με αυτές ως αφετηρία να μπορούμε να υπολογίσουμε πιθανότητες άλλων πιο σύνθετων. Απαιτούνται και κάποιες επιπλέον σχέσεις για τις $p_1, p_2, \ldots, p_N$. Στο δεύτερο παράδειγμα, μια τέτοια περίπτωση υπολογισμού πιθανοτήτων δίνουμε (Παράδειγμα 3.3.7).

Στο τρίτο παράδειγμα (Παράδειγμα 3.3.8) αποδεικνύουμε ότι αν ο δειγματικός χώρος ενός πειράματος τύχης είναι π ε π ε ρ α σ μ έ ν ο ς και αν επιπλέον τα απλά ενδεχόμενά του είναι ε ξ ί σ ο υ π ι θ α ν ά, τότε τα τρία αξιώματα οδηγούν σε έναν απλό τύπο υπολογισμού της πιθανότητας ενδεχομένου. Πρόκειται για τον ίδιο τύπο που δίνεται από τον κλασικό ορισμό. Για την εφαρμογή αυτού του τύπου, όπως ήδη διαπιστώσαμε όταν μιλήσαμε για τον κλασικό ορισμό (θυμηθείτε τα Παραδείγματα 3.3.1-3.3.3), απαιτούνται μόνο ικανότητες στην απαρίθμηση ευνοϊκών και δυνατών περιπτώσεων. Τέλος, στο Παράδειγμα 3.3.9, δίνουμε μια ακόμη εφαρμογή του τύπου αυτού σε ένα «διάσημο/πολύ γνωστό» πρόβλημα.

Παράδειγμα 3.3.6

Σε μια έρευνα που ολοκληρώθηκε πρόσφατα, μελετήθηκαν μεταξύ άλλων, οι τιμές δύο αιματολογικών δεικτών, έστω A και B, στους άνδρες ηλικίας 50 έως 70 ετών που κατοικούν μόνιμα στο βορειοανατολικό Αιγαίο. Βρέθηκε ότι αν επιλέξουμε τυχαία έναν άνδρα που μένει μόνιμα σε νησί του βορειοανατολικού Αιγαίου και είναι ηλικίας 50 έως 70 ετών, η πιθανότητα να βρίσκεται σε φυσιολογικό επίπεδο η τιμή του δείκτη A στο αίμα του είναι ίση με 0.45. Αντίστοιχα, η πιθανότητα να βρίσκεται σε φυσιολογικό επίπεδο η τιμή του δείκτη B στο αίμα του είναι ίση 0.30 και η πιθανότητα να βρίσκονται σε φυσιολογικό επίπεδο και οι δύο δείκτες είναι ίση με 0.10. Επιλέγουμε τυχαία έναν άνδρα που μένει μόνιμα σε νησί του βορειοανατολικού Αιγαίου και είναι ηλικίας 50 έως 70 ετών και ελέγχουμε τις τιμές των δεικτών A και B στο αίμα του. Ποια είναι η πιθανότητα να βρεθεί με τιμή σε φυσιολογικό επίπεδο

α) τουλάχιστον ένας από τους δύο δείκτες

β) μόνο ο δείκτης A

γ) μόνο ο δείκτης B

δ) μόνο ένας από τους δείκτες A, B

ε) κανένας από τους δείκτες A, B.

Α π ά ν τ η σ η: Έστω τα ενδεχόμενα

A: Η τιμή του δείκτη A είναι σε φυσιολογικό επίπεδο

B: Η τιμή του δείκτη B είναι σε φυσιολογικό επίπεδο.

Σύμφωνα με τα ευρήματα της έρευνας, έχουμε

$P(A) = 0.45$, $P(B) = 0.30$ και $P(AB) = 0.10$.

Με βάση όσα αναφέραμε όταν μιλήσαμε για τις πράξεις μεταξύ ενδεχομένων *(δες και το Σχήμα 3.2.13)*, οι ζητούμενες πιθανότητες, αντίστοιχα είναι

α) $P(A \cup B) = P(A) + P(B) - P(AB) = 0.45 + 0.30 - 0.10 = 0.65$
β) $P(AB') = P(A) - P(AB) = 0.45 - 0.10 = 0.35$
γ) $P(BA') = P(B) - P(AB) = 0.30 - 0.10 = 0.20$
δ) $P(AB' \cup BA') = P(AB') + P(BA') = 0.35 + 0.20 = 0.55$
ε) $1 - P(A \cup B) = 1 - 0.65 = 0.35$
ή $P(A'B') = P[(A \cup B)'] = 1 - P(A \cup B) = 0.35$.

Παράδειγμα 3.3.7. *Από παρατηρήσεις πολλών ετών έχει εκτιμηθεί ότι η πιθανότητα ένας αιμοδότης που προσέρχεται σε κάποιο κέντρο αιμοδοσίας του λεκανοπεδίου Αττικής, να έχει ομάδα αίματος A είναι ίση με την πιθανότητα να έχει ομάδα αίματος O και δεκαπλάσια από την πιθανότητα να έχει ομάδα αίματος AB ενώ η πιθανότητα να έχει ομάδα αίματος B είναι τριπλάσια από την πιθανότητα να έχει ομάδα αίματος AB. Είναι γνωστό ότι ένας ασθενής με ομάδα αίματος A μπορεί να δεχθεί αίμα μόνο από τις ομάδες O και A. Ένας εθελοντής αιμοδότης προσέρχεται σε ένα κέντρο αιμοδοσίας του λεκανοπεδίου Αττικής για να δώσει αίμα για έναν ασθενή που έχει ομάδα αίματος A. Ποια είναι η πιθανότητα το αίμα του εθελοντή να είναι συμβατό με του ασθενούς.*

Απάντηση: Ένας κατάλληλος δειγματικός χώρος του πειράματος είναι προφανώς ο

$$\Omega = \{A, B, AB, O\}.$$

Ας συμβολίσουμε τις πιθανότητες των απλών ενδεχομένων $\{A\}$, $\{B\}$, $\{AB\}$ και $\{O\}$ με p_1, p_2, p_3 και p_4, αντίστοιχα, δηλαδή,

$$p_1 = P(\{A\}),\ p_2 = P(\{B\}),\ p_3 = P(\{AB\}) \text{ και } p_4 = P(\{O\}).$$

Από το πρώτο και το τρίτο αξίωμα προκύπτουν όπως είδαμε οι σχέσεις
1. $p_i \geq 0$, *για κάθε* $i = 1, 2, 3, 4$
2. $p_1 + p_2 + p_3 + p_4 = 1$

και σύμφωνα με τα δεδομένα του προβλήματος, επιπλέον ισχύουν οι σχέσεις, $p_1 = p_4 = 10p_3$ και $p_2 = 3p_3$.

Αντικαθιστώντας από τις τελευταίες σχέσεις τα p_2, p_3 και p_4 στην $p_1 + p_2 + p_3 + p_4 = 1$, παίρνουμε

$$p_1 + \frac{3p_1}{10} + \frac{p_1}{10} + p_1 = 1 \Rightarrow \frac{24p_1}{10} = 1 \Rightarrow p_1 = \frac{5}{12} \cong 0.42$$

και επομένως

$$p_4 \cong 0.42,\ p_3 = \frac{p_1}{10} \cong 0.04 \text{ και } p_2 = 3p_3 \cong 0.12.$$

Έτσι, αξιοποιώντας τη σχέση $p_1 + p_2 + p_3 + p_4 = 1$ (που προκύπτει από το τρίτο αξίωμα του ορισμού) και τις σχέσεις που μας δόθηκαν υπολογίσαμε τις πιθανότητες όλων των απλών ενδεχομένων και πλέον μπορούμε να υπολογίσουμε την πιθανότητα οποιουδήποτε ενδεχομένου του Ω.

Εφόσον ένας ασθενής με ομάδα αίματος A μπορεί να δεχθεί αίμα μόνο από τις ομάδες O και A, η ζητούμενη πιθανότητα, προφανώς είναι η πιθανότητα του ενδεχομένου $\{O, A\}$, επομένως έχουμε

$$P(\{O, A\}) = P(\{O\}) + P(\{A\}) = p_4 + p_1 = 0.42 + 0.42 = 0.84.$$

Παράδειγμα 3.3.8. (πεπερασμένοι δειγματικοί χώροι με ισοπίθανα απλά ενδεχόμενα)

Έστω $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots, \omega_N\}$ πεπερασμένος δειγματικός χώρος με $N(\Omega) = N$ ισοπίθανα δειγματικά σημεία. Έστω επίσης A, ένα οποιοδήποτε ενδεχόμενο του Ω με $N(A)$ δειγματικά σημεία (στοιχεία). Θα δείξουμε ότι η πιθανότητα $P(A)$, πραγματοποίησης του A, δίνεται από τον τύπο

$$P(A) = \frac{N(A)}{N}.$$

Σύμφωνα με το *Πόρισμα 3.3.1(β)*, για τις πιθανότητες,

$$p_1 = P(\{\omega_1\}),\ p_2 = P(\{\omega_2\}),\ p_3 = P(\{\omega_3\}),\ \ldots,\ p_N = P(\{\omega_N\})$$

πραγματοποίησης των απλών ενδεχομένων, $\{\omega_1\}$, $\{\omega_2\}$, $\{\omega_3\}$, ..., $\{\omega_N\}$, ισχύει

$$\sum_{i=1}^{N} p_i = p_1 + p_2 + \ldots + p_N = 1.$$

Οι πιθανότητες των *απλών ενδεχομένων* θεωρούνται ίσες και έστω p η τιμή τους, έστω δηλαδή, $p_1 = p_2 = p_3 = \ldots = p_N = p$. Έτσι, έχουμε

$$\sum_{i=1}^{N} p_i = p_1 + p_2 + \ldots + p_N = 1 \Leftrightarrow Np = 1 \Leftrightarrow p = \frac{1}{N}.$$

Έστω ένα οποιοδήποτε ενδεχόμενο A του Ω με $N(A)$ *δειγματικά σημεία* (στοιχεία). Λόγω της *Πρότασης 3.3.1(γ)* για την πιθανότητα $P(A)$, πραγματοποίησης του A, έχουμε

$$P(A) = \sum_{\omega \in A} p(\omega) = \underbrace{p + p + \ldots + p}_{N(A) \text{ το πλήθος}} = N(A) \cdot p = N(A) \cdot \frac{1}{N} = \frac{N(A)}{N}.$$

Δείξαμε έτσι, πώς από τα τρία αξιώματα του *αξιωματικού ορισμού*, προκύπτει ο τύπος που δίνεται στον *κλασικό ορισμό* για τον υπολογισμό της πιθανότητας ενδεχομένου. Ας δούμε ως ένα παράδειγμα εφαρμογής του τύπου αυτού ένα ενδιαφέρον πρόβλημα, γνωστό στη βιβλιογραφία ως *πρόβλημα των γενεθλίων*. Μπορείτε επίσης να ξαναδείτε τα παραδείγματα που δώσαμε όταν μιλήσαμε για τον *κλασσικό ορισμό*.

Παράδειγμα 3.3.9. (το πρόβλημα των γενεθλίων)

Ποια είναι η πιθανότητα σε μια τάξη k φοιτητών, δύο τουλάχιστον να έχουν γενέθλια την ίδια ημέρα. Θεωρήστε ότι το έτος έχει 365 ημέρες και ότι $k \leq 365$.

Απάντηση: Ο *δειγματικός χώρος* Ω του πειράματος αποτελείται από όλες τις διατεταγμένες k-άδες $\alpha_1\alpha_2\alpha_3 \ldots \alpha_k$, όπου α_1 είναι η ημέρα γενεθλίων του *1ου* φοιτητή, α_2 είναι η ημέρα γενεθλίων του *2ου* φοιτητή, ... και α_k είναι η ημέρα γενεθλίων του *kου* φοιτητή. Πρόκειται για πεπερασμένο δειγματικό χώρο με

$$N(\Omega) = 365^k$$

δηλαδή, με 365^k δυνατά αποτελέσματα, αφού η ημερομηνία γενεθλίων του *1ου* φοιτητή μπορεί να είναι οποιαδήποτε από τις 365 ημέρες του έτους, η ημερομηνία γενεθλίων του *2ου* φοιτητή (όποια και αν είναι η ημερομηνία γενεθλίων του *1ου*) μπορεί επίσης να είναι οποιαδήποτε από τις 365 ημέρες του έτους, κ.ο.κ. Δηλαδή, για το σχηματισμό μιας διατεταγμένης *k-άδας* το στοιχείο α_1 μπορεί να πάρει οποιαδήποτε από 365 διαφορετικές τιμές, το στοιχείο α_2 μπορεί επίσης να πάρει οποιαδήποτε από 365 διαφορετικές τιμές (όποια και αν είναι η τιμή που πήρε το α_1), κ.ο.κ. και επομένως οι διαφορετικές *k-άδες* $\alpha_1\alpha_2\alpha_3 \dots \alpha_k$ που είναι δυνατόν να δημιουργηθούν σύμφωνα με την *πολλαπλασιαστική αρχή* είναι

$$\underbrace{365 \cdot 365 \cdot \ldots \cdot 365}_{k \text{ το πλήθος}} = 365^k.$$

Με την παραδοχή ότι τα 365^k δυνατά αποτελέσματα είναι ισοπίθανα θα υπολογίσουμε τη ζητούμενη πιθανότητα ως το λόγο των ευνοϊκών προς τις δυνατές περιπτώσεις. Έστω λοιπόν το ενδεχόμενο

A: Δύο τουλάχιστον από τους k φοιτητές έχουν γενέθλια την ίδια ημέρα

και το συμπλήρωμά του

A΄: οι k φοιτητές έχουν διαφορετικές ημερομηνίες γενεθλίων.

Θα υπολογίσουμε την πιθανότητα

$$P(A') = \frac{N(A')}{N(\Omega)}$$

του ενδεχομένου A' και στη συνέχεια από τον τύπο $P(A) = 1 - P(A')$ θα υπολογίσουμε τη ζητούμενη πιθανότητα.

Το πλήθος των ευνοϊκών περιπτώσεων $N(A')$ για την πραγματοποίηση του A' προφανώς είναι $365 \cdot 364 \cdot 363 \cdot \ldots \cdot [365 - (k-1)]$, δηλαδή,

$$N(A') = 365 \cdot 364 \cdot 363 \cdot \ldots \cdot (365 - k + 1)$$

αφού στα ευνοϊκά για το ενδεχόμενο A' αποτελέσματα, οι k φοιτητές πρέπει να έχουν διαφορετικές ημερομηνίες γενεθλίων, δηλαδή, το A' αποτελείται από όλες τις διατεταγμένες *k-άδες* $\alpha_1\alpha_2\alpha_3 \dots \alpha_k$ που έχουν διαφορετικά στοιχεία δηλαδή που είναι $\alpha_i \neq \alpha_j$ (για $i \neq j$). Αυτό σημαίνει ότι στις ευνοϊκές *k-άδες*, το α_2 δεν παίρνει 365 διαφορετικές τιμές αλλά $365 - 1 = 364$ γιατί εξαιρείται η τιμή που πήρε το α_1 (δηλαδή η ημερομηνία γενεθλίων του *1ου* φοιτητή), ανάλογα, το α_3 παίρνει $365 - 2 = 363$ διαφορετικές τιμές γιατί εξαιρούνται οι ημερομηνίες γενεθλίων των προηγούμενων δύο φοιτητών και τέλος το α_k παίρνει $365 - (k-1)$ διαφορετικές τιμές αφού εξαιρούνται οι ημερομηνίες γενεθλίων των προηγούμενων $k - 1$ φοιτητών.

Έτσι έχουμε

$$P(A') = \frac{365 \cdot 364 \cdot 363 \cdot \ldots \cdot (365 - k + 1)}{365^k}$$

και άρα η ζητούμενη πιθανότητα είναι

$$P(A) = 1 - P(A') = 1 - \frac{365 \cdot 364 \cdot 363 \cdot \ldots \cdot (365 - k + 1)}{365^k}, \quad 1 \leq k \leq 365.$$

Για να υπολογίσουμε τα $N(\Omega)$ και $N(A')$ μπορούμε να σκεφθούμε και ως εξής: κάθε δυνατό αποτέλεσμα είναι μια επαναληπτική διάταξη των $\nu = 365$ ημερών του έτους ανά k και επομένως με εφαρμογή του τύπου που δίνει το πλήθος των επαναληπτικών διατάξεων των ν στοιχείων ανά k έχουμε $N(\Omega) = \nu^k = 365^k$ ενώ κάθε ευνοϊκό αποτέλεσμα είναι μια διάταξη των $\nu = 365$ ημερών του έτους ανά k και επομένως, $N(A') = (\nu)_k = (365)_k = 365 \cdot 364 \cdot 363 \cdot \ldots \cdot (365 - k + 1)$. Σχόλιο 3.3.4

Στον πίνακα που ακολουθεί δίνουμε τις τιμές της πιθανότητας του ενδεχομένου A για διάφορες τιμές του k. Σχόλιο 3.3.5.

k	5	10	23	30	50	70
$P(A)$	0.027	0.117	0.507	0.706	0.970	0.999

Παρατηρείστε ότι αυξανομένου του k, η πιθανότητα από k άτομα να βρεθούν τουλάχιστον δύο που να έχουν γενέθλια την ίδια ημέρα ασφαλώς αυξάνεται και μάλιστα φθάνει πολύ γρήγορα σε τιμές κοντά στο 1. Για παράδειγμα, μεταξύ μόλις 50 ατόμων η πιθανότητα τουλάχιστον δύο να έχουν γενέθλια την ίδια ημέρα είναι πολύ μεγάλη, ίση με 97%, ενώ μεταξύ 70 ατόμων είναι σχεδόν βέβαιο ότι τουλάχιστον δύο έχουν γενέθλια την ίδια ημέρα. Παρατηρείστε επίσης, ότι για $k = 23$ είναι $P(A) = 0.507$, δηλαδή, μεταξύ 23 ατόμων[19] η πιθανότητα να υπάρχουν τουλάχιστον δύο που να έχουν γενέθλια την ίδια ημέρα είναι μεγαλύτερη από την πιθανότητα να μην υπάρχουν!!!

Ερώτηση: Στο αμφιθέατρο βρίσκονται για το μάθημα της Στατιστικής 70 φοιτητές και ο καθηγητής. Ο καθηγητής μόλις έχει εξηγήσει το πρόβλημα των γενεθλίων όταν ένας φοιτητής απευθυνόμενος στον καθηγητή και τους συμφοιτητές του λέει: έχω γενέθλια στις 24/6, υπάρχει κάποιος άλλος στο αμφιθέατρο που να έχει γενέθλια στις 24/6; Τι είναι άραγε πιο πιθανό, να βρεθεί και κάποιος άλλος που να έχει γενέθλια στις 24/6 ή να μη βρεθεί; (Υπόδειξη: Δείτε την Άσκηση 3.5)

19. Ας πούμε μεταξύ των 22 ποδοσφαιριστών και του διαιτητή σε έναν ποδοσφαιρικό αγώνα!

Μαθαίνω

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 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.

Παράδειγμα 4.1.4. (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

$$P(S) = \frac{2000}{3000} = 0.67, \quad P(S') = 1 - P(S) = 0.33,$$

$$P(E/S) = \frac{250}{2000} = 0.125 \text{ και } P(E/S') = \frac{5}{1000} = 0.005.$$

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. Ανεξαρτησία ενδεχομένων

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

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

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

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', επομένως

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

Παράδειγμα 4.5.3.

Ένα εργοστάσιο έχει δύο γραμμές παραγωγής Γ1 και Γ2. Στο τμήμα ποιοτικού ελέγχου όταν διαπιστώνεται ότι ένα προϊόν είναι ελαττωματικό κατατάσσεται σε μια από δύο κατηγορίες Ε1 και Ε2 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 Γ1, Γ2 και στις δύο κατηγορίες Ε1, Ε2.

	$Γ1$	$Γ2$	*Σύνολα*
$E1$	500	300	800
$E2$	400	240	640
Σύνολα	900	540	1440

Ας θεωρήσουμε τα ενδεχόμενα
Γ1: το προϊόν προέρχεται από τη γραμμή παραγωγής Γ1
Γ2: το προϊόν προέρχεται από τη γραμμή παραγωγής Γ2
Ε1: το προϊόν ανήκει στην κατηγορία Ε1
Ε2: το προϊόν ανήκει στην κατηγορία Ε2.

Τα ενδεχόμενα Γ2 και Ε2 είναι άραγε ανεξάρτητα ή εξαρτημένα;

Απάντηση: Με βάση το στατιστικό ορισμό της πιθανότητας έχουμε

$$P(Γ2) = \frac{540}{1440} = 0.3750, \quad P(E2) = \frac{640}{1440} = 0.4444$$

$$\text{και} \quad P(Γ2E2) = \frac{240}{1440} = 0.1666$$

και επειδή

$$P(Γ2)P(E2) = 0.3750 \cdot 0.4444 = 0.1666 = P(Γ2E2),$$

τα $Γ2$ και $E2$ είναι ανεξάρτητα.

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

$$P(Γ2|E2) = \frac{P(Γ2E2)}{P(E2)} = \frac{240/1440}{640/1440} = \frac{240}{640} = 0.3750 = P(Γ2)$$

και ασφαλώς

$$P(E2|Γ2) = \frac{P(E2Γ2)}{P(Γ2)} = \frac{240/1440}{540/1440} = \frac{240}{540} = 0.4444 = P(E2)$$

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

Εφόσον τα ενδεχόμενα $Γ2$ και $E2$ είναι ανεξάρτητα, μήπως και τα $Γ1$ 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. Παρατηρήστε ότι το $Γ1$ είναι το συμπλήρωμα του $Γ2$ (αφού ένα ελαττωματικό προϊόν είτε θα προέρχεται από τη γραμμή *Γ1* είτε από τη *Γ2*) και ότι το *Ε1* είναι το συμπλήρωμα του *Ε2* (αφού ένα ελαττωματικό προϊόν είτε θα κατατάσσεται στην κατηγορία *Ε1* είτε στην *Ε2*). Επομένως, με βάση την *Πρόταση 4.5.1*, τα *Γ1* και *Ε1* θα είναι ανεξάρτητα ως τα συμπληρώματα των ανεξάρτητων ενδεχομένων *Γ2* και *Ε2*. Ανεξάρτητα θα είναι επίσης και τα *Γ2, Ε1* αλλά και τα *Γ1, Ε2* 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 Α, Β και Γ του δειγματικού χώρου Ω ενός πειράματος τύχης λέγονται ανεξάρτητα ή τελείως ανεξάρτητα αν ισχύουν οι σχέσεις

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

$$P(A\Gamma) = P(A)P(\Gamma) \qquad (4.5.2)$$

$$P(B\Gamma) = P(B)P(\Gamma)$$

και

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

Παρατηρήστε ότι, παρότι οι συνθήκες (4.5.2) εξασφαλίζουν ότι τρία ενδεχόμενα *Α*, *Β* και *Γ* 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 *Α* είναι ανεξάρτητο από ένα ενδεχόμενο *Β* και επίσης από ένα ενδεχόμενο *Γ*, αυτό δε σημαίνει ότι είναι ανεξάρτητο και από την τομή τους *ΒΓ (δες Άσκηση 4.25)*. Δηλαδή, ενώ η πραγματοποίηση του *Β* δεν επηρεάζει την πραγματοποίηση του *Α* και επίσης ούτε η πραγματοποίηση του *Γ* επηρεάζει την πραγματοποίηση του *Α*, εντούτοις, η ταυτόχρονη πραγματοποίηση των *Β* και *Γ* μπορεί να επηρεάζει την πραγματοποίηση του *Α*!!! 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, Γ, (τελείως) ανεξάρτητα πρέπει να είναι ανεξάρτητα αφενός τα A *και* B, A *και* Γ, B *και* Γ, αλλά αφετέρου και τα A *και* $B\Gamma$, B *και* $A\Gamma$, και τέλος, τα Γ *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 Γ 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 $\nu \geq 2$ ενδεχομένων.

Ορισμός 4.5.3.
(ανεξαρτησία $\nu \geq 2$ ενδεχομένων)

Θα λέμε ότι $\nu \geq 2$ ενδεχόμενα $A_1, A_2, \ldots, A_\nu$ του δειγματικού χώρου Ω ενός πειράματος τύχης είναι ανεξάρτητα ή τελείως ανεξάρτητα αν για οποιοδήποτε σύνολο k δεικτών $i_1, i_2, \ldots, i_k$ από το σύνολο $\{1, 2, \ldots, \nu\}$ ισχύει

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

Δηλαδή, λέμε ότι $\nu \geq 2$ 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 ν ο αριθμός των συνθηκών που πρέπει να ικανοποιούνται. Για παράδειγμα, για $\nu = 4$ ανεξάρτητα ενδεχόμενα A_1, A_2, A_3, A_4, πρέπει να ικανοποιούνται

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

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

$\binom{4}{4} = 1$ συνθήκη, η $P(A_1A_2A_3A_4) = P(A_1)P(A_2)P(A_3)P(A_4)$,

δηλαδή συνολικά πρέπει να ικανοποιούνται 11 συνθήκες!

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

$\binom{\nu}{2}+\binom{\nu}{3}+\ldots+\binom{\nu}{\nu}=2^{\nu}-\nu-1$ 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 ν, π.χ. αν $\nu = 5$ ενδεχόμενα, για να ελεγχθεί αν αυτά είναι τελείως ανεξάρτητα πρέπει να ελεγχθούν $2^5 - 5 - 1 = 26$ 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 $\nu \geq 2$ ανεξάρτητων ενδεχομένων είναι ανεξάρτητα όπως και οποιαδήποτε k από αυτά με τα συμπληρώματα των υπόλοιπων $\nu - k$ 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
Α1: ο κάτοικος Αττικής το 2019 ζει
Α2: ο κάτοικος Θεσσαλονίκης το 2019 ζει
Α3: ο κάτοικος Ρόδου το 2019 ζει.

Ζητάμε την πιθανότητα του ενδεχομένου

$$E = (A_1'A_2A_3) \cup (A_1A_2'A_3) \cup (A_1A_2A_3')$$

το οποίο είναι ένωση ξένων ανά δύο ενδεχομένων και επομένως

$$P(E) = P(A_1'A_2A_3) + P(A_1A_2'A_3) + P(A_1A_2A_3').$$

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

$$P(E) = P(A_1')P(A_2)P(A_3) + P(A_1)P(A_2')P(A_3) + P(A_1)P(A_2)P(A_3') =$$
$$= 0.4 \cdot 0.9 \cdot 0.3 + 0.6 \cdot 0.1 \cdot 0.3 + 0.6 \cdot 0.9 \cdot 0.7 = 0.504.$$

αυτά, καθώς και η έκβαση οποιασδήποτε ομάδας από αυτά, δεν επηρεάζει την έκβαση των υπολοίπων. Έτσι, αν τα $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_\nu$ είναι *ανεξάρτητα πειράματα*, τότε για οποιαδήποτε ενδεχόμενα

$$A_1 \subseteq \Omega_1, A_2 \subseteq \Omega_2, \ldots, A_\nu \subseteq \Omega_\nu,$$

όπου καθένα σχετίζεται μόνο με το αντίστοιχο πείραμα $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_\nu$, ορίζουμε

$$P(A_1A_2 \ldots A_\nu) = P(A_1)P(A_2) \ldots P(A_\nu).$$

Με $A_1A_2 \ldots A_\nu$ συμβολίζουμε το ενδεχόμενο $A_1 \times A_2 \times \ldots \times A_\nu$ του δειγματικού χώρου $\Omega_1 \times \Omega_2 \times \ldots \times \Omega_\nu$ του σύνθετου πειράματος, δηλαδή, συμβολίζουμε το ενδεχόμενο το οποίο πραγματοποιείται όταν, κατά την εκτέλεση του πειράματος ε_1 πραγματοποιηθεί το A_1, κατά την εκτέλεση του πειράματος ε_2 πραγματοποιηθεί το $A_2, \ldots,$ και κατά την εκτέλεση του πειράματος ε_ν πραγματοποιηθεί το A_ν.

Παράδειγμα 4.5.7. Έστω το πείραμα ε: *«ρίχνω ένα νόμισμα δύο φορές και στη συνέχεια ένα ζάρι»*. Προφανώς το πείραμα αυτό αναλύεται στα πειράματα
ε_1: *«ρίχνω ένα νόμισμα»* με δειγματικό χώρο $\Omega_1 = \{\kappa, \gamma\}$,
ε_2: *«ρίχνω ένα νόμισμα»* με δειγματικό χώρο $\Omega_2 = \{\kappa, \gamma\}$ και
ε_3: *«ρίχνω ένα ζάρι»* με δειγματικό χώρο $\Omega_3 = \{1, 2, 3, 4, 5, 6\}$.

Ο δειγματικός χώρος Ω του (σύνθετου) πειράματος ε, 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

$$\Omega = \{\kappa\kappa1, \kappa\kappa2, \kappa\kappa3, \kappa\kappa4, \kappa\kappa5, \kappa\kappa6, \kappa\gamma1, \kappa\gamma2, \kappa\gamma3, \kappa\gamma4, \kappa\gamma5, \kappa\gamma6,$$
$$\gamma\kappa1, \gamma\kappa2, \gamma\kappa3, \gamma\kappa4, \gamma\kappa5, \gamma\kappa6, \gamma\gamma1, \gamma\gamma2, \gamma\gamma3, \gamma\gamma4, \gamma\gamma5, \gamma\gamma6\}.$$

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

$$A_1 = \{\kappa\}, A_2 = \{\gamma\} \text{ και } A_3 = \{1, 3, 5\} \text{ τότε } A_1A_2A_3 = \{\kappa\gamma1, \kappa\gamma3, \kappa\gamma5\}$$

και είναι προφανές ότι το ενδεχόμενο $A_1A_2A_3$ του $\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$ πραγματοποιείται όταν κατά την εκτέλεση του πειράματος ε_1 εμφανισθεί «κεφαλή», κατά την εκτέλεση του πειράματος ε_2 εμφανισθεί «γράμματα» και κατά την εκτέλεση του πειράματος ε_3 εμφανισθεί περιττός αριθμός. Με την παραδοχή ότι το νόμισμα και το ζάρι είναι αμερόληπτα προφανώς έχουμε

$$P(A_1A_2A_3) = \frac{3}{24} = \frac{1}{8}, \; P(A_1) = \frac{1}{2}, \; P(A_2) = \frac{1}{2} \text{ και } P(A_3) = \frac{3}{6} = \frac{1}{2}.$$

Παρατηρήστε ότι για τα $A_1 = \{\kappa\}$, $A_2 = \{\gamma\}$ και $A_3 = \{1, 3, 5\}$ η σχέση $P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3)$ επαληθεύεται. Εύκολα επίσης μπορείτε να δείξετε ότι η συνθήκη ανεξαρτησίας επαληθεύεται για οποιαδήποτε ενδεχόμενα των Ω_1, Ω_2 και Ω_3. Το ότι τα ε_1, ε_2 και ε_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 1.5.2.

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

$$P(A_1A_2 \ldots A_\nu) = P(A_1)P(A_2) \ldots P(A_\nu).$$

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 $\Omega_1 \times \Omega_2 \times \ldots \times \Omega_\nu$ 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 ν πειράματα στα οποία αναλύεται ένα σύνθετο πείραμα είναι ανεξάρτητες επαναλήψεις του ιδίου πειράματος τότε λέμε ότι το πείραμα αποτελείται από *ν ανεξάρτητες δοκιμές* ή ότι πρόκειται για *μια σειρά ή ακολουθία ν ανεξάρτητων δοκιμών* ή για *ν 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 διαδοχικές ανεξάρτητες επαναλήψεις ενός πειράματος με δύο μόνο δυνατά αποτελέσματα, λέμε ότι το πείραμα αποτελείται από *ν ανεξάρτητες δοκιμές Bernoulli* ή από μια *σειρά* ή *ακολουθία ν 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 ν ανεξάρτητες δοκιμές Bernoulli, είναι διατεταγμένες ν-άδες από «επιτυχίες» και «αποτυχίες» δηλαδή έχουν τη μορφή $\underbrace{\alpha\varepsilon\varepsilon\alpha\varepsilon \ldots \varepsilon\alpha\varepsilon}_{\nu \text{ το πλήθος}}$.

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 ιδιαιτέρως χρήσιμη για τη συνέχεια, όταν θα μιλήσουμε για μοντέλα πιθανοτήτων.

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

(α) *να εμφανισθούν ν «επιτυχίες» είναι ίση με p^ν*

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

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

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

Απόδειξη: Έστω τα ενδεχόμενα

$$A_i,\ i=1,2,\dots,\nu: \text{ στην } i \text{ δοκιμή εμφανίζεται «επιτυχία».}$$

(α) Ζητάμε την πιθανότητα του ενδεχομένου $A_1A_2\dots A_\nu$.

Από την ανεξαρτησία των δοκιμών προφανώς έχουμε

$$P(A_1A_2\dots A_\nu)=P(A_1)P(A_2)\dots P(A_\nu)=p\cdot p\cdot\dots\cdot p=p^\nu.$$

(β) Ζητάμε την πιθανότητα του ενδεχομένου $A_1'A_2'\dots A_\nu'$.

Από την ανεξαρτησία των δοκιμών προφανώς έχουμε

$$\begin{aligned}P(A_1'A_2'\dots A_\nu') &= P(A_1')P(A_2')\dots P(A_\nu') = \\ &= (1-p)\cdot(1-p)\cdot\dots\cdot(1-p)=(1-p)^\nu.\end{aligned}$$

(γ) Ζητάμε την πιθανότητα του συμπληρώματος του ενδεχομένου $A_1'A_2'\dots A_\nu'$.

Προφανώς η ζητούμενη πιθανότητα είναι

$$1-(1-p)^\nu.$$

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

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

Όπως δείξαμε στο *Παράδειγμα 3.1.1δ*, το πλήθος αυτών των διατεταγμένων ν-άδων είναι

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

Δηλαδή, το ενδεχόμενο του οποίου ζητάμε την πιθανότητα αποτελείται από

$$\binom{\nu}{r}$$

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

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

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

Η *δοκιμή Bernoulli*, όπως είδαμε, είναι ένα πείραμα με δύο μόνο (αμοιβαίως αποκλειόμενα) δυνατά αποτελέσματα, ε και α («επιτυχία» και «αποτυχία», αντίστοιχα). Αν ένα πείραμα έχει k (αμοιβαίως αποκλειόμενα) δυνατά αποτελέσματα, $r_1, r_2, \ldots, r_k$, ονομάζεται *πολυωνυμική 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 δομική* με $k = 4$ δυνατά αποτελέσματα (Α, Β, ΑΒ και Ο). 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.

Αν ένα πείραμα τύχης αποτελείται από ν ανεξάρτητες πολυωνυμικές δοκιμές, όπου σε κάθε δοκιμή τα δυνατά αποτελέσματα είναι k, έστω $r_1, r_2, \ldots, r_k$, και η πιθανότητα να εμφανισθεί το αποτέλεσμα r_i σε μια (οποιαδήποτε) δοκιμή είναι p_i, τότε η πιθανότητα, στις ν δοκιμές, να εμφανισθούν ν_1 αποτελέσματα r_1, ν_2 αποτελέσματα r_2, ... και ν_k αποτελέσματα r_k είναι ίση με Πρόταση 4.5.3.

$$\frac{\nu!}{\nu_1! \cdot \nu_2! \cdot \ldots \cdot \nu_k!} p_1^{\nu_1} p_2^{\nu_2} \ldots p_k^{\nu_k},$$

όπου

$$\sum_{i=1}^{k} \nu_i = \nu \text{ και } \sum_{i=1}^{k} p_i = 1.$$

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*).