

## ***Δεσμευμένη πιθανότητα και Ανεξαρτησία ενδεχομένων***

- 4.1 Γιατί δεσμευμένη πιθανότητα – Το νόημα της δεσμευμένης πιθανότητας*
- 4.2 Ο πολλαπλασιαστικός τύπος*
- 4.3 Το θεώρημα ολικής πιθανότητας*
- 4.4 Το θεώρημα Bayes*
- 4.5 Ανεξαρτησία ενδεχομένων*
- 4.6 Σύντομη ανασκόπηση βασικών εννοιών, προτάσεων και τύπων*
- 4.7 Προβλήματα και ασκήσεις*



#### 4.1 Γιατί δεσμευμένη πιθανότητα – Το νόημα της δεσμευμένης πιθανότητας

Η πιθανότητα, ως ένα μέτρο του βαθμού βεβαιότητας που έχουμε για την εμφάνιση ενός ενδεχομένου, είναι δυνατόν να αναθεωρηθεί και να προσαρμοσθεί κατάλληλα, αν σε κάποιο στάδιο του στοχαστικού πειράματος που μελετάμε προκύψουν πρόσθετες πληροφορίες για την έκβασή του. Μια τέτοια πληροφορία, μπορεί μάλιστα να είναι ότι κάποιο άλλο ενδεχόμενο **ήδη έχει εμφανισθεί**. Χαρακτηριστικό είναι το ακόλουθο παράδειγμα.

**Παράδειγμα 4.1.1:** Στα συμπεράσματα μιας έρευνας που πρόσφατα ολοκληρώθηκε αναφέρεται, μεταξύ άλλων, ότι το ποσοστό των φορέων του βακίλου της φυματίωσης σε μια συγκεκριμένη περιοχή είναι 0.75%. Αυτό σημαίνει ότι αν επιλέξουμε (τυχαία) από την περιοχή αυτή έναν κάτοικο, η πιθανότητα να είναι φορέας του βακίλου της φυματίωσης είναι 0.0075. Αν όμως ο κάτοικος υποβληθεί σε κάποιο σχετικό διαγνωστικό ιατρικό τεστ, για παράδειγμα στο TB tine τεστ, τότε **όταν το αποτέλεσμα του τεστ γίνει γνωστό**, η πιθανότητα να είναι φορέας προφανώς αναθεωρείται και μάλιστα περιμένουμε να αυξηθεί αν το αποτέλεσμα του τεστ είναι θετικό και να μειωθεί αν το αποτέλεσμα είναι αρνητικό. Διευκρινίζουμε ότι ακόμη και μετά το αποτέλεσμα του TB tine τεστ δεν είμαστε βέβαιοι για το αν ο εξεταζόμενος είναι ή δεν είναι φορέας του βακίλου της φυματίωσης γιατί το τεστ αυτό δεν κάνει πάντοτε σωστή διάγνωση<sup>1</sup>. Συγκεκριμένα, όταν το εξεταζόμενο άτομο είναι φορέας το TB tine τεστ κάνει σωστή διάγνωση με πιθανότητα 0.92 ενώ όταν δεν είναι φορέας κάνει σωστή διάγνωση με πιθανότητα 0.96.

Ας θεωρήσουμε τα ενδεχόμενα

$B$ : ο κάτοικος είναι φορέας του βακίλου της φυματίωσης

$\Theta$ : το αποτέλεσμα του τεστ είναι θετικό.

Δίνεται ότι  $P(B) = 0.0075$ , δηλαδή, δίνεται ότι η πιθανότητα να είναι φορέας του βακίλου της φυματίωσης ένας κάτοικος που επιλέξαμε τυχαία από τη συγκεκριμένη περιοχή, είναι 0.0075.

Άραγε πώς επηρεάζεται αυτή η πιθανότητα αν ο κάτοικος υποβληθεί στο TB tine τεστ και το αποτέλεσμα είναι θετικό; Δηλαδή, ποια είναι πλέον η πιθανότητα να πραγματοποιηθεί το ενδεχόμενο  $B$  δοθέντος ότι πραγματοποιήθηκε το ενδεχόμενο  $\Theta$ ; Αυτή η **εκ των υστέρων** πιθανότητα του  $B$ , δηλαδή, η πιθανότητα πραγματοποίησης του ενδεχομένου  $B$  **μετά** την πληροφορία ότι το αποτέλεσμα του τεστ είναι θετικό, ονομάζεται **δεσμευμένη πιθανότητα του  $B$  δοθέντος του  $\Theta$**  και συμβολίζεται με  $P(B|\Theta)$ . Αντίστοιχα, αν το αποτέλεσμα του τεστ είναι αρνητικό η **εκ των υστέρων** πιθανότητα του  $B$  ονομάζεται **δεσμευμένη πιθανότητα του  $B$  δοθέντος του  $\Theta'$**  και συμβολίζεται με  $P(B|\Theta')$ .

Παρατηρείστε ότι εκτός από την **εκ των προτέρων** πιθανότητα  $P(B)$  του  $B$  (δηλαδή την πιθανότητα πραγματοποίησης του ενδεχομένου  $B$  πριν γίνει γνωστό το αποτέλεσμα του τεστ), μας δίνονται και οι δεσμευμένες πιθανότητες  $P(\Theta|B)$  και  $P(\Theta'|B')$  που αναφέρονται στην αξιοπιστία του τεστ. Συγκεκριμένα, δίνεται ότι  $P(\Theta|B) = 0.92$  και  $P(\Theta'|B') = 0.96$ . Στη συνέχεια θα δούμε πώς με βάση αυτά τα

<sup>1</sup> όπως εξάλλου συμβαίνει σε αρκετές περιπτώσεις διαγνωστικών τεστ

δεδομένα μπορούμε να υπολογίσουμε τις εκ των υστέρων πιθανότητες  $P(B|\Theta)$  και  $P(B|\Theta')$  του  $B$  που μας ενδιαφέρουν.

Θα δούμε επίσης πώς μπορούμε να χρησιμοποιήσουμε τις δεσμευμένες πιθανότητες  $P(\Theta|B)$  και  $P(\Theta'|B')$  για να υπολογίσουμε τη μη δεσμευμένη πιθανότητα  $P(\Theta)$ , δηλαδή, την πιθανότητα το αποτέλεσμα του τεστ για έναν κάτοικο που επιλέγουμε τυχαία να είναι θετικό. ■

Στα επόμενα θα διαπιστώσουμε ότι σε αρκετές περιπτώσεις η λύση σύνθετων προβλημάτων υπολογισμού μη δεσμευμένων πιθανοτήτων απλοποιείται με την εισαγωγή και χρήση κατάλληλων δεσμευμένων πιθανοτήτων και μάλιστα όχι μόνο όταν έχουμε πρόσθετες πληροφορίες για την έκβαση του πειράματος αλλά και όταν δεν έχουμε! Χαρακτηριστικό είναι το παράδειγμα που ακολουθεί.

**Παράδειγμα 4.1.2:** Σε ένα συρτάρι φαρμακείου βρίσκονται 7 κουτιά με συγκεκριμένο φαρμακευτικό σκεύασμα. Σε 3 από αυτά, το φαρμακευτικό σκεύασμα έχει λήξει. Ο φαρμακοποιός δε γνωρίζει ότι στο συρτάρι υπάρχουν κουτιά με ακατάλληλο σκεύασμα και έτσι στον πρώτο πελάτη που του ζητάει ένα κουτί με το συγκεκριμένο σκεύασμα δίνει ένα από τα 7 που το επιλέγει τυχαία. Όταν και ένας δεύτερος πελάτης ζητάει το συγκεκριμένο σκεύασμα, ο φαρμακοποιός του δίνει ένα που τυχαία επίσης επιλέγει από τα 6 που έχουν απομείνει στο συρτάρι.

Ας θεωρήσουμε τα ενδεχόμενα

$A_i, i = 1, 2$ : για τον  $i$ -στό πελάτη επιλέγεται κουτί με σκεύασμα που έχει λήξει.

Παρατηρείστε ότι ενώ εύκολα βρίσκουμε ότι η πιθανότητα να δοθεί στον πρώτο πελάτη κουτί με σκεύασμα που έχει λήξει είναι  $P(A_1) = 3/7$  και αντίστοιχα ότι η πιθανότητα να του δοθεί κουτί με σκεύασμα που δεν έχει λήξει είναι  $P(A'_1) = 1 - P(A_1) = 4/7$ , για τον υπολογισμό της πιθανότητας  $P(A_2)$  (να δοθεί στο δεύτερο πελάτη κουτί με σκεύασμα που έχει λήξει) ενώ γνωρίζουμε ότι τα δυνατά αποτελέσματα είναι 6, δε γνωρίζουμε πόσα είναι τα ευνοϊκά αποτελέσματα γιατί αυτό εξαρτάται από το αποτέλεσμα της πρώτης επιλογής. Έτσι, αν στον πρώτο πελάτη δόθηκε κουτί με σκεύασμα που έχει λήξει τότε η πιθανότητα στο δεύτερο πελάτη να δοθεί κουτί με σκεύασμα που έχει λήξει είναι προφανώς  $2/6$ , δηλαδή, η δεσμευμένη πιθανότητα του  $A_2$  δοθέντος του  $A_1$  είναι  $P(A_2 | A_1) = 2/6$  ενώ αν στον πρώτο πελάτη δόθηκε κουτί με σκεύασμα που δεν έχει λήξει τότε  $P(A_2 | A'_1) = 3/6$ . Στη συνέχεια θα δούμε πώς μπορούμε να αξιοποιήσουμε αυτές τις δεσμευμένες πιθανότητες για να υπολογίσουμε τη μη δεσμευμένη πιθανότητα  $P(A_2)$  που μας ενδιαφέρει! (Παράδειγμα 4.3.3). ■

Ας δούμε ένα ακόμη παράδειγμα.

**Παράδειγμα 4.1.3:** Ένας φίλος μας ρίχνει ένα αμερόληπτο ζάρι μια φορά και έστω ότι ενδιαφέρεται για την πιθανότητα εμφάνισης του ενδεχομένου  $A = \{1, 4, 6\}$ .

Ο δειγματικός χώρος,  $\Omega = \{1, 2, 3, 4, 5, 6\}$ , του πειράματος είναι πεπερασμένος με ισοπίθανα απλά ενδεχόμενα και επομένως  $P(A) = 3/6 = 1/2$ , αφού τα ευνοϊκά για την πραγματοποίηση του  $A$  αποτελέσματα είναι τρία, το 1, το 4 και το 6 (Σχήμα 4.1.1α).

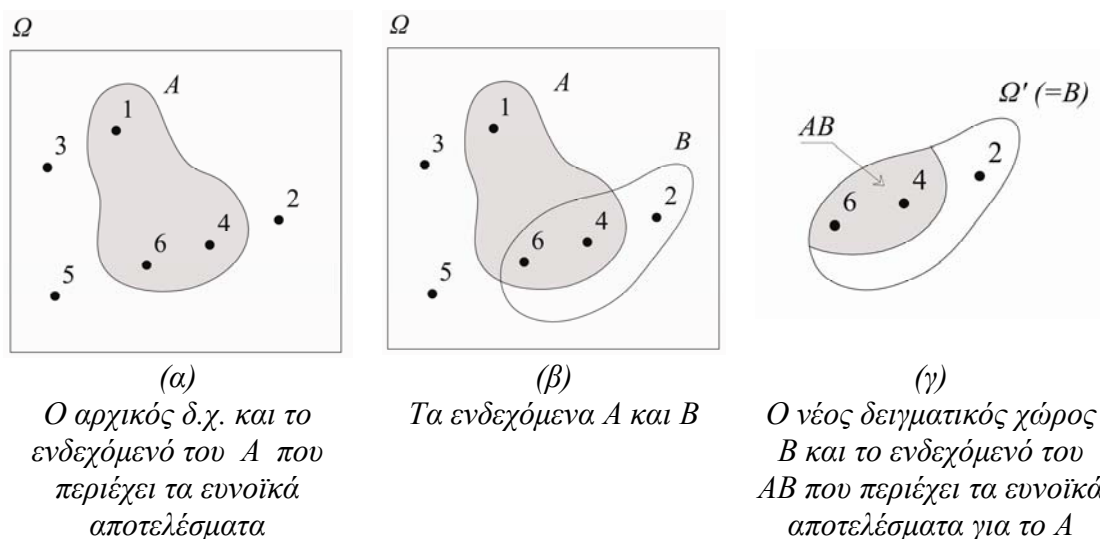
Ρωτάμε το φίλο μας αν εμφανίστηκε το  $A$  αλλά δε μας απαντάει ευθέως. Μας λέει ότι το αποτέλεσμα είναι άρτιος αριθμός, δηλαδή, για την έκβαση του πειράματος, έχουμε την πληροφορία ότι έχει πραγματοποιηθεί το ενδεχόμενο  $B = \{2, 4, 6\}$  (Σχήμα 4.1.1β). Αυτό σημαίνει ότι πλέον τα δυνατά αποτελέσματα δεν είναι έξι αλλά τρία, δηλαδή η πληροφορία που πήραμε για την έκβαση του πειράματος, **συρρικνώνει/περιορίζει** το δειγματικό χώρο  $\Omega = \{1, 2, 3, 4, 5, 6\}$  στο σύνολο  $B = \{2, 4, 6\}$ . Είναι ως να πρόκειται πλέον για ένα νέο πείραμα με δειγματικό χώρο  $\Omega' = B = \{2, 4, 6\}$  και τα ευνοϊκά αποτελέσματα για την πραγματοποίηση του  $A$  να **συρρικνώνονται/περιορίζονται** στην τομή  $AB$  (Σχήμα 4.1.1γ). Έτσι, η πιθανότητα εμφάνισης του  $A$  προκύπτει αν δούμε τον αριθμό (το πλήθος) των ευνοϊκών περιπτώσεων ως ποσοστό του αριθμού των στοιχείων του νέου δειγματικού χώρου, δηλαδή, η πιθανότητα εμφάνισης του  $A$  είναι πλέον ίση με

$$\frac{N(AB)}{N(\Omega')} = \frac{N(AB)}{N(B)} = \frac{2}{3}.$$

Για να υπολογίσουμε/αναθεωρήσουμε την πιθανότητα εμφάνισης του ενδεχομένου  $A$  με βάση την πληροφορία ότι πραγματοποιήθηκε το ενδεχόμενο  $B$ , δηλαδή, για να υπολογίσουμε τη **δεσμευμένη πιθανότητα**  $P(A|B)$ , εύλογα εργασθήκαμε στο νέο/συρρικνωμένο δειγματικό χώρο  $\Omega' = B = \{2, 4, 6\}$ . Αντίστοιχα, το ίδιο κάναμε και στο Παράδειγμα 4.1.2.

Βέβαια, έχει ενδιαφέρον να δούμε αν θα μπορούσαμε να υπολογίσουμε τη **δεσμευμένη πιθανότητα**  $P(A|B)$  στο πλαίσιο του αρχικού δειγματικού χώρου  $\Omega = \{1, 2, 3, 4, 5, 6\}$ . Πράγματι μπορούμε, αφού

$$P(A|B) = \frac{N(AB)}{N(B)} = \frac{2}{3} = \frac{2/6}{3/6} = \frac{N(AB)/N(\Omega)}{N(B)/N(\Omega)} = \frac{P(AB)}{P(B)}.$$



Σχήμα 4.1.1

Η έννοια της δεσμευμένης πιθανότητας του  $A$  δοθέντος του  $B$

Η συλλογιστική που οδήγησε στον τύπο  $P(A|B) = P(AB)/P(B)$  ισχύει και μπορεί να εφαρμοσθεί γενικότερα σε πεπερασμένους δειγματικούς χώρους με ισοπίθανα απλά ενδεχόμενα. Μπορεί επίσης να εφαρμοσθεί σε πεπερασμένους δειγματικούς χώρους με μη ισοπίθανα απλά ενδεχόμενα αλλά και σε συνεχείς δειγματικούς χώρους (συμφωνεί δηλαδή και με την έννοια/ερμηνεία της πιθανότητας ως οριακή σχετική

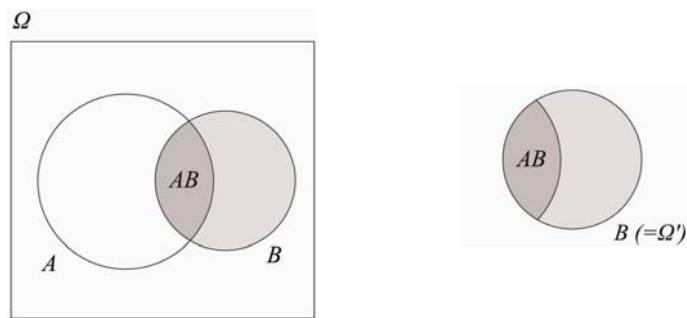
συχνότητα). Έτσι, εύλογα οδηγούμαστε στον ακόλουθο **ορισμό** της **δεσμευμένης πιθανότητας**.

**Ορισμός 4.1.1 (δεσμευμένη πιθανότητα):** Αν  $A$  και  $B$  δύο ενδεχόμενα του δειγματικού χώρου  $\Omega$  ενός πειράματος τύχης και  $P(B) > 0$ , τότε η **δεσμευμένη πιθανότητα του  $A$  δοθέντος του  $B$  (conditional probability)** συμβολίζεται με  $P(A|B)$  και δίνεται από τον τύπο

$$P(A|B) = \frac{P(AB)}{P(B)} \quad (4.1.1).$$

Αν  $P(B) = 0$ , η δεσμευμένη πιθανότητα  $P(A|B)$  δεν ορίζεται.

Η πιθανότητα  $P(A)$  ονομάζεται **εκ των προτέρων πιθανότητα (prior probability)** του  $A$ , ενώ η δεσμευμένη πιθανότητα  $P(A|B)$  ονομάζεται **εκ των υστέρων πιθανότητα (posterior probability)** του  $A$ .

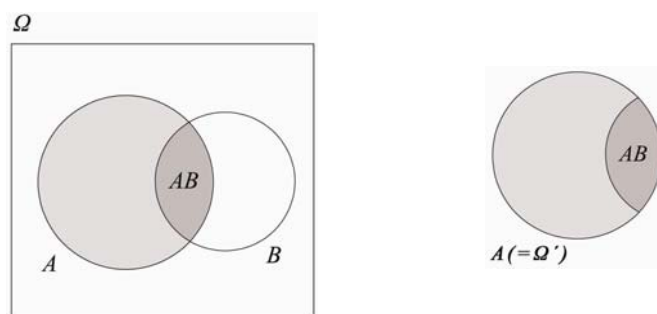


Σχήμα 4.1.2

Για τη δεσμευμένη πιθανότητα  $P(A|B)$  ο δειγματικός χώρος  $\Omega$  περιορίζεται στο ενδεχόμενο  $B$  και το ενδεχόμενο  $A$  των εννοϊκών αποτελεσμάτων, περιορίζεται στο ενδεχόμενο  $AB$

Η **δεσμευμένη πιθανότητα του  $B$  δοθέντος του  $A$**  (και εφόσον  $P(A) > 0$ ) δίνεται από τον τύπο

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



Σχήμα 4.1.3

Για τη δεσμευμένη πιθανότητα  $P(B|A)$  ο δειγματικός χώρος  $\Omega$  περιορίζεται στο ενδεχόμενο  $A$  και το ενδεχόμενο  $B$  των εννοϊκών αποτελεσμάτων, περιορίζεται στο ενδεχόμενο  $AB$

■

**Ερώτηση:** Τι διαφοροποιεί τις πιθανότητες  $P(AB)$ ,  $P(A|B)$  και  $P(B|A)$  αφού και οι τρεις αναφέρονται/αφορούν στο ίδιο σύνολο ευνοϊκών αποτελεσμάτων  $AB$ ;

**Παράδειγμα 4.1.4:** Το 51% των κατοίκων μιας συγκεκριμένης περιοχής είναι άνδρες (και το 49% γυναίκες). Από πρόσφατη έρευνα γνωρίζουμε ότι το 4.2% των κατοίκων αυτής της περιοχής πάσχει από αχρωματοψία. Επίσης από την ίδια έρευνα γνωρίζουμε ότι 4% των κατοίκων της περιοχής αυτής είναι άνδρες που πάσχουν από αχρωματοψία. Αν επιλέξουμε τυχαία ένα άτομο από αυτή την περιοχή και είναι άνδρας, ποια είναι η πιθανότητα να πάσχει από αχρωματοψία.

**Απάντηση:** Ας θεωρήσουμε τα ενδεχόμενα

$A$ : το άτομο πάσχει από αχρωματοψία

$B$ : το άτομο είναι άνδρας.

Δίνεται ότι  $P(A) = 0.042$ ,  $P(B) = 0.51$  και  $P(AB) = 0.04$ .

Ζητάμε την πιθανότητα το άτομο που επιλέξαμε να πάσχει από αχρωματοψία δοθέντος ότι είναι άνδρας, δηλαδή ζητάμε τη δεσμευμένη πιθανότητα  $P(A|B)$ . Από τον τύπο (4.1.1) παίρνουμε

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{0.04}{0.51} = 0.078.$$

Δηλαδή, αν γίνει γνωστό ότι ένα άτομο που επελέγη τυχαία από τη συγκεκριμένη περιοχή είναι άνδρας, τότε η πιθανότητα το άτομο αυτό να έχει αχρωματοψία είναι 0.078, ή αλλιώς, η πιθανότητα ένας άνδρας (από τη συγκεκριμένη περιοχή) να έχει αχρωματοψία είναι 0.078.

**Με όρους ποσοστών**, η (εκ των υστέρων) πιθανότητα  $P(A|B) = 0.078$  ερμηνεύεται ως εξής: στη συγκεκριμένη περιοχή, το 7.8% των ανδρών έχουν αχρωματοψία (δες και Σχήμα 4.1.4

Παρατηρείστε ότι  $P(A|B) \neq P(A)$  (αφού  $0.078 \neq 0.042$ ). Αυτό με όρους ποσοστών σημαίνει ότι το ποσοστό των πασχόντων από αχρωματοψία στους άνδρες της συγκεκριμένης περιοχής διαφέρει από το ποσοστό των πασχόντων από αχρωματοψία στο «γενικό πληθυσμό» (στο σύνολο των κατοίκων της συγκεκριμένης περιοχής). Δηλαδή, η πιθανότητα ένα άτομο να πάσχει από αχρωματοψία **επηρεάζεται/δεν είναι ανεξάρτητη** από το αν το άτομο είναι άνδρας και μάλιστα αυξάνεται αφού  $P(A|B) > P(A)$ .

Η ποσότητα  $|P(A|B) - P(A)|$  ονομάζεται **μεταβολή της πιθανότητας** του  $A$  δοθέντος του  $B$ .

Έτσι, η **μεταβολή της πιθανότητας** ένα άτομο να πάσχει από αχρωματοψία δοθέντος ότι είναι άνδρας είναι  $|P(A|B) - P(A)| = |0.078 - 0.042| = 0.036$ .

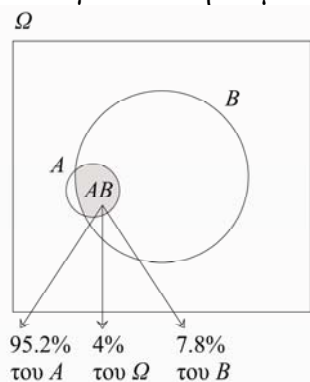
Παρατηρείστε επίσης ότι και

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{0.04}{0.042} = 0.952 > P(B) = 0.51$$

δηλαδή, ενώ στο γενικό πληθυσμό το ποσοστό των ανδρών είναι 51%, στους πάσχοντες από αχρωματοψία το ποσοστό των ανδρών είναι 95.2% (δες Σχήμα 4.1.4).

Επειδή η εμφάνιση του ενός από τα  $A$  και  $B$  αυξάνει την πιθανότητα να εμφανισθεί το άλλο, λέμε ότι τα ενδεχόμενα αυτά είναι **θετικά συσχετισμένα** (δες και το

Παράδειγμα 4.1.5 που ακολουθεί). Στο τέλος του κεφαλαίου όταν θα μιλήσουμε για την **ανεξαρτησία (και την εξάρτηση) ενδεχομένων** θα επανέλθουμε σε αυτό το θέμα. Προς το παρόν, ας ολοκληρώσουμε το παράδειγμα μας εξετάζοντας αν το ποσοστό των γυναικών που πάσχουν από αχρωματοψία διαφέρει από το ποσοστό των ατόμων που πάσχουν από αχρωματοψία στο γενικό πληθυσμό.



Σχήμα 4.1.4

Οι πιθανότητες  $P(AB)$ ,  $P(A|B)$  και  $P(B|A)$  ως ποσοστά

Ας υπολογίσουμε δηλαδή, τη δεσμευμένη πιθανότητα  $P(A|B')$ . Από τον τύπο (4.1.1) έχουμε

$$P(A|B') = \frac{P(AB')}{P(B')} = \frac{P(AB')}{0.49}.$$

Η πιθανότητα του ενδεχομένου  $AB'$  εύκολα υπολογίζεται από τη σχέση  $A = AB \cup AB'$  (δες Παρατήρηση 3.2.1). Πράγματι, επειδή τα  $AB$  και  $AB'$  είναι ξένα μεταξύ τους έχουμε  $P(A) = P(AB \cup AB') = P(AB) + P(AB')$  και επομένως  $0.042 = 0.04 + P(AB') \Rightarrow P(AB') = 0.002$ . Έτσι, η ζητούμενη πιθανότητα είναι

$$P(A|B') = \frac{P(AB')}{0.49} = \frac{0.002}{0.49} = 0.004.$$

Παρατηρείστε ότι  $P(A|B') \neq P(A)$ , δηλαδή, το ποσοστό των γυναικών που πάσχουν από αχρωματοψία διαφέρει από το ποσοστό των ατόμων που πάσχουν από αχρωματοψία στο γενικό πληθυσμό, και μάλιστα είναι μικρότερο ( $P(A|B') < P(A)$ ). Η μεταβολή αυτής της πιθανότητας, δηλαδή, η μεταβολή της πιθανότητας ένα άτομο να πάσχει από αχρωματοψία δοθέντος ότι είναι γυναίκα είναι  $|P(A|B') - P(A)| = |0.004 - 0.042| = 0.038$ .

■

**Παράδειγμα 4.1.5:** Έστω  $A$  και  $B$  δύο ενδεχόμενα ενός δειγματικού χώρου  $\Omega$  με  $P(A) > 0$  και  $P(B) > 0$ . Θα δείξουμε ότι αν  $P(A|B) > P(A)$  τότε  $P(B|A) > P(B)$  και αντιστρόφως. Δύο τέτοια ενδεχόμενα λέγονται **θετικά συσχετισμένα** αφού η εμφάνιση του ενός αυξάνει την πιθανότητα εμφάνισης του άλλου. Αντιστοίχως, αν  $P(A|B) < P(A)$  τότε  $P(B|A) < P(B)$  και αντιστρόφως. Στην περίπτωση αυτή τα ενδεχόμενα λέγονται **αρνητικά συσχετισμένα** αφού η εμφάνιση του ενός μειώνει την πιθανότητα εμφάνισης του άλλου.

Απάντηση: Αν  $P(A|B) > P(A)$ , δηλαδή, αν

$$\frac{P(AB)}{P(B)} > P(A)$$

τότε προφανώς



$$P(AB) > P(A)P(B)$$

και επομένως

$$P(B|A) = \frac{P(AB)}{P(A)} > \frac{P(A)P(B)}{P(A)} = P(B).$$

Το αντίστροφο αποδεικνύεται ομοίως, όπως και η ισοδυναμία των σχέσεων  $P(A|B) < P(A)$  και  $P(B|A) < P(B)$ .

**Παράδειγμα 4.1.6:** α) Αν  $A, B$  είναι δύο ξένα ενδεχόμενα ενός δειγματικού χώρου  $\Omega$  με  $P(A) > 0$  και  $P(B) > 0$  τότε  $P(A|B) = 0$  και  $P(B|A) = 0$ . β) Αν  $A, B$  δύο ενδεχόμενα ενός δειγματικού χώρου  $\Omega$  με  $B \subseteq A$  τότε  $P(A|B) = 1$ .

Απάντηση: α) Πράγματι εφόσον τα ενδεχόμενα  $A, B$  είναι ξένα θα έχουμε  $AB = \emptyset$  και επομένως

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{0}{P(B)} = 0 \text{ και } P(B|A) = \frac{P(AB)}{P(A)} = \frac{0}{P(A)} = 0.$$

Πρόκειται για συμπέρασμα που αναμέναμε αφού, εφόσον τα  $A, B$  είναι ξένα, η γνώση ότι εμφανίστηκε το ένα αποκλείει την εμφάνιση του άλλου.

β) Πράγματι εφόσον  $B \subseteq A$  θα έχουμε  $AB = B$  και επομένως

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Πρόκειται επίσης για ένα συμπέρασμα που αναμέναμε αφού, εφόσον  $B \subseteq A$ , αν γνωρίζουμε ότι εμφανίστηκε το  $B$  είναι βέβαιο ότι εμφανίστηκε και το  $A$ .

**Παράδειγμα 4.1.7 (συνέχεια του Παραδείγματος 3.3.6):** α) Αν για τον άνδρα που επιλέξαμε διαπιστώσουμε ότι ο δείκτης  $B$  βρίσκεται σε φυσιολογικό επίπεδο, ποια είναι η πιθανότητα και ο δείκτης  $A$  να βρίσκεται σε φυσιολογικό επίπεδο και ποια να μη βρίσκεται σε φυσιολογικό επίπεδο; β) Αν για τον άνδρα που επιλέξαμε διαπιστώσουμε ότι ο δείκτης  $B$  δε βρίσκεται σε φυσιολογικό επίπεδο ποια είναι η πιθανότητα ο δείκτης  $A$  να βρίσκεται σε φυσιολογικό επίπεδο; γ) Ποιο ποσοστό των ανδρών ηλικίας 50 έως 70 ετών που μένουν μόνιμα στο ΒΑ Αιγαίο και δεν έχουν το δείκτη  $B$  σε φυσιολογικό επίπεδο, έχουν το δείκτη  $A$  σε φυσιολογικό επίπεδο;

Απάντηση: Γνωρίζουμε ότι  $P(A) = 0.45$ ,  $P(B) = 0.30$ ,  $P(AB) = 0.10$ ,  $P(AB') = 0.35$  και  $P(BA') = 0.20$ .

α) Προφανώς ζητάμε τις δεσμευμένες πιθανότητες  $P(A|B)$  και  $P(A'|B)$ . Από τον τύπο (4.1.1) έχουμε

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{0.10}{0.30} = \frac{1}{3} \text{ και } P(A'|B) = \frac{P(A'B)}{P(B)} = \frac{0.20}{0.30} = \frac{2}{3}.$$

Παρατηρείστε ότι  $P(A'|B) = 1 - P(A|B)$ . Όπως θα δούμε στην Πρόταση 4.1.2 πρόκειται για ιδιότητα της δεσμευμένης πιθανότητας (η οποία επαληθεύεται και στο συγκεκριμένο παράδειγμα).

β) Προφανώς ζητάμε τη δεσμευμένη πιθανότητα  $P(A|B')$  επομένως έχουμε

$$P(A|B') = \frac{P(AB')}{P(B')} = \frac{0.35}{0.70} = \frac{1}{2}.$$

Παρατηρείστε ότι  $P(A|B') \neq 1 - P(A|B)$ . Δηλαδή, γενικά δεν είναι σωστή η σχέση  $P(A|B') = 1 - P(A|B)$ .

γ) Πρόκειται για το ερώτημα (β) διατυπωμένο με όρους ποσοστών, επομένως η απάντηση είναι 50%. ■

Ας συμβολίσουμε με  $P_B(\cdot)$  τη συνολοσυνάρτηση η οποία σε κάθε ενδεχόμενο  $A$  του  $\Omega$  αντιστοιχίζει τη δεσμευμένη πιθανότητα  $P(A|B)$ . Έστω δηλαδή,

$$P_B(A) = P(A|B) = \frac{P(AB)}{P(B)}.$$

Από τον τύπο (4.1.1) και τα τρία αξιώματα του αξιωματικού ορισμού της συνάρτησης πιθανότητας  $P(\cdot)$ , άμεσα προκύπτουν για την  $P_B(\cdot)$  οι ακόλουθες ιδιότητες που είναι αντίστοιχες με τα τρία αξιώματα. Δηλαδή, η  $P_B(\cdot)$  ικανοποιεί τα τρία αξιώματα που πρέπει να ικανοποιεί μια συνάρτηση πιθανότητας.

**Πρόταση 4.1.1:** Έστω  $\Omega$  ένας δειγματικός χώρος και  $B$  ένα ενδεχόμενό του με  $P(B) > 0$ . Ισχύουν οι ακόλουθες ιδιότητες.

(1)  $P(A|B) \geq 0$ , για κάθε ενδεχόμενο  $A$  του συνόλου των ενδεχομένων του  $\Omega$

(2)  $P(\Omega|B) = 1$

(3)  $P(A_1 \cup A_2 \cup \dots \cup A_n \cup \dots | B) = P(A_1|B) + P(A_2|B) + \dots + P(A_n|B) + \dots$

για οποιαδήποτε ακολουθία,  $A_1, A_2, \dots, A_n, \dots$ , ζένων ανά δύο ενδεχομένων του συνόλου των ενδεχομένων του  $\Omega$ . ■

Στην Πρόταση 4.1.2 που ακολουθεί δίνουμε κάποιες ακόμη χρήσιμες ιδιότητες της δεσμευμένης πιθανότητας αντίστοιχες με αυτές της Πρότασης 3.3.1 (και που αποδεικνύονται ανάλογα).

**Πρόταση 4.1.2:** Έστω  $\Omega$  ένας δειγματικός χώρος και  $B$  ένα ενδεχόμενό του με  $P(B) > 0$ . Για τη δεσμευμένη πιθανότητα ισχύουν οι ακόλουθες ιδιότητες.

(α)  $P(\emptyset|B) = 0$

(β)  $P(A'|B) = 1 - P(A|B)$ .

(γ)  $P(A - \Gamma|B) = P(A\Gamma'|B) = P(A|B) - P(A\Gamma|B)$ .

(δ) Αν  $\Gamma \subseteq A$ , τότε  $P(\Gamma|B) \leq P(A|B)$ .

(ε)  $P(A \cup \Gamma|B) = P(A|B) + P(\Gamma|B) - P(A\Gamma|B)$ . ■

**Παράδειγμα 4.1.8:** Έστω  $A$  και  $B$  δύο ενδεχόμενα ενός δειγματικού χώρου  $\Omega$  τέτοια ώστε (α) η πιθανότητα να εμφανισθεί το  $A$  αλλά όχι το  $B$  είναι 0.15 (β) η πιθανότητα να εμφανισθεί το  $B$  αλλά όχι το  $A$  είναι 0.1 και (γ) η πιθανότητα να μην εμφανισθεί ούτε το  $A$  ούτε το  $B$  είναι 0.7. Ζητάμε τη δεσμευμένη πιθανότητα  $P(A|B)$ .

Απάντηση: Δίνεται ότι  $P(AB') = 0.15$ ,  $P(A'B) = 0.1$  και  $P[(A \cup B)'] = 0.7$ . Προφανώς (δείτε το Σχήμα 4.1.5 και θυμηθείτε την Παρατήρηση 3.2.1)

$$A \cup B = AB' \cup AB \cup A'B$$

και επομένως

$$P(A \cup B) = P(AB') + P(AB) + P(A'B)$$

άρα

$$1 - 0.7 = 0.15 + P(AB) + 0.1 \Rightarrow P(AB) = 0.05.$$

Επίσης,

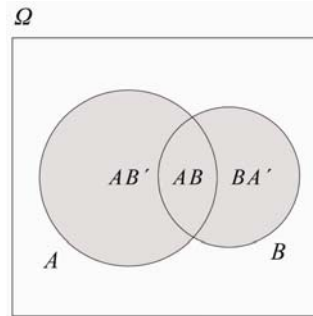
$$B = AB \cup A'B$$

και επομένως

$$P(B) = P(AB) + P(A'B) \Rightarrow P(B) = 0.05 + 0.1 = 0.15$$

άρα η ζητούμενη πιθανότητα είναι

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{0.05}{0.15} = \frac{1}{3}.$$



Σχήμα 4.1.5

Διαμέριση  $\{AB', AB, BA'\}$

της ένωσης  $A \cup B$  δύο ενδεχομένων  $A, B$

Ολοκληρώνουμε τα παραδείγματα που επιλέξαμε για εισαγωγή στην έννοια της δεσμευμένης πιθανότητας, με ένα πρόβλημα πιθανοτήτων από αυτά που εύκολα μπορεί να μας παρασύρουν σε μια «προφανή και απλή» αλλά ... λάθος απάντηση.

**Παράδειγμα 4.1.9:** Επιλέγουμε τυχαία μια οικογένεια από το σύνολο των οικογενειών που έχουν δύο παιδιά και καταγράφουμε το φύλο των παιδιών. Ποια είναι η πιθανότητα και τα δύο παιδιά της οικογένειας που επιλέξαμε να είναι αγόρια αν γνωρίζουμε ότι τουλάχιστον ένα από τα δύο παιδιά είναι αγόρι;

Απάντηση: Ένας κατάλληλος δειγματικός χώρος του πειράματος είναι ο

$$\Omega = \{κκ, ακ, κα, αα\}.$$

Ως αποτέλεσμα του πειράματος επιλέξαμε να καταγράφουμε το φύλο και τη σειρά γέννησης των παιδιών. Για παράδειγμα, το αποτέλεσμα «κα» σημαίνει ότι η οικογένεια έχει ένα αγόρι και ένα κορίτσι και ότι το πρώτο παιδί που γεννήθηκε είναι κορίτσι και το δεύτερο αγόρι.

Ας θεωρήσουμε τα ενδεχόμενα

$A$ : και τα δύο παιδιά της οικογένειας είναι αγόρια

$B$ : τουλάχιστον ένα από τα δύο παιδιά της οικογένειας είναι αγόρια.

Προφανώς,  $A = \{αα\}$  και  $B = \{ακ, κα, αα\}$ . Ζητάμε τη δεσμευμένη πιθανότητα  $P(A|B)$ . Θα την υπολογίσουμε θεωρώντας ότι τα τέσσερα απλά ενδεχόμενα του  $\Omega$  είναι ισοπίθανα.

1<sup>ος</sup> τρόπος: Εργαζόμαστε στον αρχικό δειγματικό χώρο  $\Omega$

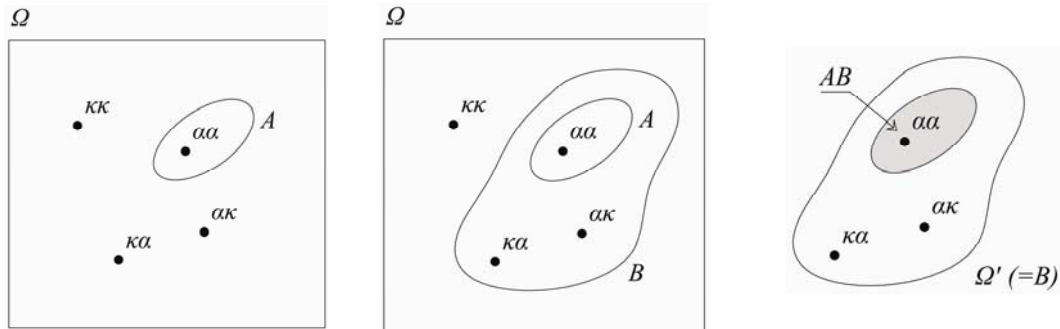
Επειδή  $A \subseteq B$  θα είναι  $AB = A$  και επομένως  $P(AB) = P(A)$ . Επίσης, επειδή ο δειγματικός χώρος του πειράματος είναι πεπερασμένος με ισοπίθανα απλά ενδεχόμενα προφανώς έχουμε,  $P(AB) = P(A) = \frac{1}{4}$  και  $P(B) = \frac{3}{4}$  οπότε

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A)}{P(B)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

2<sup>ος</sup> τρόπος: Εργαζόμαστε απευθείας στο νέο δειγματικό χώρο  $\Omega' = B$

Δεδομένου ότι τουλάχιστον ένα από τα δύο παιδιά της οικογένειας είναι αγόρι, ο δειγματικός χώρος του πειράματος συρρικνώνεται/περιορίζεται στο σύνολο  $B = \{ακ, κα, αα\}$  και οι ευνοϊκές περιπτώσεις περιορίζονται στην τομή  $AB = \{αα\}$  (δες Σχήματα 4.1.6). Τα τρία απλά ενδεχόμενα του νέου δειγματικού χώρου  $B$  είναι ισοπίθανα, επομένως

$$P(A|B) = 1/3.$$



Σχήματα 4.1.6  
 $P(A|B) = 1/3$

Η απάντηση επομένως στο ερώτημα που τέθηκε είναι 1/3 και **όχι** 1/2 που ίσως ήταν η πρώτη μας σκέψη!! Η χρήση κατάλληλης δεσμευμένης πιθανότητας μας βοήθησε να αποφύγουμε την «παγίδα» της «προφανούς και απλής» αλλά λάθος απάντησης.

Σημειώνουμε ότι θα μπορούσαμε στο αποτέλεσμα του πειράματος να μη δηλώνουμε και τη σειρά γέννησης των παιδιών, αλλά μόνο το φύλο. Έτσι, κατάλληλος δειγματικός χώρος είναι και ο  $\Omega_1 = \{κκ, κα, αα\}$  όπου το αποτέλεσμα «κα» δηλώνει μόνο το φύλο και όχι και τη σειρά γέννησης όπως και ο  $\Omega_2 = \{0, 1, 2\}$  όπου ως αποτέλεσμα καταγράφουμε τον αριθμό των αγοριών της οικογένειας. Αν εργασθούμε με το δειγματικό χώρο  $\Omega_1$  (ή με τον  $\Omega_2$ ), τι λέτε, η απάντηση θα είναι και πάλι 1/3 ή μήπως θα αλλάξει;

Επισημαίνουμε ότι για να υπολογίσουμε μια δεσμευμένη πιθανότητα  $P(A|B)$ , μπορούμε να επιλέξουμε από δύο τρόπους:  
**(α)** Να υπολογίσουμε στον αρχικό δειγματικό χώρο  $\Omega$  τις πιθανότητες των ενδεχομένων  $AB$  και  $B$  και να εφαρμόσουμε τον τύπο (4.1.1).  
**(β)** Να εργασθούμε απευθείας στο νέο δειγματικό χώρο  $\Omega' = B$ .

Ας δούμε τώρα πώς, με την εισαγωγή κατάλληλων δεσμευμένων πιθανοτήτων, μπορούμε να απλοποιήσουμε τη λύση προβλημάτων υπολογισμού μη δεσμευμένων πιθανοτήτων!

#### 4.2 Ο πολλαπλασιαστικός τύπος/νόμος

Ας θεωρήσουμε δύο ενδεχόμενα  $A, B$  ενός δειγματικού χώρου  $\Omega$ . Αν  $P(B) > 0$ , τότε από τον τύπο

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

ορισμού της δεσμευμένης πιθανότητας  $P(A|B)$  προφανώς παίρνουμε

$$P(AB) = P(B)P(A|B). \tag{4.2.1}$$

Επίσης, αν  $P(A) > 0$ , τότε από τον τύπο

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

ορισμού της δεσμευμένης πιθανότητας  $P(B | A)$  παίρνουμε

$$P(AB) = P(A)P(B | A). \quad (4.2.2)$$

Έτσι, αν για δύο ενδεχόμενα  $A, B$  μπορούμε να υπολογίσουμε (ή με κάποιο τρόπο γνωρίζουμε) τη δεσμευμένη πιθανότητα του ενός δοθέντος του άλλου, τότε από τον τύπο (4.2.1) (ή τον (4.2.2)) μπορούμε να υπολογίσουμε τη (μη δεσμευμένη) πιθανότητα  $P(AB)$  της τομής τους  $AB$  (και εφόσον βέβαια γνωρίζουμε ή μπορούμε να υπολογίσουμε την πιθανότητα  $P(B)$  ή  $P(A)$  αντίστοιχα). Ας δούμε ένα τέτοιο παραδείγματα υπολογισμού της τομής δύο ενδεχομένων.

**Παράδειγμα 4.2.1 (συνέχεια του Παραδείγματος 4.1.2):** Θα υπολογίσουμε την πιθανότητα

- α) να δοθεί και στους δύο πελάτες κουτί με σκεύασμα που έχει λήξει.
- β) να δοθεί και στους δύο πελάτες κουτί με σκεύασμα που δεν έχει λήξει
- γ) στον πρώτο πελάτη να δοθεί κουτί με σκεύασμα που έχει λήξει και στον δεύτερο κουτί με σκεύασμα που δεν έχει λήξει
- δ) στον πρώτο πελάτη να δοθεί κουτί με σκεύασμα που δεν έχει λήξει και στον δεύτερο κουτί με σκεύασμα που έχει λήξει.

Απάντηση: Ας θεωρήσουμε τα ενδεχόμενα

$A_i, i = 1, 2$ : για τον  $i$ -στό πελάτη επιλέγεται κουτί με σκεύασμα που έχει λήξει.

- α) Ζητάμε την πιθανότητα του ενδεχομένου  $A_1 A_2$ . Επειδή, όπως εξηγήσαμε,  $P(A_1) = 3/7$  και  $P(A_2 | A_1) = 2/6$ , για τη ζητούμενη πιθανότητα έχουμε

$$P(A_1 A_2) = P(A_1)P(A_2 | A_1) = \frac{3}{7} \cdot \frac{2}{6} = \frac{1}{7}.$$

- β) Ζητάμε την πιθανότητα του ενδεχομένου  $A'_1 A'_2$ . Επειδή, προφανώς  $P(A'_1) = 4/7$  και  $P(A'_2 | A'_1) = 3/6$ , για τη ζητούμενη πιθανότητα έχουμε

$$P(A'_1 A'_2) = P(A'_1)P(A'_2 | A'_1) = \frac{4}{7} \cdot \frac{3}{6} = \frac{2}{7}.$$

- γ) Ζητάμε την πιθανότητα του ενδεχομένου  $A_1 A'_2$ . Επειδή προφανώς  $P(A_1) = 3/7$  και  $P(A'_2 | A_1) = 4/6$ , για τη ζητούμενη πιθανότητα έχουμε

$$P(A_1 A'_2) = P(A_1)P(A'_2 | A_1) = \frac{3}{7} \cdot \frac{4}{6} = \frac{2}{7}.$$

- δ) Ζητάμε την πιθανότητα του ενδεχομένου  $A'_1 A_2$ . Επειδή προφανώς  $P(A'_1) = 4/7$  και  $P(A_2 | A'_1) = 3/6$ , για τη ζητούμενη πιθανότητα έχουμε

$$P(A'_1 A_2) = P(A'_1)P(A_2 | A'_1) = \frac{4}{7} \cdot \frac{3}{6} = \frac{2}{7}.$$

Παρατηρείστε ότι οι πιθανότητες που υπολογίσαμε αθροίζουν στο 1 (σκεφθείτε γιατί). Επίσης, ως άσκηση, βρείτε α) την πιθανότητα να δοθεί σε τουλάχιστον έναν από τους δύο πελάτες κουτί με σκεύασμα που έχει λήξει και β) την πιθανότητα να δοθεί μόνο σε έναν από τους δύο πελάτες κουτί με σκεύασμα που έχει λήξει.

Σημειώνουμε τέλος, ότι οι πιθανότητες που υπολογίσαμε μπορούν να υπολογισθούν και χωρίς τη χρήση δεσμευμένων πιθανοτήτων. Δείτε το ως μια απλή άσκηση. (Παρατηρείστε ότι το πείραμα έχει  $7 \cdot 6 = 42$  δυνατά αποτελέσματα και για τις

ζητούμενες πιθανότητες το πλήθος των εννοϊκών αποτελεσμάτων, αντίστοιχα, είναι  $3 \cdot 2 = 6$ ,  $4 \cdot 3 = 12$ ,  $3 \cdot 4 = 12$ ,  $4 \cdot 3 = 12$ .)

Με χρήση κατάλληλων δεσμευμένων πιθανοτήτων μπορούμε να υπολογίσουμε την πιθανότητα της τομής και περισσότερων των δύο ενδεχομένων. Ας δούμε την πρόταση που ακολουθεί.

**Πρόταση 4.2.1 (ο πολλαπλασιαστικός τύπος):** Αν  $A_1, A_2, \dots, A_n$ ,  $n$  ενδεχόμενα ενός δειγματικού χώρου  $\Omega$ , με  $P(A_1 A_2 \dots A_{n-1}) > 0$ , τότε

$$P(A_1 A_2 \dots A_n) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 A_2) \dots P(A_n | A_1 A_2 \dots A_{n-1}). \quad (4.2.3)$$

Ο τύπος αυτός ονομάζεται **πολλαπλασιαστικός τύπος ή πολλαπλασιαστικός νόμος (multiplication formula/rule)**.

Απόδειξη: Είναι απλή. Εφαρμόζοντας τον τύπο (4.1.1) για τις δεσμευμένες πιθανότητες που εμφανίζονται στο δεύτερο μέλος της (4.2.3) έχουμε

$$P(A_1)P(A_2 | A_1)P(A_3 | A_1 A_2) \dots P(A_n | A_1 A_2 \dots A_{n-1}) = P(A_1) \cdot \frac{P(A_1 A_2)}{P(A_1)} \cdot \frac{P(A_1 A_2 A_3)}{P(A_1 A_2)} \dots \frac{P(A_1 A_2 \dots A_n)}{P(A_1 A_2 \dots A_{n-1})} = P(A_1 A_2 \dots A_n).$$

Η συνθήκη  $P(A_1 A_2 \dots A_{n-1}) > 0$  εξασφαλίζει ότι οι δεσμευμένες πιθανότητες που εμφανίζονται στο δεύτερο μέλος του πολλαπλασιαστικού τύπου ορίζονται. Πράγματι,  $P(A_1) \geq P(A_1 A_2) \geq \dots \geq P(A_1 A_2 \dots A_{n-1}) > 0$  (αφού  $A_1 \supseteq A_1 A_2 \supseteq \dots \supseteq A_1 A_2 \dots A_{n-1}$ ).

Οι τύποι (4.2.1) και (4.2.2) προφανώς αποτελούν ειδική περίπτωση του τύπου (4.2.3) για  $n = 2$ .

**Παράδειγμα 4.2.2 (συνέχεια του Παραδείγματος 4.1.2):** Θα υπολογίσουμε την πιθανότητα, στους τρεις πρώτους πελάτες να δοθεί κουτί με σκεύασμα που δεν έχει λήξει.

Απάντηση: Ας θεωρήσουμε τα ενδεχόμενα

$A_i, i = 1, 2, 3$ : για τον  $i$ -στο πελάτη επιλέγεται κουτί με σκεύασμα που έχει λήξει.

Ζητάμε την πιθανότητα του ενδεχομένου  $A'_1 A'_2 A'_3$ . Από τον πολλαπλασιαστικό τύπο έχουμε

$$P(A'_1 A'_2 A'_3) = P(A'_1)P(A'_2 | A'_1)P(A'_3 | A'_1 A'_2) = \frac{4}{7} \cdot \frac{3}{6} \cdot \frac{2}{5} = \frac{4}{35}.$$

**Σχόλιο 4.2.1:** Ο πολλαπλασιαστικός τύπος, όπως φαίνεται και από τα παραδείγματα που δώσαμε προηγουμένως, βοηθάει ιδιαίτερα σε περιπτώσεις προβλημάτων που αφορούν σε οικογένειες ενδεχομένων που μπορούν να τοποθετηθούν σε μια σειρά (χρονολογική, λογική, ή άλλη). Δείτε ως επιπλέον σχετικά παραδείγματα και τα Προβλήματα 4.10, 4.11 και 4.12.

### 4.3 Το θεώρημα ολικής πιθανότητας

Όταν μιλήσαμε για τις πράξεις μεταξύ ενδεχομένων, δείξαμε ότι αν  $\{B_1, B_2, \dots, B_n\}$  είναι μια **διαμέριση** του δειγματικού χώρου  $\Omega$  ενός πειράματος τύχης τότε για

οποιοδήποτε ενδεχόμενο  $A$  του  $\Omega$ , τα ενδεχόμενα  $AB_1, AB_2, \dots, AB_\nu$  αποτελούν μια διαμέριση του  $A$ , δηλαδή

$$A = AB_1 \cup AB_2 \cup \dots \cup AB_\nu \text{ και } (AB_i)(AB_j) = \emptyset \text{ (για κάθε } i \neq j \text{)}.$$

Αυτό σημαίνει ότι όταν πραγματοποιείται το ενδεχόμενο  $A$ , πραγματοποιείται σε συνδυασμό με ένα και μάλιστα με ακριβώς ένα (μόνο ένα) από τα  $B_1, B_2, \dots, B_\nu$  (δες Σχήμα 4.3.1). Έτσι έχουμε

$$P(A) = P(AB_1 \cup AB_2 \cup \dots \cup AB_\nu)$$

και λόγω της ιδιότητας της πεπερασμένης προσθετικότητας, παίρνουμε

$$P(A) = P(AB_1) + P(AB_2) + \dots + P(AB_\nu).$$

Ο τύπος αυτός υπολογισμού της πιθανότητας ενός ενδεχομένου  $A$  ως άθροισμα των πιθανοτήτων των ξένων ανά δύο ενδεχομένων  $AB_1, AB_2, \dots, AB_\nu$  των οποίων η ένωση είναι όλο το  $A$ , ονομάζεται **τύπος της ολικής πιθανότητας (law of total probability)**.

Αν για όλα τα  $i=1,2,\dots,\nu$ , είναι  $P(B_i) > 0$ , τότε οι δεσμευμένες πιθανότητες  $P(A|B_i)$ ,  $i=1,2,\dots,\nu$  ορίζονται και ο τύπος της ολικής πιθανότητας, σε συνδυασμό με τον πολλαπλασιαστικό τύπο

$$P(AB_i) = P(A|B_i)P(B_i), \quad i=1,2,\dots,\nu$$

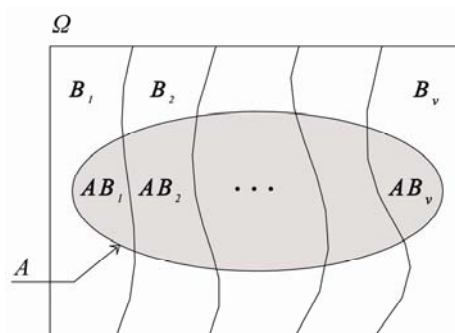
γράφεται

$$\begin{aligned} P(A) &= P(AB_1) + P(AB_2) + \dots + P(AB_\nu) = \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_\nu)P(B_\nu). \end{aligned}$$

Αποδείξαμε έτσι, το θεώρημα ολικής πιθανότητας.

**Θεώρημα 4.3.1 (το θεώρημα ολικής πιθανότητας):** Αν  $\{B_1, B_2, \dots, B_\nu\}$  είναι μια διαμέριση του δειγματικού χώρου  $\Omega$  ενός πειράματος τύχης, με  $P(B_i) > 0$  για όλα τα  $i=1,2,\dots,\nu$ , τότε, για κάθε ενδεχόμενο  $A$  του  $\Omega$  ισχύει ότι

$$P(A) = \sum_{i=1}^{\nu} P(AB_i) = \sum_{i=1}^{\nu} P(A|B_i)P(B_i). \quad (4.3.1)$$



Σχήμα 4.3.1

$$\begin{aligned} P(A) &= P(AB_1 \cup AB_2 \cup \dots \cup AB_\nu) = \\ &= P(AB_1) + P(AB_2) + \dots + P(AB_\nu) = \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_\nu)P(B_\nu) \end{aligned}$$

**Παράδειγμα 4.3.1:** Οι ενήλικες κάτοικοι ( $\geq 18$  ετών) μιας περιοχής έχουν ταξινομηθεί σε πέντε ηλικιακές ομάδες. Στον πίνακα που ακολουθεί φαίνεται το ποσοστό που κατέχει κάθε ηλικιακή ομάδα στο σύνολο των ενηλίκων. Επίσης φαίνεται το ποσοστό

των κατοίκων κάθε ηλικιακής ομάδας που πίνουν τουλάχιστον τρεις καφέδες ημερησίως. Ποια είναι η πιθανότητα ένας ενήλικας αυτής της περιοχής που επιλέγεται τυχαία από το σύνολο των ενηλίκων να πίνει περισσότερους από τρεις καφέδες ημερησίως;

	Ηλικιακή Ομάδα				
	1	2	3	4	5
	18-24	25-34	35-49	50-64	≥ 65
Ποσοστό κατοίκων με $\geq 3$ καφέδες ( επί της ηλικιακής ομάδας)	26%	35%	10%	30%	25%
Ποσοστό που κατέχει η ηλικιακή ομάδα ( επί του συνόλου των ενηλίκων)	9%	20%	31%	23%	17%

Απάντηση: Από τους ενήλικες κατοίκους της περιοχής επιλέγουμε τυχαία έναν. Ο δειγματικός χώρος  $\Omega$  αυτού του πειράματος τύχης είναι το σύνολο όλων των ενηλίκων κατοίκων αυτής της περιοχής. Ας θεωρήσουμε τα ενδεχόμενα

$B_i$ : ο κάτοικος ανήκει στην ηλικιακή ομάδα  $i, i = 1, 2, \dots, 5$ .

$A$ : ο κάτοικος πίνει τουλάχιστον 3 καφέδες ημερησίως.

Μας δίνεται ότι  $P(A|B_1) = 0.26$ ,  $P(A|B_2) = 0.35$ ,  $P(A|B_3) = 0.10$ ,  $P(A|B_4) = 0.30$ ,  $P(A|B_5) = 0.25$ ,  $P(B_1) = 0.09$ ,  $P(B_2) = 0.20$ ,  $P(B_3) = 0.31$ ,  $P(B_4) = 0.23$  και  $P(B_5) = 0.17$  και μας ζητείται η **μη δεσμευμένη** πιθανότητα  $P(A)$  του ενδεχομένου  $A$ .

Τα ενδεχόμενα  $B_1, B_2, \dots, B_5$  προφανώς αποτελούν μια διαμέριση του δειγματικού χώρου του πειράματος αφού κάθε ενήλικας κάτοικος  $\geq 18$  ετών ανήκει σε μια από τις πέντε ηλικιακές ομάδες και μάλιστα μόνο σε μια, ή αλλιώς, οι πέντε ηλικιακές ομάδες καλύπτουν όλες τις ηλικίες  $\geq 18$  ετών, δηλαδή  $B_1 \cup B_2 \cup \dots \cup B_5 = \Omega$ , και  $B_i B_j = \emptyset$  (για κάθε  $i \neq j$ ). Επομένως, για τη ζητούμενη πιθανότητα  $P(A)$ , μπορεί να εφαρμοσθεί ο τύπος (4.3.1) του **θεωρήματος ολικής πιθανότητας**. Έτσι έχουμε

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_5)P(B_5) = 0.26 \cdot 0.09 + 0.35 \cdot 0.20 + 0.10 \cdot 0.31 + 0.30 \cdot 0.23 + 0.25 \cdot 0.17 = 0.2359.$$

Επομένως, η πιθανότητα ένας ενήλικας κάτοικος της συγκεκριμένης περιοχής που επιλέγεται τυχαία, να πίνει τουλάχιστον 3 καφέδες ημερησίως είναι (περίπου) 0.24.

Ίσως παρατηρήσατε ότι η πιθανότητα του ενδεχομένου  $A$ , μέσω του τύπου της **ολικής πιθανότητας** εκφράστηκε και υπολογίστηκε ως ο **σταθμισμένος μέσος όρος** των πιθανοτήτων του  $A$  εντός κάθε ηλικιακής ομάδας, με συντελεστές στάθμισης 0.09, 0.20, 0.31, 0.23 και 0.17 που εκφράζουν αντίστοιχα, τα σχετικά μεγέθη των ηλικιακών ομάδων. Πράγματι έτσι είναι.

Εστω  $\{B_1, B_2, \dots, B_n\}$  μια διαμέριση του δειγματικού χώρου  $\Omega$  ενός πειράματος τύχης, με  $P(B_i) > 0$  για κάθε  $i = 1, 2, \dots, n$ . Με τον τύπο (4.3.1) του θεωρήματος της ολικής πιθανότητας, η μη δεσμευμένη πιθανότητα  $P(A)$  ενός οποιουδήποτε ενδεχομένου  $A$  του  $\Omega$  εκφράζεται ως ο **σταθμισμένος μέσος όρος** των δεσμευμένων πιθανοτήτων  $P(A|B_i)$ , με συντελεστές στάθμισης, αντίστοιχα, τις πιθανότητες  $P(B_i)$ .

■



**Παράδειγμα 4.3.2 (συνέχεια του Παραδείγματος 4.1.1):** Αν κάποιος υποβληθεί στο TB *tine* τεστ για να ελεγχθεί αν είναι φορέας του βακίλου της φυματίωσης, το αποτέλεσμα δεν είναι όπως είδαμε απόλυτο, δηλαδή, δεν είναι πάντοτε σωστό. Μπορεί δηλαδή, το αποτέλεσμα του τεστ να είναι θετικό ενώ ο εξεταζόμενος δεν είναι φορέας (**ψευδώς θετικό**). Επίσης, μπορεί να είναι αρνητικό ενώ ο εξεταζόμενος είναι φορέας (**ψευδώς αρνητικό**). Βέβαια, από σχετικές μελέτες, η αξιοπιστία του τεστ μας είναι γνωστή. Έτσι, μας δίνεται η πιθανότητα το αποτέλεσμα του τεστ να είναι **αληθώς θετικό**, δηλαδή να είναι θετικό όταν ο εξεταζόμενος είναι φορέας όπως και η πιθανότητα το αποτέλεσμα του τεστ να είναι **αληθώς αρνητικό**, δηλαδή να είναι αρνητικό όταν ο εξεταζόμενος δεν είναι φορέας. Συγκεκριμένα, μας δίνεται ότι

$$P(\Theta | B) = 0.92 \text{ και } P(\Theta' | B') = 0.96.$$

Υπενθυμίζουμε ότι με  $\Theta$  και  $B$  έχουμε αντίστοιχα συμβολίσει τα ενδεχόμενα «το αποτέλεσμα του τεστ είναι θετικό» και «ο εξεταζόμενος είναι φορέας του βακίλου της φυματίωσης».

Ένα πρώτο εύλογο ερώτημα είναι το εξής: ποια είναι η πιθανότητα για έναν κάτοικο που επιλέγεται τυχαία από τη συγκεκριμένη περιοχή και υποβάλλεται στο TB *tine* τεστ το αποτέλεσμα του τεστ να είναι θετικό, ή αλλιώς, τι ποσοστό των κατοίκων στη συγκεκριμένη περιοχή δίνει θετικό αποτέλεσμα.

Απάντηση: Επειδή τα ενδεχόμενα  $B$  και  $B'$  αποτελούν μια διαμέριση του δειγματικού χώρου (αφού προφανώς  $B \cup B' = \Omega$  και  $BB' = \emptyset$ ) και  $P(B) = 0.0075 > 0$  άρα και  $P(B') = 1 - P(B) = 0.9925 > 0$ , από τον τύπο (4.3.1) του θεωρήματος ολικής πιθανότητας έχουμε

$$P(\Theta) = P(\Theta | B)P(B) + P(\Theta | B')P(B')$$

όπου,  $P(\Theta | B) = 0.92$ ,  $P(B) = 0.0075$ ,  $P(B') = 0.9925$  και  $P(\Theta | B')$  η πιθανότητα το αποτέλεσμα του τεστ να είναι **ψευδώς θετικό**, δηλαδή να είναι θετικό όταν ο εξεταζόμενος δεν είναι φορέας η οποία υπολογίζεται από τη σχέση

$$P(\Theta | B') = 1 - P(\Theta' | B') = 1 - 0.96 = 0.04$$

και επομένως έχουμε

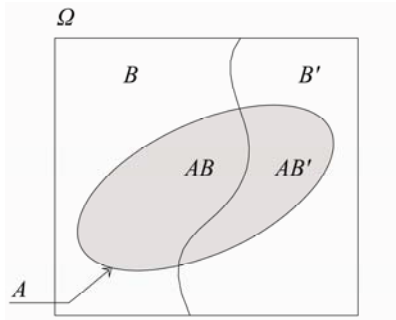
$$P(\Theta) = P(\Theta | B)P(B) + P(\Theta | B')P(B') = 0.92 \cdot 0.0075 + 0.04 \cdot 0.9925 = 0.0466.$$

Έτσι, για έναν κάτοικο από τη συγκεκριμένη περιοχή που επιλέγεται τυχαία και υποβάλλεται στο TB *tine* τεστ, η πιθανότητα το τεστ να δώσει θετικό αποτέλεσμα είναι 0.0466. Παρατηρείστε ότι η πιθανότητα αυτή είναι μεγαλύτερη (και μάλιστα αρκετά) από την  $P(B) = 0.0075$ , δηλαδή από την πιθανότητα ένας κάτοικος να είναι φορέας. Ασφαλώς, πρόκειται για κάτι αναμενόμενο. Συμφωνείτε;

■

**Παρατήρηση 4.3.1:** Έστω  $A$  και  $B$  δύο ενδεχόμενα του δειγματικού χώρου  $\Omega$  ενός πειράματος τύχης με  $0 < P(B) < 1$ . Τα ενδεχόμενα  $B$  και  $B'$  (δες και Σχήμα 4.3.2) αποτελούν διαμέριση του  $\Omega$  (αφού  $B \cup B' = \Omega$  και  $BB' = \emptyset$ ) και επομένως, ο τύπος (4.3.1) της ολικής πιθανότητας για την  $P(A)$  με βάση τη διαμέριση  $\{B, B'\}$ , γράφεται

$$P(A) = P(A | B)P(B) + P(A | B')P(B') \quad (4.3.2)$$



Σχήμα 4.3.2

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

Σημειώνουμε ότι οι δεσμευμένες πιθανότητες που εμφανίζονται στον τύπο (4.3.2) προφανώς ορίζονται αφού υποθέσαμε ότι  $P(B) > 0$  και  $P(B) < 1$  επομένως και  $P(B') = 1 - P(B) > 0$ .

**Παράδειγμα 4.3.3 (συνέχεια του Παραδείγματος 4.1.2):** Θα υπολογίσουμε την πιθανότητα,  $P(A_2)$ , να δοθεί στο δεύτερο πελάτη κουτί με σκεύασμα που έχει λήξει.

Με απλή εφαρμογή του τύπου ολικής πιθανότητας παίρνουμε

$$P(A_2) = P(A_2 | A_1)P(A_1) + P(A_2 | A_1')P(A_1') = \frac{2}{6} \cdot \frac{3}{7} + \frac{3}{6} \cdot \frac{4}{7} = \frac{3}{7}.$$

Παρατηρείστε ότι  $P(A_2) = P(A_1)$ , δηλαδή, η πιθανότητα να δοθεί στον δεύτερο πελάτη κουτί με σκεύασμα που έχει λήξει είναι ίση με την πιθανότητα να δοθεί στον πρώτο πελάτη κουτί με σκεύασμα που έχει λήξει!!!

#### 4.4 Το θεώρημα του Bayes

**Παράδειγμα 4.4.1 (συνέχεια του Παραδείγματος 4.3.2):** Στο Παράδειγμα 4.3.2 υπολογίσαμε, με βάση το θεώρημα της ολικής πιθανότητας, την πιθανότητα του ενδεχομένου  $\Theta$ : το αποτέλεσμα του τεστ είναι θετικό και βρήκαμε  $P(\Theta) = 0.0466$ . Ένα ερώτημα που επίσης (και κυρίως) ενδιαφέρει είναι το εξής: αν πραγματοποιηθεί το ενδεχόμενο  $\Theta$ , δηλαδή, αν για έναν τυχαία επιλεγμένο κάτοικο που υποβλήθηκε στο TB tine τεστ το αποτέλεσμα βρεθεί θετικό, ποια είναι η πιθανότητα ο εξεταζόμενος πράγματι να είναι φορέας του βακίλου της φυματίωσης και ποια να μην είναι;

Απάντηση: Υπενθυμίζουμε ότι έχουμε ορίσει το ενδεχόμενο  $B$ : ο εξεταζόμενος είναι φορέας του βακίλου της φυματίωσης. Ζητάμε τις δεσμευμένες πιθανότητες,  $P(B|\Theta)$  και  $P(B'|\Theta)$ .

Από τον τύπο (4.1.1) της δεσμευμένης πιθανότητας και αξιοποιώντας τον πολλαπλασιαστικό τύπο, εύκολα παίρνουμε

$$P(B|\Theta) = \frac{P(B\Theta)}{P(\Theta)} = \frac{P(\Theta|B)P(B)}{P(\Theta)} = \frac{0.92 \cdot 0.0075}{0.0466} = \frac{0.0069}{0.0466} = 0.1480$$

οπότε  $P(B'|\Theta) = 1 - P(B|\Theta) = 1 - 0.1480 = 0.8520$ .

Δηλαδή, αν για έναν τυχαία επιλεγμένο κάτοικο που υποβλήθηκε στο TB tine τεστ το αποτέλεσμα βρεθεί θετικό, η πιθανότητα ο κάτοικος αυτός να είναι πράγματι φορέας του βακίλου της φυματίωσης είναι μόλις 0.1480 και η πιθανότητα να μην είναι

φορέας είναι 0.8520. Απρόσμενο και εκ πρώτης όψεως περίεργο αποτέλεσμα, δε νομίζετε; (δείτε όμως το Σχόλιο 4.4.1 στη συνέχεια).

Τη ζητούμενη εκ των υστέρων πιθανότητα  $P(B|\Theta)$  την υπολογίσαμε εφαρμόζοντας το **θεώρημα/τύπο του Bayes (Bayes theorem/formula)** αφού πρώτα το αποδείξαμε! Πράγματι, από τον τύπο ορισμού της δεσμευμένης πιθανότητας, φθάνουμε στο **θεώρημα/τύπο του Bayes** έτσι απλά.

**Θεώρημα 4.4.1 (το θεώρημα του Bayes):** Αν  $\{B_1, B_2, \dots, B_\nu\}$  είναι μια διαμέριση του δειγματικού χώρου  $\Omega$  ενός πειράματος τύχης, με  $P(B_i) > 0$  για όλα τα  $i=1, 2, \dots, \nu$ , τότε, για κάθε ενδεχόμενο  $A$  του  $\Omega$  με  $P(A) > 0$ , ισχύει ότι

$$P(B_i | A) = \frac{P(A | B_i)P(B_i)}{\sum_{i=1}^{\nu} P(A | B_i)P(B_i)}, \quad i=1, 2, \dots, \nu. \quad (4.4.1)$$

Απόδειξη: Ο τύπος (4.4.1) προκύπτει άμεσα από τον τύπο ορισμού της δεσμευμένης πιθανότητας

$$P(B_i | A) = \frac{P(B_i A)}{P(A)}$$

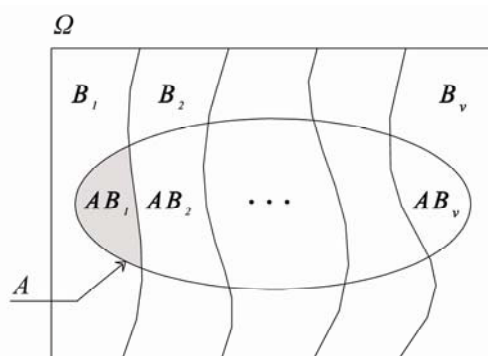
αν αναλύσουμε τον παρανομαστή σύμφωνα με τον τύπο της ολικής πιθανότητας και αντικαταστήσουμε τον αριθμητή με βάση τον πολλαπλασιαστικό τύπο

$$P(B_i A) = P(A | B_i)P(B_i).$$

Έτσι, για παράδειγμα, για τη δεσμευμένη πιθανότητα  $P(B_1 | A)$  παίρνουμε

$$P(B_1 | A) = \frac{P(A | B_1)P(B_1)}{P(A | B_1)P(B_1) + P(A | B_2)P(B_2) + \dots + P(A | B_\nu)P(B_\nu)}.$$

Παρατηρείστε ότι ο αριθμητής είναι ένας από τους όρους της ολικής πιθανότητας που εμφανίζεται στον παρανομαστή. (Σκεφθείτε πώς σχετίζεται η δεσμευμένη πιθανότητα  $P(B_1 | A)$  με την πιθανότητα του σκιαγραφημένου ενδεχομένου στο Σχήμα 4.4.1).



Σχήμα 4.4.1

Διαμέριση  $\{AB_1, AB_2, \dots, AB_\nu\}$  του  $A$

Επειδή (όπως μόλις διαπιστώσαμε) ο τύπος του Bayes προκύπτει ως μια απλή εφαρμογή του πολλαπλασιαστικού τύπου και του θεωρήματος ολικής πιθανότητας, ίσως, εκ πρώτης όψεως, φαίνεται να αποτελεί μια ήσσονος σημασίας επέκταση του τύπου ορισμού της δεσμευμένης πιθανότητας. Όμως δεν είναι έτσι. Πρόκειται για ένα αποτέλεσμα που έχει προκαλέσει μεγάλο και ζωηρό ενδιαφέρον, αφενός γιατί δίνει

απάντηση σε ενδιαφέροντα πρακτικά ερωτήματα και αφετέρου γιατί από μια ευρύτερη σκοπιά, θέτει βαθύτερα (και με φιλοσοφική διάσταση) ζητήματα και ερωτήματα που μάλιστα οδήγησαν στη δημιουργία «σχολής» και ιδιαίτερου κλάδου στη Στατιστική, τη *Μπεϋζιανή Στατιστική*. Όμως, οι σχετικές συζητήσεις προκάλεσαν και διαφωνίες και «σχίσμα» μεταξύ των στατιστικών! Στη συνέχεια, δε θα επεκταθούμε σε θέματα που τίθενται από τη σχετική συζήτηση. Θα χρησιμοποιήσουμε τον *τύπο του Bayes* για το σκοπό που αρχικά προοριζόταν από τον αιδεσιμότατο *Thomas Bayes*. Για τον υπολογισμό της «αντίστροφης πιθανότητας».

Αν για όλα τα  $i=1,2,\dots,n$  γνωρίζουμε τις δεσμευμένες πιθανότητες  $P(A|B_i)$ , το *θεώρημα του Bayes* μας επιτρέπει να υπολογίσουμε δεσμευμένες πιθανότητες στην «αντίστροφη κατεύθυνση», δηλαδή, να υπολογίσουμε τις  $P(B_i|A)$  από τις  $P(A|B_i)$ .

Πρέπει να διευκρινίσουμε ότι το *Θεώρημα του Bayes* όπως διατυπώθηκε προηγουμένως, δημοσιεύθηκε το 1812 από τον *Pierre-Simon Laplace*, στον οποίο οφείλεται και το όνομα του, προς τιμήν του Άγγλου ιερέα και μαθηματικού *Thomas Bayes (1701-1761)* του οποίου εργασία με το συγκεκριμένο αποτέλεσμα είχε δημοσιευθεί δύο χρόνια μετά το θάνατό του (*An essay towards solving a problem in the doctrine of chance, Philosophical Transactions of the Royal Society, 53, 1763*).

**Παράδειγμα 4.4.2 (συνέχεια του Παραδείγματος 4.3.1):** Υπενθυμίζουμε ότι στο Παράδειγμα 4.3.1 έχουμε ορίσει τα ενδεχόμενα

$B_i$ : ο κάτοικος ανήκει στην ηλικιακή ομάδα  $i, i=1,2,\dots,5$ .

$A$ : ο κάτοικος πίνει τουλάχιστον 3 καφέδες ημερησίως.

Για καθένα  $i=1,2,\dots,5$  γνωρίζουμε τη δεσμευμένη πιθανότητα  $P(A|B_i)$ , δηλαδή, γνωρίζουμε την πιθανότητα ένας ενήλικας που επιλέγεται τυχαία να πίνει τουλάχιστον τρεις καφέδες ημερησίως, δοθέντος ότι ανήκει στην  $i$  ηλικιακή ομάδα. Μας είναι επίσης γνωστές οι **εκ των προτέρων** πιθανότητες  $P(B_i)$  των ενδεχομένων  $B_i$ . Δηλαδή, γνωρίζουμε την εκ των προτέρων (πριν την εκτέλεση του πειράματος) πιθανότητα ένας ενήλικας που επιλέγεται τυχαία να ανήκει στην  $i$  ηλικιακή ομάδα.. Θα χρησιμοποιήσουμε τον *τύπο του Bayes* για να υπολογίσουμε για κάθε  $i=1,2,\dots,5$  την «αντίστροφη» **εκ των υστέρων** πιθανότητα  $P(B_i|A)$ , δηλαδή, την πιθανότητα ένας ενήλικας που επιλέγεται τυχαία να ανήκει στην  $i$  ηλικιακή ομάδα, δοθέντος ότι πίνει τουλάχιστον τρεις καφέδες ημερησίως.

Απάντηση: Στο Παράδειγμα 4.3.1, χρησιμοποιώντας τον τύπο της ολικής πιθανότητας, βρήκαμε

$$P(A) = \sum_{i=1}^5 P(A|B_i)P(B_i) = 0.2359.$$

Επομένως, από τον *τύπο του Bayes*

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^5 P(A|B_i)P(B_i)}, \quad i=1,2,\dots,5$$

παίρνουμε

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{\sum_{i=1}^5 P(A|B_i)P(B_i)} = \frac{0.26 \cdot 0.09}{0.2359} = 0.10$$

$$P(B_2 | A) = \frac{P(A | B_2)P(B_2)}{\sum_{i=1}^5 P(A | B_i)P(B_i)} = \frac{0.35 \cdot 0.20}{0.2359} = 0.30$$

$$P(B_3 | A) = \frac{P(A | B_3)P(B_3)}{\sum_{i=1}^5 P(A | B_i)P(B_i)} = \frac{0.10 \cdot 0.31}{0.2359} = 0.13$$

$$P(B_4 | A) = \frac{P(A | B_4)P(B_4)}{\sum_{i=1}^5 P(A | B_i)P(B_i)} = \frac{0.30 \cdot 0.23}{0.2359} = 0.29$$

$$P(B_5 | A) = \frac{P(A | B_5)P(B_5)}{\sum_{i=1}^5 P(A | B_i)P(B_i)} = \frac{0.25 \cdot 0.17}{0.2359} = 0.18.$$

Έτσι, ενώ οι εκ των προτέρων πιθανότητες να επιλεγεί ενήλικας από τις ηλικιακές ομάδες 18-24, 25-34, 35-49, 50-64 και  $\geq 65$  αντίστοιχα είναι 0.09, 0.20, 0.31, 0.23 και 0.17, οι αντίστοιχες εκ των υστέρων πιθανότητες, δοθέντος ότι ο ενήλικας που επελέγη πίνει τουλάχιστον τρεις καφέδες ημερησίως, είναι 0.10, 0.30, 0.13, 0.29 και 0.18. Παρατηρείστε τη μεγάλη μεταβολή της πιθανότητας να επιλεγεί ενήλικας από την ηλικιακή ομάδα 35-49 δοθέντος ότι πίνει τουλάχιστον τρεις καφέδες ημερησίως ( $|P(B_3 | A) - P(B_3)| = |0.13 - 0.31| = 0.18$ ). Αυτό συμβαίνει γιατί επιλέγεται ενήλικας που πίνει τουλάχιστον τρεις καφέδες ημερησίως ενώ το ποσοστό των ενηλίκων σε αυτή την ηλικιακή ομάδα που πίνει τουλάχιστον τρεις καφέδες ημερησίως, είναι μικρό σε σχέση με τις άλλες ομάδες (το μικρότερο). Επίσης, μεγάλη μεταβολή παρατηρείται στην ηλικιακή ομάδα 25-34 αλλά στην αντίθετη κατεύθυνση. Σκεφθείτε γιατί (παρατηρείστε ότι σε αυτή την ομάδα το ποσοστό των ενηλίκων που πίνει τουλάχιστον τρεις καφέδες ημερησίως είναι το μεγαλύτερο).

■

Μια ενδιαφέρουσα για τις εφαρμογές (και όχι μόνο) ερμηνεία του θεωρήματος του Bayes είναι η εξής:

*Αν τα ενδεχόμενα  $B_i, i = 1, 2, \dots, \nu$  θεωρηθούν ως «αιτίες», ο τύπος του Bayes εκφράζει την πιθανότητα, το «αποτέλεσμα»  $A$  (δηλαδή, η εμφάνιση του ενδεχομένου  $A$ ) να οφείλεται στην «αιτία»  $B_i$ . Έτσι, όταν μας είναι γνωστό το «αποτέλεσμα»  $A$ , ο τύπος του Bayes μας επιτρέπει να οδηγηθούμε σε λογικά συμπεράσματα για την «αιτία» που το προκάλεσε.*

Για παράδειγμα, η εκ των υστέρων πιθανότητα  $P(B | \Theta) = 0.1480$  που υπολογίσαμε προηγουμένως στο Παράδειγμα 4.4.1 μπορεί να ερμηνευθεί ως εξής: Υπάρχει 14.8% πιθανότητα το αποτέλεσμα του τεστ να βγήκε θετικό επειδή το άτομο είναι φορέας. Διευκρινίζουμε ότι το «αποτέλεσμα» εδώ είναι ότι το τεστ (για το άτομο που επελέγη) βγήκε θετικό και η εικαζόμενη «αιτία» είναι ότι το άτομο είναι φορέας. Σκεφθείτε (για εξάσκηση) πώς μπορούν να ερμηνευθούν, ανάλογα, οι εκ των υστέρων πιθανότητες που υπολογίσαμε στο Παράδειγμα 4.4.2.

Πριν δώσουμε μερικά ακόμη παραδείγματα εφαρμογής του τύπου του Bayes, κρίνουμε σκόπιμο να κάνουμε ένα σχόλιο για την τιμή της εκ των υστέρων πιθανότητας  $P(B | \Theta)$  που υπολογίσαμε στο Παράδειγμα 4.4.1.

**Σχόλιο 4.4.1:** Το TB tine τεστ είναι ένα πολύ αξιόπιστο τεστ, αφού κάνει σωστή διάγνωση με πιθανότητα 0.92 αν ο εξεταζόμενος είναι φορέας και με πιθανότητα 0.96 αν ο εξεταζόμενος δεν είναι φορέας. Επίσης, κάνει **ψευδώς θετική διάγνωση**, δηλαδή δίνει θετικό αποτέλεσμα ενώ ο εξεταζόμενος δεν είναι φορέας, με μικρή πιθανότητα, ίση με 0.04 ( $P(\Theta|B')=1-P(\Theta'|B')=1-0.96=0.04$ ). Εντούτοις, βρήκαμε ότι  $P(B|\Theta)=0.1480$ , δηλαδή, βρήκαμε ότι από τα άτομα που το τεστ έχει δώσει θετικό αποτέλεσμα (ή αλλιώς, που το τεστ έχει διαγνώσει ότι είναι φορείς), μόνο το 14.8% είναι πράγματι φορείς, ή ισοδύναμα, από τα άτομα που το τεστ έχει δώσει θετικό αποτέλεσμα, το 85.2% **δεν είναι** φορείς. Ομολογουμένως πολύ παράξενο αποτέλεσμα. Πώς εξηγείται τόσο μεγάλη αστοχία και φυσικά γεννάται το ερώτημα, ποια η σκοπιμότητα ενός τέτοιου διαγνωστικού τεστ με τόσο μεγάλο ποσοστό λανθασμένων διαγνώσεων. Η εξήγηση βρίσκεται στο ότι το ποσοστό των φορέων στον πληθυσμό είναι πολύ μικρό, μόλις 0.75% ( $P(B)=0.0075$ ) και επομένως το ποσοστό όσων δεν είναι φορείς είναι πολύ μεγάλο (99.25%) με αποτέλεσμα να είναι πολλές οι θετικές διαγνώσεις που οφείλονται σε λάθη του τεστ. Θεωρήστε, για παράδειγμα, ότι στη συγκεκριμένη περιοχή υπάρχουν 10000 κάτοικοι. Φορείς αναμένεται να είναι μόλις 75 ( $0.0075 \cdot 10000 = 75$ ) ενώ αναμένεται να μην είναι φορείς 9925. Έτσι, το TB tine τεστ αναμένεται να δώσει 69 σωστές θετικές διαγνώσεις (από τους 75 κατοίκους που είναι φορείς,  $0.92 \cdot 75 = 69$ ), αλλά και 397 λάθος/ψευδώς θετικές διαγνώσεις (από τους 9925 κατοίκους που δεν είναι φορείς,  $0.04 \cdot 9925 = 397$ ). Δηλαδή, από τις συνολικά  $69 + 397 = 466$  θετικές διαγνώσεις, οι συντριπτικά περισσότερες είναι ψευδώς θετικές, παρότι το τεστ δίνει ψευδώς θετικό αποτέλεσμα με πολύ μικρή πιθανότητα (0.04). Παρατηρήστε ότι αυτό συμβαίνει γιατί είναι πολύ περισσότεροι οι κάτοικοι που δεν είναι φορείς, αφού είναι σπάνιο το ενδεχόμενο να είναι κάποιος φορέας. Έτσι το ποσοστό των σωστών θετικών διαγνώσεων αναμένεται ίσο με  $69/466 = 0.1480$ , ενώ των ψευδώς θετικών  $397/466 = 0.8519$ . Δείτε επίσης στον πίνακα που ακολουθεί πόσο πολύ επηρεάζεται η πιθανότητα  $P(B|\Theta)$  (δηλαδή, το ποσοστό των σωστών θετικών διαγνώσεων) από την πιθανότητα  $P(B)$  ο εξεταζόμενος να είναι φορέας.

	$P(B)$			
	0.0075	0.01	0.02	0.05
$P(B \Theta)$	0.1480	0.1885	0.3194	0.5476

Σε ότι αφορά τη σκοπιμότητα πραγματοποίησης του τεστ επισημαίνουμε ότι η αξία του μπορεί να εκτιμηθεί καλύτερα αν παρατηρήσουμε ότι μετά το θετικό αποτέλεσμα η εκ των υστέρων πιθανότητα το άτομο να είναι φορέας, παρότι είναι μικρή (14.8%), εντούτοις είναι περίπου 20 φορές μεγαλύτερη από την εκ των προτέρων (πριν τη θετική διάγνωση) πιθανότητα να είναι φορέας (0.75%) γεγονός το οποίο μπορεί να υποδεικνύει τι είδους παρακολούθηση πρέπει να γίνει στη συνέχεια (επιπλέον εξετάσεις κτλ). Βέβαια, για τη σκοπιμότητα πραγματοποίησης διαγνωστικών τεστ για σπάνιες και ασυμπτωματικές ασθένειες λαμβάνονται υπόψη και άλλες παράμετροι όπως για παράδειγμα αν το τεστ έχει επιπτώσεις στην υγεία του εξεταζόμενου (αν γίνεται, π.χ., χρήση ακτινοβολίας) κ.ά., όμως δε θα επεκταθούμε περισσότερο.

Ολοκληρώνοντας, σημειώνουμε ότι στις επιδημιολογικές αλλά και γενικότερα στις ιατρικές έρευνες, για την αξιοπιστία και την διαγνωστική αξία των διαγνωστικών ελέγχων χρησιμοποιείται ειδική ορολογία. Έτσι, αν θεωρήσουμε τα ενδεχόμενα

$B$ : ο εξεταζόμενος είναι ασθενής

$\Theta$ : το αποτέλεσμα του τεστ είναι θετικό,

η πιθανότητα  $P(\Theta|B)$  το τεστ να κάνει σωστή διάγνωση όταν ο εξεταζόμενος είναι ασθενής ονομάζεται **ευαισθησία (sensitivity)** του τεστ και η πιθανότητα  $P(\Theta'|B')$  το τεστ να δώσει σωστό αποτέλεσμα όταν ο εξεταζόμενος δεν είναι ασθενής ονομάζεται **ειδικότητα ή πιστότητα (specificity)** του τεστ. Η **προβλεπτική ή διαγνωστική αξία (predictive value)** του τεστ ορίζεται με τις εκ των υστέρων (όταν γίνει γνωστό το αποτέλεσμα του τεστ) πιθανότητες  $P(B|\Theta)$  και  $P(B'|\Theta')$  που υπολογίζονται όπως είδαμε από το θεώρημα του Bayes με βάση την ευαισθησία και την ειδικότητα του τεστ (και τη διάδοση της ασθένειας στον πληθυσμό).

**Παράδειγμα 4.4.3:** Σε ένα θερμοκήπιο έχει εγκατασταθεί σύστημα συναγερμού έκτακτης ανάγκης. Σύμφωνα με τις προδιαγραφές του κατασκευαστή, όταν παρουσιασθεί κατάσταση έκτακτης ανάγκης ο συναγερμός χτυπά με πιθανότητα 0.90 ενώ χτυπά και όταν δεν παρουσιάζεται κατάσταση έκτακτης ανάγκης με πιθανότητα 0.01. Επίσης, έχει εκτιμηθεί ότι η πιθανότητα να παρουσιασθεί στο θερμοκήπιο κατάσταση έκτακτης ανάγκης είναι 0.002. Αν ο συναγερμός μόλις χτύπησε, ποια είναι η πιθανότητα να έχει πράγματι παρουσιασθεί κατάσταση έκτακτης ανάγκης.

Απάντηση: Ας θεωρήσουμε τα ενδεχόμενα

$A$ : ο συναγερμός χτυπάει

$B$ : στο θερμοκήπιο έχει παρουσιασθεί κατάσταση έκτακτης ανάγκης.

Δίνεται ότι  $P(B) = 0.002$ , δηλαδή, μας είναι γνωστή η **εκ των προτέρων** πιθανότητα  $P(B)$  να παρουσιασθεί κατάσταση έκτακτης ανάγκης και ζητείται να υπολογίσουμε την **εκ των υστέρων** πιθανότητα  $P(B|A)$ , δηλαδή, την πιθανότητα να έχει παρουσιασθεί κατάσταση έκτακτης ανάγκης δοθέντος ότι χτύπησε ο συναγερμός. Δίνεται επίσης ότι  $P(A|B) = 0.90$  και  $P(A|B') = 0.01$ . Από τον τύπο του Bayes παίρνουμε

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B')P(B')} = \frac{0.90 \cdot 0.002}{0.90 \cdot 0.002 + 0.01 \cdot 0.998} = 0.1528.$$

Δηλαδή, η πιθανότητα να χτύπησε ο συναγερμός επειδή υπάρχει κατάσταση έκτακτης ανάγκης, είναι μόλις 15.28%. Έτσι, όπως και στην περίπτωση του ιατρικού διαγνωστικού τεστ που είδαμε προηγουμένως (δες Σχόλιο 4.4.1), παρά τις πολύ καλές προδιαγραφές του συναγερμού, αν ακούσουμε να χτυπάει ο συναγερμός, η πιθανότητα να έχει συμβεί κατάσταση έκτακτης ανάγκης είναι μόνο 15.28%.

Άραγε, αν στο θερμοκήπιο εγκαταστήσουμε ακόμη καλύτερο συναγερμό (και μάλλον πιο ακριβό) θα αυξηθεί η  $P(B|A)$ ; Δηλαδή, θα αυξηθεί η πιθανότητα, όταν χτυπήσει ο συναγερμός η αιτία να είναι ότι υπάρχει κατάσταση έκτακτης ανάγκης; Όπως φαίνεται και στον πίνακα που ακολουθεί, αυξάνοντας την ευαισθησία μόνο του συναγερμού, η αύξηση της διαγνωστικής ικανότητάς του είναι πολύ μικρή ή μηδενική. Όπως εξηγήσαμε και στο Σχόλιο 4.4.1 το πρόβλημα βρίσκεται στο ότι το ενδεχόμενο  $B$  είναι σπάνιο.

	$P(A B)$			
	0.90	0.95	0.97	0.99
$P(B A)$	0.1538	0.1610	0.1610	0.1610

## Μια μοντέρνα εφαρμογή του Θεωρήματος του Bayes Φίλτρα Spam που βασίζονται στο Θεώρημα του Bayes (Bayesian Spam Filters)

Τα *Spam* είναι ανεπιθύμητα *e-mails* που κατακλύζουν τα ηλεκτρονικά *mailboxes* δημιουργώντας προβλήματα στους χρήστες ηλεκτρονικού ταχυδρομείου αλλά και στα συστήματα διαχείρισης ηλεκτρονικού ταχυδρομείου. Προέκυψε έτσι η ανάγκη ανάπτυξης εργαλείων λογισμικού τα οποία να φιλτράρουν τα εισερχόμενα *e-mails* και να απορρίπτουν τα *Spam*. Πολλά από τα εργαλεία που αναπτύχθηκαν για το σκοπό αυτό (όπως το *SpamAssassin* ή το *ASSP*) βασίζονται στο *Θεώρημα του Bayes*<sup>2</sup>!! Ας δούμε πώς.

Η βασική ιδέα για την ανάπτυξη αυτών των φίλτρων είναι ότι κάποιες λέξεις όπως «*opportunity*», «*offer*», «*special*», ή συνδυασμοί λέξεων όπως «*enhance performance*», μπορεί να μαρτυρούν ότι ένα μήνυμα που περιέχει κάποια ή κάποιες από αυτές τις λέξεις είναι *Spam*. Αν επομένως απαντηθεί το ερώτημα «*ποια είναι η πιθανότητα, να είναι Spam ένα μήνυμα που διαπιστώσαμε ότι περιέχει μια ή περισσότερες τέτοιες λέξεις*» και βρεθεί ότι η πιθανότητα αυτή είναι μεγάλη (μεγαλύτερη από κάποιο επίπεδο που θέτουμε π.χ. 95%) τότε ένα τέτοιο μήνυμα μπορεί να απορριφθεί από το φίλτρο, δηλαδή, να θεωρηθεί *Spam*. Βέβαια, ένα τέτοιο φίλτρο μπορεί να κάνει λάθη. Δηλαδή, μπορεί ένα μήνυμα να το θεωρήσει *Spam* ενώ δεν είναι, καθώς επίσης ένα μήνυμα να μην το θεωρήσει *Spam* ενώ είναι. Αυτό που επιδιώκεται είναι να ελαχιστοποιείται η πιθανότητα να θεωρηθεί ένα μήνυμα *Spam* ενώ δεν είναι. Είναι προφανές ότι τέτοια φίλτρα μπορούν να βασίζονται σε μία ή περισσότερες λέξεις ή σε έναν ή περισσότερους συνδυασμούς λέξεων. Ας δούμε ένα απλό φίλτρο που βασίζεται σε μια μόνο λέξη.

Έστω «*w*» μια τέτοια λέξη και ας υποθέσουμε ότι σε μια χρονική περίοδο φθάνει σε έναν *mail server* ένα σύνολο μηνυμάτων. Κάθε μήνυμα από αυτά είναι ή δεν είναι *Spam*. Έτσι, αν *S* είναι το υποσύνολο των μηνυμάτων που είναι *Spam* τότε προφανώς το *S'* είναι το υποσύνολο των μηνυμάτων που δεν είναι *Spam*. Μπορούμε να μετρήσουμε σε πόσα από τα μηνύματα του υποσυνόλου *S* και σε πόσα από τα μηνύματα του υποσυνόλου *S'* εμφανίζεται η λέξη «*w*» και έτσι να εκτιμήσουμε αφενός την πιθανότητα: *ένα μήνυμα που είναι Spam να περιέχει τη λέξη «w»* και αφετέρου την πιθανότητα: *ένα μήνυμα που δεν είναι Spam να περιέχει τη λέξη «w»*. Επίσης, μπορούμε να εκτιμήσουμε την πιθανότητα: *ένα μήνυμα που φθάνει στον mail server είναι Spam* και την πιθανότητα: *ένα μήνυμα που φθάνει στον mail server δεν είναι Spam*.

Ένα μήνυμα φθάνει στον *mail server* και διαπιστώνεται ότι περιέχει τη λέξη «*w*». Αν *E* το ενδεχόμενο: το μήνυμα περιέχει τη λέξη «*w*», τότε όπως αναφέραμε προηγουμένως οι πιθανότητες  $P(E/S)$ ,  $P(E/S')$ ,  $P(S)$  και  $P(S')$  μπορούν να εκτιμηθούν και επειδή τα ενδεχόμενα *S* και *S'* διαμερίζουν το σύνολο όλων των μηνυμάτων, από το *Θεώρημα του Bayes* έχουμε

$$P(S/E) = \frac{P(E/S)P(S)}{P(E/S)P(S) + P(E/S')P(S')}$$

Ας δούμε ένα αριθμητικό παράδειγμα.

<sup>2</sup> Η ιδέα παρουσιάστηκε για πρώτη φορά το 1998 στο συνέδριο AAI από τους Sahami, Dumais, Heckerman & Horvitz (*A Bayesian approach to filtering junk e-mail*) και προσέλκυσε το ενδιαφέρον μόλις από το 2002 με τη δημοσίευση από τον Paul Graham του άρθρου «*A Plan for Spam*» ([www.paulgraham.com/spam.html](http://www.paulgraham.com/spam.html)).



**Παράδειγμα 4.4.4 (Μπεϋζιανά φίλτρα Spam):** Βρέθηκε ότι από τα 3000 μηνύματα που έφθασαν μια χρονική περίοδο σε έναν mail server, τα 2000 είναι Spam και τα 1000 δεν είναι Spam. Βρέθηκε επίσης ότι η λέξη «Rolex» εμφανίσθηκε σε 250 από τα 2000 μηνύματα που είναι Spam και σε 5 από τα 1000 μηνύματα που δεν είναι Spam. Ένα μήνυμα φθάνει στον mail server και διαπιστώνουμε ότι περιέχει τη λέξη «Rolex». Ποια είναι η πιθανότητα το μήνυμα αυτό να είναι Spam.

Απάντηση: Θεωρώντας ότι ο αριθμός των επαναλήψεων του πειράματος (3000) είναι αρκετά μεγάλος ώστε να έχει επιτευχθεί σταθεροποίηση των σχετικών συχνοτήτων έχουμε

$$P(S) = \frac{2000}{3000} = 0.67, \quad P(S') = 1 - P(S) = 0.33, \quad P(E/S) = \frac{250}{2000} = 0.125 \quad \text{και}$$

$$P(E/S') = \frac{5}{1000} = 0.005.$$

Επομένως η ζητούμενη πιθανότητα είναι

$$P(S/E) = \frac{0.125 \cdot 0.67}{0.125 \cdot 0.67 + 0.005 \cdot 0.33} = 0.98.$$

Έτσι, αν ως επίπεδο απόρριψης ενός μηνύματος που περιέχει τη λέξη «Rolex» θέσουμε το 0.95 τότε το μήνυμα απορρίπτεται ως Spam.

Ανάλογα υπολογίζουμε τις αντίστοιχες πιθανότητες για φίλτρα που ελέγχουν περισσότερες από μια λέξεις. ■

#### 4.5 Ανεξάρτητα ενδεχόμενα

Σε όλα τα προηγούμενα παραδείγματα, όπως διαπιστώσαμε, οι δεσμευμένες πιθανότητες διαφέρουν από τις αντίστοιχες μη δεσμευμένες. Για παράδειγμα, η εκ των υστέρων (δεσμευμένη) πιθανότητα να έχει ένα άτομο αχρωματοψία δοθέντος ότι είναι άνδρας διαφέρει, και μάλιστα είναι μεγαλύτερη, από την εκ των προτέρων (μη δεσμευμένη) πιθανότητα ένα άτομο να έχει αχρωματοψία (Παράδειγμα 4.1.4). Δηλαδή, αν  $A$  και  $B$  τα ενδεχόμενα, αντίστοιχα, «το άτομο πάσχει από αχρωματοψία» και «το άτομο είναι άνδρας», τότε

$$P(A|B) > P(A)$$

ενώ

$$P(A|B') < P(A),$$

δηλαδή, η εκ των υστέρων πιθανότητα να έχει ένα άτομο αχρωματοψία δοθέντος ότι είναι γυναίκα, είναι μικρότερη από την αντίστοιχη εκ των προτέρων.

Επίσης, στο Παράδειγμα 4.1.6 είδαμε ότι για δύο ξένα ενδεχόμενα  $A, B$  (με  $P(A) > 0$  και  $P(B) > 0$ ) είναι

$$P(A|B) = 0 \quad \text{και} \quad P(B|A) = 0,$$

δηλαδή, η γνώση ότι εμφανίσθηκε το ένα από τα δύο, αποκλείει την εμφάνιση του άλλου (προφανώς αναμενόμενο για ξένα ενδεχόμενα).

Στο ίδιο παράδειγμα δείξαμε επίσης ότι αν για δύο ενδεχόμενα  $A, B$  είναι  $B \subseteq A$ , τότε

$$P(A|B) = 1$$

(αναμενόμενο, επίσης, αφού η εμφάνιση του  $B$  συνεπάγεται την εμφάνιση του  $A$ ).

Σε όλες επομένως αυτές τις περιπτώσεις, η πιθανότητα εμφάνισης του ενδεχομένου που μας ενδιαφέρει, επηρεάζεται από τη γνώση ότι κάποιο άλλο ενδεχόμενο συνέβη (ή δε συνέβη).

Υπάρχουν όμως περιπτώσεις όπου

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

δηλαδή, περιπτώσεις όπου η γνώση ότι συνέβη ή δε συνέβη ένα ενδεχόμενο  $B$  (με  $P(B) > 0$ ) δε δίνει καμία πληροφορία για την εμφάνιση ενός άλλου ενδεχομένου  $A$ , η αλλιώς, η πραγματοποίηση ή μη του  $B$  **δεν έχει καμία επίδραση** στην πραγματοποίηση του  $A$ . Λέμε τότε ότι το ενδεχόμενο  $A$  είναι **ανεξάρτητο** από το ενδεχόμενο  $B$ . Μάλιστα, εύκολα προκύπτει ότι τότε και  $P(B|A) = P(B)$  (αν  $P(A) > 0$ ), δηλαδή, και το ενδεχόμενο  $B$  είναι ανεξάρτητο από το ενδεχόμενο  $A$ .

Πράγματι

$$\begin{aligned} P(A|B) = P(A) &\Leftrightarrow \frac{P(AB)}{P(B)} = P(A) \Leftrightarrow P(AB) = P(A)P(B) \Leftrightarrow \\ &\Leftrightarrow \frac{P(AB)}{P(A)} = P(B) \Leftrightarrow P(B|A) = P(B). \end{aligned}$$

Δηλαδή, αν η πραγματοποίηση του  $B$  δεν έχει καμία επίδραση στην πραγματοποίηση του  $A$  τότε και η πραγματοποίηση του  $A$  δεν έχει καμία επίδραση στην πραγματοποίηση του  $B$ . Αυτό σημαίνει ότι η ανεξαρτησία ενδεχομένων είναι μια συμμετρική σχέση και γι' αυτό από τις παραπάνω ισοδύναμες σχέσεις, για τον ορισμό της επιλέχθηκε η  $P(AB) = P(A)P(B)$  που αποδίδει καλύτερα αυτή τη συμμετρικότητα (αλλά και γιατί αποφεύγονται οι συνθήκες  $P(A) > 0$  και  $P(B) > 0$ ).

**Ορισμός 4.5.1 (ανεξαρτησία δύο ενδεχομένων):** Δύο ενδεχόμενα  $A$  και  $B$  του δειγματικού χώρου  $\Omega$  ενός πειράματος τύχης λέγονται **ανεξάρτητα** ή **στοχαστικά/στατιστικά ανεξάρτητα (independent events)** αν ισχύει

$$P(AB) = P(A)P(B) \quad (4.5.1)$$

ενώ αν ισχύει

$$P(AB) \neq P(A)P(B),$$

τα ενδεχόμενα  $A$  και  $B$  λέγονται **εξαρτημένα (dependent events)**. ■

Με βάση το νόημα της ανεξαρτησίας, αν δύο ενδεχόμενα είναι ανεξάρτητα, λογικά, περιμένουμε και τα συμπληρώματά τους να είναι επίσης ανεξάρτητα. Πράγματι έτσι είναι.

**Πρόταση 4.5.1:** Αν  $A$  και  $B$  είναι δύο ανεξάρτητα ενδεχόμενα του δειγματικού χώρου  $\Omega$  ενός πειράματος τύχης τότε είναι ανεξάρτητα και τα ενδεχόμενα α)  $A'$  και  $B'$  β)  $A$  και  $B'$  γ)  $A'$  και  $B$ .

Απόδειξη: (α) Θα δείξουμε ότι αν δύο ενδεχόμενα  $A$  και  $B$  είναι ανεξάρτητα τότε είναι ανεξάρτητα και τα συμπληρώματά τους,  $A'$  και  $B'$ .

Για την πιθανότητα της ένωσης των  $A'$  και  $B'$  έχουμε

$$P(A' \cup B') = P(A') + P(B') - P(A'B').$$

Επίσης, (θυμηθείτε τους τύπους De Morgan) ισχύει ότι

$$P(A' \cup B') = P[(AB)'] = 1 - P(AB)$$

και επομένως

$1 - P(AB) = P(A') + P(B') - P(A'B') \Leftrightarrow P(A'B') = P(AB) - 1 + P(A') + P(B')$  και επειδή  $P(AB) = P(A)P(B)$  (αφού τα  $A$  και  $B$  είναι ανεξάρτητα), από την τελευταία σχέση παίρνουμε

$$P(A'B') = P(A)P(B) - 1 + 1 - P(A) + 1 - P(B) = [1 - P(A)][1 - P(B)] = P(A')P(B').$$

Δείξαμε δηλαδή, ότι  $P(A'B') = P(A')P(B')$  άρα πράγματι τα  $A'$  και  $B'$  είναι ανεξάρτητα.

Την απόδειξη των (β) και (γ) δείτε την ως μια απλή άσκηση (χρησιμοποιείστε τις σχέσεις  $A = AB \cup AB'$  και  $B = BA \cup BA'$ , αντίστοιχα).

**Σημείωση 4.5.1:** α) Σημειώνουμε ότι αν  $P(A|B) = P(A|B')$  τότε προφανώς θα είναι και  $P(A|B) = P(A|B') = P(A)$ . Πράγματι,

$$P(A) = P(A|B)P(B) + P(A|B')P(B') = P(A|B)[P(B) + P(B')] = P(A|B).$$

β) Είναι προφανές ότι το αδύνατο ενδεχόμενο,  $\emptyset$ , είναι ανεξάρτητο από οποιοδήποτε άλλο ενδεχόμενο  $A$  του  $\Omega$ . Πράγματι,

$$P(A\emptyset) = P(A)P(\emptyset) \Leftrightarrow P(\emptyset) = P(A)P(\emptyset) \Leftrightarrow 0 = P(A) \cdot 0 \Leftrightarrow 0 = 0.$$

Επίσης, και το βέβαιο ενδεχόμενο  $\Omega$  είναι ανεξάρτητο από οποιοδήποτε ενδεχόμενο  $A$ . Πράγματι,

$$P(A\Omega) = P(A)P(\Omega) \Leftrightarrow P(A) = P(A) \cdot 1 \Leftrightarrow P(A) = P(A).$$

(Δείτε και την Άσκηση 4.22).

Η έννοια της ανεξαρτησίας ενδεχομένων επεκτείνεται για τρία ή περισσότερα ενδεχόμενα. Όμως ας δούμε πρώτα κάποια παραδείγματα για την ανεξαρτησία δύο ενδεχομένων.

**Παράδειγμα 4.5.1:** Ρίχνουμε ένα αμερόληπτο νόμισμα δύο φορές και έστω τα ενδεχόμενα

$A$ : η ένδειξη στη δεύτερη ρίψη είναι «κεφαλή»

$B$ : η ένδειξη στην πρώτη ρίψη είναι «γράμματα».

Ένας κατάλληλος δειγματικός χώρος του πειράματος είναι προφανώς το σύνολο  $\Omega = \{κκ, κγ, γκ, γγ\}$  όπου, με «κκ» συμβολίζουμε το δειγματικό σημείο «στην πρώτη και στη δεύτερη ρίψη η ένδειξη είναι «κεφαλή»», με «κγ» το δειγματικό σημείο «στην πρώτη ρίψη η ένδειξη είναι «κεφαλή» και στη δεύτερη ρίψη «γράμματα»» κ.ο.κ.

Ας δούμε αν τα ενδεχόμενα  $A$  και  $B$  είναι ανεξάρτητα ή εξαρτημένα. Επειδή το νόμισμα είναι αμερόληπτο, τα απλά ενδεχόμενα του  $\Omega$  μπορούμε να τα θεωρήσουμε ισοπίθανα και επομένως έχουμε

$$P(A) = P(\{κκ, γκ\}) = \frac{2}{4} = \frac{1}{2},$$

$$P(B) = P(\{γκ, γγ\}) = \frac{2}{4} = \frac{1}{2} \text{ και}$$

$$P(AB) = P(\{γκ\}) = \frac{1}{4}.$$

Δηλαδή,

$$P(A)P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = P(AB)$$

και επομένως τα ενδεχόμενα  $A$  και  $B$  είναι ανεξάρτητα.

**Σχόλιο 4.5.1:** Το ότι τα ενδεχόμενα  $A$  και  $B$  στο προηγούμενο παράδειγμα είναι ανεξάρτητα, είναι βέβαια προφανές και προκύπτει αβίαστα από τις συνθήκες του πειράματος και το νόημα της ανεξαρτησίας, αφού το  $A$  συνδέεται μόνο με τη δεύτερη ρίψη και το  $B$  μόνο με την πρώτη, επομένως είναι λογικό το αποτέλεσμα της δεύτερης ρίψης να μην επηρεάζεται από το αποτέλεσμα της πρώτης (αλλά και αντίστροφα), ή αλλιώς, η πραγματοποίηση ή μη του  $B$  δεν επηρεάζει την πραγματοποίηση του  $A$ . Δηλαδή, από τα χαρακτηριστικά, τις συνθήκες και τους περιορισμούς του πειράματος είναι προφανές ότι

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

Αξίζει να επισημάνουμε ότι, στην πράξη, αυτό συμβαίνει συχνά σε πολλά προβλήματα και εφαρμογές. Δηλαδή, αν  $A$  και  $B$  δύο ενδεχόμενα, είναι δυνατόν από τους περιορισμούς και τις συνθήκες του πειράματος να προκύπτει, πέραν κάθε αμφιβολίας, ότι η πραγματοποίηση ή μη του ενός δεν έχει καμία επίδραση στην πραγματοποίηση ή μη του άλλου. Σε αυτές τις περιπτώσεις αναγνωρίζουμε και δεχόμαστε την ανεξαρτησία των  $A$  και  $B$  από τις συνθήκες του πειράματος και δεν ελέγχουμε τη σχέση ανεξαρτησίας  $P(AB) = P(A)P(B)$ . Μάλιστα, δεχόμενοι από τις συνθήκες του πειράματος ότι τα  $A$  και  $B$  είναι ανεξάρτητα, μπορούμε να χρησιμοποιήσουμε τη σχέση αυτή για τον υπολογισμό της πιθανότητας της τομής  $AB$  των  $A$  και  $B$  που πιθανόν μας ενδιαφέρει (δείτε και το παράδειγμα που ακολουθεί).

**Παράδειγμα 4.5.2:** Δύο φοιτητές, ας πούμε ο  $A$  και ο  $B$ , κάνουν την καλοκαιρινή πρακτική τους άσκηση ως διορθωτές κειμένων σε έναν εκδοτικό οίκο. Σύμφωνα με ένα τεστ ικανοτήτων, ο  $A$  εντοπίζει ένα λάθος συλλαβισμού (χωρισμού λέξης) με πιθανότητα 0.65 ενώ ο  $B$  εντοπίζει ένα λάθος συλλαβισμού με πιθανότητα 0.80. Σε καθένα φοιτητή δόθηκε να ελέγξει ένα κείμενο που είχε ένα λάθος συλλαβισμού (το κείμενο ήταν ίδιο και για τους δύο φοιτητές). Ποια είναι η πιθανότητα να μην εντοπισθεί το λάθος;

Απάντηση: Ας θεωρήσουμε τα ενδεχόμενα

$A$ : ο φοιτητής  $A$  εντοπίζει το λάθος

$B$ : ο φοιτητής  $B$  εντοπίζει το λάθος.

Ζητάμε την πιθανότητα  $P(A'B')$ .

Επειδή οι διορθωτές κατά κανόνα εργάζονται μόνοι και όχι σε συνεργασία, είναι λογικό να θεωρήσουμε ότι το αν θα εντοπίσει ή όχι ο ένας φοιτητής το λάθος δεν έχει καμία συνέπεια/επίδραση στο αν ο άλλος φοιτητής θα εντοπίσει ή όχι το λάθος, ή αλλιώς, είναι λογικό να θεωρήσουμε ότι η πραγματοποίηση ή μη του ενός από τα ενδεχόμενα  $A$  και  $B$  δε μπορεί να έχει κάποια επίδραση στην πραγματοποίηση ή μη του άλλου. Πρόκειται επομένως για ανεξάρτητα ενδεχόμενα και συνεπώς είναι ανεξάρτητα και τα  $A'$  και  $B'$ , άρα

$$P(A'B') = P(A')P(B') = 0.35 \cdot 0.20 = 0.07.$$

**Παράδειγμα 4.5.3:** Ένα εργοστάσιο έχει δύο γραμμές παραγωγής  $\Gamma 1$  και  $\Gamma 2$ . Στο τμήμα ποιοτικού ελέγχου όταν διαπιστώνεται ότι ένα προϊόν είναι ελαττωματικό κατατάσσεται σε μια από δύο κατηγορίες  $E1$  και  $E2$  ανάλογα με τη σοβαρότητα του ελαττώματος και καταγράφεται από ποια γραμμή παραγωγής προέρχεται. Στον πίνακα που ακολουθεί φαίνεται πώς κατανέμονται 1440 ελαττωματικά προϊόντα στις δύο γραμμές παραγωγής  $\Gamma 1$ ,  $\Gamma 2$  και στις δύο κατηγορίες  $E1$ ,  $E2$ .

	$\Gamma 1$	$\Gamma 2$	Σύνολο
$E1$	500	300	800
$E2$	400	240	640
Σύνολο	900	540	1440

Ας θεωρήσουμε τα ενδεχόμενα

$\Gamma 1$ : το προϊόν προέρχεται από τη γραμμή παραγωγής  $\Gamma 1$

$\Gamma 2$ : το προϊόν προέρχεται από τη γραμμή παραγωγής  $\Gamma 2$

$E1$ : το προϊόν ανήκει στην κατηγορία  $E1$

$E2$ : το προϊόν ανήκει στην κατηγορία  $E2$ .

Τα ενδεχόμενα  $\Gamma 2$  και  $E2$  είναι άραγε ανεξάρτητα ή εξαρτημένα;

Απάντηση: Με βάση το στατιστικό ορισμό της πιθανότητας έχουμε

$$P(\Gamma 2) = \frac{540}{1440} = 0.3750, \quad P(E2) = \frac{640}{1440} = 0.4444 \quad \text{και} \quad P(\Gamma 2 E2) = \frac{240}{1440} = 0.1666$$

και επειδή

$$P(\Gamma 2)P(E2) = 0.3750 \cdot 0.4444 = 0.1666 = P(\Gamma 2 E2),$$

τα  $\Gamma 2$  και  $E2$  είναι ανεξάρτητα.

Αν εργασθούμε με δεσμευμένες πιθανότητες, ασφαλώς οδηγούμαστε στο ίδιο συμπέρασμα. Πράγματι,

$$P(\Gamma 2 | E2) = \frac{P(\Gamma 2 E2)}{P(E2)} = \frac{240/1440}{640/1440} = \frac{240}{640} = 0.3750 = P(\Gamma 2)$$

και ασφαλώς

$$P(E2 | \Gamma 2) = \frac{P(E2 \Gamma 2)}{P(\Gamma 2)} = \frac{240/1440}{540/1440} = \frac{240}{540} = 0.4444 = P(E2)$$

δηλαδή, η γνώση της πραγματοποίησης του ενός ενδεχομένου δεν επηρεάζει την πραγματοποίηση του άλλου και επομένως αυτά είναι ανεξάρτητα.

Εφόσον τα ενδεχόμενα  $\Gamma 2$  και  $E2$  είναι ανεξάρτητα, μήπως και τα  $\Gamma 1$  και  $E1$  είναι επίσης ανεξάρτητα; Η απάντηση είναι ναι και μπορούμε να το τεκμηριώσουμε χωρίς να χρειασθεί να υπολογίσουμε τις κατάλληλες πιθανότητες από τα δεδομένα του πίνακα. Παρατηρείστε ότι το  $\Gamma 1$  είναι το συμπλήρωμα του  $\Gamma 2$  (αφού ένα ελαττωματικό προϊόν είτε θα προέρχεται από τη γραμμή  $\Gamma 1$  είτε από τη  $\Gamma 2$ ) και ότι το  $E1$  είναι το συμπλήρωμα του  $E2$  (αφού ένα ελαττωματικό προϊόν είτε θα κατατάσσεται στην κατηγορία  $E1$  είτε στην  $E2$ ). Επομένως, με βάση την Πρόταση 4.5.1, τα  $\Gamma 1$  και  $E1$  θα είναι ανεξάρτητα ως τα συμπληρώματα των ανεξάρτητων ενδεχομένων  $\Gamma 2$  και  $E2$ . Ανεξάρτητα θα είναι επίσης και τα  $\Gamma 2$ ,  $E1$  αλλά και τα  $\Gamma 1$ ,  $E2$  λόγω της Πρότασης 4.5.1 επίσης. Όπως θα δούμε σε επόμενο κεφάλαιο, η ανεξαρτησία αυτών των ζευγών ενδεχομένων σημαίνει (και συνεπάγεται) ότι είναι **ανεξάρτητα τα χαρακτηριστικά «γραμμή παραγωγής από την οποία προέρχεται το προϊόν» και «κατηγορία που ανήκει το προϊόν»**. Σημειώνουμε επίσης ότι ένας διδιάστατος πίνακας όπως ο παραπάνω, όπου τα δεδομένα στις στήλες αφορούν ένα χαρακτηριστικό και τα δεδομένα στις γραμμές ένα άλλο, ονομάζεται **πίνακας συνάφειας (contingency table)**. Στο Β' Μέρος θα αναφερθούμε στους πίνακες συνάφειας πιο αναλυτικά. ■

**Ορισμός 4.5.2 (ανεξαρτησία τριών ενδεχομένων):** Τρία ενδεχόμενα  $A$ ,  $B$  και  $\Gamma$  του δειγματικού χώρου  $\Omega$  ενός πειράματος τύχης λέγονται **ανεξάρτητα** ή **τελείως ανεξάρτητα** αν ισχύουν οι σχέσεις

$$\begin{aligned} P(AB) &= P(A)P(B) \\ P(A\Gamma) &= P(A)P(\Gamma) \\ P(B\Gamma) &= P(B)P(\Gamma) \end{aligned} \quad (4.5.2)$$

και

$$P(AB\Gamma) = P(A)P(B)P(\Gamma) \quad (4.5.3).$$

■ Παρατηρείστε ότι, παρότι οι συνθήκες (4.5.2) εξασφαλίζουν ότι τρία ενδεχόμενα  $A$ ,  $B$  και  $\Gamma$  είναι ανεξάρτητα ανά δύο, εντούτοις, δεν αρκούν για τον ορισμό της ανεξαρτησίας τριών ενδεχομένων. Απαιτείται επιπλέον και η (4.5.3).

Σκεφθείτε ότι αν ένα ενδεχόμενο  $A$  είναι ανεξάρτητο από ένα ενδεχόμενο  $B$  και επίσης από ένα ενδεχόμενο  $\Gamma$ , αυτό δε σημαίνει ότι είναι ανεξάρτητο και από την τομή τους  $B\Gamma$  (δες Άσκηση 4.25). Δηλαδή, ενώ η πραγματοποίηση του  $B$  δεν επηρεάζει την πραγματοποίηση του  $A$  και επίσης ούτε η πραγματοποίηση του  $\Gamma$  επηρεάζει την πραγματοποίηση του  $A$ , εντούτοις, η ταυτόχρονη πραγματοποίηση των  $B$  και  $\Gamma$  μπορεί να επηρεάζει την πραγματοποίηση του  $A$ !!! Αυτό δε μπορεί βέβαια να είναι επιθυμητό για ανεξάρτητα ενδεχόμενα, γιατί προφανώς, με βάση το νόημα της ανεξαρτησίας δύο ενδεχομένων, η γενίκευσή της για τρία ενδεχόμενα, πρέπει να διασφαλίζει ότι η πραγματοποίηση οποιουδήποτε από τα τρία ενδεχόμενα δεν επηρεάζει τις πιθανότητες πραγματοποίησης των υπολοίπων αλλά πρέπει επίσης να διασφαλίζει ότι και η πραγματοποίηση οποιωνδήποτε δύο από τα τρία δεν επηρεάζει τις πιθανότητες πραγματοποίησης των υπολοίπων. Δηλαδή, για να είναι τρία ενδεχόμενα, έστω  $A$ ,  $B$ ,  $\Gamma$ , (τελείως) ανεξάρτητα πρέπει να είναι ανεξάρτητα αφενός τα  $A$  και  $B$ ,  $A$  και  $\Gamma$ ,  $B$  και  $\Gamma$ , αλλά αφετέρου και τα  $A$  και  $B\Gamma$ ,  $B$  και  $A\Gamma$ , και τέλος, τα  $\Gamma$  και  $AB$ . Έτσι, προκύπτει η αναγκαιότητα και της (4.5.3) η οποία εκφράζει τη δεύτερη ομάδα συνθηκών (σκεφθείτε γιατί) και δεν προκύπτει από τις (4.5.2) (δες και Άσκηση 4.26).

Επίσης, ούτε η συνθήκη (4.5.3) αρκεί παρότι αποτελεί επέκταση για τρία ενδεχόμενα της συνθήκης ανεξαρτησίας (4.5.1) δύο ενδεχομένων γιατί δε διασφαλίζει ότι ισχύουν οι (4.5.2) (δες Άσκηση 4.27). Επομένως, όταν ελέγχεται η ανεξαρτησία τριών ενδεχομένων, πρέπει να ελέγχονται τόσο οι (4.5.2) όσο και η (4.5.3).

Σημειώνουμε επίσης, ότι αν για τρία ενδεχόμενα  $A$ ,  $B$  και  $\Gamma$  ικανοποιούνται οι συνθήκες (4.5.2) τότε λέμε ότι αυτά είναι **ανεξάρτητα ανά δύο** ή **κατά ζεύγη**. Από όσα προηγουμένως αναφέραμε, είναι προφανές ότι στα **ανεξάρτητα κατά ζεύγη** ενδεχόμενα, η γνώση για την ταυτόχρονη πραγματοποίηση δύο εξ αυτών μπορεί να επηρεάσει την πιθανότητα πραγματοποίησης του τρίτου ενώ στα **(τελείως) ανεξάρτητα** αυτό δε μπορεί να συμβεί.

Ας δούμε τώρα πώς ορίζεται η ανεξαρτησία  $n \geq 2$  ενδεχομένων.

**Ορισμός 4.5.3 (ανεξαρτησία  $n \geq 2$  ενδεχομένων):** Θα λέμε ότι  $n \geq 2$  ενδεχόμενα  $A_1, A_2, \dots, A_n$  του δειγματικού χώρου  $\Omega$  ενός πειράματος τύχης είναι **ανεξάρτητα** ή **τελείως ανεξάρτητα** αν για οποιοδήποτε σύνολο  $k$  δεικτών  $i_1, i_2, \dots, i_k$  από το σύνολο  $\{1, 2, \dots, n\}$  ισχύει

$$P(A_{i_1} A_{i_2} \cdots A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}).$$

■  
 Δηλαδή, λέμε ότι  $n \geq 2$  ενδεχόμενα του δειγματικού χώρου ενός πειράματος τύχης είναι *ανεξάρτητα* ή *τελείως ανεξάρτητα* αν η γνώση ότι πραγματοποιήθηκε ταυτόχρονα οποιοσδήποτε αριθμός από αυτά, δεν επηρεάζει τις πιθανότητες πραγματοποίησης των υπολοίπων ενδεχομένων. Αυτό βέβαια έχει ως συνέπεια να αυξάνει ραγδαία με το  $n$  ο αριθμός των συνθηκών που πρέπει να ικανοποιούνται. Για παράδειγμα, για  $n=4$  ανεξάρτητα ενδεχόμενα  $A_1, A_2, A_3, A_4$ , πρέπει να ικανοποιούνται

$$\binom{4}{2} = 6 \text{ συνθήκες της μορφής } P(A_i A_j) = P(A_i)P(A_j)$$

$$\binom{4}{3} = 4 \text{ συνθήκες της μορφής } P(A_i A_j A_k) = P(A_i)P(A_j)P(A_k) \text{ και}$$

$$\binom{4}{4} = 1 \text{ συνθήκη, η } P(A_1 A_2 A_3 A_4) = P(A_1)P(A_2)P(A_3)P(A_4), \text{ δηλαδή, συνολικά}$$

πρέπει να ικανοποιούνται 11 συνθήκες!

Γενικά, για  $n \geq 2$  ενδεχόμενα πρέπει να ικανοποιούνται

$$\binom{n}{2} + \binom{n}{3} + \dots + \binom{n}{n} = 2^n - n - 1 \text{ συνθήκες (θυμηθείτε το σχετικό παράδειγμα}$$

*εφαρμογής του τύπου του Νεύτωνα που δώσαμε όταν μιλήσαμε για τους διωνυμικούς συντελεστές).* Προφανώς ο αριθμός αυτός είναι μεγάλος ακόμη και για μικρά  $n$ , π.χ. αν  $n=5$  ενδεχόμενα, για να ελεγχθεί αν αυτά είναι τελείως ανεξάρτητα πρέπει να ελεγχθούν  $2^5 - 5 - 1 = 32 - 5 - 1 = 26$  συνθήκες!

Βέβαια, όπως εξηγήσαμε προηγουμένως (Σχόλιο 4.5.1), στην πράξη αναγνωρίζουμε και δεχόμαστε την ανεξαρτησία ενδεχομένων από τις συνθήκες και τους περιορισμούς του πειράματος και έτσι δεν κάνουμε έλεγχο των συνθηκών ανεξαρτησίας αλλά αντίθετα τις χρησιμοποιούμε για να υπολογίσουμε πιθανότητες τομών ενδεχομένων που πιθανόν μας ενδιαφέρουν.

Σημειώνουμε τέλος, ότι εύκολα προκύπτει ότι τα συμπληρώματα  $n \geq 2$  ανεξάρτητων ενδεχομένων είναι ανεξάρτητα όπως και οποιαδήποτε  $k$  από αυτά με τα συμπληρώματα των υπολοίπων  $n - k$  είναι επίσης ανεξάρτητα.

**Παράδειγμα 4.5.4:** Μια ασφαλιστική εταιρεία, στο πλαίσιο μιας πιλοτικής μελέτης για τις μελλοντικές υποχρεώσεις της σε αποζημιώσεις ασφαλιστήριων συμβολαίων ζωής, επέλεξε τρεις πελάτες από τους οποίους ένας είναι κάτοικος του λεκανοπεδίου Αττικής, ένας κάτοικος Θεσσαλονίκης και ο τρίτος κάτοικος Ρόδου. Σύμφωνα με τις εκτιμήσεις της εταιρείας, η πιθανότητα να ζήσει μέχρι το 2020 ο πρώτος εκτιμάται ίση με 0.6 και αντίστοιχα 0.9 και 0.3 ο δεύτερος και ο τρίτος. Ποια είναι η πιθανότητα, μέχρι το τέλος του 2019 η εταιρεία να πρέπει να πληρώσει αποζημίωση λόγω θανάτου ενός (ακριβώς) από τους τρεις πελάτες.

Απάντηση: Ας θεωρήσουμε τα ενδεχόμενα

$A_1$ : ο κάτοικος Αττικής το 2019 ζει

$A_2$ : ο κάτοικος Θεσσαλονίκης το 2019 ζει

$A_3$ : ο κάτοικος Ρόδου το 2019 ζει.

Ζητάμε την πιθανότητα του ενδεχομένου

$$E = (A'_1 A_2 A_3) \cup (A_1 A'_2 A_3) \cup (A_1 A_2 A'_3)$$

το οποίο είναι ένωση ξένων ανά δύο ενδεχομένων και επομένως

$$P(E) = P(A'_1 A_2 A_3) + P(A_1 A'_2 A_3) + P(A_1 A_2 A'_3).$$

Επειδή δεν έχουμε κάποιο λόγο να πιστεύουμε ότι τα ενδεχόμενα  $A_1$ ,  $A_2$  και  $A_3$  δεν είναι ανεξάρτητα και επειδή οι πιθανότητες των τομών ανεξάρτητων ενδεχομένων εκφράζονται ως γινόμενα πιθανοτήτων, έχουμε

$$\begin{aligned} P(E) &= P(A'_1)P(A_2)P(A_3) + P(A_1)P(A'_2)P(A_3) + P(A_1)P(A_2)P(A'_3) = \\ &= 0.4 \cdot 0.9 \cdot 0.3 + 0.6 \cdot 0.1 \cdot 0.3 + 0.6 \cdot 0.9 \cdot 0.7 = 0.504. \end{aligned}$$

■

**Παράδειγμα 4.5.5:** Έστω  $n \geq 2$  ανεξάρτητα ενδεχόμενα,  $A_1, A_2, \dots, A_n$ , του δειγματικού χώρου  $\Omega$  ενός πειράματος τύχης με  $P(A_i) = p_i$ ,  $i = 1, 2, \dots, n$ . Ας δούμε πώς υπολογίζουμε συναρτήσει των  $p_i$  την πιθανότητα α) να πραγματοποιηθούν όλα τα ενδεχόμενα  $A_i$ ,  $i = 1, 2, \dots, n$  β) να πραγματοποιηθεί τουλάχιστον ένα από τα  $A_i$ ,  $i = 1, 2, \dots, n$  γ) να μην πραγματοποιηθεί κανένα από τα  $A_i$ ,  $i = 1, 2, \dots, n$ .

Απάντηση: α) Ζητάμε την πιθανότητα του ενδεχομένου  $A_1 A_2 \dots A_n$ .

Επειδή τα  $A_1, A_2, \dots, A_n$  είναι ανεξάρτητα, έχουμε

$$P(A_1 A_2 \dots A_n) = P(A_1)P(A_2) \dots P(A_n) = p_1 p_2 \dots p_n = \prod_{i=1}^n p_i.$$

β) Ζητάμε την πιθανότητα του ενδεχομένου  $A_1 \cup A_2 \cup \dots \cup A_n$ .

Επειδή

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = 1 - P[(A_1 \cup A_2 \cup \dots \cup A_n)']$$

και επίσης επειδή

$$(A_1 \cup A_2 \cup \dots \cup A_n)' = A'_1 A'_2 \dots A'_n \quad (\text{De Morgan})$$

έχουμε

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = 1 - P(A'_1 A'_2 \dots A'_n).$$

Γνωρίζουμε επίσης, ότι τα ενδεχόμενα  $A_1, A_2, \dots, A_n$  είναι ανεξάρτητα, επομένως είναι ανεξάρτητα και τα  $A'_1, A'_2, \dots, A'_n$ , άρα η ζητούμενη πιθανότητα είναι

$$\begin{aligned} P(A_1 \cup A_2 \cup \dots \cup A_n) &= 1 - P(A'_1)P(A'_2) \dots P(A'_n) = 1 - (1 - p_1)(1 - p_2) \dots (1 - p_n) = \\ &= 1 - \prod_{i=1}^n (1 - p_i). \end{aligned}$$

γ) Ζητάμε την πιθανότητα του ενδεχομένου  $A'_1 A'_2 \dots A'_n$ , ή ισοδύναμα, του ενδεχομένου  $(A_1 \cup A_2 \cup \dots \cup A_n)'$

Προφανώς έχουμε

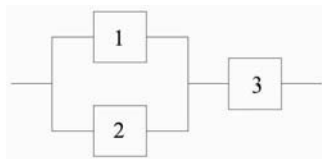
$$P[(A_1 \cup A_2 \cup \dots \cup A_n)'] = 1 - P(A_1 \cup A_2 \cup \dots \cup A_n) = \prod_{i=1}^n (1 - p_i).$$

■

**Παράδειγμα 4.5.6 (συνέχεια του Παραδείγματος 3.2.4):** Στο διάγραμμα που ακολουθεί, φαίνεται σχηματικά το σύστημα λήψης απόφασης σε μια γενική διεύθυνση ενός οργανισμού. Ο γενικός διευθυντής εγκρίνει μια εισήγηση εφόσον συμφωνήσει τουλάχιστον μια από τις διευθύνσεις 1, 2 καθώς και η διεύθυνση 3. Η πιθανότητα να εγκριθεί μια εισήγηση από τη διεύθυνση 1 εκτιμάται ίση με 0.7, από τη διεύθυνση 2 ίση



με 0.8 και από την διεύθυνση 3 ίση με 0.5. Ποια είναι η πιθανότητα να εγκριθεί από το γενικό διευθυντή η τελευταία εισήγηση που κατατέθηκε αν είναι γνωστό ότι οι διευθύνσεις του οργανισμού αποφασίζουν ανεξάρτητα;



Απάντηση: Ας θεωρήσουμε τα ενδεχόμενα

$\Sigma_i, i = 1, 2, 3$ : η διεύθυνση  $i$  συμφωνεί και  $E$ : η εισήγηση εγκρίνεται.

Ζητάμε την πιθανότητα του ενδεχομένου  $E$ . Προφανώς

$$E = (\Sigma_1 \cup \Sigma_2)\Sigma_3$$

επομένως για τη ζητούμενη πιθανότητα έχουμε

$$P(E) = P[(\Sigma_1 \cup \Sigma_2)\Sigma_3] = P(\Sigma_1 \cup \Sigma_2)P(\Sigma_3)$$

αφού τα  $\Sigma_1, \Sigma_2$  και  $\Sigma_3$  είναι ανεξάρτητα και επομένως θα είναι ανεξάρτητα και τα  $(\Sigma_1 \cup \Sigma_2)$  και  $\Sigma_3$  (δες Άσκηση 4.28).

Επίσης,

$$P(\Sigma_1 \cup \Sigma_2) = 1 - P[(\Sigma_1 \cup \Sigma_2)'] = 1 - P(\Sigma_1'\Sigma_2') = 1 - P(\Sigma_1')P(\Sigma_2') = 1 - 0.3 \cdot 0.2 = 0.94$$

επομένως, η ζητούμενη πιθανότητα είναι

$$P(E) = P(\Sigma_1 \cup \Sigma_2)P(\Sigma_3) = 0.94 \cdot 0.5 = 0.47.$$

Σημειώνουμε ότι την πιθανότητα του ενδεχομένου  $(\Sigma_1 \cup \Sigma_2)$  μπορούσαμε να την υπολογίσουμε εφαρμόζοντας απευθείας το αποτέλεσμα

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = 1 - \prod_{i=1}^n (1 - p_i)$$

του ερωτήματος (β) του Παραδείγματος 4.5.5 ή και από τον τύπο

$$P(\Sigma_1 \cup \Sigma_2) = P(\Sigma_1) + P(\Sigma_2) - P(\Sigma_1 \Sigma_2).$$

■

Θα ολοκληρώσουμε την ενότητα αυτή με την έννοια της **ανεξαρτησίας πειραμάτων** και των **επαναλαμβανόμενων ανεξάρτητων δοκιμών**. Τα ανεξάρτητα πειράματα και ειδικότερα οι επαναλαμβανόμενες ανεξάρτητες δοκιμές, βρίσκονται στον πυρήνα πολλών πιθανοθεωρητικών μοντέλων και (γι' αυτό) κατέχουν σημαντική θέση στη **Θεωρία Πιθανοτήτων**.

Αρκετά πειράματα, όπως ήδη έχουμε αναφέρει (δες Παράδειγμα 3.1.1δ), μπορούν να θεωρηθούν ως σύνθετα πειράματα που αναλύονται σε  $n$  επιμέρους απλούστερα ή σε μια σειρά  $n$  επαναλήψεων του ίδιου πειράματος. Για παράδειγμα, το πείραμα «ρίχνω ένα νόμισμα δύο φορές και στη συνέχεια ένα ζάρι» προφανώς είναι ένα πείραμα σύνθετο που αποτελείται από τρία απλούστερα πειράματα. Επίσης, το πείραμα «ελέγχεται καθένα από 10 προϊόντα για να διαπιστωθεί αν είναι ελαττωματικό ή όχι» είναι επίσης ένα πείραμα σύνθετο το οποίο αναλύεται σε 10 επαναλαμβανόμενες εκτελέσεις του ίδιου πειράματος, «ελέγχεται ένα προϊόν αν είναι ελαττωματικό ή όχι».

Έστω ότι ένα σύνθετο πείραμα αναλύεται στα  $\nu$  πειράματα  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_\nu$ , με δειγματικούς χώρους,  $\Omega_1, \Omega_2, \dots, \Omega_\nu$ , αντίστοιχα. Λέμε ότι τα  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_\nu$  είναι **ανεξάρτητα πειράματα (independent experiments)** αν δεν «αλληλοεπηρεάζονται», δηλαδή, αν η έκβαση οποιουδήποτε από αυτά, καθώς και η έκβαση οποιασδήποτε ομάδας από αυτά, δεν επηρεάζει την έκβαση των υπολοίπων. Έτσι, αν τα  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_\nu$  είναι **ανεξάρτητα πειράματα**, τότε για οποιαδήποτε ενδεχόμενα  $A_1 \subseteq \Omega_1$ ,  $A_2 \subseteq \Omega_2$ , ...,  $A_\nu \subseteq \Omega_\nu$ , όπου καθένα σχετίζεται μόνο με το αντίστοιχο πείραμα  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_\nu$ , ορίζουμε

$$P(A_1 A_2 \dots A_\nu) = P(A_1)P(A_2) \dots P(A_\nu).$$

Με  $A_1 A_2 \dots A_\nu$  συμβολίζουμε το ενδεχόμενο  $A_1 \times A_2 \times \dots \times A_\nu$  του δειγματικού χώρου  $\Omega_1 \times \Omega_2 \times \dots \times \Omega_\nu$  του σύνθετου πειράματος, δηλαδή, συμβολίζουμε το ενδεχόμενο το οποίο πραγματοποιείται όταν, κατά την εκτέλεση του πειράματος  $\varepsilon_1$  πραγματοποιηθεί το  $A_1$ , κατά την εκτέλεση του πειράματος  $\varepsilon_2$  πραγματοποιηθεί το  $A_2$ , ... και κατά την εκτέλεση του πειράματος  $\varepsilon_\nu$  πραγματοποιηθεί το  $A_\nu$ .

**Παράδειγμα 4.5.7:** Έστω το πείραμα  $\varepsilon$ : «ρίχνω ένα νόμισμα δύο φορές και στη συνέχεια ένα ζάρι». Προφανώς το πείραμα αυτό αναλύεται στα πειράματα

- $\varepsilon_1$ : «ρίχνω ένα νόμισμα» με δειγματικό χώρο  $\Omega_1 = \{\kappa, \gamma\}$ ,  
 $\varepsilon_2$ : «ρίχνω ένα νόμισμα» με δειγματικό χώρο  $\Omega_2 = \{\kappa, \gamma\}$  και  
 $\varepsilon_3$ : «ρίχνω ένα ζάρι» με δειγματικό χώρο  $\Omega_3 = \{1, 2, 3, 4, 5, 6\}$ .

Ο δειγματικός χώρος  $\Omega$  του (σύνθετου) πειράματος  $\varepsilon$ , προφανώς αποτελείται από όλες τις δυνατές διατεταγμένες τριάδες  $\alpha_1 \alpha_2 \alpha_3$ , όπου  $\alpha_i \in \{\kappa, \gamma\}$  για  $i=1, 2$  και  $\alpha_3 \in \{1, 2, 3, 4, 5, 6\}$ , δηλαδή,  $\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$ . Πρόκειται για **πεπερασμένο δειγματικό χώρο** με  $N(\Omega) = 24$  (σκεφθείτε γιατί). Πράγματι

$$\Omega = \{\kappa\kappa 1, \kappa\kappa 2, \kappa\kappa 3, \kappa\kappa 4, \kappa\kappa 5, \kappa\kappa 6, \kappa\gamma 1, \kappa\gamma 2, \kappa\gamma 3, \kappa\gamma 4, \kappa\gamma 5, \kappa\gamma 6, \gamma\kappa 1, \gamma\kappa 2, \gamma\kappa 3, \gamma\kappa 4, \gamma\kappa 5, \gamma\kappa 6, \gamma\gamma 1, \gamma\gamma 2, \gamma\gamma 3, \gamma\gamma 4, \gamma\gamma 5, \gamma\gamma 6\}.$$

Έτσι, αν θεωρήσουμε τρία ενδεχόμενα  $A_1 \subseteq \Omega_1$ ,  $A_2 \subseteq \Omega_2$  και  $A_3 \subseteq \Omega_3$ , έστω για παράδειγμα, τα ενδεχόμενα

$$A_1 = \{\kappa\}, A_2 = \{\gamma\} \text{ και } A_3 = \{1, 3, 5\} \text{ τότε } A_1 A_2 A_3 = \{\kappa\gamma 1, \kappa\gamma 3, \kappa\gamma 5\}$$

είναι προφανές ότι το ενδεχόμενο  $A_1 A_2 A_3$  του  $\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$  πραγματοποιείται όταν κατά την εκτέλεση του πειράματος  $\varepsilon_1$  εμφανισθεί «κεφαλή», κατά την εκτέλεση του πειράματος  $\varepsilon_2$  εμφανισθεί «γράμματα» και κατά την εκτέλεση του πειράματος  $\varepsilon_3$  εμφανισθεί περιττός αριθμός. Με την παραδοχή ότι το νόμισμα και το ζάρι είναι αμερόληπτα προφανώς έχουμε

$$P(A_1 A_2 A_3) = \frac{3}{24} = \frac{1}{8}, \quad P(A_1) = \frac{1}{2}, \quad P(A_2) = \frac{1}{2} \text{ και } P(A_3) = \frac{3}{6} = \frac{1}{2}.$$

Παρατηρείστε ότι για τα  $A_1 = \{\kappa\}$ ,  $A_2 = \{\gamma\}$  και  $A_3 = \{1, 3, 5\}$  η σχέση  $P(A_1 A_2 A_3) = P(A_1)P(A_2)P(A_3)$  επαληθεύεται. Εύκολα επίσης μπορείτε να δείξετε ότι η συνθήκη ανεξαρτησίας επαληθεύεται για οποιαδήποτε ενδεχόμενα των  $\Omega_1$ ,  $\Omega_2$

και  $\Omega_3$ . Το ότι τα  $\varepsilon_1$ ,  $\varepsilon_2$  και  $\varepsilon_3$  είναι ανεξάρτητα, είναι βέβαια προφανές αφού η έκβαση οποιουδήποτε από αυτά ή η έκβαση οποιωνδήποτε δύο από αυτά δεν είναι λογικό να δεχθούμε ότι επηρεάζει την έκβαση των υπολοίπων. ■

**Σημείωση 4.5.2:** Σε πρακτικές εφαρμογές, την ανεξαρτησία πειραμάτων συνήθως την ελέγχουμε με την εξέταση και τον έλεγχο των συνθηκών κάτω από τις οποίες αυτά εκτελούνται. Αυτό μας δίνει τη δυνατότητα, όταν διαπιστώνουμε ότι τα πειράματα είναι ανεξάρτητα, να μπορούμε να αξιοποιούμε (για τον υπολογισμό πιθανοτήτων) τη σχέση

$$P(A_1 A_2 \dots A_n) = P(A_1)P(A_2) \dots P(A_n).$$

Βέβαια, ο έλεγχος υποθέσεων και παραδοχών, όπως αυτός της ανεξαρτησίας, αποτελεί όπως θα δούμε στο Β' Μέρος, αντικείμενο της στατιστικής συμπερασματολογίας. Σημειώνουμε επίσης, ότι δεν γράφονται όλα τα ενδεχόμενα του δειγματικού χώρου  $\Omega_1 \times \Omega_2 \times \dots \times \Omega_n$  ως καρτεσιανά γινόμενα. Για το πώς υπολογίζουμε πιθανότητες τέτοιων ενδεχομένων όπως και για άλλα θέματα που τίθενται για τους «χώρους-γινόμενα» δε θα επεκταθούμε. ■

Εάν τα  $n$  πειράματα στα οποία αναλύεται ένα σύνθετο πείραμα είναι ανεξάρτητες επαναλήψεις του ίδιου πειράματος τότε λέμε ότι το πείραμα αποτελείται από  $n$  ανεξάρτητες δοκιμές ή ότι πρόκειται για μια σειρά ή ακολουθία  $n$  ανεξάρτητων δοκιμών ή για  $n$  επαναλαμβανόμενες ανεξάρτητες δοκιμές (*repeated independent trials*).

Όπως αναφέραμε στο προηγούμενο κεφάλαιο (δες Παράδειγμα 3.1.1α,δ), ένα πείραμα τύχης με δύο μόνο δυνατά αποτελέσματα (αμοιβαίως αποκλειόμενα) ονομάζεται *δοκιμή Bernoulli*. Το ένα αποτέλεσμα έχει επικρατήσει να ονομάζεται «επιτυχία» και το άλλο «αποτυχία». Έτσι, αν ένα πείραμα αναλύεται σε  $n$  διαδοχικές ανεξάρτητες επαναλήψεις ενός πειράματος με δύο μόνο δυνατά αποτελέσματα, λέμε ότι το πείραμα αποτελείται από  $n$  ανεξάρτητες δοκιμές *Bernoulli* ή από μια σειρά ή ακολουθία  $n$  ανεξάρτητων δοκιμών *Bernoulli*. Για παράδειγμα, η ρήψη ενός νομίσματος 10 φορές είναι ένα τυχαίο πείραμα που αναλύεται σε 10 ανεξάρτητες δοκιμές *Bernoulli*.

Αν συμβολίσουμε με «ε» το αποτέλεσμα «επιτυχία» και με «α» το αποτέλεσμα «αποτυχία» μιας δοκιμής *Bernoulli*, είναι προφανές (δες και Παράδειγμα 3.1.1δ), ότι τα στοιχεία του δειγματικού χώρου ενός πειράματος που αποτελείται από  $n$  ανεξάρτητες δοκιμές *Bernoulli*, είναι διατεταγμένες  $n$ -άδες από «επιτυχίες» και «αποτυχίες» δηλαδή έχουν τη μορφή  $\underline{\alpha \varepsilon \alpha \varepsilon \dots \varepsilon \alpha \varepsilon}$ .  
 $n$  το πλήθος

Στη συνέχεια, όταν θα μιλήσουμε για κατανομές τυχαίων μεταβλητών, θα διαπιστώσουμε ότι οι διαδοχικές ανεξάρτητες δοκιμές *Bernoulli* εμφανίζονται στην περιγραφή και την κατασκευή πολλών μοντέλων πιθανοτήτων με ευρύτατο φάσμα εφαρμογών. Προς το παρόν, ας δούμε μια ενδιαφέρουσα πρόταση ιδιαίτερα χρήσιμη για τη συνέχεια, όταν θα μιλήσουμε για μοντέλα πιθανοτήτων.

**Πρόταση 4.5.2:** Αν ένα πείραμα τύχης αποτελείται από  $n$  ανεξάρτητες δοκιμές *Bernoulli* όπου η πιθανότητα επιτυχίας σε κάθε μια δοκιμή είναι ίση με  $p$ , τότε η πιθανότητα, στις  $n$  δοκιμές

α) να εμφανισθούν  $v$  «επιτυχίες» είναι ίση με  $p^v$

β) να μην εμφανισθεί «επιτυχία» είναι ίση με  $(1-p)^v$

γ) να εμφανισθεί τουλάχιστον μια «επιτυχία» είναι ίση με  $1-(1-p)^v$

δ) να εμφανισθούν ακριβώς  $r$  «επιτυχίες» είναι ίση με  $\binom{v}{r} p^r (1-p)^{v-r}$ .

Απόδειξη: Έστω τα ενδεχόμενα

$A_i, i=1, 2, \dots, v$ : στην  $i$  δοκιμή εμφανίζεται «επιτυχία».

α) Ζητάμε την πιθανότητα του ενδεχομένου  $A_1 A_2 \dots A_v$ .

Από την ανεξαρτησία των δοκιμών προφανώς έχουμε

$$P(A_1 A_2 \dots A_v) = P(A_1)P(A_2) \dots P(A_v) = p \cdot p \cdot \dots \cdot p = p^v.$$

β) Ζητάμε την πιθανότητα του ενδεχομένου  $A'_1 A'_2 \dots A'_v$ .

Από την ανεξαρτησία των δοκιμών προφανώς έχουμε

$$P(A'_1 A'_2 \dots A'_v) = P(A'_1)P(A'_2) \dots P(A'_v) = (1-p) \cdot (1-p) \cdot \dots \cdot (1-p) = (1-p)^v.$$

γ) Ζητάμε την πιθανότητα του συμπληρώματος του ενδεχομένου  $A'_1 A'_2 \dots A'_v$ .

Επομένως η ζητούμενη πιθανότητα είναι

$$1 - (1-p)^v.$$

δ) Ο δειγματικός χώρος του πειράματος αποτελείται όπως είδαμε από διατεταγμένες  $v$ -άδες της μορφής  $\underbrace{\alpha \varepsilon \alpha \varepsilon \dots \varepsilon \alpha \varepsilon}_{v \text{ το πλήθος}}$ .

Ζητάμε την πιθανότητα του ενδεχομένου που αποτελείται από εκείνες τις διατεταγμένες  $v$ -άδες αυτής της μορφής, που περιέχουν  $r$  (ακριβώς) το πλήθος « $\varepsilon$ » και  $v-r$  το πλήθος « $\alpha$ ».

Όπως δείξαμε στο Παράδειγμα 3.1.1δ, το πλήθος αυτών των διατεταγμένων  $v$ -άδων είναι

$$\frac{v!}{r!(v-r)!} = \binom{v}{r}.$$

Δηλαδή, το ενδεχόμενο του οποίου ζητάμε την πιθανότητα αποτελείται από

$$\binom{v}{r}$$

$v$ -άδες που περιέχουν  $r$  (ακριβώς) το πλήθος « $\varepsilon$ » και  $v-r$  το πλήθος « $\alpha$ ». Η πιθανότητα εμφάνισης κάθε μιας τέτοιας  $v$ -άδας είναι, λόγω της ανεξαρτησίας των δοκιμών, προφανώς ίση με  $p^r (1-p)^{v-r}$  και επομένως η ζητούμενη πιθανότητα είναι πράγματι ίση με

$$\binom{v}{r} p^r (1-p)^{v-r}.$$

■

Ολοκληρώνουμε με μια ακόμη κατηγορία πειραμάτων που, όπως θα διαπιστώσουμε στο Β' Μέρος, συναντάται σε μεγάλο φάσμα εφαρμογών. Πρόκειται για τις **πολωνομικές δοκιμές (multinomial trials)**.

Η δοκιμή *Bernoulli*, όπως είδαμε, είναι ένα πείραμα με δύο μόνο (αμοιβαίως αποκλειόμενα) δυνατά αποτελέσματα,  $\varepsilon$  και  $\alpha$  («επιτυχία» και «αποτυχία», αντίστοιχα). Αν ένα πείραμα έχει  $k$  (αμοιβαίως αποκλειόμενα) δυνατά αποτελέσματα,  $r_1, r_2, \dots, r_k$ , ονομάζεται **πολωνομική δοκιμή**.

Είναι προφανές ότι η πολυωνυμική δοκιμή αποτελεί γενίκευση, μάλιστα εύλογη, της δοκιμής *Bernoulli* (μια δοκιμή *Bernoulli* είναι μια πολυωνυμική δοκιμή με  $k=2$  δυνατά αποτελέσματα). Έτσι, ενώ για παράδειγμα, η εξέταση ενός ατόμου για να διαπιστωθεί αν έχει ή δεν έχει προσβληθεί από ένα συγκεκριμένο ιό είναι μια δοκιμή *Bernoulli*, η εξέταση ενός ατόμου για να διαπιστωθεί η ομάδα αίματός του είναι μια πολυωνυμική δοκιμή με  $k=4$  δυνατά αποτελέσματα (A, B, AB και O). Επίσης, η υποβολή ερώτησης σε ένα άτομο για το αν συμφωνεί, διαφωνεί ή το αφήνει αδιάφορο μια συγκεκριμένη άποψη, είναι μια πολυωνυμική δοκιμή με  $k=3$  δυνατά αποτελέσματα όπως και ο έλεγχος ενός ελαττωματικού προϊόντος για να διαπιστωθεί από ποια από τις πέντε γραμμές παραγωγής παρήχθη είναι μια πολυωνυμική δοκιμή με  $k=5$  δυνατά αποτελέσματα.

**Πρόταση 4.5.3:** Αν ένα πείραμα τύχης αποτελείται από  $n$  ανεξάρτητες πολυωνυμικές δοκιμές, όπου σε κάθε δοκιμή τα δυνατά αποτελέσματα είναι  $k$ , έστω  $r_1, r_2, \dots, r_k$ , και η πιθανότητα να εμφανισθεί το αποτέλεσμα  $r_i$  σε μια (οποιαδήποτε) δοκιμή είναι  $p_i$ , τότε η πιθανότητα, στις  $n$  δοκιμές, να εμφανισθούν  $v_1$  αποτελέσματα  $r_1$ ,  $v_2$  αποτελέσματα  $r_2$ , ... και  $v_k$  αποτελέσματα  $r_k$  είναι ίση με

$$\frac{n!}{v_1! \cdot v_2! \cdot \dots \cdot v_k!} p_1^{v_1} p_2^{v_2} \dots p_k^{v_k},$$

όπου,

$$\sum_{i=1}^k v_i = n \text{ και } \sum_{i=1}^k p_i = 1.$$

■

Όπως θα διαπιστώσουμε όταν θα μιλήσουμε για τους στατιστικούς ελέγχους  $X^2$ , πρόκειται για ένα πολύ χρήσιμο αποτέλεσμα. Την απόδειξη δείτε την ως μια απλή άσκηση (αποτελεί πολύ απλή (ευθεία) γενίκευση της απόδειξης της Πρότασης 4.5.2δ).

#### 4.6 Σύντομη ανασκόπηση βασικών εννοιών, προτάσεων και τύπων

Δεσμευμένη πιθανότητα του $A$ δοθέντος του $B$	$P(A B) = \frac{P(AB)}{P(B)}$ , $P(B) > 0$ (δηλ. $P(B) \neq 0$ )
Ιδιότητες της δεσμευμένης πιθανότητας	<ol style="list-style-type: none"> <li><math>P(A B) \geq 0</math></li> <li><math>P(\Omega B) = 1</math></li> <li><math>P(A_1 \cup A_2 \cup \dots   B) = P(A_1 B) + P(A_2 B) + \dots</math> για <math>A_1, A_2, \dots</math> ξένα ανά δύο ενδεχόμενα.</li> </ol>
Άλλες ιδιότητες της δεσμευμένης πιθανότητας	$P(\emptyset B) = 0$ $P(A' B) = 1 - P(A B)$ $P(AG' B) = P(A B) - P(AG B)$ Αν $\Gamma \subseteq A$ τότε $P(\Gamma B) \leq P(A B)$ $P(A \cup \Gamma B) = P(A B) + P(\Gamma B) - P(AG B)$ Όταν $B \subseteq A$ τότε $P(A B) = 1$ .
Ο πολλαπλασιαστικός τύπος	$P(AB) = P(A)P(B A) = P(B)P(A B)$ αν $P(A) > 0$ , $P(B) > 0$ . Γενικότερα: αν $P(A_1 A_2 \dots A_{v-1}) > 0$ , $P(A_1 A_2 \dots A_v) = P(A_1)P(A_2 A_1) \dots P(A_v A_1 A_2 \dots A_{v-1})$
Το θεώρημα ολικής πιθανότητας	Αν $\{B_1, B_2, \dots, B_v\}$ μια διαμέριση του $\Omega$ με $P(B_i) > 0$ , $i = 1, 2, \dots, v$ , τότε $P(A) = P(A B_1)P(B_1) + P(A B_2)P(B_2) + \dots + P(A B_v)P(B_v)$
Ο τύπος του Bayes	Αν $\{B_1, B_2, \dots, B_v\}$ διαμέριση του $\Omega$ με $P(B_i) > 0$ για κάθε $i$ , τότε $P(B_i A) = \frac{P(A B_i)P(B_i)}{P(A)}$ , $i = 1, 2, \dots, v$ όπου $P(A)$ υπολογίζεται από το θεώρημα ολικής πιθανότητας.
Ανεξάρτητα ενδεχόμενα $A, B$	$P(AB) = P(A)P(B)$ . Αν $A, B$ ανεξάρτητα και $P(A) > 0$ , $P(B) > 0$ τότε $P(A B) = P(A)$ και $P(B A) = P(B)$
Εξαρτημένα ενδεχόμενα $A, B$	$P(AB) \neq P(A)P(B)$
Ανεξάρτητα ενδεχόμενα $A, B, \Gamma$	$P(AB) = P(A)P(B)$ , $P(A\Gamma) = P(A)P(\Gamma)$ $P(B\Gamma) = P(B)P(\Gamma)$ και $P(AB\Gamma) = P(A)P(B)P(\Gamma)$
Ανεξαρτησία και συμπληρωματικά ενδεχόμενα	Αν $A, B$ ανεξάρτητα τότε είναι ανεξάρτητα και τα ζεύγη α) $A, B'$ β) $A', B$ γ) $A', B'$
Ανεξάρτητα ενδεχόμενα και ξένα ενδεχόμενα	Αν $A, B$ ξένα (με $P(A) > 0$ και $P(B) > 0$ ) τότε: $P(A B) = 0 \neq P(A)$ και $P(B A) = 0 \neq P(B)$ συνεπώς τα $A, B$ είναι <b>εξαρτημένα</b> . Αν $A, B$ ξένα τότε: $P(AB) = 0$ και $P(A \cup B) = P(A) + P(B)$ . Αν $A, B$ <b>ανεξάρτητα</b> τότε: $P(AB) = P(A)P(B)$ και $P(A \cup B) = P(A) + P(B) - P(A)P(B)$
Ανεξάρτητα πειράματα $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_v$ με δειγματικούς χώρους, αντίστοιχα $\Omega_1, \Omega_2, \dots, \Omega_v$	Τα πειράματα $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_v$ είναι ανεξάρτητα αν η έκβαση οποιουδήποτε από αυτά, καθώς και η έκβαση οποιασδήποτε ομάδας από αυτά, δεν επηρεάζει την έκβαση των υπολοίπων. Για ανεξάρτητα πειράματα $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_v$ ισχύει ότι αν $A_1 \subseteq \Omega_1, \dots, A_v \subseteq \Omega_v$ τότε $P(A_1 A_2 \dots A_v) = P(A_1)P(A_2) \dots P(A_v)$ . (Το $A_1 A_2 \dots A_v$ πραγματοποιείται όταν κατά την εκτέλεση του πειράματος $\varepsilon_1$ πραγματοποιηθεί το $A_1, \dots$ και κατά την εκτέλεση του πειράματος $\varepsilon_v$ πραγματοποιηθεί το $A_v$ )

#### 4.7 Προβλήματα και Ασκήσεις

- 1 Έστω  $A$  και  $B$  δύο ενδεχόμενα ενός δειγματικού χώρου  $\Omega$  με  $P(B) > 0$ ,  $P(AB') = 0.3$ ,  $P[(A \cup B)'] = 0.2$  και  $P(AB) = 0.1$ . Να υπολογισθεί η δεσμευμένη πιθανότητα  $P(A|B)$ .
- 2 Γνωρίζουμε ότι το 8% των εργαζομένων σε μια εταιρεία είναι ανώτερα στελέχη. Επίσης γνωρίζουμε ότι το 30% των εργαζομένων στην εταιρεία είναι γυναίκες και ότι το 1.5% των εργαζομένων είναι γυναίκες ανώτερα στελέχη. Μεταξύ των εργαζομένων υπάρχει η εντύπωση ότι στο θέμα της επαγγελματικής εξέλιξής τους υπάρχει διάκριση εις βάρος των γυναικών. Τι λέτε, είναι βάσιμη αυτή η εντύπωση;
- 3 Μια κάλπη περιέχει ένα μαύρο και ένα άσπρο μπαλάκι. Εξάγουμε τυχαία ένα. Αν είναι άσπρο το επανατοποθετούμε στην κάλπη. Αν είναι μαύρο το επανατοποθετούμε στην κάλπη μαζί με άλλα δύο επίσης μαύρα μπαλάκια. Στη συνέχεια εξάγουμε ένα δεύτερο μπαλάκι. Ποια είναι η πιθανότητα και στις δύο εξαγωγές να εξαχθεί μαύρο μπαλάκι;
- 4 Συνέχεια του Προβλήματος 3.11: α) Από τα ζώα που έχουν προσβληθεί από μια τουλάχιστον ασθένεια ποιο ποσοστό έχει προσβληθεί και από τις δύο; β) Από τα ζώα που δεν έχουν προσβληθεί από την ασθένεια  $A$  ποιο ποσοστό έχει προσβληθεί από την ασθένεια  $B$ ;
- 5 Συνέχεια του Προβλήματος 3.12: α) Αν διαπιστωθεί ότι ένα ζώο είναι υγιές ποια είναι η πιθανότητα να είναι αρσενικό και ποια να είναι θηλυκό; β) Ποιο ποσοστό των άρρωστων ζώων είναι αρσενικά; γ) Να ελέγξετε αν ισχύει η ισότητα  $P(B|A') = 1 - P(B|A)$ .
- 6 Το 10% ενός πληθυσμού είναι άνεργοι. Επίσης, το 8% του πληθυσμού είναι γυναίκες άνεργες. Αν επιλέξουμε στην τύχη (από τον πληθυσμό αυτό) ένα άτομο και διαπιστώσουμε ότι είναι άνεργος, ποια είναι η πιθανότητα να είναι α) γυναίκα β) άνδρας.
- 7 Σε μια χώρα, η πιθανότητα να ζήσει ένας άνδρας τουλάχιστον 70 χρόνια είναι 0.85, ενώ η πιθανότητα να ζήσει τουλάχιστον 75 χρόνια είναι 0.80. Αν επιλέξουμε τυχαία έναν 70-χρονο άντρα από τη χώρα αυτή, ποια είναι η πιθανότητα να ζήσει τουλάχιστον άλλα 5 χρόνια.
- 8 **Το δίλημμα του φυλακισμένου:** Φυλακισμένος που έχει υποβάλει μαζί με άλλους δύο συγκρατούμενους του αίτηση απονομής χάριτος, πληροφορείται από ένα φίλο του φρουρό ότι δύο από τους τρεις πρόκειται να αποφυλακισθούν. Επειδή ο φρουρός δε θέλει να αποκαλύψει στον φυλακισμένο αν αυτός είναι ο ένας από τους δύο που αποφυλακίζονται, ο φυλακισμένος σκέπτεται να του ζητήσει να του αποκαλύψει ποιος από τους άλλους δύο πρόκειται να αποφυλακισθεί. Όμως διστάζει γιατί σκέπτεται ότι με την απάντηση του φρουρού μειώνεται η πιθανότητα αποφυλάκισής του από  $2/3$  σε  $1/2$ . Είναι οι δισταγμοί του φυλακισμένου δικαιολογημένοι;
- 9 Ένας φοιτητής πρέπει στο εξάμηνο που φοιτά να επιλέξει ως μάθημα επιλογής ή Χημεία ή Βιολογία. Αν επιλέξει Χημεία, εκτιμάει ότι η πιθανότητα να περάσει είναι  $1/2$  ενώ αν επιλέξει Βιολογία εκτιμάει ότι η πιθανότητα να περάσει είναι  $2/3$ .

Ποια είναι η πιθανότητα ο φοιτητής, στο εξάμηνο που φοιτά, να περάσει τη Βιολογία;

- 10 Σε ένα αγρόκτημα υπάρχουν 10 κουνέλια από τα οποία τα 3 είναι θηλυκά. Για τον έλεγχο του πληθυσμού των κουνελιών κρίθηκε σκόπιμο να απομακρυνθούν δύο από τα θηλυκά. Έτσι, στήθηκε μια παγίδα όπου πιάνονταν τα κουνέλια το ένα μετά το άλλο έως ότου πιαστούν 2 θηλυκά. Ποια είναι η πιθανότητα να συμβεί αυτό όταν πιαστεί το τέταρτο στη σειρά κουνέλι;
- 11 Από επτά όμοια κλειδιά ένα μόνο ανοίγει μια κλειδαριά. Δοκιμάζουμε χωρίς επανάθεση ένα-ένα τα κλειδιά μέχρι η κλειδαριά να ανοίξει. Ποια είναι η πιθανότητα να ανοίξει η κλειδαριά στην τρίτη δοκιμή; Γενικότερα στην  $k$  δοκιμή; ( $k = 1, 2, \dots, 7$ ).
- 12 Σε μια παρτίδα  $n$  προϊόντων υπάρχουν  $\alpha$  ελαττωματικά. Επιλέγουμε τυχαία δύο από αυτά τα προϊόντα το ένα μετά το άλλο και χωρίς επανάθεση και έστω το ενδεχόμενο  $A_i$ : το προϊόν  $i$  είναι ελαττωματικό,  $i = 1, 2$ . Να αποδείξετε ότι  $P(A_1) = P(A_2) = \alpha/n$ .
- 13 **Πολλαπλές γραμμές παραγωγής:** Σε ένα εργοστάσιο υπάρχουν τρεις διαφορετικές γραμμές παραγωγής στις οποίες κατασκευάζεται το 40%, 35% και 25% των προϊόντων του εργοστασίου αντίστοιχα. Το 2% των προϊόντων της πρώτης γραμμής είναι ελαττωματικά, ενώ τα αντίστοιχα ποσοστά για τις άλλες δύο γραμμές είναι 3% και 5%. Τα προϊόντα των τριών γραμμών παραγωγής αναμιγνύονται δημιουργώντας μια ενιαία σειρά και στη συνέχεια προωθούνται στο τμήμα ποιοτικού ελέγχου. α) Στο τμήμα ποιοτικού ελέγχου επιλέγεται τυχαία ένα προϊόν. Ποια είναι η πιθανότητα το προϊόν αυτό να είναι ελαττωματικό; β) Στο τμήμα ποιοτικού ελέγχου επιλέγεται τυχαία ένα προϊόν και διαπιστώνεται ότι είναι ελαττωματικό. Ποια είναι η πιθανότητα το προϊόν αυτό να προέρχεται από την πρώτη γραμμή παραγωγής; Ερμηνεύστε τις πιθανότητες που υπολογίσατε στα α) και β) με όρους ποσοστών (δηλαδή τι εκφράζει ως ποσοστό η κάθε πιθανότητα και επί ποίου συνόλου).
- 14 **Διαγνωστικά τεστ:** Το 2% ενός πληθυσμού πάσχει από AIDS. Η εξέταση που εφαρμόζεται για τη διάγνωση της ασθένειας δίνει σωστή διάγνωση στο 90% των περιπτώσεων, όταν το εξεταζόμενο άτομο πάσχει από AIDS, και στο 95% των περιπτώσεων όταν δεν πάσχει από AIDS. Επιλέγεται ένα άτομο από τον πληθυσμό αυτό στην τύχη και υποβάλλεται στην εξέταση. α) Ποια είναι η πιθανότητα η εξέταση να βγει θετική, δηλαδή να δείξει ότι πάσχει από AIDS. β) Ποια είναι η πιθανότητα λανθασμένης διάγνωσης γ) Ποια είναι η πιθανότητα να πάσχει πράγματι από AIDS ένα άτομο για το οποίο η εξέταση ήταν θετική δ) Ποια είναι η πιθανότητα να είναι υγιές ένα άτομο για το οποίο η εξέταση ήταν θετική.
- 15 **Διαγνωστικά τεστ:** Από μελέτες που έγιναν σε μια χώρα, διαπιστώθηκε ότι το ποσοστό των γυναικών που πάσχουν από καρκίνο της μήτρας είναι 1‰. Ένα από τα πλέον δημοφιλή τεστ για τη διάγνωση της ασθένειας, το τεστ Παπανικολάου, έχει ευαισθησία 98%, δηλαδή όταν μια γυναίκα πάσχει από καρκίνο της μήτρας το τεστ κάνει ορθή διάγνωση με πιθανότητα 0.98. Επίσης, και η ειδικότητα του τεστ Παπανικολάου είναι 98%, δηλαδή αν μια γυναίκα δεν πάσχει από καρκίνο της μήτρας, το τεστ κάνει ορθή διάγνωση με πιθανότητα επίσης 0.98. α) Αν μια



γυναίκα αυτής της χώρας υποβληθεί στο τεστ και βγει θετικό, ποια είναι η πιθανότητα η γυναίκα να έχει πράγματι καρκίνο της μήτρας. Είναι δικαιολογημένος ο υπερβολικός φόβος της κυρίας μετά το αποτέλεσμα του τεστ; Σχολιάστε επίσης, τη διαγνωστική αξία του τεστ σε σχέση με την τιμή της πιθανότητας που υπολογίσατε. β) Το ενδεχόμενο το τεστ να βγει θετικό και το ενδεχόμενο η εξεταζόμενη γυναίκα να έχει καρκίνο της μήτρας είναι ανεξάρτητα ή εξαρτημένα;

- 16 **Ερωτήσεις πολλαπλής επιλογής:** Σε μια εξέταση δίδονται τέσσερις απαντήσεις σε κάθε ερώτηση και σωστή είναι μόνο μια από τις τέσσερις. Η πιθανότητα να γνωρίζει ο εξεταζόμενος την απάντηση μιας ερώτησης είναι 80%. Στις περιπτώσεις που ο εξεταζόμενος δεν γνωρίζει την απάντηση σε μια ερώτηση, απαντάει εντελώς τυχαία διαλέγοντας μια από τις τέσσερις απαντήσεις που δίδονται. Αν ο εξεταζόμενος απαντήσει σωστά σε μια ερώτηση, ποια είναι η πιθανότητα να γνώριζε την απάντηση;
- 17 Μια ασφαλιστική εταιρεία ταξινομεί κάθε πελάτη της, ανάλογα με το ετήσιο ύψος των ζημιών που προκαλεί, σε μία από δύο κατηγορίες: είτε στην κατηγορία υψηλού κινδύνου είτε στην κατηγορία χαμηλού κινδύνου. Από τα στατιστικά στοιχεία που διατηρεί η εταιρεία προκύπτει ότι το 35% των πελατών της ανήκει στην κατηγορία υψηλού κινδύνου και το 65% στην κατηγορία χαμηλού κινδύνου. Επίσης, από τα στατιστικά στοιχεία, έχει εκτιμηθεί ότι ένας ασφαλισμένος που ανήκει στην κατηγορία υψηλού κινδύνου έχει 0.40 πιθανότητα να προκαλέσει ατύχημα κάποια χρονική στιγμή εντός ενός έτους ενώ για έναν ασφαλισμένο που ανήκει στην κατηγορία χαμηλού κινδύνου η αντίστοιχη πιθανότητα έχει εκτιμηθεί ίση με 0.2. α) Ποια είναι η πιθανότητα ένας νέος ασφαλισμένος στην εταιρεία να προκαλέσει ατύχημα κάποια χρονική στιγμή εντός ενός έτους β) Αν ένας νέος ασφαλισμένος στην εταιρεία προκαλέσει ατύχημα κάποια χρονική στιγμή εντός ενός έτους, ποια είναι η πιθανότητα ο ασφαλισμένος να ανήκει στην κατηγορία υψηλού κινδύνου. γ) Το ενδεχόμενο ένας ασφαλισμένος να ανήκει στην κατηγορία υψηλού κινδύνου και το ενδεχόμενο να προκαλέσει ατύχημα κάποια χρονική στιγμή εντός ενός έτους είναι εξαρτημένα ή ανεξάρτητα;
- 18 Η πιθανότητα να υπάρχει στο υπέδαφος μιας συγκεκριμένης περιοχής εκμεταλλεύσιμο κοιτάσμα πετρελαίου είναι 50%. Μια εταιρεία που πρόκειται να κάνει σχετική έρευνα στην περιοχή, προβληματίζεται αν θα εφαρμόσει, μεταξύ άλλων, και μια δαπανηρή μέθοδο της οποίας όμως το αποτέλεσμα δεν είναι πάντα σωστό. Αν πράγματι υπάρχει εκμεταλλεύσιμο κοιτάσμα, η μέθοδος αυτή δίνει ευοίωνο αποτέλεσμα (δηλαδή ανιχνεύει το κοιτάσμα) με πιθανότητα 80% ενώ αν δεν υπάρχει, δίνει ευοίωνο αποτέλεσμα (λανθασμένα εν προκειμένω) με πιθανότητα 30%. α) Αν η μέθοδος εφαρμοσθεί στην περιοχή, ποια είναι η πιθανότητα να δώσει ευοίωνο αποτέλεσμα; β) Αν η μέθοδος εφαρμοσθεί στην περιοχή και δώσει ευοίωνο αποτέλεσμα, ποια είναι η πιθανότητα να υπάρχει κοιτάσμα στην περιοχή; γ) Να συγκρίνετε την εκ των προτέρων πιθανότητα να υπάρχει κοιτάσμα στην περιοχή με την αντίστοιχη εκ των υστέρων πιθανότητα που υπολογίσατε στο ερώτημα (β) και να σχολιάσετε ως προς την ανεξαρτησία τα σχετικά ενδεχόμενα.
- 19 Ζητήσατε από έναν γείτονά σας να ποτίσει κατά τη διάρκεια των διακοπών σας ένα φυτό εσωτερικού χώρου που διατηρείτε στο σπίτι σας. Αν το φυτό δεν ποτισθεί, κινδυνεύει να μη ζήσει με πιθανότητα 0.8. Επειδή το φυτό είναι

ασθενικό-καχεκτικό, ακόμη και αν ποτισθεί, κινδυνεύει να μη ζήσει με πιθανότητα 0.15. Αν η πιθανότητα να θυμηθεί ο γείτονάς σας να ποτίσει το φυτό είναι 0.9 α) ποια είναι η πιθανότητα το φυτό να ζήσει β) ποια είναι η πιθανότητα ο γείτονας να ξέχασε να ποτίσει το φυτό δεδομένου ότι αυτό δεν επέζησε.

- 20 Το 60% του πληθυσμού μιας χώρας είναι καπνιστές. Από μια ασθένεια των πνευμόνων πάσχει το 80% των καπνιστών και το 30% των μη καπνιστών. α) Ποιο ποσοστό του πληθυσμού πάσχει από αυτή την ασθένεια. β) Αν ένα άτομο από τον πληθυσμό πάσχει από αυτή την ασθένεια, ποια είναι η πιθανότητα να είναι καπνιστής. Ερμηνεύστε την πιθανότητα αυτή ως ποσοστό. γ) Το ενδεχόμενο ένα άτομο να είναι καπνιστής και το ενδεχόμενο ένα άτομο να πάσχει από τη συγκεκριμένη ασθένεια είναι ανεξάρτητα ή εξαρτημένα;
- 21 Σε μια κτηνοτροφική μονάδα της Λέσβου υπάρχουν 160 αγελάδες. Από αυτές οι 40 είναι φυλής Α, οι 48 φυλής Β και οι 72 φυλής Γ. Αν από τις αγελάδες φυλής Α έχει προσβληθεί από αφθώδη πυρετό ποσοστό 15%, από τις αγελάδες φυλής Β, 20% και από τις αγελάδες φυλής Γ, 10% α) ποια είναι η πιθανότητα, μια αγελάδα που θα επιλέξουμε τυχαία να έχει προσβληθεί από αφθώδη πυρετό β) ποια είναι η πιθανότητα μια αγελάδα που επιλέξαμε τυχαία και βρέθηκε να έχει προσβληθεί από αφθώδη πυρετό, να είναι φυλής Α. Να ερμηνεύσετε τις πιθανότητες που υπολογίσατε στα (α) και (β) με όρους ποσοστών.
- 22 Να δείξετε ότι αν για ένα ενδεχόμενο  $A$  του δειγματικού χώρου  $\Omega$  ενός πειράματος τύχης είναι  $P(A)=0$  τότε αυτό είναι ανεξάρτητο από οποιοδήποτε ενδεχόμενο  $B$  του  $\Omega$ . Επίσης, να δείξετε ότι αν  $P(A)=1$  τότε το  $A$  είναι ανεξάρτητο από οποιοδήποτε ενδεχόμενο  $B$  του  $\Omega$ .
- 23 **Αξιοπιστία:** Σε ένα εργαστηριακό όργανο χρησιμοποιούνται δύο εξαρτήματα τα οποία λειτουργούν ανεξάρτητα το ένα από το άλλο. Στο 80% του χρόνου λειτουργίας του οργάνου, και τα δύο εξαρτήματα λειτουργούν χωρίς να παρουσιάζουν κάποια βλάβη. Όμως, στο 1% του χρόνου λειτουργίας, παρουσιάζουν βλάβη και τα δύο εξαρτήματα (ταυτόχρονα). Για να λειτουργήσει το όργανο πρέπει να λειτουργεί τουλάχιστον ένα από τα δύο εξαρτήματα. Να υπολογισθεί α) η αξιοπιστία του οργάνου (δηλαδή η πιθανότητα λειτουργίας του) και β) η αξιοπιστία καθενός από τα δύο εξαρτήματα (δηλαδή η πιθανότητα λειτουργίας καθενός εξαρτήματος).
- 24 Έστω  $A, B$  ενδεχόμενα του δειγματικού χώρου  $\Omega$  ενός πειράματος τύχης. Να συμπληρώσετε τον ακόλουθο πίνακα.

$P(A)$	$P(B)$	Συνθήκη για τα $A$ και $B$	$P(AB)$	$P(A \cup B)$	$P(A/B)$
0.3	0.4	Ξένα			
0.3	0.4	Ανεξάρτητα			
0.1	0.5			0.6	
0.2	0.5		0.1		

- 25 Ρίχνουμε ένα αμερόληπτο ζάρι δύο φορές και έστω τα ενδεχόμενα  $A$ : η ένδειξη στην πρώτη ρίψη είναι 4,  $B$ : η ένδειξη στη δεύτερη ρίψη είναι 3 και  $E$ : το άθροισμα των δύο ενδείξεων είναι 7. Να δείξετε ότι α) το  $E$  είναι ανεξάρτητο από το  $A$  β) το  $E$  είναι ανεξάρτητο από το  $B$  γ) το  $E$  δεν είναι ανεξάρτητο από το  $AB$ .

- 26 Έστω μια ρουλέτα που έχει 36 αριθμούς από τους οποίους οι 18 είναι κόκκινοι και οι 18 είναι μαύροι σύμφωνα με το ακόλουθο μοντέλο.

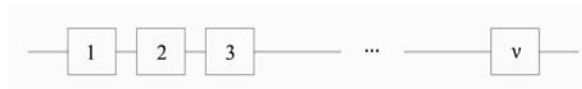
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
K	K	K	K	K	B	B	B	B	K	K	K	K	M	M	M	M	M
36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19

Έστω τα ενδεχόμενα A: εμφανίζεται κόκκινος αριθμός, B: εμφανίζεται άρτιος αριθμός και Γ: εμφανίζεται αριθμός μικρότερος ή ίσος του 18. Να δείξετε ότι α) τα ενδεχόμενα A, B και Γ είναι ανεξάρτητα κατά ζεύγη β)  $P(AB\Gamma) \neq P(A)P(B)P(\Gamma)$ .

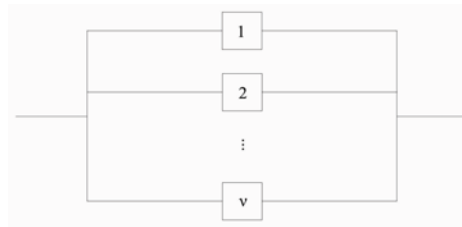
- 27 Ρίχνουμε δύο αμερόληπτα ζάρια (ένα κόκκινο και ένα μαύρο) μια φορά και έστω τα ενδεχόμενα A: στο κόκκινο ζάρι εμφανίζεται 1 ή 2, B: στο μαύρο ζάρι εμφανίζεται 3, 4 ή 5 και Γ: το άθροισμα των δύο ενδείξεων είναι 4, 11 ή 12. Να δείξετε ότι α)  $P(AB\Gamma) = P(A)P(B)P(\Gamma)$  β) τα ενδεχόμενα A, B και Γ δεν είναι ανεξάρτητα κατά ζεύγη.

- 28 Αν A, B και Γ τρία ανεξάρτητα ενδεχόμενα του δειγματικού χώρου  $\Omega$  ενός πειράματος τύχης να δείξετε ότι είναι ανεξάρτητα και τα ζεύγη  $\{A, B \cup \Gamma\}$ ,  $\{B, A \cup \Gamma\}$  και  $\{A, B \cap \Gamma\}$ .

- 29 **Αξιοπιστία:** Ένα **σειριακό σύστημα** n μονάδων λειτουργεί αν και μόνο αν λειτουργούν και οι n μονάδες του. Αντίστοιχα, ένα **παράλληλο σύστημα** n μονάδων λειτουργεί αν και μόνο αν λειτουργεί μια τουλάχιστον από τις n μονάδες του.



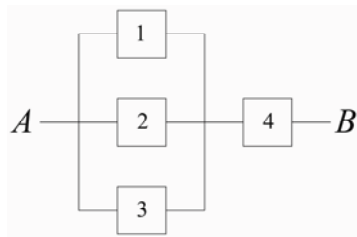
Σειριακό σύστημα



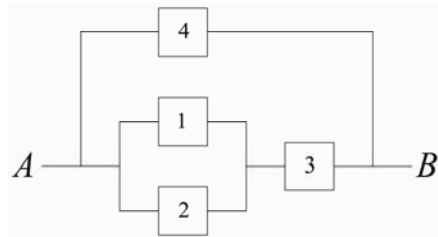
Παράλληλο σύστημα

Αν όλες οι n μονάδες ενός συστήματος έχουν την ίδια αξιοπιστία  $p$  ( $0 < p < 1$ ) και λειτουργούν ανεξάρτητα να δείξετε ότι αν είναι σειριακό τότε η αξιοπιστία του είναι  $p^n$  ενώ αν είναι παράλληλο τότε η αξιοπιστία του είναι  $1 - (1 - p)^n$ .

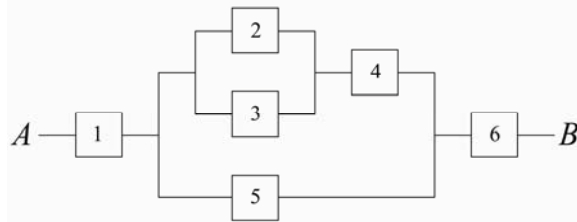
- 30 **Αξιοπιστία:** Στα σχήματα που ακολουθούν φαίνονται τρία συστήματα μεταφοράς υδροποιημένου φυσικού αερίου. Κάθε αντλία (σημειώνονται με αριθμούς 1, 2, ...) λειτουργεί ανεξάρτητα από τις άλλες και έχει δυο καταστάσεις, είτε λειτουργεί με πιθανότητα  $p_i$ ,  $i = 1, 2, \dots$  είτε δε λειτουργεί με πιθανότητα  $1 - p_i$ ,  $i = 1, 2, \dots$ . Το σύστημα έχει επίσης δύο καταστάσεις, και λειτουργεί, δηλαδή το φυσικό αέριο μεταφέρεται από το A στο B, αν και μόνο αν υπάρχει τουλάχιστον ένα μονοπάτι από το A στο B μέσω αντλιών που λειτουργούν. Να υπολογίσετε την πιθανότητα λειτουργίας (δηλαδή την αξιοπιστία) καθενός από τα τρία συστήματα (α), (β) και (γ).



(α)



(β)

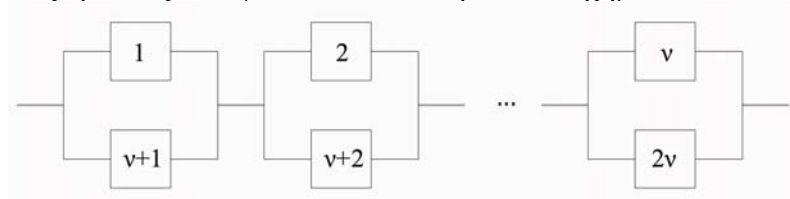


(γ)

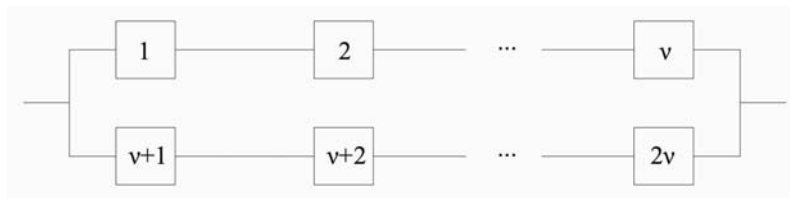
- 31 **Αξιοπιστία (Συνέχεια του Προβλήματος 3.3):** Αν τα εξαρτήματα του συστήματος λειτουργούν ανεξάρτητα και  $P(A_1) = 0.95$ ,  $P(A_2) = 0.90$ ,  $P(A_3) = 0.95$  να βρεθεί η αξιοπιστία του συστήματος (δηλαδή, η πιθανότητα λειτουργίας του συστήματος).
- 32 Ένα εξάρτημα μπορεί να παρουσιάσει είτε βλάβη τύπου I με πιθανότητα 10% είτε βλάβη τύπου II με πιθανότητα 8%. Αν οι δύο τύποι βλαβών εμφανίζονται ανεξάρτητα, βρείτε την πιθανότητα α) να παρουσιασθεί βλάβη τύπου II αν είναι γνωστό ότι ήδη έχει παρουσιασθεί βλάβη τύπου I β) να παρουσιασθεί μία τουλάχιστον από τις δύο βλάβες γ) να παρουσιασθούν συγχρόνως και οι δύο βλάβες.
- 33 Σε μια έρευνα που ολοκληρώθηκε πρόσφατα, μελετήθηκαν μεταξύ άλλων, οι τιμές δύο αιματολογικών δεικτών, έστω A και B, στους κατοίκους ηλικίας άνω των 30 ετών της επαρχίας Γορτυνίας του νομού Αρκαδίας. Σύμφωνα με τα ευρήματα αυτής της έρευνας, το ποσοστό των κατοίκων που έχουν εκτός φυσιολογικών ορίων μόνο την τιμή του δείκτη A (δηλαδή δεν έχουν εκτός φυσιολογικών ορίων και την τιμή του δείκτη B) είναι 40%, το ποσοστό των κατοίκων που έχουν εκτός φυσιολογικών ορίων μόνο την τιμή του δείκτη B (δηλαδή δεν έχουν εκτός φυσιολογικών ορίων και την τιμή του δείκτη A) είναι 60%, ενώ το ποσοστό των κατοίκων που έχουν τις τιμές και των δύο δεικτών εκτός φυσιολογικών ορίων είναι 10%. Το ενδεχόμενο ένας κάτοικος της Γορτυνίας ηλικίας άνω των 30 ετών να έχει την τιμή του ενός δείκτη εκτός φυσιολογικών ορίων είναι άραγε ανεξάρτητο από το ενδεχόμενο να έχει εκτός φυσιολογικών ορίων την τιμή του άλλου δείκτη; Αν όχι, ποια η μεταβολή της πιθανότητας να έχει ένας κάτοικος την τιμή του ενός δείκτη εκτός φυσιολογικών ορίων αν ήδη είναι γνωστό ότι έχει την τιμή του άλλου δείκτη εκτός φυσιολογικών ορίων;
- 34 Σε μια συγκεκριμένη περιοχή, το 10% των κατοίκων ανήκει στην κατηγορία υψηλού κινδύνου καρδιακού εμφράγματος. α) Επιλέγουμε τυχαία τρεις κατοίκους από αυτή την περιοχή. Ποια είναι η πιθανότητα ακριβώς ένας από τους τρεις να ανήκει στην κατηγορία υψηλού κινδύνου (καρδιακού εμφράγματος). β) Γνωρίζουμε ότι το 49% των κατοίκων είναι γυναίκες και επίσης ότι το 8% των γυναικών ανήκει στην κατηγορία υψηλού κινδύνου (καρδιακού εμφράγματος). Επιλέγουμε τυχαία ένα άτομο από την περιοχή αυτή. Ποια είναι η πιθανότητα να

είναι γυναίκα που ανήκει στην κατηγορία υψηλού κινδύνου. γ) Γνωρίζουμε επίσης ότι το 12% των ανδρών ανήκει στην κατηγορία υψηλού κινδύνου (καρδιακού εμφράγματος). Επιλέγουμε τυχαία ένα άτομο από την περιοχή αυτή και διαπιστώνουμε ότι ανήκει στην κατηγορία υψηλού κινδύνου. Ποια είναι η πιθανότητα το άτομο αυτό να είναι άνδρας.

- 35 **Αξιοπιστία:** Η αξιοπιστία κάθε μίας από τις  $n$  μονάδες ενός σειριακού συστήματος είναι ίση με  $p$  ( $0 < p < 1$ ). Ένας τεχνικός, για να αυξήσει την αξιοπιστία του συστήματος, χρησιμοποιεί επιπλέον  $n$  μονάδες με την ίδια αξιοπιστία  $p$  τις οποίες σκέφτεται να συνδέσει στο υπάρχον σύστημα με έναν από τους δύο διαφορετικούς τρόπους που φαίνονται στα παρακάτω σχήματα.



*Συνδεσμολογία I*



*Συνδεσμολογία II*

Να δείξετε ότι η αξιοπιστία του συστήματος με τη *Συνδεσμολογία I* είναι  $p^v (2 - p)^v$  και του συστήματος με τη *Συνδεσμολογία II* είναι  $p^v (2 - p^v)$ . Ποια από τις δύο συνδεσμολογίες πρέπει να χρησιμοποιήσει; (Υποθέστε ότι όλες οι μονάδες που χρησιμοποιούνται λειτουργούν ανεξάρτητα.)