} \\ x_2 - \mu_{x_2} \\ \vdots \\ x_n - \mu_{x_n} \end{bmatrix} [x_1 - \mu_{x_1}, x_2 - \mu_{x_2}, ..., x_n - \mu_{x_n}]\right) =$$

[1] Το διάνυσμα στήλης $\mathbf{x}$ έχει διαστάσεις $n \times 1$ και το ανάστροφο (transpose) διάνυσμα $\mathbf{x}^T$ έχει διαστάσεις $1 \times n$.

$$E\begin{bmatrix} (x_1-\mu_{x_1})^2 & (x_1-\mu_{x_1})(x_2-\mu_{x_2}) & \dots & (x_1-\mu_{x_1})(x_n-\mu_{x_n}) \\ (x_2-\mu_{x_2})(x_1-\mu_{x_1}) & (x_2-\mu_{x_2})^2 & \dots & (x_2-\mu_{x_2})(x_n-\mu_{x_n}) \\ \vdots & \vdots & & \vdots \\ (x_n-\mu_{x_n})(x_1-\mu_{x_1}) & (x_n-\mu_{x_n})(x_2-\mu_{x_2}) & \dots & (x_n-\mu_{x_n})^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} & \dots & \sigma_{x_1x_n} \\ \sigma_{x_2x_1} & \sigma_{x_2}^2 & \dots & \sigma_{x_2x_n} \\ \vdots & \vdots & & \vdots \\ \sigma_{x_nx_1} & \sigma_{x_nx_1} & \dots & \sigma_{x_n}^2 \end{bmatrix} = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} & \dots & \sigma_{x_1x_n} \\ \sigma_{x_1x_2} & \sigma_{x_2}^2 & \dots & \sigma_{x_2x_n} \\ \vdots & \vdots & & \vdots \\ \sigma_{x_1x_n} & \sigma_{x_2x_n} & \dots & \sigma_{x_n}^2 \end{bmatrix}$$

όπου $\sigma_{x_i}^2$ είναι η διασπορά της τ.μ. x_i και $\sigma_{x_ix_j}$ είναι η συνδιασπορά των τ.μ. x_i, x_j. Παρατηρούμε ότι επειδή εξ ορισμού $\sigma_{x_ix_j} = \sigma_{x_jx_i}$ ο πίνακας συνδιασποράς είναι πάντα συμμετρικός (και τετραγωνικός με διάσταση $n \times n$), δηλαδή ο πίνακας συνδιασποράς ισούται με τον ανάστροφό του: $\mathbf{\Sigma_x} = \mathbf{\Sigma_x}^T$.

Όταν οι τυχαίες μεταβλητές x_i, $i = 1, 2, ..., n$ είναι γραμμικά ανεξάρτητες (ασυσχέτιστες) τότε ο πίνακας συνδιακύμανσης είναι διαγώνιος. Υπενθυμίζουμε ότι για γραμμικά ανεξάρτητες τ.μ. x_i, x_j, ισχύει

$$\sigma_{x_ix_j} = E((x_i - \mu_{x_i})(x_j - \mu_{x_j})) = E(x_i - \mu_{x_i})E(x_j - \mu_{x_j}) = 0$$

οπότε όλα τα εκτός διαγωνίου στοιχεία είναι 0. Προφανώς το παραπάνω ισχύει και όταν οι τ.μ. x_i, $i = 1, 2, ..., n$ είναι ανεξάρτητες (ή απλά ανεξάρτητες ανά ζεύγη) καθώς η γραμμική ανεξαρτησία είναι υποπερίπτωση της στατιστικής ανεξαρτησίας. Εν κατακλείδι, όταν οι τ.μ. x_i, $i = 1, 2, ..., n$ είναι γραμμικά ανεξάρτητες (ή ανεξάρτητες ή ανεξάρτητες ανά ζεύγη), τότε:

$$\mathbf{\Sigma_x} = \begin{bmatrix} \sigma_{x_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{x_2}^2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sigma_{x_n}^2 \end{bmatrix}$$

Τέλος όπως και για τ.μ., μπορούμε να γράψουμε για το τ.δ. $\mathbf{x}$ διάστασης n:

$$E(g(\mathbf{x})) = \int_{\mathcal{R}^n} g(\mathbf{x})\ f_{\mathbf{x}}(\mathbf{x})\ d\mathbf{x}$$

$$E(g(\mathbf{x})|\mathbf{y}) = \int_{\mathcal{R}^n} g(\mathbf{x})\ f_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y})\ d\mathbf{x}$$

όπου $\mathbf{y}$ τ.δ. και $g()$ συνάρτηση του τ.δ. $\mathbf{x}$.

6.2 Πολυδιάστατη Κανονική Κατανομή

Το τ.δ. $\mathbf{x} = [x_1, x_2, ..., x_n]^T$ που ακολουθεί πολυδιάστατη κανονική κατανομή[2] (multivariate Gaussian) έχει Σ.Π.Π.

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{\Sigma_x}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} [\mathbf{x} - \mu_{\mathbf{x}}]^T \mathbf{\Sigma_x}^{-1} [\mathbf{x} - \mu_{\mathbf{x}}] \right\}$$

όπου $|\Sigma_x|$ υποδηλώνει την (μη μηδενική) ορίζουσα του πίνακα συνδιασποράς και $\mathbf{\Sigma_x}^{-1}$ είναι ο αντίστροφος (inverse) πίνακας του πίνακα συνδιασποράς, έτσι ώστε το γινόμενό τους να δίνει τον μοναδιαίο πίνακα $\mathbf{\Sigma_x \Sigma_x}^{-1} = \mathbf{\Sigma_x}^{-1}\mathbf{\Sigma_x} = \mathbf{I}$. Παρατηρούμε ότι ο εκθέτης της Σ.Π.Π. πολλαπλασιάζει ένα διάνυσμα γραμμής διαστάσεων $1 \times n$ με ένα συμμετρικό $n \times n$ πίνακα και μετά με ένα διάνυσμα στήλης διαστάσεων $n \times 1$, οπότε το αποτέλεσμα είναι ένας πίνακας 1×1 (δηλαδή ένας αριθμός που είναι το μέτρο πιθανότητας για το ενδεχόμενο $(x_1, x_2, ..., x_n)$).

Η αναμενόμενη τιμή και ο πίνακας συνδιασποράς του τ.δ. $\mathbf{x}$ που ακολουθεί πολυδιάστατη κανονική τ.μ. (σύμφωνα με την παραπάνω Σ.Π.Π.) είναι $\mu_{\mathbf{x}}$ και $\mathbf{\Sigma_x}$, αντίστοιχα. Συχνά γράφουμε $\mathbf{x} \sim \mathcal{N}(\mu_{\mathbf{x}}, \mathbf{\Sigma_x})$ για να υποδηλώσουμε ότι η τ.δ. $\mathbf{x}$ ακολουθεί κανονική τ.μ. με αναμενόμενη τιμή $\mu_{\mathbf{x}}$ και πίνακα συνδιασποράς $\mathbf{\Sigma_x}$. Στη συνέχεια χρησιμοποιούμε τον όρο 'από κοινού κανονικό τ.δ.' (jointly Gaussian) για να υποδηλώσουμε τ.δ. που ακολουθούν την πολυδιάστατη κανονική κατανομή.

Εφόσον οι τυχαίες μεταβλητές x_i, $i = 1, 2, ..., n$ που αποτελούν το τ.δ. $\mathbf{x}$ είναι γραμμικά ανεξάρτητες (ασυσχέτιστες) τότε η κανονική Σ.Π.Π. απλοποιείται γιατί ο πίνακας συνδιασποράς είναι διαγώνιος. Η ορίζουσα και ο αντίστροφος του διαγώνιου πίνακα $\mathbf{\Sigma_x}$ είναι:

$$\mathbf{\Sigma_x} = \begin{vmatrix} \sigma_{x_1}^2 & 0 & ... & 0 \\ 0 & \sigma_{x_2}^2 & ... & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & ... & \sigma_{x_n}^2 \end{vmatrix} = \prod_{i=1}^{n} \sigma_{x_i}^2 \qquad \mathbf{\Sigma_x^{-1}} = \begin{bmatrix} \frac{1}{\sigma_{x_1}^2} & 0 & ... & 0 \\ 0 & \frac{1}{\sigma_{x_2}^2} & ... & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & ... & \frac{1}{\sigma_{x_n}^2} \end{bmatrix}$$

οπότε ο εκθέτης της Σ.Π.Π. ισούται με (χωρίς τον όρο $-(1/2)$) :

$$[x_1 - \mu_{x_1}, x_2 - \mu_{x_2}, ..., x_n - \mu_{x_n}] \begin{bmatrix} \frac{1}{\sigma_{x_1}^2} & 0 & ... & 0 \\ 0 & \frac{1}{\sigma_{x_2}^2} & ... & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & ... & \frac{1}{\sigma_{x_n}^2} \end{bmatrix} \begin{bmatrix} x_1 - \mu_{x_1} \\ x_2 - \mu_{x_2} \\ \vdots \\ x_n - \mu_{x_n} \end{bmatrix} =$$

[2]Προσοχή: το να ακολουθεί κάθε μια από τις τ.μ. x_i, $i = 1, 2, ..., n$ κανονική κατανομή είναι αναγκαία αλλά όχι ικανή συνθήκη για να ακολουθούν οι τ.μ. $x_1, x_2, ..., x_n$ από κοινού πολυδιάστατη κανονική κατανομή. Για ανεξάρτητες τ.μ. x_i η συνθήκη είναι και ικανή.

$$= \begin{bmatrix} \frac{x_1-\mu_{x_1}}{\sigma^2_{x_1}} & 0 & ... & 0 \\ 0 & \frac{x_2-\mu_{x_2}}{\sigma^2_{x_2}} & ... & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & ... & \frac{x_n-\mu_{x_n}}{\sigma^2_{x_n}} \end{bmatrix} \begin{bmatrix} x_1-\mu_{x_1} \\ x_2-\mu_{x_2} \\ \vdots \\ x_n-\mu_{x_n} \end{bmatrix} = \sum_{i=1}^{n} \frac{(x_i-\mu_{x_i})^2}{\sigma^2_{x_i}}$$

Οπότε για τ.μ. x_i, $i = 1, 2, ..., n$ γραμμικά ανεξάρτητες τότε η Σ.Π.Π. του τ.δ. $\mathbf{x}$ ισούται με

$$f_{\mathbf{x}}(x_1, x_2, ..., x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \prod_{i=1}^{n} \sigma_{x_i}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} \frac{(x_i-\mu_{x_i})^2}{\sigma^2_{x_i}}\right\}$$

Χρησιμοποιώντας την ιδιότητα $\exp\{\sum_i x_i\} = \prod_i \exp\{x_i\}$

$$f_{\mathbf{x}}(\mathbf{x}) = f_{\mathbf{x}}(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma_{x_i}} \prod_{i=1}^{n} \exp\left\{-\frac{(x_i-\mu_{x_i})^2}{2\sigma^2_{x_i}}\right\} =$$

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_{x_i}} \exp\left\{-\frac{(x_i-\mu_{x_i})^2}{2\sigma^2_{x_i}}\right\} = \prod_{i=1}^{n} f_{x_i}(x_i) = f_{x_1}(x_1) f_{x_2}(x_2)...f_{x_n}(x_n)$$

δηλαδή η από κοινού Σ.Π.Π. των x_i ισούται με το γινόμενο των οριακών. Συμπεραίνουμε ότι οι τ.μ. x_i, $i = 1, 2, ..., n$ είναι στατιστικά ανεξάρτητες (σύμφωνα με τον ορισμό της ανεξαρτησίας τ.μ. στο Κεφ. 3). Άρα όταν οι τ.μ. x_i, $i = 1, 2, ..., n$ είναι γραμμικά ανεξάρτητες και ακολουθούν πολυδιάστατη κανονική κατανομή είναι και στατιστικά ανεξάρτητες (για την ακρίβεια δείξαμε ότι αρκεί να είναι οι x_i ασυσχέτιστες ανά ζεύγη για να είναι ανεξάρτητες).

6.3 Γραμμικές Συναρτήσεις Τ.Δ.

Η διανυσματική απεικόνιση είναι πολύ βολική όταν συναντάμε γραμμικούς μετασχηματισμούς ομάδας τ.μ. Έστω το τ.δ. $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ που είναι γραμμικός μετασχηματισμός (affine transformation) του τ.δ. $\mathbf{x} = [x_1, x_2, ..., x_n]^T$, όπου $\mathbf{A}$ είναι πίνακας πραγματικών αριθμών με διάσταση $m \times n$ και $\mathbf{b}$ διάνυσμα στήλης πραγματικών αριθμών με διαστάσεις $m \times 1$. Οι στατιστικές ιδιότητες του τ.δ. $\mathbf{y}$ μπορούν να υπολογιστούν ως συνάρτηση της διανύσματος αναμενόμενης τιμής $\mu_{\mathbf{x}}$ και του πίνακα συνδιασποράς $\mathbf{\Sigma_x}$ του $\mathbf{x}$ ως εξής:

$$\mu_{\mathbf{y}} \equiv E(\mathbf{y}) = E(\mathbf{A}\mathbf{x} + \mathbf{b}) = \mathbf{A}E(\mathbf{x}) + \mathbf{b} = \mathbf{A}\mu_{\mathbf{x}} + \mathbf{b}$$

$$\mathbf{\Sigma_y} \equiv E([\mathbf{x}-\mu_{\mathbf{x}}][\mathbf{x}-\mu_{\mathbf{x}}]^T) = E([\mathbf{A}\mathbf{x}+\mathbf{b}-\mathbf{A}\mu_{\mathbf{x}}-\mathbf{b}][\mathbf{A}\mathbf{x}+\mathbf{b}-\mathbf{A}\mu_{\mathbf{x}}-\mathbf{b}]^T)$$
$$= E(\mathbf{A}[\mathbf{x}-\mu_{\mathbf{x}}][\mathbf{A}(\mathbf{x}-\mu_{\mathbf{x}})]^T) = E(\mathbf{A}[\mathbf{x}-\mu_{\mathbf{x}}][\mathbf{x}-\mu_{\mathbf{x}}]^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}) =$$

$$= \mathbf{A}\ E([\mathbf{x}-\mu_{\mathbf{x}}][\mathbf{x}-\mu_{\mathbf{x}}]^{\mathbf{T}})\ \mathbf{A}^{\mathbf{T}} = \mathbf{A}\mathbf{\Sigma}_{\mathbf{x}}\mathbf{A}^{\mathbf{T}}$$

όπου χρησιμοποιήσαμε την ιδιότητα της γραμμικής άλγεβρας: $(Ax)^T = x^T A^T$. Βλέπουμε ότι η αναμενόμενη τιμή και συνδιασπορά του τ.δ. $\mathbf{y}$ μπορούν να υπολογιστούν ως γραμμικός μετασχηματισμός των αντίστοιχων του τ.δ. $\mathbf{x}$.

Η Σ.Π.Π. του τ.δ. $\mathbf{y}$ μπορεί να υπολογιστεί ως συνάρτηση της Σ.Π.Π. του τ.δ. $\mathbf{x}$. Γενικεύοντας τη μέθοδο των ριζών για τον υπολογισμό Σ.Π.Π. συναρτήσεων n τ.μ.:

Έστω οι τ.μ. $y_1 = g_1(x_1, x_2, ..., x_n)$, $y_2 = g_2(x_1, x_2, ..., x_n)$, ..., $y_n = g_n(x_1, x_2, ..., x_n)$ συναρτήσεις των τ.μ. $x_1, x_2, ..., x_n$. Η από κοινού Σ.Π.Π. των τ.μ. $y_1, y_2, ..., y_n$ μπορεί να υπολογιστεί ως συνάρτηση της Σ.Π.Π. $f_{x_1,x_2,...,x_n}()$ ως εξής: Έστω ότι οι $n-$άδες $(x_1^{(1)}, x_2^{(1)}, ..., x_n^{(1)})$, ..., $(x_1^{(m)}, x_2^{(m)}, ..., x_n^{(m)})$ είναι οι m το πλήθος ρίζες του συστήματος εξισώσεων $y_i = g_i(x_1, x_2, ..., x_n)$, $i = 1, 2, ..., n$. Τότε η από κοινού Σ.Π.Π. $f_{y_1,y_2,...,y_n}()$ μπορεί να εκφραστεί ως:

$$f_{y_1,y_2,...,y_n}(y_1, y_2, ..., y_n) = \sum_{i=1}^{m} \frac{f_{x_1,x_2,...,x_n}(x_1^{(i)}, x_2^{(i)}, ..., x_n^{(i)})}{|J(x_1^{(i)}, x_2^{(i)}, ..., x_n^{(i)})|}$$

όπου ο παρανομαστής είναι η ορίζουσα του $n \times n$ πίνακα μερικών παραγώγων:

$$J(x_1, x_2, ..., x_n) = \begin{vmatrix} \frac{\partial g_1(\mathbf{x})}{\partial x_1} & \frac{\partial g_1(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial g_1(\mathbf{x})}{\partial x_n} \\ \frac{\partial g_2(\mathbf{x})}{\partial x_1} & \frac{\partial g_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial g_2(\mathbf{x})}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial g_n(\mathbf{x})}{\partial x_1} & \frac{\partial g_n(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial g_n(\mathbf{x})}{\partial x_n} \end{vmatrix}$$

υπολογισμένη στις ρίζες του συστήματος εξισώσεων.

Στην ειδική περίπτωση όπου $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ και $\mathbf{x} = [x_1, x_2, ...x_n]^T$, $\mathbf{y} = [y_1, y_2, ...y_n]^T$ τ.δ., το σύστημα εξισώσεων είναι $y_i = g_i(\mathbf{x}) = \sum_{j=1}^{n} a_{ij} x_j + b_i$, για $i = 1, 2, ..., n$, όπου a_{ij} είναι το στοιχείο της γραμμής i στήλης j του πίνακα $\mathbf{A}$. Άρα η ορίζουσα του πίνακα Jacobi είναι:

$$J(\mathbf{x}) = \begin{vmatrix} \frac{\partial g_1(\mathbf{x})}{\partial x_1} & \frac{\partial g_1(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial g_1(\mathbf{x})}{\partial x_n} \\ \frac{\partial g_2(\mathbf{x})}{\partial x_1} & \frac{\partial g_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial g_2(\mathbf{x})}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial g_n(\mathbf{x})}{\partial x_1} & \frac{\partial g_n(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial g_n(\mathbf{x})}{\partial x_n} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = |\mathbf{A}|$$

Εφόσον ο πίνακας $\mathbf{A}$ είναι (τετραγωνικός και) αντιστρέψιμος το σύστημα εξισώσεων έχει μόνο μια λύση:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b} \Rightarrow \mathbf{y} - \mathbf{b} = \mathbf{A}\mathbf{x} \Rightarrow \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}) = \mathbf{A}^{-1}\mathbf{A}\mathbf{x} \Rightarrow \mathbf{x} = \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})$$

Άρα η Σ.Π.Π. του τ.δ. $\mathbf{y}$ είναι:

$$f_{\mathbf{y}}(\mathbf{y}) = \frac{f_{\mathbf{x}}(\mathbf{A}^{-1}(\mathbf{y}-\mathbf{b}))}{|\mathbf{A}|}$$

6.4 Υποδιανύσματα Τ.Δ.

Έστω το τ.δ. $\mathbf{x}$ που αποτελείται από δύο υποδιανύσματα $\mathbf{x_1}$, $\mathbf{x_2}$, δηλαδή, $\mathbf{x}^T = [\mathbf{x_1}^T, \mathbf{x_2}^T]$. Εφόσον ξέρουμε την Σ.Π.Π. του τ.δ. (δηλαδή με άλλα λόγια την από κοινού των τ.δ. $\mathbf{x_1}$, $\mathbf{x_2}$) μπορούμε να υπολογίσουμε την Σ.Π.Π. των τ.δ. $\mathbf{x_1}$ και $\mathbf{x_2}$ (δηλαδή τις οριακές). Επίσης εφόσον γνωρίζω τις στατιστικές ιδιότητες του τ.δ. $\mathbf{x}$ τότε ξέρω και τις στατιστικές ιδιότητες των υποδιανυσμάτων $\mathbf{x_1}$ και $\mathbf{x_2}$. Στην πράξη απλά γενικεύουμε για τ.δ. αυτό που γνωρίζουμε ότι ισχύει για τ.μ. (βλ. Κεφ. 2, 3).

Για παράδειγμα έστω το τυχαίο διάνυσμα $\mathbf{x} = [x_1, x_2, x_3, x_4]^T$ και τα υποδιανύσματα $\mathbf{x_1} = [x_1, x_2]^T$, $\mathbf{x_2} = [x_3, x_4]^T$. Προφανώς τα $\mathbf{x_1}$, $\mathbf{x_2}$ είναι επίσης τ.δ. Έστω η Σ.Π.Π. του $\mathbf{x}$ είναι η $f_{\mathbf{x}}() \equiv f_{x_1,x_2,x_3,x_4}()$. Τότε οι Σ.Π.Π. των $\mathbf{x_1}$, $\mathbf{x_2}$ προκύπτουν ως:

$$f_{\mathbf{x_1}}(\mathbf{x_1}) \equiv f_{x_1,x_2}(x_1,x_2) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{x_1,x_2,x_3,x_4}(x_1,x_2,x_3,x_4)\, dx_3\, dx_4$$

$$f_{\mathbf{x_2}}(\mathbf{x_2}) \equiv f_{x_3,x_4}(x_3,x_4) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{x_1,x_2,x_3,x_4}(x_1,x_2,x_3,x_4)\, dx_1\, dx_2$$

Επίσης μπορούμε να υπολογίσουμε την υπό συνθήκη Σ.Π.Π. $f_{\mathbf{x_1}|\mathbf{x_2}}(\mathbf{x_1}|\mathbf{x_2})$ ως:

$$f_{\mathbf{x_1}|\mathbf{x_2}}(\mathbf{x_1}|\mathbf{x_2}) = \frac{f_{\mathbf{x_1},\mathbf{x_2}}(\mathbf{x_1},\mathbf{x_2})}{f_{\mathbf{x_2}}(\mathbf{x_2})} \equiv \frac{f_{\mathbf{x}}(\mathbf{x})}{f_{\mathbf{x_2}}(\mathbf{x_2})}$$

όπου εξ' ορισμού η $f_{\mathbf{x}}()$ είναι η από κοινού Σ.Π.Π. των τ.δ. $\mathbf{x_1}$, $\mathbf{x_2}$ (ή αλλιώς η από κοινού Σ.Π.Π. των τ.μ. x_1, x_2, x_3, x_4).

Έστω το διάνυσμα αναμενόμενων τιμών $\mu_{\mathbf{x}}$ και ο πίνακας συνδιασποράς Σ_x του τ.δ. $\mathbf{x}$. Τότε μπορούμε να γράψουμε ότι:

$$\mu_{\mathbf{x}} = \begin{bmatrix} \mu_{x_1} \\ \mu_{x_2} \\ --- \\ \mu_{x_3} \\ \mu_{x_4} \end{bmatrix} = \begin{bmatrix} \mu_{\mathbf{x_1}} \\ \mu_{\mathbf{x_2}} \end{bmatrix}$$

$$\mathbf{\Sigma_x} = \left[\begin{array}{cc|cc} \sigma_{x_1}^2 & \sigma_{x_1x_2} & \sigma_{x_1x_3} & \sigma_{x_1x_4} \\ \sigma_{x_1x_2} & \sigma_{x_2}^2 & \sigma_{x_2x_3} & \sigma_{x_2x_4} \\ \hline \sigma_{x_1x_3} & \sigma_{x_2x_3} & \sigma_{x_3}^2 & \sigma_{x_3x_4} \\ \sigma_{x_1x_4} & \sigma_{x_2x_4} & \sigma_{x_3x_4} & \sigma_{x_4}^2 \end{array}\right] = \left[\begin{array}{cc} \mathbf{\Sigma_{x_1}} & \mathbf{\Sigma_{x_1x_2}} \\ \mathbf{\Sigma_{x_2x_1}} & \mathbf{\Sigma_{x_2}} \end{array}\right]$$

όπου $\mu_{\mathbf{x_1}}$, $\mu_{\mathbf{x_2}}$ οι αναμενόμενες τιμές και $\mathbf{\Sigma_{x_1}}$, $\mathbf{\Sigma_{x_2}}$, οι πίνακες συνδιασποράς των τ.δ. $\mathbf{x_1}$, $\mathbf{x_2}$, αντίστοιχα. Οι πίνακες $\mathbf{\Sigma_{x_1x_2}}$, $\mathbf{\Sigma_{x_2x_1}}$ είναι οι πίνακες συνδιασποράς[3] των $\mathbf{x_1}$, $\mathbf{x_2}$ και $\mathbf{x_2}$, $\mathbf{x_1}$, αντίστοιχα. Βλέπουμε ότι η συνδιασπορά υποδιανύσματος του $\mathbf{x}$ είναι υποπίνακας του πίνακα συνδιασποράς $\mathbf{\Sigma_x}$.

6.5 Παραδείγματα

Παράδειγμα 1: Έστω το τ.δ. $\mathbf{x} = [x_1, x_2, x_3, x_4]^T$ που ακολουθεί πολυδιάστατη κανονική τιμή με αναμενόμενη τιμή $\mu_{\mathbf{x}}$ και πίνακα συνδιασποράς $\mathbf{\Sigma_x}$, δηλαδή $\mathbf{x} \sim \mathcal{N}(\mu_{\mathbf{x}}, \mathbf{\Sigma_x})$:

$$\mu_{\mathbf{x}} = \left[\begin{array}{r} 1 \\ 2 \\ -1 \\ 0 \end{array}\right] \qquad \mathbf{\Sigma_x} = \left[\begin{array}{rrrr} 4 & 1 & -1 & 0 \\ 1 & 4 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array}\right]$$

Υπολογίστε την από κοινού Σ.Π.Π. και τον συντελεστή συσχέτισης των τ.μ. x_2, x_3, των x_1, x_4 και των x_1, x_3.

Η αναμενόμενη τιμή των τ.μ. x_1, x_2, x_3, x_4 δίδεται και είναι $1, 2, -1, 0$, αντίστοιχα. Οι πίνακες συνδιασποράς προκύπτουν από τον $\mathbf{\Sigma_x}$ και είναι:

$$\mathbf{\Sigma}_{[x_2,x_3]} = \left[\begin{array}{cc} 4 & 0 \\ 0 & 1 \end{array}\right] \quad \mathbf{\Sigma}_{[x_1,x_4]} = \left[\begin{array}{cc} 4 & 0 \\ 0 & 1 \end{array}\right] \quad \mathbf{\Sigma}_{[x_1,x_3]} = \left[\begin{array}{rr} 4 & -1 \\ -1 & 1 \end{array}\right]$$

Οι συντελεστές συσχέτισης είναι:

$$\rho_{x_2x_3} = \frac{\sigma_{x_2x_3}}{\sigma_{x_2}\sigma_{x_3}} = \frac{0}{\sqrt{4}\sqrt{2}} = 0 \quad \rho_{x_1x_4} = \frac{0}{\sqrt{4}\sqrt{1}} = 0 \quad \rho_{x_1x_3} = \frac{-1}{\sqrt{4}\sqrt{1}} = -0.5$$

Παρατηρούμε ότι οι τ.μ. x_2, x_3 είναι ασυσχέτιστες όπως και οι x_1, x_4. Οι τ.μ. x_1, x_3 έχουν βαθμό γραμμικής εξάρτησης -0.5 (δηλαδή αύξηση του x_1 είναι αρκετά πιθανό να οδηγήσει σε μείωση του x_3).

[3] Ο πίνακας συνδιασποράς δύο τ.δ. ΔΕΝ είναι συμμετρικός. Ισχύει ότι $\mathbf{\Sigma_{x_1x_2}} = \mathbf{\Sigma_{x_2x_1}^T}$. Επίσης όταν τα τ.δ. $\mathbf{x_1}$, $\mathbf{x_2}$ έχουν διαφορετικό πλήθος τ.μ., τότε οι πίνακες $\mathbf{\Sigma_{x_1x_2}}$, $\mathbf{\Sigma_{x_2x_1}}$ δεν είναι τετραγωνικοί.

Η από κοινού Σ.Π.Π. των x_2, x_3 προκύπτει ως η οριακή της $f_{\mathbf{x}}()$.

$$f_{x_2,x_3}(x_2,x_3) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{x_1,x_2,x_3,x_4}(x_1,x_2,x_3,x_4)\; dx_1\; dx_4$$

Ξεκινώντας με το ολοκλήρωμα ως προς x_1, ο εκθέτης της πολυδιάστατης κανονικής κατανομής μπορεί να εκφραστεί ως:

$$-\frac{1}{2}\;\sum_{i=1}^{4}\sum_{j=1}^{4}\;(x_i-\mu_{x_i})\; w_{ij}\; (x_j-\mu_{x_j})$$

όπου w_{ij} το στοιχείο της γραμμής i, στήλης j του συμμετρικού πίνακα ${\mathbf{\Sigma}_{\mathbf{x}}}^{-1}$. Διαχωρίζοντας τους όρους που περιέχουν το x_1 (και αγνοώντας τον όρο $-(1/2)$):

$$\sum_{i=2}^{4}\sum_{j=2}^{4}(x_i-\mu_{x_i})w_{ij}(x_j-\mu_{x_j})+w_{11}(x_1-\mu_{x_1})^2+2(x_1-\mu_{x_1})\sum_{i=2}^{4}w_{1i}(x_i-\mu_{x_i})$$

Παρατηρούμε ότι ως προς x_1 ο εκθέτης είναι ένα πολυώνυμο δεύτερης τάξης, δηλαδή, ο εκθέτης είναι ανάλογος του $x_1^2+2ax_1+b=(x_1-a)^2+(b-a^2)$ όπου τα a, b είναι συναρτήσεις των x_2, x_3, x_4 (και άρα 'σταθερές' ως προς το ολοκλήρωμα στο x_1). Άρα το ολοκλήρωμα:

$$\int_{-\infty}^{+\infty} f_{\mathbf{x}}(\mathbf{x})\; dx_1 = c\int_{-\infty}^{+\infty} e^{-d(x_1-a)^2}\; dx_1 = c'$$

επειδή η $e^{-d(x_1-a)^2}$ είναι ανάλογη μιας κανονικής Σ.Π.Π. και άρα το ολοκλήρωμα ισούται με μία σταθερά (ή 1 αν η Σ.Π.Π. είναι σωστά κανονικοποιημένη). Σημειώνουμε ότι c, c', d είναι συναρτήσεις των x_2, x_3, x_4. Το αποτέλεσμα c' είναι άμα κάνουμε τις πράξεις:

$$c' \sim \exp\left\{-\frac{1}{2}\sum_{i=2}^{4}\sum_{j=2}^{4}(x_i-\mu_{x_i})\left(w_{ij}-\frac{w_{1i}w_{1j}}{w_{11}}\right)(x_j-\mu_{x_j})\right\}$$

Οπότε οι x_2, x_3, x_4 ακολουθούν από κοινού κανονική κατανομή. Αντίστοιχα, μπορούμε να δείξουμε ότι και μετά την ολοκλήρωση ως προς x_4, οι x_2, x_3 ακολουθούν από κοινού κανονική κατανομή. Εν γένει οποιαδήποτε υποδιάνυσμα από κοινού κανονικού τ.δ. ακολουθεί επίσης πολυδιάστατη κανονική κατανομή. Οι αναμενόμενες τιμές και πίνακες συνδιακύμανσης έχουνε ήδη υπολογιστεί και άρα:

$$\begin{bmatrix} x_2 \\ x_3 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}\right) \quad \begin{bmatrix} x_1 \\ x_4 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

$$\begin{bmatrix} x_1 \\ x_3 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 4 & -1 \\ -1 & 1 \end{bmatrix}\right)$$

Παράδειγμα 2: Έστω το τ.δ. $\mathbf{y} = [y_1, y_2, y_3]^T$ που προκύπτει ως μετασχηματισμός του τ.δ. $\mathbf{b}$ στο προηγούμενο παράδειγμα ως εξής:

$$y_1 = x_1 + x_3 \qquad y_2 = x_2 + x_4 \qquad y_3 = x_1 + 2x_2 - 3x_3$$

Υπολογίστε την αναμενόμενη τιμή, τον πίνακα συνδιασποράς και την Σ.Π.Π. του τ.δ. $\mathbf{y}$.

Το τ.δ. $\mathbf{y}$ είναι γραμμικός μετασχηματισμός $\mathbf{y} = \mathbf{Ax}$ του τ.δ. $\mathbf{x} = [x_1, x_2, x_3, x_4]^T$ όπου ο πίνακας Α είναι:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 2 & -3 & 0 \end{bmatrix}$$

Άρα η αναμενόμενη τιμή και ο πίνακας συνδιασποράς είναι:

$$\mu_{\mathbf{y}} \equiv E(\mathbf{y}) = \mathbf{A}\mu_{\mathbf{x}} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 2 & -3 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 8 \end{bmatrix}$$

$$\mathbf{\Sigma_y} = \mathbf{A\Sigma_x A^T} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 2 & -3 & 0 \end{bmatrix} \begin{bmatrix} 4 & 1 & -1 & 0 \\ 1 & 4 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & -3 \\ 0 & 1 & 0 \end{bmatrix} \Rightarrow$$

$$\mathbf{\Sigma_y} = \begin{bmatrix} 3 & 1 & 5 \\ 1 & 7 & 11 \\ 5 & 11 & 39 \end{bmatrix}$$

Για να υπολογίσουμε την Σ.Π.Π. του τ.δ. $\mathbf{y}$ προσθέτουμε την τ.μ. $y_4 = x_4$ και δημιουργούμε το νέο τ.δ. $\mathbf{z} = [y_1, y_2, y_3, y_4]^T$. Ο γραμμικός μετασχηματισμός $\mathbf{z} = \mathbf{Bx}$ έχει τετραγωνικό πίνακα B που είναι αναστρέψιμος. Η Σ.Π.Π. της τ.μ. $\mathbf{z}$ προκύπτει με την μέθοδο των ριζών ως συνάρτηση της Σ.Π.Π. $f_{\mathbf{x}}()$:

$$f_{\mathbf{z}}(\mathbf{z}) = \frac{f_{\mathbf{x}}(\mathbf{B^{-1}z})}{|\mathbf{B}|}$$

Άρα ο εκθέτης της Σ.Π.Π. του τ.δ. z είναι (εκτός του όρου $-(1/2)$):

$$(\mathbf{B^{-1}z} - \mu_{\mathbf{x}})^{\mathbf{T}} \mathbf{\Sigma_x^{-1}} (\mathbf{B^{-1}z} - \mu_{\mathbf{x}}) = (\mathbf{B^{-1}z} - \mu_{\mathbf{x}})^{\mathbf{T}} \mathbf{I \Sigma_x^{-1} I} (\mathbf{B^{-1}z} - \mu_{\mathbf{x}}) =$$

$$= (\mathbf{B^{-1}z} - \mu_\mathbf{x})^\mathbf{T}\ \mathbf{B^T}(\mathbf{B^T})^{-1}\ \mathbf{\Sigma_x^{-1}}\ \mathbf{B^{-1}B}\ (\mathbf{B^{-1}z} - \mu_\mathbf{x}) =$$
$$= [\mathbf{B}(\mathbf{B^{-1}z} - \mu_\mathbf{x})]^\mathbf{T}\ (\mathbf{B\Sigma_x B^T})^{-1}[\mathbf{B}(\mathbf{B^{-1}z} - \mu_\mathbf{x})] =$$
$$= (\mathbf{z} - \mathbf{B}\mu_\mathbf{x})^\mathbf{T}\ (\mathbf{B\Sigma_x B^T})^{-\mathbf{1}}(\mathbf{z} - \mathbf{B}\mu_\mathbf{x}) = (\mathbf{z} - \mu_\mathbf{z})^\mathbf{T}\mathbf{\Sigma_z^{-1}}(\mathbf{z} - \mu_\mathbf{z})$$

όπου υποθέσαμε ότι οι πίνακες $\mathbf{B}$, $\mathbf{\Sigma_x}$ είναι αντιστρέψιμοι και χρησιμοποιήσαμε τις ιδιότητες $(A\Sigma A^T)^{-1} = (A^T)^{-1}\Sigma^{-1}A^{-1}$ και $(Ax)^T = x^T A^T$. Βλέπουμε ότι το τ.δ. z ακολουθεί πολυδιάστατη κανονική κατανομή. Άρα και το υποδιάνυσμα y επίσης ακολουθεί πολυδιάστατη κανονική κατανομή (βλ. Παράδειγμα 1). Εν γένει ο γραμμικός μετασχηματισμός από κοινού κανονικού τ.δ. είναι επίσης από κοινού κανονικό[4] τ.δ. Οι στατιστικές ιδιότητες του τ.δ. $\mathbf{y}$ έχουν ήδη υπολογιστεί και άρα:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 2 \\ 8 \end{bmatrix}, \begin{bmatrix} 3 & 1 & 5 \\ 1 & 7 & 11 \\ 5 & 11 & 39 \end{bmatrix}\right)$$

<u>Παράδειγμα 3</u>: Υπολογίστε την υπό συνθήκη Σ.Π.Π. $f_{x_1|x_2}()$ για το τ.δ. $\mathbf{x} = [x_1, x_2, x_3, x_4]^T$ από το Παράδειγμα 1.

Ξέρουμε ότι τόσο η Σ.Π.Π. $f_{x_1,x_2}()$ όσο και η $f_{x_2}()$ είναι κανονικές, αφού πρόκειται για υποδιανύσματα κανονικού τ.δ. $\mathbf{y}$. Συγκεκριμένα:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}\right) \qquad x_2 \sim \mathcal{N}(2, 4)$$

Άρα η υπό συνθήκη Σ.Π.Π. $f_{x_1|x_2}()$ δίδεται ως:

$$f_{x_1|x_2}(x_1|x_2) = \frac{f_{x_1,x_2}(x_1, x_2)}{f_{x_2}(x_2)}$$

Στον εκθέτη της $f_{x_1|x_2}()$ έχουμε την διαφορά ανάμεσα στον εκθέτη της $f_{x_1,x_2}()$ και της $f_{x_2}()$ δηλαδή (αγνοώντας τον όρο $-(1/2)$):

$$\sum_{i=1}^{2}\sum_{j=1}^{2}(x_i - \mu_{x_i})w'_{ij}(x_j - \mu_{x_j}) - \frac{(x_2 - \mu_{x_2})^2}{\sigma_{x_2}^2}$$

[4] Αν ο πίνακας μετασχηματισμού έχει περισσότερες γραμμές από στήλες (δηλαδή προσπαθήσουμε να δημιουργήσουμε μεγαλύτερο αριθμό τ.μ. από αυτές που μας δίνονται) τότε η ορίζουσα του πίνακα συνδιακύμανσης που προκύπτει έχει τιμή 0. Στην περίπτωση αυτή το νέο τ.δ. ακολουθεί πολυδιάστατη κανονική κατανομή αλλά όχι με την γνωστή Σ.Π.Π. (γιατί ο παρανομαστής θα ήταν 0). Το ίδιο ισχύει όταν, π.χ., δυο γραμμές του πίνακα μετασχηματισμού είναι ίδιες.

όπου w'_{ij} το στοιχείο της γραμμής i, στήλης j του αντίστροφου πίνακα συνδιασποράς των x_1, x_2. Ο εκθέτης είναι ένα πολυώνυμο δεύτερου βαθμού ως προς την τ.μ. x_1:

$$w'_{11}(x_1-\mu_{x_1})^2+2w'_{12}(x_1-\mu_{x_1})(x_2-\mu_{x_2})+\left(w'_{22}-\frac{1}{\sigma^2_{x_2}}\right)(x_2-\mu_{x_2})^2$$

που μπορεί να εκφραστεί στην μορφή $(x_1-\mu_{x_1|x_2})^2/\sigma^2_{x_1|x_2}$. Συμπεραίνουμε ότι η υπό συνθήκη Σ.Π.Π. $f_{x_1|x_2}()$ είναι κανονική με αναμενόμενη τιμή και διασπορά:

$$\mu_{x_1|x_2}=\mu_{x_1}-\frac{w'_{12}}{w'_{11}}\,(x_2-\mu_{x_2})\qquad \sigma^2_{x_1|x_2}=\frac{1}{w'_{11}}$$

Εν γένει για υποδιανύσματα $\mathbf{x_1}$, $\mathbf{x_2}$ από κοινού κανονικού τ.δ.: η υπό συνθήκη Σ.Π.Π. $f_{\mathbf{x_1}|\mathbf{x_2}}(\mathbf{x_1}|\mathbf{x_2})$ του τ.δ. $\mathbf{x_1}$ δεδομένου του $\mathbf{x_2}$ ακολουθεί επίσης πολυδιάστατη κανονική κατανομή. Αντικαθιστώντας τις τιμές $w'_{11}=4/15$, $w'_{12}=-1/15$ και τις αναμενόμενες τιμές:

$$x_1|x_2\ \sim\ \mathcal{N}(1+0.25\ (x_2-2),\ 3.75)$$

Κατ' αρχήν βλέπουμε ότι η γνώση του x_2 μειώνει κατάτι την διασπορά του x_1 (από 4 σε 3.75). Αυτό είναι λογικό καθότι τα x_1 και x_2 είναι εξαρτημένα και η γνώση του x_2 μειώνει την αβεβαιότητα για το x_1. Η μείωση είναι μικρή γιατί ο συντελεστής γραμμικής συσχέτισης είναι μόνο 0.25. Δεύτερον, βλέπουμε ότι η υπό συνθήκη αναμενόμενη τιμή του x_1 είναι συνάρτηση του x_2. Όταν η συνθήκη είναι η $x_2=\mu_{x_2}=2$ η γνώση του x_2 δεν επηρεάζει την αναμενόμενη τιμή του x_1 (γιατί;). Η υπό συνθήκη αναμενόμενη τιμή του x_1 αυξάνει όταν η συνθήκη είναι $x_2=x_{02}>\mu_{x_2}$, λόγω της θετικής γραμμικής εξάρτησης (αύξηση του x_2 είναι πιθανότερο να οδηγήσει σε αύξηση του x_1).

Μπορούμε να επιβεβαιώσουμε αριθμητικά τα αποτελέσματά μας δημιουργώντας γεννήτριες των x_1, x_2. Δημιουργούμε πρώτα την $x_1\sim\mathcal{N}(1,4)$. Στη συνέχεια θέλουμε να δημιουργήσουμε την $x_2\sim\mathcal{N}(2,4)$ με συνδιασπορά 1 με την x_1. Για τον σκοπό αυτό δημιουργούμε μια βοηθητική κανονική τ.μ. $x_3'\sim\mathcal{N}(0,1)$ έτσι ώστε η x_2 να είναι συνάρτηση των x_1, x_3 ως εξής: $x_2=ax_1+bx_3'+c$, όπου $a,b,c\in\mathcal{R}$. Οι x_1, x_3 είναι ανεξάρτητες οπότε για να υπολογίσουμε τα a, b, c πρέπει να λύσουμε το σύστημα:

$$\left\{\begin{array}{c}\mu_{x_2}=a\mu_{x_1}+b\mu_{x_3'}+c\\ \sigma^2_{x_2}=a^2\sigma^2_{x_1}+b^2\sigma^2_{x_3'}\\ a\sigma^2_{x_1}+b\sigma_{x_1x_3'}=\sigma_{x_1x_2}\end{array}\right\}\Rightarrow\left\{\begin{array}{c}2=a+c\\ 4=4a^2+b^2\\ a=\frac{1}{4}\end{array}\right\}\Rightarrow\left\{\begin{array}{c}a=\frac{1}{4}\\ b=\sqrt{\frac{15}{4}}\\ c=\frac{7}{4}\end{array}\right\}$$

Άρα η από κοινού γεννήτρια x_1, x_2 μέσω της βοηθητικής x_3':

```
n = 10000;
x1 = randn(1,n)*2 + 1;
x3 = randn(1,n);
x2 = (1/4)*x1 + sqrt(15/4)*x3 + 7/4;
...
```

Η x_2 ακολουθεί κανονική κατανομή ως άθροισμα κανονικών τ.μ. Ο κώδικας που εκτιμά την αναμενόμενη τιμή και διασπορά της τ.μ. x_1 υπό συνθήκη $x_2 = x_{02}$, όπου η x_{02} παίρνει τιμές στο διάστημα $[-4, 8]$ είναι:

```
n = 10000;
x2 = randn(1,n)*2 + 2;
x3 = randn(1,n);
x1 = (1/4)* x2 + sqrt(15/4)*x3 + 1/2;
...
x2_condition = linspace(-4,8,100);
for i = 1:100;
  x1_cond_x2 = (1/4)* x2_condition(i) + sqrt(15/4)*x3 + 1/2;
  mn(i) = mean(x1_cond_x2);
  vr(i) = var(x1_cond_x2);
end
plot(x2_condition,mn);
plot(x2_condition,vr);
```

Όπου εδώ δημιουργήσαμε αρχικά την τ.μ. x_2 (συνθήκη) και υπολογίσαμε την x_1 ως συνάρτηση της x_2 και της βοηθητικής x_3'. Επιβεβαιώνουμε γραφικά ότι υπάρχει γραμμική σχέση ανάμεσα στην υπό συνθήκη αναμενόμενη τιμή και την τιμή της συνθήκης, ενώ η υπό συνθήκη διασπορά είναι ανεξάρτητη της τιμής της συνθήκης. Σημείωση: η μέθοδος που ακολουθήσαμε για την δημιουργία εξαρτημένων τ.μ. ισχύει μόνο για την κανονική κατανομή (ή εν γένει για κατανομές για τις οποίες το άθροισμα τ.μ. ακολουθεί την ίδια κατανομή με τα μέλη του αθροίσματος).

6.6 Σύνοψη

- Τα τυχαία διανύσματα είναι συναρτήσεις από τον χώρο ενδεχομένων στον $n-$διάστατο χώρο των πραγματικών αριθμών, δηλαδή, $\Omega \rightarrow \mathcal{R}^n$.
- Ένα $n-$διάστατο τυχαίο διάνυσμα αποτελείται από n τ.μ.
- Είναι σημαντικό να θυμόμαστε ότι τα τ.δ. δεν εισάγουν καινούργιες έννοιες, απλά έναν καινούργιο φορμαλισμό.

- Ομαδοποιώντας τ.μ. σε τ.δ. μπορούμε να χρησιμοποιήσουμε γνώσεις γραμμικής άλγεβρας, π.χ., για τον υπολογισμό στατιστικών ιδιοτήτων.
- Η Σ.Π.Π. τ.δ. είναι η από κοινού Σ.Π.Π. των τ.μ. που το αποτελούν.
- Οι πιο σημαντικές στατιστικές ιδιότητες τ.δ. $\mathbf{x}$ είναι το διάνυσμα αναμενόμενης τιμής $\mu_{\mathbf{x}}$ και ο πίνακας συνδιασποράς $\mathbf{\Sigma_x}$. Ο πίνακας συνδιασποράς είναι τετραγωνικός και συμμετρικός και περιέχει την διασπορά των τ.μ. στη διαγώνιο και την συνδιασπορά των ζευγών τ.μ. εκτός διαγωνίου.
- Όταν τα μέλη ενός τ.δ. που ακολουθεί την πολυδιάστατη κανονική κατανομή είναι ασυσχέτιστα (ανά ζεύγη), τότε είναι και ανεξάρτητα.
- Η αναμενόμενη τιμή, πίνακας συνδιασποράς και Σ.Π.Π. γραμμικού μετασχηματισμού $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ τ.δ. $\mathbf{x}$ προκύπτει ως:

$$\mu_{\mathbf{y}} = \mathbf{A}\mu_{\mathbf{x}} + \mathbf{b} \quad \mathbf{\Sigma_y} = \mathbf{A}\mathbf{\Sigma_x}\mathbf{A^T} \quad f_{\mathbf{y}}(\mathbf{y}) = \frac{f_{\mathbf{x}}(\mathbf{A^{-1}}(\mathbf{y}-\mathbf{b}))}{|\mathbf{A}|}$$

- Εν γένει η αναμενόμενη τιμή και πίνακας συνδιασποράς υποδιανύσματος τ.δ. προκύπτουν ως υποδιανύσματα και υποπίνακες των αντίστοιχων του τ.δ. Η Σ.Π.Π. υποδιανύσματος υπολογίζεται μέσω της ιδιότητας του διαμερισμού (οριακή) από την Σ.Π.Π. του τ.δ.
- Υποδιανύσματα από κοινού κανονικών τ.δ. ακολουθούν πολυδιάστατη κανονική κατανομή.
- Γραμμικοί μετασχηματισμοί από κοινού κανονικών τ.δ. ακολουθούν πολυδιάστατη κανονική κατανομή.
- Για τ.δ. $\mathbf{x_1}$, $\mathbf{x_2}$: Το υπό συνθήκη τ.δ. $\mathbf{x_1}|\mathbf{x_2}$ ακολουθεί πολυδιάστατη κανονική κατανομή, εάν τα τ.δ. $\mathbf{x_1}, \mathbf{x_2}$ είναι από κοινού κανονικά.

Ασκήσεις

1. Έστω $\mathbf{\Sigma_x}$ ο πίνακας συνδιασποράς του τ.δ. $\mathbf{x} = [x_1, x_2, ..., x_n]^T$ και $\sigma_\mathbf{x} = [\sigma^2_{x_1}, \sigma^2_{x_2}, ..., \sigma^2_{x_n}]^T$ το διάνυσμα διασπορών. Εκφράστε τον πίνακα συντελεστών γραμμικής συσχέτισης $\mathbf{P_x}$ με στοιχείο i, j το $\rho_{x_i x_j}$ ως συνάρτηση των $\mathbf{\Sigma_x}$, $\sigma_\mathbf{x}$.

2. Υλοποιήστε μια συνάρτηση που λαμβάνει ως είσοδο τον πίνακα συνδιασποράς $\mathbf{\Sigma_x}$ και (χρησιμοποιώντας πολλαπλασιασμό πινάκων) υπολογίζει τον πίνακα συντελεστών γραμμικής συσχέτισης $\mathbf{P_x}$. Συγκρίνεται τα αποτελέσματά σας με την συνάρτηση corrcoef() του MATLAB.

3. Δίδονται οι πίνακες $\begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}$, $\begin{bmatrix} -1 & 2 \\ 2 & 2 \end{bmatrix}$, $\begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix}$. Ποιοι από τους παραπάνω μπορεί να είναι πίνακες συνδιασποράς; Ποιοι από τους παραπάνω μπορεί να είναι πίνακες συνδιασποράς πολυδιάστατης κανονικής κατανομής;

4. Γενικεύστε τα κριτήρια που πρέπει να πληρεί ένας πίνακας για να είναι πίνακας συνδιασποράς πολυδιάστατης κανονικής κατανομής. Τι τιμές πρέπει να παίρνει η ορίζουσα του πίνακα συνδιασποράς (βλ. Σ.Π.Π.);

5. Βρείτε ένα παράδειγμα δύο (ή τριών) κανονικών τ.μ. που δεν είναι από κοινού κανονικές.

6. Έστω το τ.δ. $\mathbf{x} = [x_1, x_2, x_3, x_4, x_5]^T$ που ακολουθεί την πολυδιάστατη κανονική κατανομή. Κάθε μια από τις τ.μ. του τ.δ. ακολουθεί την τυπική κανονική κατανομή $x_i \sim \mathcal{N}(0, 1)$, $i = 1, ..., 5$. Ο συνδιασπορά των τ.μ. x_i, x_j είναι $2 \min\{i, j\}/(i + j)$. Υπολογίστε την από κοινού Σ.Π.Π. $f_{x_1,x_3}()$ και την υπό συνθήκη $f_{x_1,x_3|x_4}()$

7. Υπολογίστε την Σ.Π.Π., αναμενόμενη τιμή και συνδιασπορά του τ.δ. $\mathbf{y} = [y_1, y_2, y_3, y_4, y_5]^T$, όπου $y_i = \sum_{j=1}^{i} x_j$, $i = 1, ..., 5$ (όπου το τ.δ. $\mathbf{x}$ δίδεται στην προηγούμενη άσκηση).

8. Υπολογίστε την υπό συνθήκη Σ.Π.Π. $f_{y_1,y_2|y_3,y_4}()$ για την δεδομένα της προηγούμενης άσκησης.

9. Υπολογίστε την γενική μορφή την αναμενόμενης τιμής και διασποράς τ.δ. $f_{\mathbf{x_1}|\mathbf{x_2}}()$ για τ.δ. $\mathbf{x_1}$, $\mathbf{x_2}$, εάν η $f_{\mathbf{x_1},\mathbf{x_2}}()$ είναι κανονική.

10. Δείξτε ότι $\sigma^2_{x_1|x_2} \leq \sigma^2_{x_1}$ όπου x_1, x_2 κανονικές τ.μ. Ισχύει το παραπάνω όταν οι τ.μ. x_1, x_2 δεν ακολουθούν κανονική κατανομή;

11. Δημιουργήστε μια γεννήτρια τ.δ. $\mathbf{y}$ που ακολουθούν την Σ.Π.Π. του Παραδείγματος 2 και υπολογίστε το διάνυσμα αναμενόμενης τιμής και πίνακα συνδιασποράς αριθμητικά.

12. Έστω ότι στο παράδειγμα 2 δημιουργούσαμε επίσης τις τ.μ. $y_4 = x_4$, $y_5 = x_3 - x_4$. Ακολουθεί το τ.δ. $[y_1, y_2, y_3, y_4, y_5]^T$ την πολυδιάστατη κανονική κατανομή; Ποια είναι η Σ.Π.Π.;