

2

Στοιχεία πιθανοτήτων και στοχαστικών διαδικασιών

2.1 Εισαγωγικά

Το παρόν παράρτημα δεν έχει σαν στόχο να καλύψει αναλυτικά την ύλη της Θεωρίας Πιθανοτήτων και Στοχαστικών Διαδικασιών. Υπάρχει στη βιβλιογραφία πληθώρα σχετικών συγγραμμάτων όπου οι δύο περιοχές αναπτύσσονται εκτεταμένα και με διαφορετικό βαθμό δυσκολίας και έμφασης¹. Έτσι, περιοριζόμαστε σε συνοπτική παρουσίαση των βασικών εννοιών και αποτελεσμάτων που είναι απαραίτητα για την κατανόηση του αντικειμένου του παρόντος βιβλίου.

2.2 Χώρος πιθανότητας

Καλούμε *χώρο πιθανότητας* μια τριάδα οντοτήτων, που θα συμβολίσουμε με (Θ, \mathcal{G}, P) , με τις παρακάτω ιδιότητες

- Το Θ είναι ένα οποιοδήποτε σύνολο στοιχείων, το οποίο καλείται *δειγματοχώρος*.
- Το \mathcal{G} είναι ένα σύνολο από υποσύνολα του Θ , τα οποία συνιστούν μια σ -άλγεβρα. Συγκεκριμένα, τα στοιχεία του \mathcal{G} πρέπει να ικανοποιούν τις ακόλουθες ιδιότητες:
 - Τα δύο σύνολα Θ και \emptyset είναι στοιχεία του \mathcal{G} .
 - Εάν $A \in \mathcal{G}$, τότε και $A^c \in \mathcal{G}$, όπου A^c το συμπληρωματικό του A .
 - Εάν $A_1, A_2 \in \mathcal{G}$, τότε και $A_1 \cup A_2 \in \mathcal{G}$.
 - Εάν μια ακολουθία από σύνολα $A_1, A_2, \dots \in \mathcal{G}$, τότε και $\bigcup_{i=1}^{\infty} A_i \in \mathcal{G}$.

Τα στοιχεία του \mathcal{G} καλούνται *γεγονότα*.

¹Σαν καταλληλότερο σύγγραμμα για Μηχανικούς προτείνεται το βιβλίο [PA1999] το οποίο διακρίνεται για τη μαθηματική του αυστηρότητα παράλληλα με την απλότητα παρουσίασης της ύλης.

- Το P είναι μια απεικόνιση από το σύνολο \mathcal{G} στο διάστημα $[0,1]$ με τις εξής ιδιότητες:
 - $P(\emptyset) = 0, P(\Theta) = 1$.
 - Εάν $A_1, A_2 \in \mathcal{G}$ και $A_1 \cap A_2 = \emptyset$, τότε $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.
 - Εάν μια ακολουθία από σύνολα $A_1, A_2, \dots \in \mathcal{G}$ και κάθε $A_i \cap A_j = \emptyset$ για $i \neq j$, τότε

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

Η συνάρτηση P καλείται *συνάρτηση πιθανότητας* και, όπως παρατηρούμε, ορίζει πιθανότητες μόνο για τα στοιχεία του συνόλου \mathcal{G} , δηλαδή τα γεγονότα.

2.3 Δεσμευμένη ή υπό συνθήκη πιθανότητα

Η έννοια της δεσμευμένης πιθανότητας αποτελεί σημαντική ανακάλυψη για τη θεωρία πιθανοτήτων. Έστω γεγονός $B \in \mathcal{G}$ με $P(B) > 0$ τότε θα καλούμε δεσμευμένη (ή υπό συνθήκη) πιθανότητα ενός γεγονότος $A \in \mathcal{G}$ με δεδομένο το B την εξής ποσότητα:

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

όπου για ευκολία συμβολίζουμε την τομή $A \cap B$ των δύο συνόλων σαν το “γινόμενο” $A \cdot B$. Από τον ορισμό της δεσμευμένης πιθανότητας προκύπτει ότι

$$P(A \cdot B) = P(A|B)P(B) = P(B|A)P(A). \quad (2.1)$$

Είναι επίσης εύκολο, χρησιμοποιώντας τον ορισμό, να γενικεύσουμε τις προηγούμενες ισότητες και να δείξουμε για τρία γεγονότα A, B, C ότι μπορούμε να γράψουμε

$$P(A \cdot B|C) = P(A|B \cdot C)P(B|C) = P(B|A \cdot C)P(A|C). \quad (2.2)$$

Πράγματι

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

αποδεικνύοντας την πρώτη ισότητα. Με τον ίδιο τρόπο αποδεικνύεται και η δεύτερη.

2.4 Τυχαίες μεταβλητές

Είναι δυνατό να ορίσουμε συναρτήσεις που να απεικονίζουν στοιχεία του δειγματοχώρου Θ στους πραγματικούς αριθμούς \mathbb{R} . Έστω $\chi(\theta)$ μια τέτοια συνάρτηση, δηλαδή $\theta \in \Theta$ και $\chi(\theta) \in \mathbb{R}$. Η πλέον στοιχειώδης πράξη στους πραγματικούς αριθμούς είναι η πράξη

της σύγκρισης. Εάν επομένως $x \in \mathbb{R}$, μας ενδιαφέρει να διαπιστώσουμε πόσο συχνά συμβαίνει $\chi(\theta) \leq x$, με άλλα λόγια να ανακαλύψουμε το σύνολο (υποσύνολο του Θ)

$$A_x = \{\theta : \chi(\theta) \leq x\},$$

το οποίο επιθυμούμε να “μετρήσουμε” με τη βοήθεια της συνάρτησης πιθανότητας P . Για να μπορέσουμε να δώσουμε πιθανότητα στο εν λόγω σύνολο είναι απαραίτητο το $A_x \in \mathcal{G}$, αφού το \mathcal{G} , εξ ορισμού, περιέχει όλα τα δυνατά σύνολα στα οποία μπορούμε να δώσουμε πιθανότητα. Έχουμε επομένως τον ακόλουθο ορισμό.

Μια συνάρτηση $\chi(\theta)$ από το Θ στους πραγματικούς αριθμούς, θα καλείται *μετρήσιμη ή τυχαία μεταβλητή*, εάν για κάθε πραγματικό x το σύνολο

$$A_x = \{\theta : \chi(\theta) \leq x\} \in \mathcal{G},$$

δηλαδή το A_x είναι ένα γεγονός.

Για μια τυχαία μεταβλητή $\chi(\theta)$ μια πολύ σημαντική ποσότητα είναι η συνάρτηση

$$F_\chi(x) = P(\chi(\theta) \leq x),$$

η οποία καλείται *συνάρτηση κατανομής* της $\chi(\theta)$ και είναι αύξουσα ως προς x με ιδιότητες

$$F_\chi(-\infty) = 0, \quad F_\chi(\infty) = 1.$$

Η παράγωγος της συνάρτησης κατανομής (όταν υπάρχει) καλείται *συνάρτηση πυκνότητας πιθανότητας* και ικανοποιεί

$$f_\chi(x) = \frac{dF_\chi(x)}{dx} \geq 0, \quad \int_{-\infty}^{\infty} f_\chi(x) dx = 1.$$

Η πυκνότητα πιθανότητας δεν εκφράζει πιθανότητα για κανένα γεγονός. Παρατηρούμε ωστόσο ότι μπορούμε να γράψουμε

$$P(x < \chi \leq x + dx) = F_\chi(x + dx) - F_\chi(x) = f_\chi(x) dx. \quad (2.3)$$

Με άλλα λόγια η πυκνότητα πιθανότητας $f_\chi(x)$ επί το διαφορικό dx εκφράζει ουσιαστικά την πιθανότητα η τυχαία μας μεταβλητή χ να πάρει τιμή μέσα στο διαφορικό διάστημα $(x, x + dx]$, που αποτελεί φυσικά έναν έμμεσο τρόπο να δηλώσουμε ότι η χ παίρνει την τιμή x . **Να μιλήσω για δέλτα συναρτήσεις!!!!**

2.4.1 Πείραμα

Με τις τυχαίες μεταβλητές μοντελοποιούμε φαινόμενα τα οποία είναι δύσκολο να περιγράψουμε με ντετερμινιστικό τρόπο, είτε διότι είναι εξαιρετικά πολύπλοκα, είτε διότι δεν υπάρχει η απαραίτητη πληροφορία.

Θα επιχειρήσουμε να δώσουμε στις τυχαίες μεταβλητές, πέρα από το μαθηματικό ορισμό, κάποια φυσική σημασία, η οποία να είναι σύμφωνη με τον τρόπο που οι οντότητες αυτές χρησιμοποιούνται στην πράξη. Όπως είδαμε, μια τυχαία μεταβλητή $\chi(\theta)$ είναι ουσιαστικά μια συνάρτηση από τον δειγματοχώρο στους πραγματικούς. Υπάρχει επομένως μια διαδικασία επιλογής στοιχείων του δειγματοχώρου και απεικόνισής τους στους πραγματικούς. Η διαδικασία αυτή καλείται *πείραμα* και ο πραγματικός αριθμός $\chi(\theta)$ που προκύπτει καλείται *υλοποίηση* της τυχαίας μεταβλητής.

Στα περισσότερα πρακτικά προβλήματα για τα πειράματα και τις υλοποιήσεις θεωρείται ότι ευθύνεται η “Φύση” ή η “Τυχαιότητα”, αφού ο Μελετητής δεν έχει συνήθως κανένα έλεγχο. Επιπλέον, ο Μελετητής είναι δυνατό να μην γνωρίζει τον δειγματοχώρο αλλά ούτε και τη συνάρτηση $\chi(\cdot)$. Π.χ. στη διαδικασία ρίψης ενός ζαριού η Φύση επιλέγει τις συνθήκες κάτω από τις οποίες εκτελείται η ρίψη και το αποτέλεσμα είναι ένας ακέραιος από ένα έως έξι. Στο παράδειγμα αυτό παρατηρούμε ότι είναι άγνωστος ο δειγματοχώρος καθώς και ο τρόπος αντιστοίχισης με τους πραγματικούς αριθμούς (που στο συγκεκριμένο παράδειγμα είναι μόνο το σύνολο $\{1, 2, \dots, 6\}$).

Προφανώς η Φύση μπορεί να επαναλάβει το ίδιο πείραμα πολλές φορές (π.χ. τη ρίψη ζαριού) και κάθε φορά να επιλέγει διαφορετικό στοιχείο του δειγματοχώρου, το οποίο απεικονίζεται σε διαφορετικό πραγματικό αριθμό. Αποτελέσματα πειραμάτων, δηλαδή διαφορετικές υλοποιήσεις, θα τα συμβολίζουμε με $\chi(\theta_1), \chi(\theta_2), \dots$, ώστε να γίνεται σαφής η διαφορετική επιλογή της Φύσης στα στοιχεία θ του δειγματοχώρου.

2.4.2 Μέσος όρος και διασπορά

Καλούμε *στοχαστικό μέσον όρο* της τυχαίας μεταβλητής χ το ολοκλήρωμα²

$$\bar{\chi} = E[\chi] = \int_{-\infty}^{\infty} x f_{\chi}(x) dx.$$

Ο στοχαστικός μέσος όρος, αποτελεί μια αντιπροσωπευτική ντετερμινιστική τιμή της τυχαίας μεταβλητής (συνάρτησης) χ .

Ο παραπάνω ορισμός του μέσου όρου προϋποθέτει γνώση της συνάρτησης κατανομής της τυχαίας μεταβλητής. Ένας άλλος τρόπος υπολογισμού, ή ακριβέστερα εκτίμησης, του μέσου όρου είναι ο *αριθμητικός μέσος όρος*

$$\bar{\chi} \approx \frac{\chi(\theta_1) + \chi(\theta_2) + \dots + \chi(\theta_n)}{n},$$

ο οποίος απαιτεί πολλαπλές υλοποιήσεις της τυχαίας μεταβλητής.

Ένας άμεσος τρόπος δημιουργίας τυχαίων μεταβλητών είναι μέσω ντετερμινιστικών μετασχηματισμών. Έστω επομένως ότι $G(x)$ είναι μ

² Στο εξής η εξάρτηση από την μεταβλητή θ θα προσδιορίζεται εφόσον είναι απολύτως αναγκαίο.

και διασπορά

$$\sigma_{\chi}^2 = E[(\chi - \bar{\chi})^2] = \int_{-\infty}^{\infty} (x - \bar{\chi})^2 f_{\chi}(x) dx.$$

, ενώ η διασπορά υποδηλώνει το πόσο “παίζει” η συνάρτηση χ γύρω από την αντιπροσωπευτική της τιμή $\bar{\chi}$. Παρατηρούμε ότι, όταν $\sigma_{\chi} = 0$, τότε η τυχαία μεταβλητή είναι μια σταθερή συνάρτηση (ίση προς τη μέση της τιμή $\bar{\chi}$).

Οι έννοιες που ορίσαμε για μια τυχαία μεταβλητή εύκολα επεκτείνονται και σε περισσότερες. Εάν χ_1, χ_2 δύο τυχαίες μεταβλητές (δηλαδή για κάθε επιλογή του θ μας διατίθενται δύο πραγματικοί αριθμοί), τότε είναι δυνατό να ορίσουμε την *από κοινού συνάρτηση κατανομής*

$$F_{\chi_1, \chi_2}(x_1, x_2) = P(\chi_1 \leq x_1, \chi_2 \leq x_2),$$

την πιθανότητα δηλαδή να έχουμε συγχρόνως $\chi_1 \leq x_1$ και $\chi_2 \leq x_2$. Είναι πολύ εύκολο να διαπιστώσουμε ότι εάν χ_i , $i = 1, 2$, είναι μετρήσιμες, τότε η εν λόγω πιθανότητα υπάρχει (γιατί).

Η μερική παράγωγος της (από κοινού) συνάρτησης κατανομής $F_{\chi_1, \chi_2}(x_1, x_2)$ ως προς x_1 και x_2

$$f_{\chi_1, \chi_2}(x_1, x_2) = \frac{\partial^2 F_{\chi_1, \chi_2}(x_1, x_2)}{\partial x_1 \partial x_2}$$

καλείται *από κοινού συνάρτηση πυκνότητας πιθανότητας*. Για τις δύο συναρτήσεις ισχύουν οι εξής ιδιότητες

$$F_{\chi_1}(x_1) = F_{\chi_1, \chi_2}(x_1, \infty), \quad F_{\chi_2}(x_2) = F_{\chi_1, \chi_2}(\infty, x_2),$$

$$f_{\chi_1, \chi_2}(x_1, x_2) \geq 0$$

$$f_{\chi_1}(x_1) = \int_{-\infty}^{\infty} f_{\chi_1, \chi_2}(x_1, x_2) dx_2, \quad f_{\chi_2}(x_2) = \int_{-\infty}^{\infty} f_{\chi_1, \chi_2}(x_1, x_2) dx_1$$

$$\iint_{-\infty}^{\infty} f_{\chi_1, \chi_2}(x_1, x_2) dx_1 dx_2 = 1.$$

Δύο τυχαίες μεταβλητές χ_1, χ_2 θα καλούνται *ανεξάρτητες* όταν

$$F_{\chi_1, \chi_2}(x_1, x_2) = F_{\chi_1}(x_1)F_{\chi_2}(x_2) \quad \text{ή} \quad f_{\chi_1, \chi_2}(x_1, x_2) = f_{\chi_1}(x_1)f_{\chi_2}(x_2).$$

Καλούμε *συσχέτιση* δύο τυχαίων μεταβλητών χ_1, χ_2 την ποσότητα

$$\begin{aligned} \text{cov}\{\chi_1, \chi_2\} &= E[(\chi_1 - \bar{\chi}_1)(\chi_2 - \bar{\chi}_2)] \\ &= \iint (x_1 - \bar{\chi}_1)(x_2 - \bar{\chi}_2) f_{\chi_1, \chi_2}(x_1, x_2) dx_1 dx_2 \\ &= E[\chi_1 \chi_2] - \bar{\chi}_1 \bar{\chi}_2. \end{aligned}$$

Όταν η συσχέτιση δύο τυχαίων μεταβλητών είναι μηδέν, τότε οι τυχαίες μεταβλητές καλούνται *ασυσχέτιστες*.

Οι παραπάνω ορισμοί επεκτείνονται φυσικά σε περισσότερες από δύο τυχαίες μεταβλητές κατά τον προφανή τρόπο. Στην περίπτωση των περισσότερων της μιας τυχαίων μεταβλητών είναι προτιμότερο να θεωρούμε τις τυχαίες μεταβλητές σαν όρους ενός (τυχαίου) διανύσματος. Για την περίπτωση επομένως K τυχαίων μεταβλητών μπορούμε να γράψουμε

$$\mathcal{X} = [\chi_1 \ \chi_2 \ \cdots \ \chi_K]^t$$

και να ορίσουμε το μέσο διάνυσμα σαν

$$\bar{\mathcal{X}} = \mathbb{E}[\mathcal{X}] = \int X f_{\mathcal{X}}(X) dX$$

και την μήτρα συνδιασποράς

$$\Sigma_{\mathcal{X}} = \mathbb{E}[(\mathcal{X} - \bar{\mathcal{X}})(\mathcal{X} - \bar{\mathcal{X}})^t] = \mathbb{E}[\mathcal{X} \mathcal{X}^t] - \bar{\mathcal{X}} \bar{\mathcal{X}}^t.$$

Από τον ορισμό εύκολα διαπιστώνουμε ότι το στοιχείο i, j της μήτρας συνδιασποράς είναι ίσο προς τη συσχέτιση των τυχαίων μεταβλητών χ_i, χ_j , ως εκ τούτου η μήτρα $\Sigma_{\mathcal{X}}$ είναι *συμμετρική*. Το i -οστό διαγώνιο στοιχείο της μήτρας είναι ίσο προς τη διασπορά της τυχαίας μεταβλητής χ_i . Τέλος η μήτρα συνδιασποράς, μπορεί εύκολα να αποδειχθεί ότι είναι *μη αρνητικά ορισμένη*, μια σημαντική και πολύ χρήσιμη ιδιότητα.

Ένα κλασικό και πρακτικά χρήσιμο παράδειγμα από κοινού συνάρτησης πυκνότητας πιθανότητας αποτελεί η περίπτωση K Gaussian τυχαίων μεταβλητών. Εάν $\bar{\mathcal{X}}, \Sigma_{\mathcal{X}}$ το διάνυσμα των μέσων όρων και η μήτρα συνδιασποράς των εν λόγω μεταβλητών τότε

$$f_{\mathcal{X}}(X) = \frac{1}{\sqrt{(2\pi)^K |\Sigma_{\mathcal{X}}|}} e^{-\frac{1}{2}(X - \bar{\mathcal{X}})^t \Sigma_{\mathcal{X}}^{-1} (X - \bar{\mathcal{X}})},$$

όπου $X = [x_1 \ x_2 \ \cdots \ x_K]^t$ (διάνυσμα πραγματικών τυχαίων μεταβλητών) και $|\Sigma_{\mathcal{X}}|$ συμβολίζει την ορίζουσα της μήτρας $\Sigma_{\mathcal{X}}$.

Ιδιότητες Gaussian μεταβλητών: Οι Gaussian μεταβλητές διαθέτουν τις ακόλουθες δύο πολύ σημαντικές ιδιότητες

- Αποτελεί ενδιαφέρουσα (και απλή) άσκηση η απόδειξη της πρότασης ότι όταν Gaussian τυχαίες μεταβλητές είναι ασυσχέτιστες είναι υποχρεωτικά και ανεξάρτητες.
- Αποτελεί ενδιαφέρουσα (και όχι ιδιαίτερα δύσκολη) άσκηση η απόδειξη της πρότασης ότι γραμμικός συνδυασμός από Gaussian τυχαίες μεταβλητές δημιουργεί πάλι Gaussian τυχαίες μεταβλητές.

Η δεύτερη ιδιότητα είναι εξαιρετικά χρήσιμη επειδή, ως γνωστόν, για να καθοριστούν οι Gaussian τυχαίες μεταβλητές αρκεί να υπολογιστούν οι μέσοι όροι και η μήτρα συνδιασποράς, πράγμα απλό για την περίπτωση των γραμμικών συνδυασμών.

2.4.3 Δεσμευμένη πυκνότητα πιθανότητας

Εάν έχουμε μια τυχαία μεταβλητή $\chi(\theta)$ με πυκνότητα πιθανότητας $f_\chi(x)$, τότε μπορούμε στην περίπτωση αυτή να ορίσουμε τη δεσμευμένη πυκνότητα πιθανότητας $f_\chi(x|\chi \in B)$. Με άλλα λόγια ενδιαφερόμαστε να δούμε με ποιο τρόπο αλλάζει η πυκνότητα πιθανότητας της τυχαίας μεταβλητής όταν μας δίνεται η επιπλέον πληροφορία ότι η τυχαία μεταβλητή παρατηρήθηκε στο εσωτερικό ενός συνόλου B .

Προκειμένου να χρησιμοποιήσουμε τον ορισμό της δεσμευμένης πιθανότητας, κάνουμε χρήση της (2.3), συγκεκριμένα

$$\begin{aligned} f_\chi(x|\chi \in B)dx &= P(x < \chi \leq x + dx | \chi \in B) = \frac{P(x < \chi \leq x + dx \ \& \ \chi \in B)}{P(\chi \in B)} \\ &= \frac{f_\chi(x)\mathbb{1}_B(x) dx}{\int_{x \in B} f_\chi(x) dx}, \end{aligned}$$

από τη οποία συμπεραίνουμε ότι

$$f_\chi(x|\chi \in B) = \frac{f_\chi(x)}{\int_{x \in B} f_\chi(x) dx} \mathbb{1}_B(x).$$

2.4.4 Βασικές ιδιότητες για γεγονότα

Με τη βοήθεια της δεσμευμένης πιθανότητας και συγκεκριμένα με χρήση της (2.1) είναι δυνατό να αποδειχθεί ένας αριθμός από πολύ ενδιαφέρουσες ιδιότητες οι οποίες παρατίθενται στη συνέχεια.

Ισότητα 1: Αμεση γενίκευση της (2.1) αποτελεί η εξής περίπτωση: έστω γεγονότα $A_1, A_2, \dots, A_K \in \mathcal{G}$, τότε

$$\begin{aligned} P(A_K \cdot A_{K-1} \cdots A_1) &= P(A_K | A_{K-1} \cdots A_1) P(A_{K-1} \cdots A_1) = \cdots \\ &= P(A_K | A_{K-1} \cdots A_1) P(A_{K-1} | A_{K-2} \cdots A_1) \cdots P(A_2 | A_1) P(A_1). \end{aligned}$$

Η πρώτη ισότητα είναι ουσιαστικά η (2.1) με $B = A_{K-1} \cdots A_1$. Στη συνέχεια επαναλαμβάνεται η ισότητα αυτή για $K-1, K-2, \dots, 2$.

Ισότητα 2: (Ολική Πιθανότητα) Έστω γεγονότα $A_1, A_2, \dots, A_K \in \mathcal{G}$ για τα οποία ισχύει $A_1 \cup A_2 \cup \cdots \cup A_K = \Theta$ με $A_i \cdot A_j = \emptyset$ για $i \neq j$, καθώς και γεγονός $B \in \mathcal{G}$, τότε

$$\begin{aligned} P(B) &= P(B \cdot \Theta) = P(B \cdot \cup_{i=1}^K A_i) = P(\cup_{i=1}^K (B \cdot A_i)) = \sum_{i=1}^K P(B \cdot A_i) \\ &= \sum_{i=1}^K P(B | A_i) P(A_i), \end{aligned}$$

όπου η πρώτη ισότητα της τελευταίας σχέσης προκύπτει από το γεγονός ότι τα σύνολα $B \cdot A_i$ είναι μεταξύ τους ξένα.

Ισότητα 3: Με χρήση της προηγούμενης ισότητας μπορούμε να δείξουμε για γεγονότα A_1, \dots, A_K, B όπως παραπάνω ότι ισχύει

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

Οι πιθανότητες $P(A_i)$ καλούνται *εκ των προτέρων* (ή αρχικές) πιθανότητες των γεγονότων A_i , ενώ οι $P(A_i|B)$ *εκ των υστέρων* με δεδομένο το γεγονός B . Οι εκ των προτέρων πιθανότητες εκφράζουν την αρχική γνώση που υπάρχει για τα γεγονότα A_i ενώ οι εκ των υστέρων το πως διαμορφώνονται οι πιθανότητες *μετά* την εμφάνιση του γεγονότος B .

2.4.5 Βασικές ισότητες για πυκνότητες πιθανότητας

Οι ισότητες που παρουσιάστηκαν για γεγονότα έχουν τα ισοδύναμά τους και στην περίπτωση των πυκνοτήτων πιθανότητας τυχαίων μεταβλητών. Έστω δύο τυχαία διανύσματα \mathcal{X}, \mathcal{Y} με αντίστοιχη από κοινού πυκνότητα πιθανότητας $f_{\mathcal{X}, \mathcal{Y}}(X, Y)$, τότε η πυκνότητα πιθανότητας του \mathcal{X} με δεδομένο ότι $\mathcal{Y} \in B$ είναι

$$f_{\mathcal{X}|\mathcal{Y} \in B}(X|\mathcal{Y} \in B) = \frac{\int_{Y \in B} f_{\mathcal{X}, \mathcal{Y}}(X, Y) dY}{\int_{Y \in B} f_{\mathcal{Y}}(Y) dY} = \frac{\int_{Y \in B} f_{\mathcal{X}, \mathcal{Y}}(X, Y) dY}{\int_{-\infty}^{\infty} \int_{Y \in B} f_{\mathcal{X}, \mathcal{Y}}(X, Y) dY dX}.$$

Είναι επίσης δυνατό να θεωρήσουμε για το \mathcal{Y} το διαφορικό γεγονός $\mathcal{Y} = Y$ (δηλαδή $Y < \mathcal{Y} \leq Y + dY$), οπότε η δεσμευμένη πυκνότητα πιθανότητας γράφεται

$$f_{\mathcal{X}|\mathcal{Y}=Y}(X|Y) = \frac{f_{\mathcal{X}, \mathcal{Y}}(X, Y)}{f_{\mathcal{Y}}(Y)} = \frac{f_{\mathcal{X}, \mathcal{Y}}(X, Y)}{\int_{-\infty}^{\infty} f_{\mathcal{X}, \mathcal{Y}}(X, Y) dX}.$$

Εύκολα διαπιστώνουμε ότι η συνάρτηση $f_{\mathcal{X}|\mathcal{Y}=Y}(X|Y)$ αποτελεί όντως πυκνότητα πιθανότητας αφού είναι μη αρνητική και εάν ολοκληρωθεί ως προς X το αποτέλεσμα είναι μονάδα. Η παραπάνω σχέση αποτελεί το ισοδύναμο της δεσμευμένης πιθανότητας για συναρτήσεις πυκνότητας πιθανότητας. Κατ' αναλογία με την (2.1) μπορούμε επίσης να γράψουμε

$$f_{\mathcal{X}, \mathcal{Y}}(X, Y) = f_{\mathcal{X}|\mathcal{Y}=Y}(X|Y)f_{\mathcal{Y}}(Y).$$

Από την ύπαρξη της υπό συνθήκη πυκνότητας πιθανότητας απορρέει και η ύπαρξη της *υπό συνθήκη μέσης τιμής*

$$E[\mathcal{X}|\mathcal{Y} = Y] = \int_{-\infty}^{\infty} X f_{\mathcal{X}|\mathcal{Y}=Y}(X|Y) dX = G(Y)$$

το οποίο είναι φυσικά μια (διανυσματική) συνάρτηση του Y .

Για την υπό συνθήκη πυκνότητα πιθανότητα ισχύει ένας αριθμός από ενδιαφέρουσες ιδιότητες η απόδειξη των οποίων είναι απλή και επαφίεται στον αναγνώστη. Το ισοδύναμο της Ισότητας 1 για πυκνότητες πιθανότητας είναι η ακόλουθη σχέση.

Ισότητα 4: Έστω τυχαίες μεταβλητές $\chi_1, \chi_2, \dots, \chi_n$ τότε

$$\begin{aligned} f_{\chi_n, \dots, \chi_1}(x_n, \dots, x_1) = \\ f_{\chi_n | \chi_{n-1}, \dots, \chi_1}(x_n | x_{n-1}, \dots, x_1) \times f_{\chi_{n-1} | \chi_{n-2}, \dots, \chi_1}(x_{n-1} | x_{n-2}, \dots, x_1) \times \dots \\ \times f_{\chi_2 | \chi_1}(x_2 | x_1) \times f_{\chi_1}(x_1). \end{aligned}$$

Για την ειδική περίπτωση που

$$f_{\chi_n | \chi_{n-1}, \dots, \chi_1}(x_n | x_{n-1}, \dots, x_1) = f_{\chi_n | \chi_{n-1}}(x_n | x_{n-1})$$

τότε η ακολουθία $\chi_1, \chi_2, \dots, \chi_n$ καλείται *Markov*. Συνδυασμός γεγονότων και πυκνοτήτων πιθανότητας οδηγεί στην ακόλουθη ισότητα.

Ισότητα 5: Έστω τυχαία διανύσματα \mathcal{X}, \mathcal{Y} και ας υποθέσουμε ότι το \mathcal{Y} παίρνει τιμές μέσα στο σύνολο Ω όπου υποθέτουμε ότι $P(\mathcal{Y} \in \Omega) = 1$. Έστω επίσης ότι ισχύει $\Omega = \cup_{i=1}^K A_i$, όπου τα γεγονότα A_i είναι ανά δύο ξένα μεταξύ τους. Τότε

$$\begin{aligned} f_{\mathcal{X}}(X) &= f_{\mathcal{X}, \mathcal{Y}}(X, \mathcal{Y} \in \Omega) = f_{\mathcal{X}, \mathcal{Y}}(X, \mathcal{Y} \in \cup_{i=1}^K A_i) = \sum_{i=1}^K f_{\mathcal{X}, \mathcal{Y}}(X, \mathcal{Y} \in A_i) \\ &= \sum_{i=1}^K f_{\mathcal{X} | \mathcal{Y}}(X | \mathcal{Y} \in A_i) P(\mathcal{Y} \in A_i). \end{aligned}$$

Ισότητα 6: Το αντίστοιχο της Ισότητας 3 με την βοήθεια της Ισότητας 5, γράφεται

$$\begin{aligned} P(\mathcal{Y} \in A_i | \mathcal{X} = X) &= \frac{f_{\mathcal{X} | \mathcal{Y}}(X | \mathcal{Y} \in A_i) P(\mathcal{Y} \in A_i)}{f_{\mathcal{X}}(X)} \\ &= \frac{f_{\mathcal{X} | \mathcal{Y}}(X | \mathcal{Y} \in A_i) P(\mathcal{Y} \in A_i)}{\sum_{i=1}^K f_{\mathcal{X} | \mathcal{Y}}(X | \mathcal{Y} \in A_i) P(\mathcal{Y} \in A_i)}, \end{aligned}$$

όπου ισχύουν οι υποθέσεις της προηγούμενης ισότητας. Η ισότητα αυτή χρησιμοποιείται για τον υπολογισμό της εκ των υστέρων πιθανότητας του A_i με δεδομένο ότι το τυχαίο διάνυσμα \mathcal{X} έλαβε τη τιμή $\mathcal{X} = X$.

2.4.6 Ιδιότητα της κλιμάκωσης του μέσου όρου

Μια εξαιρετικά χρήσιμη ιδιότητα η οποία βασίζεται στην υπό συνθήκη μέση τιμή είναι η εξής. Έστω $G(X, Y)$ συνάρτηση των διανυσμάτων X, Y , τότε μπορούμε να γράψουμε

$$\begin{aligned} \mathbb{E}[G(\mathcal{X}, \mathcal{Y})] &= \iint_{-\infty}^{\infty} G(X, Y) f_{\mathcal{X}, \mathcal{Y}}(X, Y) dX dY \\ &= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} G(X, Y) f_{\mathcal{X}|\mathcal{Y}}(X|Y) dX \right\} f_{\mathcal{Y}}(Y) dY \\ &= \mathbb{E}[\mathbb{E}[G(\mathcal{X}, \mathcal{Y})|\mathcal{Y}]]. \end{aligned} \quad (2.4)$$

Με άλλα λόγια ο μέσος όρος μιας τυχαίας ποσότητας είναι δυνατό να υπολογιστεί κλιμακωτά, υπολογίζοντας δηλαδή αρχικά τον υπό συνθήκη μέσος όρο ως προς κάποιες τυχαίες μεταβλητές και κατόπιν, τον μέσο όρο της τυχαίας ποσότητας που προκύπτει.

Παράδειγμα 2.1: Έστω δύο στοχαστικά διανύσματα \mathcal{X}, \mathcal{Y} τα οποία είναι από κοινού Gaussian με μέσες τιμές $\bar{\mathcal{X}}, \bar{\mathcal{Y}}$, μήτρες συνδυασποράς $\mathbb{E}[(\mathcal{X} - \bar{\mathcal{X}})(\mathcal{X} - \bar{\mathcal{X}})^t] = \Sigma_{\mathcal{X}}, \mathbb{E}[(\mathcal{Y} - \bar{\mathcal{Y}})(\mathcal{Y} - \bar{\mathcal{Y}})^t] = \Sigma_{\mathcal{Y}}$ και μήτρα ετεροσυσχέτισης $\mathbb{E}[(\mathcal{X} - \bar{\mathcal{X}})(\mathcal{Y} - \bar{\mathcal{Y}})^t] = \Sigma_{\mathcal{X}, \mathcal{Y}}$. Δείξτε ότι η δεσμευμένη πυκνότητα πιθανότητας $f_{\mathcal{X}|\mathcal{Y}}(X|Y)$ είναι Gaussian με μέση τιμή και διασπορά που δίνονται από τις σχέσεις

$$\begin{aligned} \mathbb{E}[\mathcal{X}|\mathcal{Y}] &= \bar{\mathcal{X}} + \Sigma_{\mathcal{X}, \mathcal{Y}} \Sigma_{\mathcal{Y}}^{-1} (\mathcal{Y} - \bar{\mathcal{Y}}) \\ \mathbb{E}[(\mathcal{X} - \mathbb{E}[\mathcal{X}|\mathcal{Y}])(\mathcal{X} - \mathbb{E}[\mathcal{X}|\mathcal{Y}])^t | \mathcal{Y}] &= \Sigma_{\mathcal{X}|\mathcal{Y}} = \Sigma_{\mathcal{X}} - \Sigma_{\mathcal{X}, \mathcal{Y}} \Sigma_{\mathcal{Y}}^{-1} \Sigma_{\mathcal{Y}, \mathcal{X}}. \end{aligned}$$

Για την απόδειξη της πρότασης αρκεί να υπολογίσουμε τη συνάρτηση

$$f_{\mathcal{X}|\mathcal{Y}}(X|Y) = f_{\mathcal{X}, \mathcal{Y}}(X, Y) / f_{\mathcal{Y}}(Y).$$

Για ευκολία θα θεωρήσουμε ότι οι δύο μέσοι όροι είναι μηδέν. Έχουμε τότε ότι (βλέπε Εδάφιο A.5)

$$\begin{aligned} f_{\mathcal{X}, \mathcal{Y}}(X, Y) &= \frac{1}{\sqrt{(2\pi)^{N_x + N_y} |\Sigma|}} e^{-\frac{1}{2} [X^t \ Y^t] \Sigma^{-1} [X^t \ Y^t]^t} \\ f_{\mathcal{Y}}(Y) &= \frac{1}{\sqrt{(2\pi)^{N_y} |\Sigma_{\mathcal{Y}}|}} e^{-\frac{1}{2} Y^t \Sigma_{\mathcal{Y}}^{-1} Y} \end{aligned}$$

όπου N_x, N_y τα μήκη των διανυσμάτων \mathcal{X}, \mathcal{Y} αντίστοιχα, $\Sigma = [\Sigma_{\mathcal{X}} \ \Sigma_{\mathcal{X}, \mathcal{Y}}; \Sigma_{\mathcal{Y}, \mathcal{X}}^t \ \Sigma_{\mathcal{Y}}]$ είναι η μήτρα συνδυασποράς του εννιαίου διανύσματος $[\mathcal{X}^t \ \mathcal{Y}^t]^t$ και $|A|$ συμβολίζει την ορίζουσα της μήτρας A . Χρησιμοποιώντας τις δύο αυτές σχέσεις υπολογίζεται ότι

$$f_{\mathcal{X}|\mathcal{Y}}(X|Y) = \frac{1}{\sqrt{(2\pi)^{N_x} \frac{|\Sigma|}{|\Sigma_{\mathcal{Y}}|}}} e^{-\frac{1}{2} [X^t \ Y^t] \Sigma^{-1} [X^t \ Y^t]^t + \frac{1}{2} Y^t \Sigma_{\mathcal{Y}}^{-1} Y}. \quad (2.5)$$

Με τη βοήθεια της ταυτότητας αντιστροφής του Shur

$$\begin{bmatrix} \Sigma_{\mathcal{X}} & \Sigma_{\mathcal{X}, \mathcal{Y}} \\ \Sigma_{\mathcal{Y}, \mathcal{X}}^t & \Sigma_{\mathcal{Y}} \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_{\mathcal{Y}}^{-1} \end{bmatrix} + \begin{bmatrix} \mathbf{I} \\ -\Phi_{\mathcal{X}, \mathcal{Y}}^t \end{bmatrix} \Sigma_{\mathcal{X}|\mathcal{Y}}^{-1} [\mathbf{I} - \Phi_{\mathcal{X}, \mathcal{Y}}]$$

όπου $\Phi_{\mathcal{X}, \mathcal{Y}} = \Sigma_{\mathcal{X}, \mathcal{Y}} \Sigma_{\mathcal{Y}}^{-1}$, και την ταυτότητα ορίζοντας μητρώων σε μπλοκ μορφή

$$\left\| \begin{bmatrix} \Sigma_{\mathcal{X}} & \Sigma_{\mathcal{X}, \mathcal{Y}} \\ \Sigma_{\mathcal{X}, \mathcal{Y}}^t & \Sigma_{\mathcal{Y}} \end{bmatrix} \right\| = |\Sigma_{\mathcal{Y}}| |\Sigma_{\mathcal{X}|\mathcal{Y}}|,$$

μετά από αντικατάσταση στη Σχέση (2.5), καταλήγουμε

$$f_{\mathcal{X}|\mathcal{Y}}(X|Y) = \frac{1}{\sqrt{(2\pi)^{N_x} |\Sigma_{\mathcal{X}|\mathcal{Y}}|}} e^{-\frac{1}{2}(X - \Phi_{\mathcal{X}|\mathcal{Y}} Y)^t \Sigma_{\mathcal{X}|\mathcal{Y}}^{-1} (X - \Phi_{\mathcal{X}|\mathcal{Y}} Y)}.$$

Από την προηγούμενη σχέση συμπεραίνουμε ότι η δεσμευμένη πιθανότητα της \mathcal{X} είναι όντως Gaussian με τον ζητούμενο μέσον όρο και μήτρα συνδιασπορά. ■

2.5 Ιδιότητα της αλλαγής μέτρου

Στη συνέχεια θα αναφερθούμε σε μια πολύ ενδιαφέρουσα ιδιότητα του μέσου όρου την οποία θα χρησιμοποιήσουμε σε επόμενο εδάφιο.

Έστω τυχαίο διάνυσμα \mathcal{X} το οποίο έχει πυκνότητα πιθανότητας $f_{\mathcal{X}}(X)$. Ας υποθέσουμε επίσης ότι $\tilde{f}_{\mathcal{X}}(X)$ αποτελεί εναλλακτική πυκνότητα πιθανότητας για το ίδιο τυχαίο διάνυσμα. Μπορούμε τώρα να ορίσουμε το λόγο πιθανοφάνειας $L(X) = \frac{f_{\mathcal{X}}(X)}{\tilde{f}_{\mathcal{X}}(X)}$ το οποίο είναι *βαθμωτή* ποσότητα και αποτελεί ένα (εν γένει μη γραμμικό) μετασχηματισμό του διανύσματος X . Έστω τέλος μη γραμμική συνάρτηση $G(X)$ για την οποία ενδιαφερόμαστε να υπολογίσουμε τον μέσον όρο $E[G(\mathcal{X})]$ όπου $E[\cdot]$ συμβολίζει μέσον όρο ως προς την πυκνότητα πιθανότητας $f_{\mathcal{X}}(X)$. Έχουμε τον εξής απλό υπολογισμό³

$$\begin{aligned} E[G(\mathcal{X})] &= \int G(X) f_{\mathcal{X}}(X) dX = \int G(X) \frac{f_{\mathcal{X}}(X)}{\tilde{f}_{\mathcal{X}}(X)} \tilde{f}_{\mathcal{X}}(X) dX \\ &= \int G(X) L(X) \tilde{f}_{\mathcal{X}}(X) dX \\ &= \tilde{E}[G(\mathcal{X}) L(\mathcal{X})], \end{aligned}$$

όπου $\tilde{E}[\cdot]$ εκφράζει μέσον όρο ως προς την εναλλακτική πυκνότητα πιθανότητας $\tilde{f}_{\mathcal{X}}(X)$. Παρατηρούμε ότι είναι δυνατό να υπολογίσουμε τον μέσον όρο μιας τυχαίας ποσότητας αλλάζοντας (μέτρο) πυκνότητα πιθανότητας, αρκεί να εφαρμόσουμε τη σχετική διόρθωση με τη βοήθεια του λόγου πιθανοφάνειας. Η προφανής αυτή ιδιότητα της αλλαγής μέτρου έχει πολλές και σημαντικές εφαρμογές στη Θεωρία Πιθανοτήτων και στη Στατιστική.

³ Στην ανάλυση που ακολουθεί έχουν παραληφθεί ορισμένες τεχνικές λεπτομέρειες. Π.χ. θεωρούμε ότι στα σημεία X για τα οποία $\tilde{f}_{\mathcal{X}}(X) = 0$, πρέπει να ισχύει ότι $f_{\mathcal{X}}(X) = 0$. Με τον περιορισμό αυτό αποφεύγεται ο λόγος πιθανοφάνειας να παίρνει άπειρη τιμή. Το γεγονός ότι η τιμή του λόγου πιθανοφάνειας είναι απροσδιόριστη δεν αποτελεί πρόβλημα αφού η συνεισφορά των σημείων αυτών στο συνολικό ολοκλήρωμα είναι μηδενική (γιατί;).