κεφάλαιο 7

ΔΕΣΜΕΥΜΕΝΗ ΠΙΘΑΝΟΤΗΤΑ ΣΤΟΧΑΣΤΙΚΗ ΑΝΕΞΑΡΤΗΣΙΑ ΤΟ ΘΕΩΡΗΜΑ ΤΟΥ BAYES

- Υπολογισμός της Δεσμευμένης Πιθανότητας
- Στοχαστική Ανεξαρτησία
- Δένδρα Πιθανότητας
- Υπολογισμός της Πιθανότητας στην Απλή Τυχαία Δειγματοληψία
- Περιθώρια ή Ολική Πιθανότητα
- Το θεώρημα του Bayes
- ΑΣΚΗΣΕΙΣ

7.1 Η Έννοια της Δεσμευμένης Πιθανότητας

Όταν υπολογίζουμε την πιθανότητα του ενδεχομένου A ρωτάμε πόσο είναι πιθανό να πραγματοποιηθεί το A, γνωρίζοντας ότι βρισκόμαστε στο δειγματικό χώρο S. Είναι όμως δυνατό να έχουμε εκ των προτέρων την πληροφορία ότι το αποτέλεσμα του πειράματος θα βρίσκεται σε ένα υποσύνολο B του S. Στην περίπτωση αυτή μπορεί να βελτιώσουμε τις προβλέψεις μας για το αποτέλεσμα, αν ξαναϋπολογίσουμε τις πιθανότητες, παίρνοντας υπόψη μας τη γνώση αυτή.

Έτσι π.χ. η πιθανότητα για άνδρα ηλικίας 48 χρόνων να πεθάνει τα επόμενα 10 χρόνια μπορεί να είναι διαφορετική σε ένα υποσύνολο B του δειγματικού χώρου ας πούμε σ' αυτούς που έχουν βάρος μεγαλύτερο από το κανονικό κατά 15 τουλάχιστον κιλά. Επομένως, όταν η ασφαλιστική εταιρία πουλάει ασφάλεια ζωής σε άνδρα 48 χρόνων και γνωρίζει ότι είναι υπέρβαρος κατά 15 κιλά τουλάχιστον, θα υπολογίσει την πιθανότητα αποζημίωσης (λόγω) θανάτου στα επόμενα 10 χρόνια, παίρνοντας υπόψη της τη γνώση αυτή.

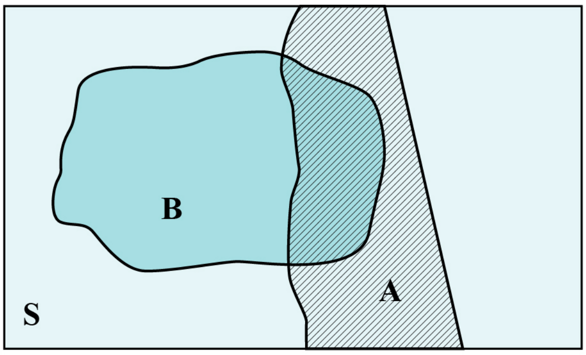
Η αναθεωρημένη αυτή πιθανότητα ονομάζεται δεσμευμένη ή πιθανότητα υπό συνθήκη του B, συμβολίζεται με P(A | B) και απαντά το ερώτημα "πόσο πιθανό είναι να πραγματοποιηθεί το A δεδομένου ότι πραγματοποιείται το B"

7.2 Υπολογισμός της Δεσμευμένης Πιθανότητας

Η **δεσμευμένη πιθανότητα** *(conditional probability)* ενός ενδεχομένου A ή η πιθανότητα του A υπό συνθήκη B συμβολίζεται ως P(A | B) και υπολογίζεται με βάση τον τύπο:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0 \tag{7.2.1}$$

Όπως φαίνεται και στο παρακάτω βέννειο διάγραμμα, η P(A | B) είναι η πιθανότητα του A όταν ο δειγματικός χώρος S έχει περιοριστεί στο B.



Για να διακρίνουμε την P(A) από την P(A | B) λέμε την P(A) **περιθώρια** ή **χωρίς συνθήκη** ή **απόλυτη πιθανότητα** *(marginal or unconditional or absolute probability)*. Σημειώνεται ότι και οι περιθώριες πιθανότητες είναι υπό συνθήκη του αρχικού δειγματικού χώρου αφού $P(A \cap S) = P(A)$ και $P(S) = 1$, οπότε έχουμε:

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

Παράδειγμα 7.2.1 Σε τυχερό παιχνίδι ρίχνουμε δύο ζάρια οπότε ο δειγματικός χώρος του πειράματος είναι ο

$$S = \begin{Bmatrix} (1,1) & (1,2) & \cdots & (1,6) \\ (2,1) & (2,2) & \cdots & (2,6) \\ \vdots & \vdots & & \vdots \\ (6,1) & (6,2) & \cdots & (6,6) \end{Bmatrix}$$

και αποτελείται από 36 αποτελέσματα, το καθένα από τα οποία έχει πιθανότητα να πραγματοποιηθεί ίση με 1/36. Αν A και B τα ακόλουθα δύο ενδεχόμενα του S

$$A = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$$

$$B = \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\}$$

Τότε, $P(A) = 5/36$ και $P(B) = 6/36$

Ας υποθέσουμε ότι ρίξαμε το πρώτο ζάρι και πήραμε την ένδειξη “3”, δηλαδή πραγματοποιήθηκε το B. Ο δειγματικός χώρος του πειράματος έχει τώρα περιοριστεί στα 6 αποτελέσματα του Β απο τα οποία μόνον στο (3,3) πραγματοποιείται και το Α. Επομένως έχουμε:

$$P(A \mid B) = 1/6$$

Τα δεδομένα του παραδείγματος αυτού δίνονται στον πίνακα που ακολουθεί, όπου οι γραμμές παριστάνουν το αποτέλεσμα της πρώτης ρίψης και οι στήλες το αποτέλεσμα της δεύτερης ρίψης. Στο εσωτερικό του πίνακα σημειώνονται οι πιθανότητες για τα 36 ισοπίθανα σημεία. Μετά την πραγματοποίηση του Β ο αρχικός δειγματικός χώρος του πειράματος, περιορίζεται στο Β και P(A | B) = 1/6.

		1	2	3	4	5	6	Άθροισμα
	1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
B	3	1/36	1/36	1/36	1/36	1/36	1/36	1/6=P(B)
	4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
		A						
	1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
Άθροισμα		1/6	1/6	1/6	1/6	1/6	1/6	

Παράδειγμα 7.2.2 Έστω ότι το 38% των τροχαίων ατυχημάτων γίνεται την νύχτα (Ν), το 53% γίνεται υπό την επήρεια αλκοόλ (Α), και το 24% γίνεται την νύχτα υπό την επήρεια αλκοόλ. Ποια είναι η πιθανότητα για ένα ατύχημα να έγινε υπό την επήρεια του αλκοόλ, δεδομένου ότι έγινε τη νύχτα;

Δίνεται ότι:

$$P(N) = 0.38, \quad P(A) = 0.53, \quad P(A \cap N) = 0.24$$

και η ζητούμενη πιθανότητα είναι η

$$P(A \mid N) = \frac{P(A \cap N)}{P(N)} = \frac{0.24}{0.38} = 0.63$$

Παράδειγμα 7.2.3 Η πιθανότητα για αγόρι σε μια γέννηση ισούται με 0.512. Σε τρεις εγκύους η πιθανότητα του ενδεχομένου Ε "η μία τουλάχιστον να γεννήσει αγόρι" ισούται με

$$\begin{aligned} P(E) \equiv P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) \\ &\quad - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3) \\ &= 3(0.512) - 3(0.512)(0.512) - (0.512)(0.512)(0.512) \\ &= 0.615 \end{aligned}$$

- Αν είναι γνωστό ότι η μία τουλάχιστον θα γεννήσει αγόρι, ποιά είναι η πιθανότητα του ενδεχομένου Β = "και οι τρεις να γεννήσουν αγόρι;"

Επειδή $B = A_1 \cap A_2 \cap A_3$ ισχύει $P(B) = P(A_1)P(A_2)P(A_3) = (0.512)^3 = 0.00134$ και $P(B \cap E) = P(B)$, οπότε η ζητούμενη πιθανότητα είναι η

$$P(B \mid E) = \frac{P(B \cap E)}{P(E)} = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cup A_2 \cup A_3)} = \frac{0.001342}{0.615} = 0.218$$

Το θεώρημα του πολλαπλασιασμού για την υπό συνθήκη πιθανότητα

Από τις σχέσεις $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$ και $P(B \mid A) = \dfrac{P(A \cap B)}{P(A)}$ παίρνουμε:

$$P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

που είναι γνωστή ως **θεώρημα πολλαπλασιασμού για την υπό συνθήκη πιθανότητα** *(multiplication theorem for conditional probability)*.

Για τρία ενδεχόμενα Α, Β και Γ του S η σχέση (7.2.1) γίνεται

$$P(\Gamma \mid A \cap B) = \frac{P(A \cap B \cap \Gamma)}{P(A \cap B)}$$

$$\Leftrightarrow P(A \cap B \cap \Gamma) = P(\Gamma \mid A \cap B)P(A \cap B) \qquad (7.2.4)$$

Με αντικατάσταση της $P(A \cap B) = P(A)P(B \mid A)$ στην (7.2.4) παίρνουμε:

$$P(A \cap B \cap \Gamma) = P(\Gamma \mid A \cap B)P(B \mid A)P(A) \qquad (7.2.5)$$

Με τρόπο ανάλογο, το θεώρημα του πολλαπλασιασμού της υπό συνθήκη πιθανότητας επεκτείνεται σε περισσότερα από τρία ενδεχόμενα.

Παράδειγμα 7.2.4 Ένα κουτί περιέχει 20 προϊόντα από τα οποία τα 5 είναι ελαττωματικά. Τρία προϊόντα λαμβάνονται με τυχαίο τρόπο το ένα μετά το άλλο. Να υπολογιστεί η πιθανότητα και τα τρία να είναι μη ελαττωματικά.

Έστω E_i = {i = το i προϊόν είναι μη ελαττωματικό}, i = 1,2,3. Τότε η πιθανότητα $P(E_1)$ το πρώτο προϊόν να είναι μη ελαττωματικό είναι 15/20 = ¾, αφού τα 15 από τα 20 προϊόντα είναι μη ελαττωματικά.
Αν το πρώτο είναι μη ελαττωματικό τότε η πιθανότητα $P(E_2 | E_1)$ ότι το δεύτερο είναι μη ελαττωματικό είναι 14/19, επειδή μόνο 14 από τα 19 εναπομείναντα είναι μη ελαττωματικά.
Αν τα πρώτα δύο προϊόντα είναι μη ελαττωματικά, τότε η πιθανότητα $P(E_3 | E_1 \cap E_2)$ ότι το τρίτο προϊόν είναι μη ελαττωματικό είναι 13/18 επειδή μόνο 13 από τα εναπομείναντα 18 είναι μη ελαττωματικά.
Εφαρμόζοντας την (7.2.5) παίρνουμε:

$$\begin{aligned} P(E_1 \cap E_2 \cap E_3) &= P(E_1)P(E_2 | E_1)P(E_3 | E_1 \cap E_2) \\ &= \frac{3}{4}\cdot\frac{14}{19}\cdot\frac{13}{18} = 0.40 \end{aligned}$$

ως άσκηση: Να δειχτεί ότι ισχύει επίσης

$$P(E_1 \cap E_2 \cap E_3) = \frac{\binom{15}{3}}{\binom{20}{3}}$$

7.3 Στοχαστική Ανεξαρτησία

Είπαμε ότι P(A|B) είναι η πιθανότητα του ενδεχομένου A η οποία υπολογίζεται παίρνοντας υπόψη ότι πραγματοποιήθηκε το B. Αν η πιθανότητα του A δεν επηρεάζεται από την πραγματοποίηση του B, τότε έχουμε:

$$P(A | B) = P(A) \tag{7.3.1}$$

και λέμε ότι το A είναι **στοχαστικά** ή **στατιστικά ανεξάρτητο** *(stochastically or statistically independent)* ή απλώς **ανεξάρτητο** του B.

Όταν αντικαταστήσουμε την (7.3.1) στην (7.2.2) παίρνουμε:

$$P(A \cap B) = P(A)P(B) \tag{7.3.2}$$

και επομένως, για $P(A) \neq 0$, ισχύει:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B) \tag{7.3.3}$$

Από το συνδυασμό των (7.3.1) και (7.3.3) προκύπτει ότι, όταν το ενδεχόμενο Α είναι ανεξάρτητο του Β τότε και το Β είναι ανεξάρτητο του Α. Επομένως για να δούμε αν δύο ενδεχόμενα Α και Β είναι ανεξάρτητα, αρκεί να ελέγξουμε αν ισχύει μία από τις σχέσεις (7.3.1), (7.3.2), (7.3.3).

Τρία ενδεχόμενα E_1, E_2 και E_3 είναι στοχαστικώς ανεξάρτητα αν είναι ανά δύο ανεξάρτητα, δηλαδή αν ισχύουν:

(α) $P(E_1 \cap E_2) = P(E_1)P(E_2)$

$P(E_1 \cap E_3) = P(E_1)P(E_3)$,

$P(E_2 \cap E_3) = P(E_2)P(E_3)$

και

(β) $P(E_1 \cap E_2 \cap E_3) = P(E_1)P(E_2)P(E_3)$

Θα πρέπει να τονίσουμε ότι η (β) δεν έπεται από την (α). Δηλαδή μπορεί τρία ενδεχόμενα να είναι ανά δύο ανεξάρτητα και να μην είναι όλα μαζί ανεξάρτητα. Η έννοια της ανεξαρτησίας γενικεύεται εύκολα στα n ενδεχόμενα

Στις περισσότερες πρακτικές εφαρμογές η ανεξαρτησία ή μη των ενδεχομένων ορίζεται από τις συνθήκες κάτω από τις οποίες εκτελείται το πείραμα.

Παράδειγμα 7.3.1 Έστω ότι ρίχνουμε ένα ζεύγος απολύτως συμμετρικών νομισμάτων. Αν συμβολίσουμε με Κ την όψη "κορώνα" και με Γ την όψη "γράμματα" έχουμε τον δειγματικό χώρο

$$S = \{KK, K\Gamma, \Gamma K, \Gamma\Gamma\}$$

Έστω τώρα τα ενδεχόμενα

E_1 ={"γράμματα" στο πρώτο νόμισμα} = {ΓΚ, ΓΓ}

E_2 ={"γράμματα" στο δεύτερο νόμισμα} = {ΚΓ, ΓΓ}

E_3 ={"γράμματα" σε ακριβώς ένα νόμισμα} = {ΚΓ, ΓΚ}

Τότε $P(E_1) = P(E_2) = P(E_3) = \frac{2}{4} = \frac{1}{2}$.

Έχουμε ακόμη

$$P(E_1 \cap E_2) = P(\{\Gamma\Gamma\}) = \frac{1}{4} = P(E_1)P(E_2)$$

$$P(E_1 \cap E_3) = P(\{\Gamma K\}) = \frac{1}{4} = P(E_1)P(E_3)$$

$$P(E_2 \cap E_3) = P(\{K\Gamma\}) = \frac{1}{4} = P(E_2)P(E_3)$$

οπότε η συνθήκη (α) ικανοποιείται, δηλαδή τα ενδεχόμενα είναι ανά δύο ανεξάρτητα. Όμως $E_1 \cap E_2 \cap E_3 = \varnothing$ οπότε

$$P(E_1 \cap E_2 \cap E_3) = P(\varnothing) = 0 \neq P(E_1)P(E_2)P(E_3) = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

Δηλαδή τα τρία ενδεχόμενα ΔΕΝ είναι ανεξάρτητα.

Παράδειγμα 7.3.2 Τα μηχανήματα ενός εργοστασίου χωρίζονται στα τμήματα Α, Β και Γ. Ο υπεύθυνος μηχανικός, από την προηγούμενη πείρα του και τις πληροφορίες του σχετικά με τη συντήρηση των μηχανών, δίνει πιθανότητα εμφάνισης βλάβης στη διάρκεια ενός 24ώρου λειτουργίας, για τα τρία τμήματα, αντίστοιχα, P(A) = 0.3, P(B) = 0.1 και P(Γ) = 0.07. Η οργάνωση της παραγωγής είναι τέτοια ώστε η παραγωγική διαδικασία διακόπτεται μόνον αν εμφανιστεί βλάβη και στα τρία τμήματα. Έστω ότι η βλάβη που παρουσιάζεται σε ένα τμήμα δεν επηρεάζει την πιθανότητα εμφάνισης βλάβης στα υπόλοιπα. Ποιά είναι η πιθανότητα να μη σταματήσει η παραγωγική διαδικασία λόγω βλάβης στη διάρκεια ενός 24ώρου;

Αξιοπιστία (reliability) ενός συστήματος είναι η πιθανότητα να λειτουργήσει ικανοποιητικά, κάτω από ορισμένες συνθήκες και για προκαθορισμένο χρονικό διάστημα.

Αφού η πιθανότητα εμφάνισης βλάβης σε ένα τμήμα δεν επηρεάζεται από την εμφάνιση βλάβης στα υπόλοιπα, έχουμε:

$$P(A \cap B \cap \Gamma) = P(A)P(B)P(\Gamma) = (0.3)(0.1)(0.07) = 0.0021$$

Άρα η πιθανότητα να μη σταματήσει η παραγωγική διαδικασία λόγω βλάβης είναι ίση με $1 - 0.0021 = 0.9979$. Η πιθανότητα αυτή ονομάζεται **αξιοπιστία.**

Παράδειγμα 7.3.3 Σε σοκολατένια αυγουλάκια ορισμένης μάρκας υπάρχουν 16 διαφορετικές φιγούρες σε ίσες αναλογίες και τα αυγουλάκια είναι καλά ανακατεμένα. Αν αγοράσω 6 τέτοια αυγουλάκια ποιά είναι η πιθανότητα να μην έχω καμια επανάληψη;

Η ζητούμενη πιθανότητα ισούται με

$$P(A) = \frac{15}{16}\frac{14}{16}\frac{13}{16}\frac{12}{16}\frac{11}{16} = 0.3437$$

○○● **Σημείωση** Στην περίπτωση που θέλουμε να υπολογίσουμε την πιθανότητα πραγματοποίησης ενός τουλάχιστον από $k > 2$ ανεξάρτητα αλλά όχι ασυμβίβαστα ενδεχόμενα, χρήσιμη είναι η ακόλουθη σχέση

$$\begin{aligned} P(E_1 \cup E_2 \cup ....E_k) &= 1 - P(\overline{E_1 \cap E_2 \cap ....\cap E_k}) \\ &= 1 - P(\overline{E_1})P(\overline{E_2}).....P(\overline{E_K}) \end{aligned}$$

Παράδειγμα 7.3.4 Η πιθανότητα P(E) να συμβεί σοβαρό τροχαίο σε ορισμένο δρόμο ισούται με 0.02. Ποιά είναι η πιθανότητα, για κάποιον που κάνει τη διαδρομή 50 φορές, να έχει σοβαρό τροχαίο;

Αν $P(E_i)$ είναι το ενδεχόμενο τροχαίου στη i διαδρομή, ζητούμε την πιθανότητα του ενδεχομένου "ο οδηγός, στις 50 διαδρομές να πάθει ατύχημα μια τουλάχιστον φορά" δηλαδή το $E_1 \cup E_2 \cup ....E_{50}$.

Το συμπλήρωμά του είναι το ενδεχόμενο "να μην πάθει καμμιά φορά ατύχημα στις 50 διαδρομές" δηλαδή το $\overline{E_1 \cap E_2 \cap ....\cap E_{50}}$, οπότε θα υπολογίσουμε τη ζητούμενη πιθανότητα ως εξής:

$$P(E_1 \cup E_2 \cup ....E_{50}) = 1 - P(\overline{E_1 \cap E_2 \cap ....\cap E_{50}})$$

Υποθέτοντας ότι τα ενδεχόμενα ατυχήματος σε διαφορετικές επαναλήψεις της διαδρομής είναι ανεξάρτητα, η ζητούμενη πιθανότητα ισούται με

$$1 - P(\overline{E}_1)P(\overline{E}_2),.......,P(\overline{E}_{50})$$

ή $$1 - (0.98)^{50} \approx 1 - 0.364 = 0.636$$

○○● **Σημείωση** Είναι εύκολο να πει κάποιος ότι σε 50 διαδρομές είναι βέβαιο ότι θα συμβεί ατύχημα, αφού $(0.02) \times 50 = 1$. Σε έναν τέτοιο συλλογισμό δεν παίρνεται υπόψη ότι τα ενδεχόμενα δεν είναι ασυμβίβαστα και άρα δεν μπορούμε απλώς να προσθέσουμε τις πιθανότητες που αντιστοιχούν στις μεμονωμένες διαδρομές. Πρέπει να τονιστεί οτι, γενικά, ο εμπειρικός τρόπος σκέψης στον υπολογισμό των πιθανοτήτων πρέπει να αποφεύγεται!

Γενέθλια

Απλή στατιστική ανάλυση δείχνει ότι αυτοί που έχουν γιορτάσει τα πιο πολλά γενέθλια είναι οι πιο ηλικιωμένοι. Συμπέρασμα: με το να γιορτάζεις τα γενέθλιά σου εξασφαλίζεις μακροβιότητα!

Με αφορμή τα γενέθλια ας ξεκαθαρίσουμε άλλη μια φορά το πώς υπολογίζονται οι πιθανότητες κάτω από συνθήκες στατιστικής ανεξαρτησίας ή εξάρτησης ακολουθώντας σταδιακά το ακόλουθο παράδειγμα:

Παράδειγμα 7.3.5

▪ **Ποιά είναι η πιθανότητα για δυό αγνώστους να έχουν γενέθλια τον ίδιο μήνα με σένα;**

Έστω E_1 και E_2, τα ενδεχόμενα ο πρώτος και ο δεύτερος άγνωστος, αντίστοιχα, να έχουν γενέθλια τον ίδιο μήνα με σένα. Στο καθένα απο τα ενδεχόμενα αυτά αντιστοιχεί πιθανότητα ιση με (1/12) και επειδή είναι ανεξάρτητα ισχύει:

$$P(E_1 \cap E_2) = P(E_1)P(E_2) = \frac{1}{12}\frac{1}{12} = \frac{1}{144} = 0.007$$

▪ **Ποιά είναι η πιθανότητα για μια ομάδα τριών αγνώστων να έχουν γενέθλια τον ίδιο μήνα;**

$$P(E_1 \cap E_2 \cap E_3) = P(E_1)P(E_2)P(E_3) = \frac{1}{12}\frac{1}{12}\frac{1}{12} = \frac{1}{1728} = 0.00058$$

▪ **Σε μια ομάδα k<12 ατόμων ποιά είναι η πιθανότητα να βρούμε τουλάχιστον δύο με γενέθλια τον ίδιο μήνα**

Θα υπολογίσουμε αρχικά την συμπληρωματική πιθανότητα δηλαδή την πιθανότητα, έστω P(A), για τα k άτομα να έχουν γενέθλια διαφορετικό μήνα. Επειδή

P(για τον δεύτερο να έχει γενέθλια διαφορετικό μήνα από τον πρώτο) = 11/12
P(για τον τρίτο να έχει γενέθλια διαφορετικό μήνα από δύο πρώτους) = 10/12

Υπολογίζουμε:

$$P(A)=\left(\frac{11}{12}\right)\left(\frac{10}{12}\right)\left(\frac{9}{12}\right)......\left(\frac{12-k+1}{12}\right)$$

π.χ. για k=4

$$P(A)=\frac{11}{12}\frac{10}{12}\frac{9}{12}=0.57$$

Οπότε η πιθανότητα για τα 4 άτομα να έχουν γενέθλια τον ίδιο μήνα είναι ίση με $1-P(A)=0.43$.

◘ Σε μια παρέα k ατόμων, με $2 \le k \le 365$, ποια είναι η πιθανότητα να βρεθούν τουλάχιστον δύο που να έχουν γενέθλια την ίδια μέρα του χρόνου;

Θα υπολογίσουμε αρχικά την συμπληρωματική πιθανότητα δηλαδή την πιθανότητα, έστω P(A), για τα k άτομα να έχουν γενέθλια διαφορετική ημέρα. Υποθέτουμε ότι, για κάθε άτομο οι 365 ημέρες του έτους έχουν την ίδια πιθανότητα να είναι ημέρες γενεθλίων και αγνοούμε τα δίσεκτα χρόνια. Τότε το πρώτο άτομο μπορεί να έχει γενέθλια μία από τις 365 ημέρες. Η πιθανότητα για το δεύτερο άτομο να έχει διαφορετική ημέρα γενεθλίων ισούται προς 364/365. Δεδομένου ότι τα δύο πρώτα άτομα έχουν διαφορετική ημέρα γενεθλίων η πιθανότητα για το τρίτο άτομο να έχει διαφορετική ημέρα γενεθλίων ισούται προς 363/365 και ούτω καθεξής.

Οι στατιστικοί αλιεύουν μαργαριτάρια

"Για να βρούμε την πιθανότητα σε μια ομάδα 30 ατόμων δύο τουλάχιστον να έχουν γενέθλια την ίδια μέρα υπολογίζουμε:
$1/366+2/366+.....+29/266 \;=\; 435/366$.
Αυτό το πολύ μεγάλο ποσοστό σημαίνει ότι είναι σχεδόν βέβαιο ότι σε οποιαδήποτε ομάδα 30 ατόμων δύο τουλάχιστον άτομα θα έχουν γενέθλια την ίδια μέρα. Άρα μη παραξενευτείτε αν το παρατηρήσετε-θα έπρεπε να το περιμένετε" *(αλιεύτηκε από την Utts, 2003)*

Έτσι η πιθανότητα, τα k άτομα να έχουν διαφορετική ημερομηνία γενεθλίων είναι ίση προς

$$P(A)=\frac{364}{365}\frac{363}{365}\cdots\frac{365-k+1}{365}=\frac{(364)(363)...(365-k+1)}{365^{k-1}},\quad k\le 365$$

Επομένως η πιθανότητα τουλάχιστον δύο από τα k άτομα να έχουν την ίδια ημέρα γενεθλίων (συμβολίζουμε με P_k) ισούται προς την πιθανότητα του συμπληρώματος του A, δηλαδή

$$P_k=1-P(\bar{A})=1-\frac{(364)(363)...(365-k+1)}{365^{k-1}}$$

Ετσι π.χ. σε μια παρέα k = 5 ατόμων η πιθανότητα να έχουν τουλάχιστον δύο την ίδια ημέρα γενεθλίων υπολογίζεται ως

$$P_5 = 1 - \frac{364}{365}\,\frac{363}{365}\,\frac{362}{365}\,\frac{361}{365} = 1 - 0.9730 = 0.027$$

Με τη βοήθεια Η/Υ υπολογίζεται ότι για k = 33, παίρνουμε $P_k = 0.75$ περίπου. Ομοίως υπολογίζεται ότι ο αριθμός k των ατόμων της παρέας στον οποίο αντιστοιχεί $P_k = 0.51$, ισούται με k = 23. Δηλαδή σε μια παρέα 23 ατόμων η πιθανότητα τουλάχιστον δύο από αυτά να έχουν την ίδια ημέρα γενεθλίων είναι 0.51.
*(*Το πρόβλημα των γενεθλίων σ'αυτή την τέταρτη εκδοχή τέθηκε και αναλύθηκε για πρώτη φορά από τους Kemeny, Snell και Thompson στο βιβλίο τους με τίτλο *Introduction to Finite Mathematics*, Prentice Hall, 1957*)*.

Ασυμβίβαστα και ανεξάρτητα ενδεχόμενα

Όπως είδαμε και στο προηγούμενο παράδειγμα είναι σημαντικό να κάνουμε σαφή διάκριση ανάμεσα στις έννοιες "ασυμβίβαστα" και "ανεξάρτητα" ενδεχόμενα. Ειδικότερα, αν δύο ενδεχόμενα Α και Β είναι ασυμβίβαστα, τότε $A \cap B = \varnothing$ και $P(A \cap B) = 0$ και το αντίστροφο.

Τα ασυμβίβαστα όμως ενδεχόμενα δεν είναι αναγκαστικά και ανεξάρτητα. Είναι δυνατόν $P(A \mid B) = 0$ και $P(A) \neq 0$, οπότε $P(A \mid B) \neq P(A)$ και τα Α και Β δεν είναι ανεξάρτητα. Και αντίστροφα, τα ανεξάρτητα ενδεχόμενα δεν είναι αναγκαστικά και ασυμβίβαστα.
Συνοψίζοντας, έχουμε:

- Για τα **ασυμβίβαστα ενδεχόμενα** Α, Β ισχύει:

i) $P(A \cap B) = 0$, ii) $P(AUB) = P(A)+P(B)$, iii) $P(A \mid B)=0$

- Για τα **ανεξάρτητα ενδεχόμενα** Α, Β ισχύει:

i) $P(A \cap B) = P(A)P(B)$, ii) $P(AUB)=P(A)+P(B) - P(A \cap B)$, iii) $P(A \mid B)=P(A)$

Παράδειγμα 7.3.6 Από μια καλά ανακατωμένη τράπουλα παίρνουμε ένα χαρτί. Αν Α είναι το ενδεχόμενο να είναι "άσσος" και Β το ενδεχόμενο να είναι "σπαθί", τότε τα Α και Β δεν είναι ασυμβίβαστα διότι

$$P(A \cap B) = 1/52$$

Επειδή $P(A) = 4/52$, $P(B) = 13/52$ και ισχύει

$$P(A \cap B) = P(A)P(B) = 1/52$$

τα Α και Β είναι ανεξάρτητα.

Παράδειγμα 7.3.7 Ρίχνουμε ένα ζάρι και έστω A το ενδεχόμενο να πάρουμε την όψη 1, 2, 3 και B το ενδεχόμενο να πάρουμε την όψη 4, 5, 6. Δηλαδή έχουμε $P(A) = 3/6$ και $P(B) = 3/6$.

Επειδή τα A και B είναι ασυμβίβαστα $P(A \cap B) = 0$. Εξάλλου ισχύει $P(A \mid B) = 0$, επειδή η πραγματοποίηση του B συνεπάγεται τη μη πραγματοποίηση του A.
Τέλος επειδή $P(A) \neq P(A \mid B)$ τα A και B δεν είναι ανεξάρτητα.

Παράδειγμα 7.3.8 Ένα ζευγάρι κάνει 3 παιδιά. Ποια είναι η πιθανότητα να είναι όλα του ίδιου φύλου;

Επειδή οι πιθανότητες στις διαδοχικές γεννήσεις είναι ανεξάρτητες, η πιθανότητα να κάνουν τρία κορίτσια

$$P(K_1 \cap K_2 \cap K_3) = \frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{6}$$

και ομοίως η πιθανότητα να κάνουν τρία αγόρια ισούται με

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{6}$$

Επειδή τα δύο ενδεχόμενα είναι ασυμβίβαστα, η πιθανότητα να κάνουν τρία παιδιά του ίδιου φύλου (δηλαδή είτε τρία κορίτσια είτε τρία αγόρια) ισούται με 2/6.

Η Πλάνη του Τζογαδόρου

Όταν, σε μια πεπερασμένη ακολουθία ανεξάρτητων δοκιμών ενός πειράματος τύχης, οι παρατηρούμενες σχετικές συχνότητες αποκλίνουν από τις πιθανότητες, δημιουργείται η πεποίθηση ότι οι επόμενες δοκιμές θα είναι εξισορροπιστικές. Έτσι, π.χ. αν σε 5 διαδοχικές κληρώσεις ενός λαχείου δεν προέκυψε κανένας αριθμός που να λήγει σε ζυγό, πολλοί συστηματικού αγοραστές λαχείου πιστεύουν ότι στην επόμενη κλήρωση η πιθανότητα για ζυγό λήγοντα θα είναι μεγαλύτερη. Επειδή αυτή είναι μια συχνή πλάνη σε όσους παίζουν συστηματικά τυχερά παιχνίδια, στη διεθνή βιβλιογραφία έχει καθιερωθεί ο όρος **πλάνη του τζογαδόρου** (*the gambler's fallacy*) να σημαίνει, γενικά, την λανθασμένη πεποίθηση ότι η πιθανότητα ενός ενδεχομένου είναι μεγαλύτερη όταν έχει πραγματοποιηθεί κάποιο άλλο, παρόλο που τα δύο ενδεχόμενα είναι ανεξάρτητα.

Ένας γιατρός παρηγορεί τον άρρωστο με το ακόλουθο σχόλιο.
"Έχεις μια πολλή σοβαρή αρρώστια και εννιά στους δέκα πάσχοντες πεθαίνουν. Μη στεναχωριέσαι όμως. Είσαι τυχερός που ήρθες σε μένα γιατί πρόσφατα είχα εννιά τέτοιους αρρώστους και όλοι πέθαναν!" (Rao, 1989).

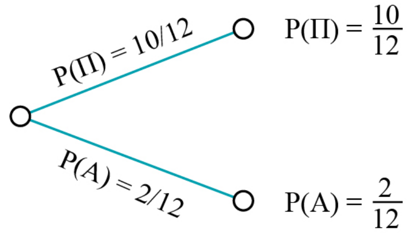
7.4 Δένδρα Πιθανότητας

Όταν ο δειγματικός χώρος έχει λίγα δυνατά αποτελέσματα, η διαδικασία απαρίθμησης και συνακόλουθα ο προσδιορισμός των αντίστοιχων πιθανοτήτων διευκολύνεται σημαντικά με το λεγόμενο **δέντρο πιθανότητας** (*probability tree*). Σε κάθε δέντρο πιθανότητας έχουμε κορυφές και κλαδιά που ορίζουν σύνθετα ενδεχόμενα. Ισχύουν τα εξής:

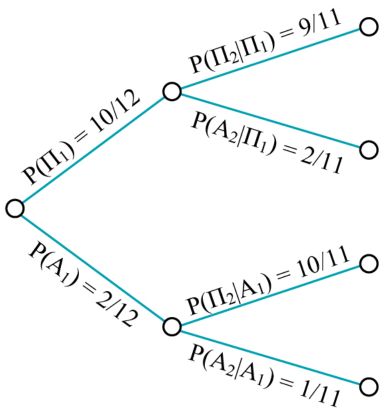
- Η κοινή πιθανότητα κάθε μονοπατιού του δένδρου ισούται με το γινόμενο των πιθανοτήτων όλων των κλαδιών του
- Το άθροισμα των πιθανοτήτων που αντιστοιχούν σε όλα τα κλαδιά που ξεκινούν από μια κορυφή ισούται με 1. Επίσης το άθροισμα των κοινών πιθανοτήτων όλων των μονοπατιών ισούται με 1.

Θα παρουσιάσουμε τα δέντρα πιθανότητας με παραδείγματα, οι αναγνώστες πάντως θα πρέπει να έχουν υπόψη τους ότι ένα δένδρο πιθανότητας μπορεί γρήγορα να γίνει εξαιρετικά δύσχρηστο με την αύξηση των σταδίων του πειράματος.

Παράδειγμα 7.4.1 Από ένα κλειστό κουτί με 10 Πράσινες μπάλες και 2 Άσπρες τραβούμε μία. Τότε ο δειγματικός χώρος έχει δύο αποτελέσματα και οι αντίστοιχες πιθανότητες παριστάνονται με τα δύο κλαδιά του ακόλουθου δένδρου.



Αν τραβήξουμε δύο μπάλες διαδοχικά, τότε ο δειγματικός χώρος έχει τέσσερα αποτελέσματα και οι αντίστοιχες πιθανότητες παριστάνονται με το ακόλουθο δένδρο πιθανότητας:



$$P(\Pi_1\Pi_2) = \frac{9}{11}\cdot\frac{10}{12} = 0.681$$

$$P(\Pi_1 A_2) = \frac{2}{11}\cdot\frac{10}{12} = 0.152$$

$$P(A_1\Pi_2) = \frac{10}{11}\cdot\frac{2}{12} = 0.152$$

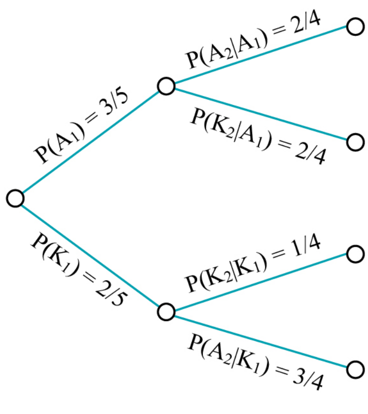
$$P(A_1 A_2) = \frac{1}{11}\cdot\frac{2}{12} = 0.015$$

Άθροισμα 1.000

Ο δείκτης i = 1,2 στο Π και το Α δηλώνει πρώτη και δεύτερη λήψη αντίστοιχα. Έτσι π.χ. $P(\Pi_2 \,|\, A_1)$ δηλώνει την πιθανότητα να πάρουμε πράσινη μπάλα στη δεύτερη λήψη, όταν είχαμε πάρει άσπρη μπάλα στην πρώτη.

Παράδειγμα 7.4.2 Μια παρέα από τρία αγόρια και δύο κορίτσια, μετά το μάθημα της Στατιστικής, έφαγαν στο φαστφούντ γνωστής αλυσίδας και κέρδισαν με το λογαριασμό τους ένα ταξίδι για δύο άτομα στο Λονδίνο. Αποφάσισαν να κληρώσουν το λαχείο βάζοντας 5 χαρτάκια με τα ονόματά τους σε ένα καπέλο. Ποια είναι η πιθανότητα να τύχει να ταξιδέψουν μαζί αγόρι και κορίτσι;

Το δένδρο πιθανότητας που αντιστοιχεί στη διαδικασία αυτή είναι το εξής:



$$P(A_1\cap A_2) = \frac{3}{5}\cdot\frac{2}{4} = 0.30$$

$$P(A_1\cap K_2) = \frac{3}{5}\cdot\frac{2}{4} = 0.30$$

$$P(K_1\cap K_2) = \frac{2}{5}\cdot\frac{1}{4} = 0.10$$

$$P(K_1\cap A_2) = \frac{2}{5}\cdot\frac{3}{4} = 0.30$$

Άθροισμα 1.00

Ο λαχνός θα τύχει σε αγόρι και κορίτσι αν πραγματοποιηθεί το ενδεχόμενο $K_2 \cap A_1$ ή $A_2 \cap K_1$. Επειδή τα ενδεχόμενα αυτά είναι ασυμβίβαστα η πιθανότητα να συμβεί κάποιο απ' αυτά ισούται με:

$$P(K_2 \cap A_1) + P(A_2 \cap K_1) = 0.30 + 0.30 = 0.60$$

Πίνακες Διπλής Εισόδου

Όταν δε γνωρίζουμε αν υπάρχουν συνθήκες στατιστικής ανεξαρτησίας αρκεί να διαθέτουμε τις κοινές πιθανότητες. Είναι σύνηθες στην πράξη οι πιθανότητες αυτές να δίνονται με τη μορφή ενός **πίνακα διπλής εισόδου** *(two-way or contingency table)* όπως στο ακόλουθο παράδειγμα:

Παράδειγμα 7.4.3 Διαφημιστική εταιρεία με μακρόχρονη πείρα στην αποστολή ερωτηματολογίων υπολόγισε τον ακόλουθο πίνακα για τις πιθανότητες να είναι ο παραλήπτης άνδρας (Α) ή γυναίκα (Γ) και να απαντήσει (ΝΑΙ) ή να μην απαντήσει (ΟΧΙ).

Απάντηση / Φύλο		ΝΑΙ	ΟΧΙ
Άνδρες	Α	0.10	0.30
Γυναίκες	Γ	0.25	0.35

Έτσι π.χ. η πιθανότητα για κάποιον παραλήπτη να είναι άνδρας και να απαντήσει ισούται με 0.10 ενώ η πιθανότητα να είναι γυναίκα και να μην απαντήσει ισούται με 0.35. Η πιθανότητα να είναι ο παραλήπτης γυναίκα ισούται με:

$$P(\Gamma) = P[\Gamma \cap N \cup \Gamma \cap O] = P(\Gamma \cap N) + P(\Gamma \cap O)$$

αφού τα ενδεχόμενα $\Gamma \cap N$ και $\Gamma \cap O$ είναι ασυμβίβαστα.
Με τρόπο ανάλογο υπολογίζονται όλες οι απόλυτες πιθανότητες και γράφονται στο περιθώριο του πίνακα (γι' αυτό και ονομάζονται περιθώριες πιθανότητες). Έτσι ο προηγούμενος πίνακας θα πάρει τη μορφή

Απάντηση / Φύλο	Ν	Ο	Άθροισμα
Α	0.10	0.30	0.40
Γ	0.25	0.35	0.60
Άθροισμα	0.35	0.65	1.00

Για τα δεδομένα του πίνακα αυτού έχουμε (όπου π.χ. AN να διαβάζεται A∩N,):

$$P(N\,|\,A)=\frac{P(AN)}{P(A)}=\frac{0.10}{0.40}=0.25$$

$$P(O\,|\,A)=\frac{P(OA)}{P(A)}=\frac{0.30}{0.40}=0.75$$

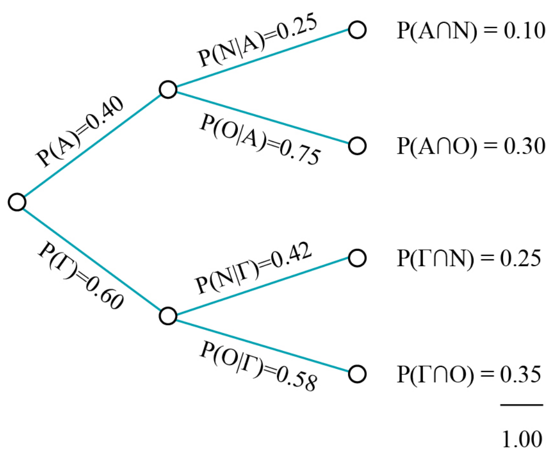
$$P(N\,|\,\Gamma)=\frac{P(N\Gamma)}{P(\Gamma)}=\frac{0.25}{0.60}=0.42$$

$$P(O\,|\,\Gamma)=\frac{P(O\Gamma)}{P(\Gamma)}=\frac{0.35}{0.60}=0.58$$

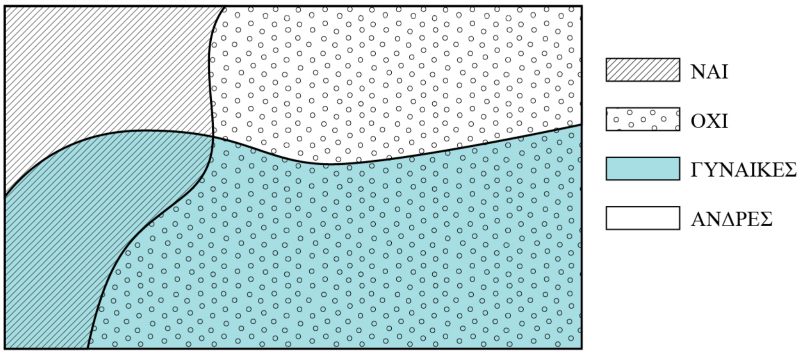
Παρατηρούμε ότι το φύλο και η συμπεριφορά ως προς τα ερωτηματολόγια δεν είναι ανεξάρτητα διότι έχουμε π.χ.

$$P(N\,|\,A) = 0.25 \neq P(N) = 0.35$$

Οι πληροφορίες αυτές μπορούν να δοθούν και με τη μορφή ενός δένδρου πιθανοτήτων ως εξής:



Σημείωση Το φύλο αποτελεί ένα διαμερισμό του δειγματικού χώρου η στάση ως προς το ερωτηματολόγιο αποτελεί άλλον διαμερισμό του δειγματικού χώρου. Έτσι, με ένα βέννειο διάγραμμα μπορούμε να παρουσιάσουμε γραφικά τις περιθώριες πιθανότητες ως εξής:



7.5 Οι Πιθανότητες στην Απλή Τυχαία Δειγματοληψία από Απαριθμήσιμο Πληθυσμό

Έστω ότι από έναν απαριθμήσιμο πληθυσμό N στοιχείων θέλουμε να πάρουμε ένα δείγμα μεγέθους n. Όταν τα στοιχεία του δείγματος παίρνονται με τέτοιο τρόπο ώστε κάθε δυνατή n–ιάδα να έχει την ίδια πιθανότητα να περιληφθεί στο δείγμα, λέμε ότι έχουμε **απλή τυχαία δειγματοληψία** *(simple random sampling)* και το δείγμα ονομάζεται **απλό τυχαίο δείγμα** *(simple random sample)* ή **τυχαίο δείγμα**.

Σε επόμενα κεφάλαια αναλύεται διεξοδικά το πώς εφαρμόζεται στην πράξη αυτή αλλά και άλλες μέθοδοι δειγματοληψίας. Στο σημείο αυτό είναι αρκετό να κατανοήσουμε την απλή τυχαία δειγματοληψία ως μια διαδικασία κλήρωσης και θα υπολογίσουμε τις πιθανότητες κάτω απο διαφορετικά σχήματα λήψης των στοιχείων. Διακρίνουμε τις ακόλουθες περιπτώσεις:

- **Παίρνουμε με τυχαίο τρόπο ένα στοιχείο του πληθυσμού, το εξετάζουμε και το επαναθέτουμε στον πληθυσμό πριν να πάρουμε το επόμενο στοιχείο.**

Στην περίπτωση αυτή, το πλήθος των δυνατών δειγμάτων που μπορεί να προκύψουν ισούται με

$$\underbrace{N.N.N.....N}_{(n \text{ φορές})} = N^n$$

δηλαδή όσες οι *μεταθέσεις* των N στοιχείων ανά n.

Είναι προφανές ότι, στη δειγματοληψία με επανάθεση, η πιθανότητα πραγματοποίησης ενός αποτελέσματος σε μια λήψη είναι ανεξάρτητη από το αποτέλεσμα των προηγούμενων λήψεων. Το μέγεθος του δείγματος μπορεί να είναι, θεωρητικά, απεριόριστο γι' αυτό και η δειγματοληψία **με επανάθεση** λέγεται και **άπειρη δειγματοληψία** *(infinite sampling)*.

Σε κάθε δείγμα αντιστοιχει η ίδια πιθανότητα η οποία στην δειγματοληψία με επανάθεση ισούται με

$$P = 1/N^n$$

Παράδειγμα 7.5.1 Από 4 φοιτητές, τους Άλφα, Βήτα, Γάμα και Δέλτα, θέλουμε να πάρουμε ένα τυχαίο δείγμα μεγέθους n = 2 με απλή τυχαία δειγματοληψία με επανάθεση. Τα δυνατά δείγματα είναι τα εξής: $4^2 = 16$

{ΑΑ, ΑΒ, ΑΓ, ΑΔ, ΒΑ, ΒΒ, ΒΓ, ΒΔ, ΓΑ, ΓΒ, ΓΓ, ΓΔ, ΔΑ, ΔΒ, ΔΓ, ΔΔ}

Απλή τυχαία είναι η δειγματοληψία που δίνει σε κάθε μία από τις δυάδες αυτές ίση πιθανότητα. Ένας τρόπος για να το πετύχουμε αυτό είναι ο εξής: θα γράψουμε σε 4 ίδια χαρτιά τα 4 ονόματα, θα τα ρίξουμε σε ένα καπέλο αφού πρώτα τα τυλίξουμε και θα τραβήξουμε ένα. Στη συνέχεια επαναφέρουμε αυτό που προέκυψε στην πρώτη λήψη στο καπέλο και τραβούμε το δεύτερο. Και στις δύο ρίψεις κάθε στοιχείο έχει την ίδια πιθανότητα να εμφανιστεί και η πιθανότητα μιας δυάδας έστω της ΒΓ ισούται με:

$$P(\text{ΒΓ}) = P(\text{Β})P(\text{Γ}) = \left(\frac{1}{4}\right)\left(\frac{1}{4}\right) = \frac{1}{16}$$

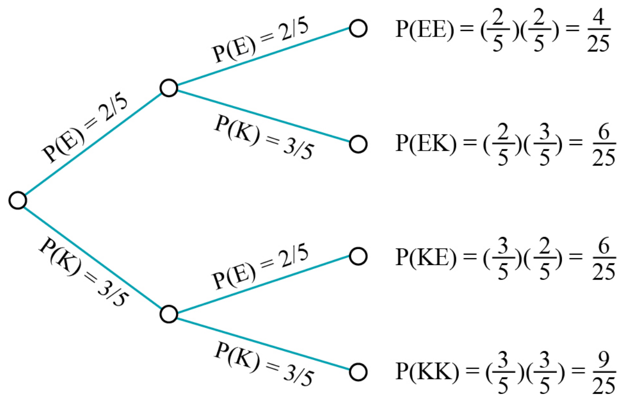
Παράδειγμα 7.5.2 Σε ένα κουτί περιέχονται 5 προϊόντα από τα οποία τα δύο είναι ελαττωματικά. Παίρνουμε διαδοχικά δύο προϊόντα από το κουτί με επανάθεση. Να βρεθεί η πιθανότητα να πάρουμε ένα ελαττωματικό (Ε) προϊόν και ένα καλό (Κ) αν η σειρά εμφάνισης δεν έχει σημασία.

Επειδή η δειγματοληψία γίνεται με επανάθεση το πλήθος των δυνατών δειγμάτων ισούται με (5)(5) = 25. Μας ενδιαφέρει το ενδεχόμενο να πάρουμε:

(i) είτε στην πρώτη λήψη ένα καλό (Κ) και στη δεύτερο ένα ελαττωματικό (Ε)
(ii) είτε στην πρώτη λήψη ένα Ε και στη δεύτερη ένα Κ.

Οι δυνατοί τρόποι με τους οποίους μπορεί να πραγματοποιηθεί το (i) είναι 3×2=6, και για το (ii) οι δυνατοί τρόποι είναι 2×3=6. Επειδή η σειρά εμφάνισης δεν έχει σημασία το ενδεχόμενο που μας ενδιαφέρει μπορεί να πραγματοποιηθεί με 12 τρόπους οπότε η ζητούμενη πιθανότητα ισούται με 12/25.

Η πιθανότητα αυτή προσδιορίζεται ευκολότερα με το ακόλουθο δένδρο πιθανότητας. Στο δένδρο αυτό ΕΕ σημαίνει το πρώτο ελαττωματικό και το δεύτερο ελαττωματικό, ΕΚ το πρώτο ελαττωματικό και το δεύτερο καλό και ούτω καθεξής.



Παράδειγμα 7.5.3 Σε λαχείο που κυκλοφορεί σε όλη τη χώρα, ο κάθε λαχνός έχει 9 ψηφία και είναι γνωστό ότι κυκλοφόρησαν τόσοι λαχνοί όσες οι δυνατές 9-άδες. Ποιά είναι η πιθανότητα που αντιστοιχεί σε κάθε λαχνό υποθέτοντας οτι υπάρχουν επαναλαμβανόμενα ψηφία;;

Η κλήρωση ενός λαχείου είναι μια τυπική διαδικασία απλής τυχαίας δειγματοληψίας με επανάθεση. Απο τα 10 ψηφία {0,1,2,3,....8,9} μπορούν να δημιουργηθούν 10^9 = 1.000.000.000 λαχνοί οπότε η πιθανότητα κλήρωσης για τον κάθε έναν απο αυτούς ισούται με 0.000000001!

Παράδειγμα 7.5.4 Αν για τις ταμπέλες αυτοκινήτου, χρησιμοποιούμε 4 ψηφία από το 0 ως το 9, τότε i) πόσα αυτοκίνητα μπορούμε να καταγράψουμε; ii) πόσα γίνονται τα αυτοκίνητα βάζοντας και δύο από τα 24 γράμματα μπροστά από τους αριθμούς; iii) αν κάποιος θυμάται μόνο τα δύο τελευταία ψηφία ενός αριθμού κυκλοφορίας πόσα αυτοκίνητα πρέπει να ελεγχθούν για να το βρούμε;

Στην ουσία πρέπει να καταγράψουμε τα δυνατά αποτελέσματα στις 3 περιπτώσεις. Έχουμε:

(i) μπορούμε να καταγράψουμε 10^4=10000 αυτοκίνητα
(ii) τα αυτοκίνητα γίνονται $(24)^2$10000=5760000
(iii) $(24)^2 \cdot 10^2 = 57600$

◘ Παίρνουμε με τυχαίο τρόπο ένα στοιχείο του πληθυσμού και, χωρίς να το επανατοποθετήσουμε, παίρνουμε το επόμενο. Συνεχίζουμε μέχρι να πάρουμε n στοιχεία.

Σε 2 λήψεις το πλήθος των δυνατών αποτελεσμάτων είναι

$$N \cdot (N-1)$$

Σε n λήψεις το πλήθος των δυνατών αποτελεσμάτων είναι

$$N \cdot (N-1) \cdot \ldots\ldots\ldots \cdot (N-n+1)$$

δηλαδή όσες είναι οι *διατάξεις* των N στοιχείων ανά n. Αν n λήψεις γίνονται με κλήρωση, τότε σε κάθε δείγμα αντιστοιχεί ίδια πιθανότητα η οποία στην δειγματοληψία χωρίς επανάθεση ισούται με

$$P' = \frac{1}{N \cdot (N-1) \cdot \ldots\ldots \cdot (N-n+1)}$$

Παράδειγμα 7.5.4 Σε ένα κουτί υπάρχουν 8 όμοια χαρτιά τυλιγμένα στα οποία είναι σημειωμένα τα γράμματα Α, Β, Γ, Δ, Ε, Ζ, Η, Θ και θέλουμε να πάρουμε ένα δείγμα τριών χαρτιών, χωρίς επανάθεση. Ποια είναι η πιθανότητα το πρώτο γράμμα να είναι το Α;

Ο αριθμός των δυνατών δειγμάτων ισούται με 8·7·6=336. Απ’ αυτά υπάρχουν 7·6=42 δείγματα στα οποία το πρώτο γράμμα είναι το Α. Επομένως η ζητούμενη πιθανότητα είναι ίση με 42/336 = 1/8.

◘ Τα n στοιχεία του δείγματος παίρνονται ταυτόχρονα και δεν μας ενδιαφέρει η σειρά εμφάνισης.

Σε 2 λήψεις το πλήθος των δυνατών αποτελεσμάτων είναι

$$\frac{N \cdot (N-1)}{2}$$

Σε n λήψεις το πλήθος των δυνατών αποτελεσμάτων είναι

$$C_n^N = \frac{N \cdot (N-1) \cdot \ldots\ldots \cdot (N-n+1)}{n \cdot (n-1) \cdot \ldots\ldots \cdot 1} = \frac{N!}{(N-n)!\,n!}$$

δηλαδή όσοι οι *συνδυασμοί* των N στοιχείων ανα n και η πιθανότητα για το καθένα από αυτά είναι

$$\frac{1}{C_n^N}$$

Στην απλη τυχαία δειγματοληψία με ταυτόχρονη λήψη στοιχείων σε κάθε δείγμα αντιστοιχει πιθανότητα ίση με

$$P'' = n!(N-n)!/N!$$

Παράδειγμα 7.5.5 Από έναν πληθυσμό 4 στοιχείων, των Α,Β,Γ και Δ, θα πάρουμε δείγμα 2 στοιχείων ταυτόχρονα (η σειρά εμφάνισης δεν έχει σημασία). Τα δυνατά αποτελέσματα είναι:

{ΑΒ, ΑΓ, ΑΔ, ΒΓ, ΒΔ, ΓΔ}

Δηλαδή όσοι είναι και οι συνδυασμοί των τεσσάρων πραγμάτων ανά δύο

$$C_2^4 = \frac{4!}{2!\cdot 2!} = 6$$

○○● **Σημείωση** Όταν το μέγεθος N του πληθυσμού είναι πολύ μεγάλο σε σχέση με το μέγεθος του δείγματος n, ισχύει:

$$P \cong P' \cong P''$$

οπότε δεν έχει σημασία αν η δειγματοληψία γίνεται με επανάθεση ή χωρίς. Αυτό είναι πολύ σημαντικό γιατί, όπως θα δούμε στα επόμενα κεφάλαια, η θεωρία κάνει την υπόθεση οτι οι λήψεις των στοιχείων είναι ανεξάρτητες πράγμα που ισχύει μόνο στην δειγματοληψία με επανάθεση. Συνήθως όμως δεν μας ενδιαφέρει να εξετάσουμε κάποιο στοιχείο του πληθυσμού περισσότερες απο μια φορές οπότε, στην πράξη, η δειγματοληψία γίνεται χωρίς επανάθεση και συνηθέστερα με ταυτόχρονη λήψη στοιχείων.

Παράδειγμα 7.5.6 Οι 4 ρίψεις ενός νομίσματος είναι ένα τυχαίο δείγμα από τον πληθυσμό όλων των δυνατών ρίψεων που είναι άπειρος. Σε 4 ρίψεις θα προκύψει ένα από τα δυνατά αποτελέσματα που συνολικά είναι όσες και οι μεταθέσεις των 2 ανα 4 δηλαδή με $2^4 = 16$. Τα αποτελέσματα αυτά, δηλαδή οι 16 ακολουθίες των δύο όψεων δίνονται στην συνέχεια χωρισμένες σε 5 ομάδες

α. {ΓΓΓΓ}
β. {ΓΓΓΚ, ΓΓΚΓ, ΓΚΓΓ, ΚΓΓΓ}
γ. {ΓΓΚΚ, ΓΚΚΓ, ΓΚΓΚ, ΚΓΓΚ, ΚΚΓΓ, ΚΓΚΓ}
δ. {ΚΚΚΓ, ΚΚΓΚ, ΚΓΚΚ, ΓΚΚΚ}
ε. {ΚΚΚ}

Τονίζεται ότι στη απλή τυχαία δειγματοληψία σε κάθε μία από τις 16 ακολουθίες αντιστοιχεί πιθανότητα ίση με $(0.5)^4=1/2^4$. Αν μας ενδιαφέρει να προσδιορίσουμε το πλήθος των δειγμάτων στα οποία ο αριθμός των όψεων "κορώνα" (Κ) ισούται με 2 αυτός είναι ίσος με το πλήθος των συνδυασμών των 4 πραγμάτων ανά 2 δηλαδή με

$$\frac{4!}{2!2!}=6$$

Οι ακολουθίες που ικανοποιούν το κριτήριο αυτό είναι αυτές που περιέχονται στην ομάδα γ.

Παράδειγμα 7.5.7 Στο συρτάρι της υπεύθυνης στην Βιβλιοθήκη υπάρχουν 6 φοιτητικά βιβλιάρια από τα οποία τα 4 ανήκουν σε φοιτητές και τα 2 σε φοιτήτριες. Αν πάρουμε συγχρόνως τρία απ' αυτά με τρόπο τυχαίο ποια είναι η πιθανότητα να ανήκουν και τα τρία σε φοιτητή;

Επειδή παίρνουμε συγχρόνως τα 3 βιβλιάρια, ο αριθμός των δυνατών δειγμάτων ισούται με $\binom{6}{3}=\frac{6!}{(3!)(3!)}=20$. Εξάλλου, υπάρχουν $\binom{4}{3}=4$ δυνατοί τρόποι να πάρουμε 3 βιβλιάρια φοιτητών από τα 4 που υπάρχουν. Έτσι η ζητούμενη πιθανότητα ισούται με 4/20 = 1/5.

7.6 Περιθώρια ή Ολική Πιθανότητα

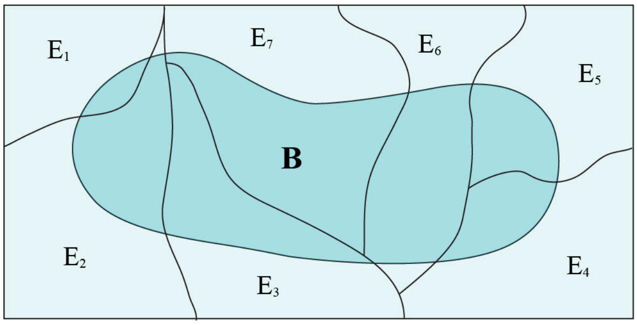
Η $P(A \cap B)$ ονομάζεται **κοινή πιθανότητα** *(joint probability)* του Α και του Β, η οποία για λόγους ευκολίας θα συμβολίζεται με P(AB) και θα διαβάζεται ως πιθανότητα του Α και Β.
Όταν τα Α και Β δεν είναι ανεξάρτητα τότε από τις (7.2.2) και (7.2.3) παίρνουμε:

$$\begin{aligned} P(A\cap B) \equiv P(AB) &= P(A\,|\,B)P(B) \\ &= P(B\,|\,A)P(A) \end{aligned}$$

Στο κεφάλαιο 6 είδαμε ότι ένας διαμερισμός έστω $B_1, B_2, ..., B_k$ του δειγματικού χώρου S διαμερίζει και οποιοδήποτε άλλο ενδεχόμενο Α του S. Έτσι, το Α μπορεί να γραφεί ως

$$\begin{aligned} A = A\cap S &= A\cap(B_1\cup B_2\cup ...\cup B_k) \\ &= (A\cap B_1)\cup(A\cap B_2)\cup ...\cup(A\cap B_k) \end{aligned}$$

όπου τα ενδεχόμενα $A \cap B_1, A \cap B_2, ..., A \cap B_k$ είναι ανά δύο ασυμβίβαστα και η ένωσή τους είναι το Α. Το σχήμα που ακολουθεί είναι επανάληψη του αντίστοιχου σχήματος του κεφαλαίου 6 και παρουσιάζει έναν τέτοιο διαμερισμό του Β με k=7.



Εφαρμόζοντας το αξίωμα της πρόσθεσης υπολογίζουμε

$$P(A) = P(AB_1) + P(AB_2) + ... + P(AB_k) \qquad (7.6.1)$$

όπου κάθε όρος $P(A B_i)$ μπορεί να αντικατασταθεί με $P(A \mid B_i)P(B_i)$ οπότε προκύπτει

$$P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + ... + P(A \mid B_\kappa)P(B_\kappa) \qquad (7.6.2)$$

Η P(A) ονομάζεται **περιθώρια πιθανότητα** *(marginal probability)* ή **ολική πιθανότητα** *(total probability)* και οι (7.6.1) και (7.6.2) είναι πάρα πολύ χρήσιμες για τον υπολογισμό της

Η Τεχνική της Τυχαιοποιημένης Απάντησης[1]

Στις κοινωνικές έρευνες συχνά υπάρχουν ερωτήσεις “ευαίσθητες” που έχουν σχέση π.χ. με φοροκλοπή, χειροδικία συζύγου ή παιδιών, ναρκωτικά ή ερωτική συμπεριφορά. Ένας τρόπος να πάρουμε απάντηση σε μια τέτοια ερώτηση παρουσιάζεται στο εξής παράδειγμα:

---

[1] ή **“***Πώς να πάρετε μια απάντηση χωρίς να είστε βέβαιοι ότι έχετε κάνει την ερώτηση***”** **(**βλέπε Campbell C. – Joiner B. ,1973)

Παράδειγμα 7.6.1 Μαζί με την "ευαίσθητη" ερώτηση βάζουμε και μια άλλη άσχετη ερώτηση για την οποία γνωρίζουμε τις πιθανότητες των δυνατών απαντήσεων. Π.χ.

Α. *απατήσατε τον (την) σύντροφό σας έστω και μια φορά*;
Β. *είναι ο αριθμός της ταυτότητάς σας περιττός αριθμός*;

και ζητούμε από τον ερωτώμενο να ρίξει ένα μη δολιευμένο νόμισμα και αν εμφανιστεί "κορώνα" να απαντήσει στην ερώτηση Α ενώ αν εμφανιστεί "γράμματα" να απαντήσει στην ερώτηση Β. Αν σε μία τέτοια έρευνα απάντησαν "ναι" (Ν) ποσοστό 36% των απαντησάντων τότε διαθέτουμε τις εξής πληροφορίες:

$$P(N) = 0.36 \quad P(A) = 0.5 \quad P(B) = 0.5 \quad P(N \mid B) = 0.5$$

όπου π.χ. P(A) είναι η πιθανότητα κάποιος να έχει απαντήσει στην πρώτη ερώτηση. Από τα δεδομένα αυτά και τον τύπο (7.6.2) παίρνουμε:

$$\begin{aligned} P(N) &= P(NA) + P(NB) \\ &= P(N \mid A)P(A) + P(N \mid B)P(B) \end{aligned}$$

ή

$$\begin{aligned} &0.36 = P(N \mid A)0.5 + (0.5)(0.5) \\ &\Leftrightarrow P(N \mid A) = \frac{0.36 - 0.25}{0.5} = 0.22 \end{aligned}$$

*(*Η τεχνική της τυχαιοποιημένης απάντησης πρωτοεφαρμόστηκε από τους*)*

Περιθώριες Πιθανότητες σε συνθήκες στατιστικής ανεξαρτησίας

Όταν η πραγματοποίηση ενός ενδεχομένου δεν επηρεάζει την πραγματοποίηση οποιουδήποτε άλλου ενδεχομένου, τότε έχουμε συνθήκες στατιστικής ανεξαρτησίας. Στην περίπτωση αυτή είδαμε ότι ισχύει:

- Η κοινή πιθανότητα πραγματοποίησης δύο ή περισσότερων ενδεχομένων συγχρόνως ή διαδοχικά εμφανιζομένων ισούται με το γινόμενο των περιθώριων πιθανοτήτων τους.
- Η δεσμευμένη πιθανότητα κάθε ενδεχομένου ισούται με την περιθώρια πιθανότητά του.

Τα δένδρα πιθανότητας και οι πίνακες διπλής εισόδου είναι και εδώ χρήσιμοι τρόποι παρουσίασης των διαθέσιμων πληροφοριών για όλα τα δυνατά αποτελέσματα του δειγματικού χώρου.

Παράδειγμα 7.6.2 Ένας κύβος έχει 4 κόκκινες και 2 άσπρες πλευρές και είναι συμμετρικός έτσι ώστε σε μία ρίψη κάθε πλευρά του έχει την ίδια πιθανότητα εμφάνισης. Αν ρίξουμε τον κύβο τρεις φορές τότε ποια είναι η πιθανότητα να πάρουμε τουλάχιστον μια φορά μία άσπρη πλευρά;

Οι συνδυασμοί ενδεχομένων που ικανοποιούν τη συνθήκη αυτή είναι

$$K_1K_2A_3 \quad K_1A_2K_3 \quad K_1A_2A_3 \quad A_1K_2K_3 \quad A_1K_2A_3 \quad A_1A_2K_3 \quad A_1A_2A_3$$

Τα ενδεχόμενα όμως αυτά είναι ασυμβίβαστα και επομένως η πιθανότητα να συμβεί τουλάχιστον ένα από αυτά ισούται με το άθροισμα των πιθανοτήτων τους. Δηλαδή έχουμε:

$$\left(\frac{4}{6}\right)\left(\frac{4}{6}\right)\left(\frac{2}{6}\right)+\left(\frac{4}{6}\right)\left(\frac{2}{6}\right)\left(\frac{4}{6}\right)+\left(\frac{2}{6}\right)\left(\frac{4}{6}\right)\left(\frac{2}{6}\right)+\left(\frac{2}{6}\right)\left(\frac{4}{6}\right)\left(\frac{4}{6}\right)+$$
$$+\left(\frac{2}{6}\right)\left(\frac{4}{6}\right)\left(\frac{2}{6}\right)+\left(\frac{2}{6}\right)\left(\frac{2}{6}\right)\left(\frac{4}{6}\right)+\left(\frac{2}{6}\right)\left(\frac{2}{6}\right)\left(\frac{2}{6}\right)=0.704$$

Επειδή υπάρχει μόνον ένα αποτέλεσμα στο οποίο δεν υπάρχει η λευκή όψη, η ζητούμενη πιθανότητα υπολογίζεται ευκολότερα ως (βλέπε και σημείωση στο 7.3)

$$1-P(K_1K_2K_3)=1-\left(\frac{4}{6}\right)^3=0.704$$

Μυστικά και ψέματα

Την παραμονή των εξετάσεων στο μάθημα της Στατιστικής, δύο συμφοιτητές πήγαν στο κλαμπ απ' όπου δεν τα κατάφεραν να φύγουν έγκαιρα γιατί εκεί συνάντησαν δύο όμορφες συμφοιτήτριές τους-καταπολεμούσαν και αυτές το άγχος για το ίδιο μάθημα! Οι δύο καλοί συμφοιτητές το πρωΐ δυσκολεύτηκαν να ξυπνήσουν με αποτέλεσμα να φτάσουν στις εξετάσεις μισή ώρα πριν το τέλος**.** Είπαν στον υπεύθυνο καθηγητή ότι έπαθαν λάστιχο καθώς ερχόταν με αυτοκίνητο από τη διπλανή πόλη και τον παρακάλεσαν να τους εξετάσει. Επειδή ήταν τακτικοί στις παρακολουθήσεις, ο καθηγητής ξαναετοίμασε ένα τεστ και τους έβαλε να γράψουν σε διαφορετικά γραφεία (και χωρίς κινητό!). Το τεστ είχε δύο μέρη: Το πρώτο ήταν μάλλον εύκολο και έπαιρνε 3 μονάδες. Στο δεύτερο μέρος που βρισκόταν στη μέσα σελίδα, υπήρχε μόνο η ερώτηση “ποιό ήταν το λάστιχο;” και έπιανε 7 μονάδες. Μπορείτε να ελέγξετε το πόσο καλά κατανοήσατε αυτό και το προηγούμενο κεφάλαιο απαντώντας στις ακόλουθες ερωτήσεις: i) Ποιά είναι η πιθανότητα να δώσουν την ίδια απάντηση; ii) Πόση θα ήταν η ζητούμενη πιθανότητα αν αντί να αφήσουν τα κορίτσια να πάνε σπίτι τους τα έπειθαν να πάνε μαζί τους και το πρωΐ οργάνωναν το ψέμα προς τον καθηγητή και οι τέσσερις μαζί;

7.7 Το Θεώρημα του Bayes

Έστω $B_1, B_2, ..., B_\kappa$ ένας διαμερισμός του δειγματικού χώρου S και A ένα οποιοδήποτε άλλο ενδεχόμενο του S για το οποίο P(A) > 0. Τότε σύμφωνα με τις (7.6.1) και (7.6.2) ισχύει:

$$P(A) = \sum_{j=1}^{\kappa} P(AB_j) \Leftrightarrow P(A) = \sum_{j=1}^{\kappa} P(A \mid B_j)P(B_j) \qquad (7.7.1)$$

Εξάλλου, από τον ορισμό της δεσμευμένης πιθανότητας έχουμε:

$$P(A \mid B_j) = \frac{P(AB_j)}{P(B_j)}, \quad P(B_j) > 0 \qquad (7.7.2)$$

και

$$P(B_j \mid A) = \frac{P(AB_j)}{P(A)}, \quad P(A) > 0 \qquad (7.7.3)$$

Από το συνδυασμό των (7.7.2) και (7.7.3) παίρνουμε:

$$P(AB_j) = P(A \mid B_j)P(B_j) = P(B_j \mid A)P(A) \qquad (7.7.4)$$

Τέλος με αντικατάσταση των (7.7.4) και (7.7.1) στην (7.7.3) παίρνουμε:

$$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{\sum_{j=1}^{\kappa} P(A \mid B_j)P(B_j)}$$

ή

$$P(B_j \mid A) = P(B_j)\frac{P(A \mid B_j)}{\sum_{j=1}^{\kappa} P(A \mid B_j)P(B_j)} \qquad (7.7.5)$$

Η εξίσωση (7.7.5) είναι γνωστή ως **εξίσωση του Bayes** *(Bayesian equation)* ή **θεώρημα του Bayes** *(Bayes' theorem)* και οφείλει το όνομά του στον Άγγλο υπουργό και μαθηματικό Rev. Thomas Bayes (1702–1761).

Το θεώρημα του Bayes αποτελεί τη βάση ενός ολόκληρου κλάδου της στατιστικής συμπερασματολογίας που ονομάζεται **Bayesian Στατιστική** και που παρουσιάζει πολλές και ενδιαφέρουσες εφαρμογές όχι μόνο σε κοινωνικές επιστήμες, αλλά και σε κάθε επιστημονική περιοχή.

Η αρχή της Στατιστικής αυτής είναι ο υποκειμενικός προσδιορισμός της $P(B_j)$ η οποία ονομάζεται **αρχική** ή **εκ των προτέρων πιθανότητα** *(initial or prior probability)* του ενδεχομένου B_j. Μετά την πραγματοποίηση του A η πιθανότητα του B_j αναθεωρείται σύμφωνα με την εξίσωση (7.6.5). Η $P(B_j \mid A)$ ονομάζεται **αναθεωρημένη** ή **εκ των υστέρων πιθανότητα** *(revised or posterior probability)* του B_j.

Η εξίσωση του Bayes λοιπόν μπορεί να θεωρηθεί ως ένας μηχανισμός αναπροσαρμογής της αρχικής πιθανότητας σε αναθεωρημένη πιθανότητα μετά την απόκτηση πληροφορίας για την πραγματοποίηση του A. Ο λόγος $P(A \mid B_j) / P(A)$ που πολλαπλασιάζει την αρχική πιθανότητα $P(B_j)$ είναι το μέτρο της σχετικότητας της πληροφορίας αυτής. Έτσι όταν ο λόγος αυτός ισούται με 1 τότε η πληροφορία είναι άσχετη με το B_j και $P(B_j \mid A) = P(B_j)$. Οποιαδήποτε άλλη τιμή μεγαλύτερη ή μικρότερη δηλώνει σχετικότητα της πληροφορίας και επομένως

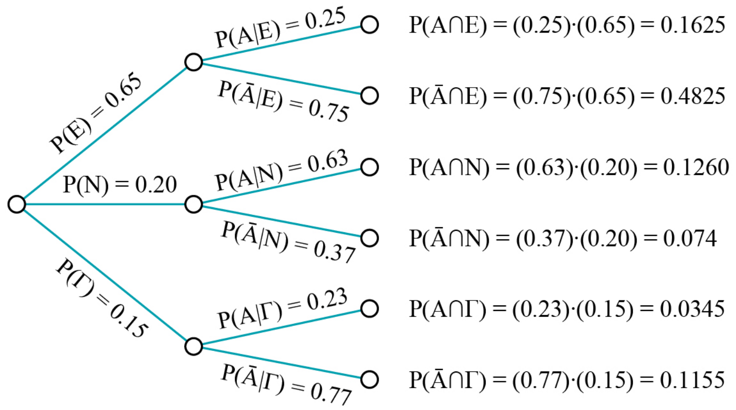
$$P(B_j \mid A) \neq P(B_j)$$

Παράδειγμα 7.7.1 Ένα οικονομικό μάθημα διδάσκεται από κοινού σε φοιτητές του Οικονομικού, της Νομικής και της Γεωπονικής. Το 65% των φοιτητών που έχουν γραφεί στο μάθημα αυτό προέρχεται από το Οικονομικό, το 20% από τη Νομική και το 15% από τη Γεωπονική. Από τους εγγεγραμμένους φοιτητές του Οικονομικού δίνουν εξετάσεις στην περίοδο του Ιουνίου το 25% ενώ οι αναλογίες για τους φοιτητές της Νομικής και της Γεωπονικής είναι αντίστοιχα, 63% και 23%. Ζητείται να υπολογιστούν:(i) Η πιθανότητα P(A), ένας φοιτητής που έχει επιλεγεί στην τύχη να δώσει το μάθημα αυτό τον Ιούνιο, (ii) Δεδομένου ότι ένας φοιτητής έρχεται να δώσει το μάθημα τον Ιούνιο να υπολογιστεί η πιθανότητα αυτός να προέρχεται από το Οικονομικό.

(i) Έστω Ε, Ν, Γ τα ενδεχόμενα ένας φοιτητής να προέρχεται από το Οικονομικό, τη Νομική και τη Γεωπονική, αντίστοιχα. Τότε έχουμε:

$$\begin{aligned} P(A) &= P(A \cap E) + P(A \cap N) + P(A \cap \Gamma) \\ &= P(A \mid E)P(E) + P(A \mid N)P(N) + P(A \mid \Gamma)P(\Gamma) \\ &= (0.25)(0.65) + (0.63)(0.20) + (0.23)(0.15) \\ &= 0.323 \end{aligned}$$

Η ανάλυση αυτή μπορεί να παρασταθεί με το εξής δένδρο πιθανότητας:



$$\text{(ii)}\ \ P(E\,|\,A)=\frac{P(E)P(A\,|\,E)}{P(A)}=\frac{P(E)P(A\,|\,E)}{P(E)P(A\,|\,E)+P(N)P(A\,|\,N)+P(\Gamma)P(A\,|\,\Gamma)}$$

$$=\frac{0.1625}{0.323}=0.5031$$

Σημείωση Από το προηγούμενο παράδειγμα, βλέπουμε ότι το θεώρημα του Bayes μας δίνει την πιθανότητα ενός ενδεχομένου Ε υπό συνθήκη του ενδεχομένου Α, ως πηλίκο της πιθανότητας ενός υποσυνόλου του Α προς την πιθανότητα του συνόλου Α. Ειδικότερα παρατηρούμε ότι στον αριθμητή υπάρχει η πιθανότητα που αντιστοιχεί στη μία τελική κορυφή του αντίστοιχου δένδρου και στον παρονομαστή το άθροισμα των πιθανοτήτων στις τρεις κορυφές του δένδρου στις οποίες υπάρχουν οι τρεις ασυμβίβαστοι τρόποι με τους οποίους μπορεί να συμβεί το Α.

Παράδειγμα 7.7.2 Δύο ίδιες μηχανές M_1 και M_2 παράγουν ακριβώς τα ίδια προϊόντα. Η παραγόμενη ποσότητα προϊόντων είναι ίδια για τις δύο μηχανές. Είναι όμως γνωστό, από προηγούμενη πείρα, ότι η M_1 παράγει 2% ελαττωματικά προϊόντα ενώ η M_2 παράγει 5%.Αν πάρουμε ένα προϊόν με τρόπο τυχαίο, να βρεθεί η πιθανότητα

(i) να είναι ελαττωματικό,

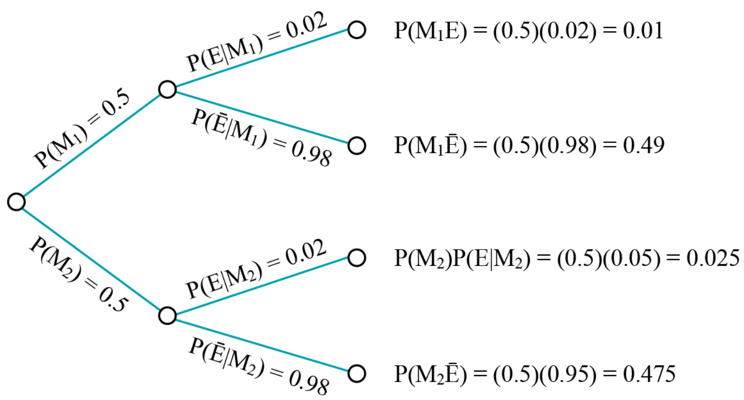
(ii) να προέρχεται από την M_1, αν γνωρίζουμε ότι είναι ελαττωματικό.

Έστω Ε είναι το ενδεχόμενο ένα προϊόν να είναι ελαττωματικό. Αν πάρουμε τυχαία ένα προϊόν, τότε αυτό μπορεί να προέρχεται είτε από την M_1, με πιθανότητα $P(M_1)=0.5$, είτε από την M_2 με πιθανότητα $P(M_2)=0.5$. Έχουμε δηλαδή:

(i) $E = (E \cap M_1) \cup (E \cap M_2)$ και

$$
\begin{aligned}
P(E) &= P(E \cap M_1) + P(E \cap M_2) \\
&= P(E \mid M_1)P(M_1) + P(E \mid M_2)P(M_2) \\
&= (0.02)(0.50) + (0.05)(0.50) \\
&= 0.010 + 0.025 = 0.035
\end{aligned}
$$

Η ανάλυση αυτή μπορεί να παρασταθεί με το εξής δένδρο πιθανότητας:



(ii) Για να βρούμε την πιθανότητα $P(M_1 \mid E)$ εφαρμόζουμε τον τύπο του Bayes και έχουμε:

$$P(M_1 \mid E) = \frac{P(M_1)P(E \mid M_1)}{P(E)} = \frac{0.010}{0.035} = 0.29$$

Παράδειγμα 7.7.3 Έστω A_1 και A_2 τα ενδεχόμενα για ένα παιδί που γεννιέται να είναι αγόρι και κορίτσι, αντίστοιχα και $P(A_1) = 0.55$. Τα A_1 και A_2 αποτελούν έναν διαμερισμό του δειγματικού χώρου αφού $A_1 \cup A_2 = S$ και $A_1 \cap A_2 = \varnothing$. Αν Ε το ενδεχόμενο για κάποιο νεογέννητο να είναι φορέας κληρονομικής ασθένειας και ισχύει $P(E \mid A_1) = 0.02$), $P(E \mid A_2) = 0.042$), να υπολογιστούν:

i) $P(E)$ ii) $P(\overline{E})$ και iii) $P(A_1 \mid E)$.

i)
$$
\begin{aligned}
P(E) &= P(E \mid A_1)P(A_1) + P(E \mid E_2)P(E_2) \\
&= (0.55)(0.02) + (0.45)(0.042) \\
&= 0.0299
\end{aligned}
$$

ii) $P(\bar{E}) = 0.971$

iii) $P(A_1 | E) = \dfrac{P(E | A_1)P(A_1)}{P(E)} = \dfrac{(0.55)(0.02)}{0.029} = 0.379$

Το Θεώρημα του Bayes στη διάγνωση μιας ασθένειας

Τα τεστ τα οποία προτείνονται για την διάγνωση μιας ασθένειας αξιολογούνται σχεδον πάντα πιθανοκρατικά.

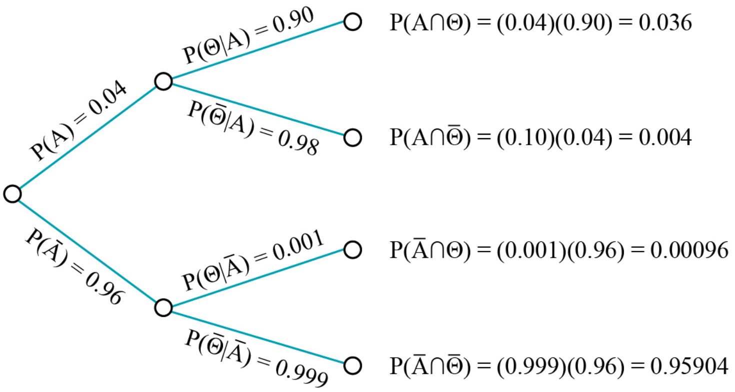
Ονομάζεται **ευαισθησία** *(sensitivity)* ενός τεστ η πιθανότητα να είναι θετικό όταν υπάρχει ασθένεια δηλαδή η

ευαισθησία = P(τεστ θετικό | υπάρχει ασθένεια)

και **ειδικότητα** (*specificity*) η πιθανότητα να είναι αρνητικό όταν το άτομο είναι υγιές δηλαδή η

ειδικότητα = P(τεστ αρνητικό | ΔΕΝ υπάρχει ασθένεια)

Παράδειγμα 7.7.4 Ορισμένο είδος φυματίωσης προσβάλει τα βοοειδή μιας χώρας σε ποσοστό 4% και στην περίπτωση αυτή τα αντίστοιχα ζώα θανατώνονται. Για την ανίχνευση της ασθένειας χρησιμοποιείται ένα τεστ το οποίο έχει ευαισθησία 0.90 και ειδικότητα 0.999. Αν συμβολίσουμε με A το ενδεχόμενο το ζώο να πάσχει και με Θ το ενδεχόμενο το τεστ να είναι θετικό, οι πληροφορίες αυτές παριστάνονται στο ακόλουθο δένδρο πιθανότητας



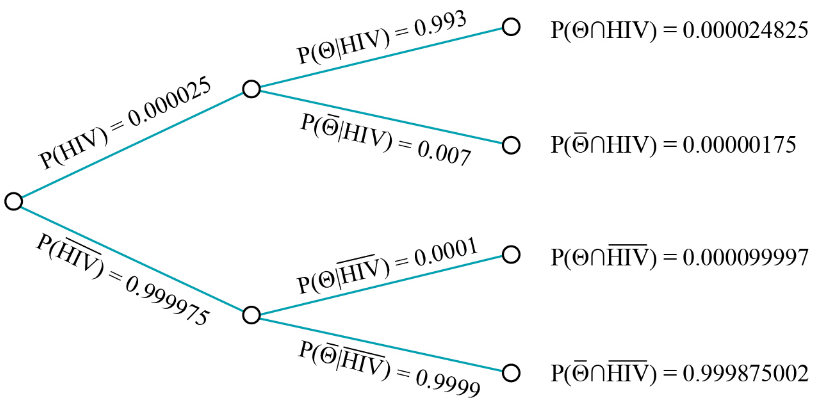
Για να υπολογίσουμε την πιθανότητα για κάποιο ζώο που το τεστ του ήταν θετικό (Θ) να πάσχει (Α), θα εφαρμόσουμε το Θεώρημα του Bayes ως εξής:

$$P(A|\Theta) = \frac{P(\Theta\,|\,A)P(A)}{P(\Theta\,|\,A)P(A) + P(\Theta\,|\,\bar{A})P(\bar{A})}$$
$$= \frac{(0.90)(0.004)}{(0.90)(0.004) + (0.001)(0.996)} = \frac{0.0036}{0.0000996} = 0.75$$

Παράδειγμα 7.7.5 Οι Johnson W.O. και Gastwirth J.L. σε ένα άρθρο τους το 1991 στο περιοδικό Journal of the Royal Statistical Society, Series B, εκτίμησαν ότι στο γενικό πληθυσμό, χωρίς γνωστούς παράγοντες κινδύνου, η πιθανότητα για κάποιον να είναι φορέας του ιού HIV είναι ίση με 0.000025. Δηλαδή σε πληθυσμό 10 εκατομμυρίων υπάρχουν 250 φορείς. Για την ανίχνευση του ιού αυτού στο αίμα χρησιμοποιείται διαγνωστικό τεστ με το όνομα Ελίζα. Τα χαρακτηριστικά του τεστ Ελίζα, δίνονται από τις ακόλουθες δύο δεσμευμένες πιθανότητες:

P(τεστ Θετικό | το άτομο είναι φορέας HIV) = 0.993
P(τεστ Αρνητικό | το άτομο ΔΕΝ είναι φορέας HIV) = 0.9999

Οι πληροφορίες αυτές παριστάνονται στο ακόλουθο δένδρο πιθανότητας:



Επειδή ένα τεστ θετικό σε ένα άτομο που δεν είναι φορέας έχει τρομαχτικές συνέπειες για τη ζωή του, μας ενδιαφέρει ιδιαίτερα η πιθανότητα του ενδεχομένου αυτού.

$$P(HIV \mid \Theta) = \frac{P(\Theta \mid HIV)P(HIV)}{P(\Theta \mid HIV)P(HIV) + P(\Theta \mid \overline{HIV})P(\overline{HIV})}$$

$$= \frac{(0.993)(0.000025)}{(0.993)(0.000025) + (0.999975)(0.0001)} = 0.1989$$

Δηλαδή, αν σε κάποιον το τεστ Ελίζα βγει θετικό, αυτός έχει πιθανότητα να μην είναι φορέας περίπου 80% (Chatterjee C. – Handcock M. – Simonoff J., 1995).

❒ ΑΣΚΗΣΕΙΣ

1. Θέλετε να αγοράσετε λαχείο και διστάζετε ανάμεσα σε τρεις λαχνούς με αριθμούς αντίστοιχα 573104, 007007 και 777777. Κατά τη γνώμη σας ποιός απο τους τρεις αριθμούς είναι πιθανότερο να κληρωθεί;

2. Συμφοιτητής σας, συστηματικός παίκτης του Λότο επέλεξε να παίζει έξι αριθμούς-τους ίδιους κάθε εβδομάδα μέχρι να κερδίσει. Το ερώτημα που θέτει στον/στη καθηγητή/τρια της Στατιστικής είναι το εξής: Πόσο πρέπει να ζήσει ώστε να εξασφαλίσει μαθηματικά την βεβαιότητα οτι θα κερδίσει; Ποιά είναι η δική σας γνώμη;

3. Αν A, B δύο ενδεχόμενα και ισχύει $P(A) = 0.65$, $P(B) = 0.10$ και $P(A \cap B) = 0.065$ απαντήστε στα ακόλουθα ερωτήματα:

i) Είναι τα A, B ασυμβίβαστα και γιατί;
ii) Είναι τα A, B ανεξάρτητα και γιατί;

4. Παρέα 4 ατόμων, οι Α, Β, Γ και Δ αποκλείστηκαν από τα χιόνια σε καταφύγιο και κάποιος θα πρέπει να φροντίσει να φέρει τα ξύλα της ημέρας από την αποθήκη που βρίσκεται 200 μέτρα μακριά από το καταφύγιο. Ο “τυχερός” επιλέγεται με κλήρωση. Ας υποθέσουμε ότι η “επιχείρηση ξύλα” θα χρειαστεί να επαναληφθεί 2 φορές. Να υπολογιστεί για τον Α:

i) Η πιθανότητα να κληρωθεί για ξύλα αν συμφωνήθηκε ότι αυτός που κληρώθηκε την πρώτη φορά απαλλάσσεται στη δεύτερη κλήρωση

ii) Η πιθανότητα να κληρωθεί τουλάχιστον μία φορά αν συμφωνήθηκε ότι αυτός που κληρώθηκε την πρώτη φορά συμμετέχει ισότιμα και στη δεύτερη κλήρωση.

5. Το ποσοστό των εφήβων αγοριών που παίζουν συστηματικά ΛΟΤΟ ισούται με 0.33 ενώ το αντίστοιχο ποσοστό στα κορίτσια ισούται με 0.07. Υποθέτοντας ότι τα ποσστά των αγοριών και κοριτσιών στην εφηβεία είναι ίσα ποιά είναι η πιθανότητα για ένα τυχαίο παιδί στην εφηβεία να είναι κορίτσι και συστηματική παίκτρια του ΛΟΤΟ;

6. Σε ποδοσφαιρικό ματς μεταξύ ΠΑΝΑΘΗΝΑΪΚΟΥ και ΟΛΥΜΠΙΑΚΟΥ σημειώθηκε σκορ 3-1. αλλά δεν γνωρίζετε τη σειρά με την οποία "μπήκαν" τα γκολ. Δώστε με ένα δεντρο όλες τις δυνατές περιπτώσεις και με τη βοήθειά του προσδιορίστε την πιθανότητα να σημειώθηκε σε κάποιο στάδιο του παιχνιδιού το σκορ 1-1.

7. Ένα αεροπλάνο σε περίπτωση βλάβης μπορεί να συνεχίσει με ασφάλεια την πτήση του αρκεί, στο τετρακινητήριο, να δουλεύουν οι δύο κινητήρες του ενώ στο δικινητήριο, αρκει να δουλεύει η μία. Αν η πιθανότητα για έναν κινητήρα αεροπλάνου να παρουσιάσει σοβαρό πρόβλημα εν πτήσει ισούται με 0.001 τοτε ποιός τύπος αεροπλάνου είναι πιο ασφαλής και γιατί; Ποιά είναι η κρίσιμη υπόθεση που κάνετε για να υπολογίσετε την πιθανότητα ασφαλούς πτήσης στους δύο τύπους αεροπλάνου;

8. Στο πιο ψηλό ράφι της ντουλάπας σας έχετε δύο ζευγάρια παπούτσια-ένα κόκκινα και ένα μαύρα. Τραβάτε στα τυφλά ένα παπούτσι με κάθε χέρι. Ποιά είναι η πιθανότητα να τραβήξετε ζευγάρι; (Lomax and Moosavi,1998).

9. Ένας άνδρας και μια γυναίκα, άσχετοι μεταξύ τους έχουν απο δυό παιδιά ο καθένας. Ένα τουλάχιστον απο τα παιδιά της γυναίκας είναι αγόρι και το μεγαλύτερο παιδί του άντρα είναι επίσης αγόρι. Ποιός από τους δύο έχει μεγαλύτερη πιθανότητα να έχει δύο αγόρια; (από το *stats*, nr 18, winter 1997).

10. Παραγωγική διαδικασία περιλαμβάνει τα στάδια Α, Β, Γ και Δ και η πιθανότητα για κάθε στάδιο να σταματήσει είναι, αντίστοιχα, $P(A)=0.02$, $P(B)=0.01$, $P(\Gamma)=0.12$, $P(\Delta)=0.05$. Αν τα τρία στάδια είναι ανεξάρτητα και για να σταματήσει η διαδικασία αρκεί να εμφανιστεί βλάβη σε ένα τουλάχιστον, ζητείται:

i) η πιθανότητα να σταματήσει η διαδικασία

ii) Αν σταματήσει η διαδικασία ποιά είναι η πιθανότητα να υπάρχει βλάβη και στα τέσσερα στάδια;

(**καθοδήγηση:** επειδή τα επιμέρους ενδεχόμενα δεν είναι ασυμβίβαστα, είναι προτιμότερο να προσδιορίσετε την $P(A \cup B \cup \Gamma \cup \Delta)$ από το συμπλήρωμά της).

11. Σε μια εκτεταμένη κοινωνική έρευνα πήραμε ένα τυχαίο δείγμα παντρεμένων ζευγαριών και τους δώσαμε να απαντήσουν σ' ένα ερωτηματολόγιο. Σ' αυτό, υπάρχει η ερώτηση α: "απατήσατε» τον(την) σύντροφό σας έστω και μία φορά;" Μαζί με την ερώτηση αυτή βάζουμε και την ερώτηση β: "ο αριθμός της ταυτότητάς σας είναι ζυγός;" Ζητάμε από τον ερωτώμενο να ρίξει ένα νόμισμα και αν εμφανιστεί η όψη «κεφαλή» να απαντήσει την ερώτηση α, διαφορετικά, την β. Έστω ότι ποσοστό 52% απάντησε "ναι". Αν A είναι το ενδεχόμενο ο ερωτώμενος να απαντήσει "ναι", B_1 και B_2 τα ενδεχόμενα να απαντήσει την ερώτηση α και β, αντίστοιχα, έχουμε $P(A) = 0.52$, $P(B_1) = 0.50$, $P(B_2) = 0.50$. Ακόμη γνωρίζουμε ότι $P(A \mid B_2) = 0.5$. Να υπολογιστεί η πιθανότητα $P(A \mid B_1)$.

12. Θέλετε να μάθετε την αναλογία των σπουδαστών που δοκίμασαν ναρκωτικά και επειδή θεωρείτε πολύ πιθανόν ότι μια ευθεία ερώτηση δεν θα δώσει ειλικρινείς απαντήσεις εφαρμόζετε την ακόλουθη τεχνική τυχαιοποιημένης απάντησης: Σε κάθε σπουδαστή δίνεται ένα κουτί με 6 κόκκινα χαρτάκια, 3 κίτρινα και 3 πράσινα και ο σπουδαστής καλείται να τραβήξει ένα κρυφά. Αν τραβήξει κόκκινο (Κ) θα απαντήσει στην ερώτηση: «χρησιμοποίησες ποτέ ναρκωτικές ουσίες»; Αν τραβήξει πράσινο (Π) θα απαντήσει «ΝΑΙ» και αν τραβήξει κίτρινο (ΚΙ) θα απαντήσει «ΟΧΙ». Αν το 48% των σπουδαστών απάντησε «ΝΑΙ», ποιο είναι το ποσοστό που χρησιμοποίησε ναρκωτικές ουσίες;

13. Στο ΠΡΟ-ΠΟ έχουμε για κάθε αγώνα τρία δυνατά αποτελέσματα 1, 2 ή Χ. Για 13 αγώνες το πλήθος των διαφορετικών στηλών οι οποίες είναι δυνατόν να σχηματιστούν ισούται με $3^{13} = 1594323$. Για μια στήλη που συμπληρώνεται με τρόπο τυχαίο, η πιθανότητα να πιάσει 13άρι ισούται με 1/1594323. Αν κάποιος παίκτης θέλει να παίξει 8 αγώνες "στάνταρντ" τότε έχει 3^5=243 δυνατούς τρόπους να συμπληρώσει τους υπόλοιπους 5 αγώνες. Συμβολίζοντας με A_5 το ενδεχόμενο σωστής πρόβλεψης αυτών των 5 αγώνων και B_8 το ενδεχόμενο σωστής πρόβλεψης των 8 "στάνταρντ" αγώνων να υπολογιστεί η πιθανότητα του A_5 δεδομένου του B_8.

14. Στο τμήμα ελέγχου της ποιότητας εργάζονται δύο ειδικευμένοι εργάτες, οι A_1 και A_2. Ο A_1 ελέγχει το 60% των προϊόντων που φθάνουν στο τμήμα και ο A_2 το 40%. Το 5% από τα προϊόντα που ο A_1 κατατάσσει στα "μη ελαττωματικά" είναι στην πραγματικότητα ελαττωματικά, ενώ το αντίστοιχο ποσοστό για τον A_2 είναι 4%. Ποια είναι η πιθανότητα, ένα προϊόν που θα πάρουμε στην τύχη από τα χαρακτηρισμένα ως μη ελαττωματικά,

- **i)** να είναι ελαττωματικό και να προέρχεται από τον A_1,
- **ii)** να είναι ελαττωματικό και να προέρχεται από τον A_2,
- **iii)** να είναι ελαττωματικό,
- **iv)** αν ένα προϊόν έχει χαρακτηριστεί ως ελαττωματικό, ποια είναι η πιθανότητα να έχει ελεγχθεί από τον A_1 και ποια να έχει ελεγχθεί από τον A_2;

15. Ο ακόλουθος πίνακας δείχνει το σχετικό μερίδιο της αγοράς και το ποσοστό των ελαττωματικών συσκευών που πουλούν 5 μάρκες ηλεκτρικού πλυντηρίου ρούχων.

	Ποσοστό αγοράς	Ποσοστό ελαττωματικών
A_1	0.62	0.15
A_2	0.16	0.11
A_3	0.15	0.07
A_4	0.05	0.04
A_5	0.02	0.01

Να υπολογιστεί:

- **i)** Ποια είναι η πιθανότητα σε ένα σπίτι που διαλέξαμε τυχαία από έναν μεγάλο αριθμό σπιτιών που έχουν πλυντήριο, να βρούμε μια συσκευή ελαττωματική (Ε);
- **ii)** Αν είναι γνωστό ότι ένα νοικοκυριό αγόρασε μια συσκευή ελαττωματική, ποια είναι η πιθανότητα η μάρκα της να είναι A_4;

16. Στον ακόλουθο πίνακα δίνονται οι αναλογίες των επιχειρηματιών που κατατάσσονται σύμφωνα με τα προσδοκώμενα κέρδη τους και τον χαρακτήρα τους. Ο πίνακας προέκυψε από μεγάλο και τυχαίο δείγμα έτσι ώστε οι αναλογίες μπορούν να θεωρηθούν ότι αντιπροσωπεύουν πιθανότητες για το σύνολο των επιχειρηματιών. Ο χαρακτηρισμός ενός επιχειρηματία σε "αισιόδοξο", "συγκρατημένο" ή "απαισιόδοξο", έγινε από τον ίδιο.

Χαρακτήρας \ Προσδοκώμενα κέρδη		Ασυνήθιστα υψηλά A_1	Κανονικά A_2	Ασυνήθιστα χαμηλά A_3
Αισιόδοξος	B_1	0.11	0.12	0.07
Συγκρατημένος	B_2	0.09	0.17	0.21
Απαισιόδοξος	B_3	0.05	0.03	0.15

i) Ποια είναι η πιθανότητα ένας επιχειρηματίας να είναι αισιόδοξος;

ii) Αν είναι αισιόδοξος ποια είναι η πιθανότητα να προσδοκά ασυνήθιστα υψηλά κέρδη;

iii) Ποια είναι η πιθανότητα για έναν επιχειρηματία να προσδοκά ασυνήθιστα υψηλά κέρδη;

17. Μία τράπεζα ταξινομεί τους πελάτες της σε δύο κατηγορίες: σε δανειολήπτες υψηλού κινδύνου (Υ) και σε δανειολήπτες χαμηλού κινδύνου (Χ) και μόνον το 20% των δανείων της χορηγούνται σε πελάτες της πρώτης κατηγορίας. Από τα δάνεια που χορηγεί ποσοστό 5% είναι επισφαλή (Ε) και μόνον το 40% αυτών των επισφαλών δανείων έχουν δοθεί σε δανειολήπτες υψηλού κινδύνου. Ποια είναι η πιθανότητα για έναν δανειολήπτη υψηλού κινδύνου να βρεθεί επισφαλής;

18. Για τη ζήτηση ενός νέου προϊόντος ο υπεύθυνος της επιχείρησης επισυνάπτει πιθανότητα 0.60 να είναι μεγάλη, 0.25 να είναι μέτρια και 0.15 να είναι χαμηλή. Σχετική έρευνα της αγοράς έδειξε ότι η ζήτηση θα είναι υψηλή. Από προηγούμενη πείρα είναι γνωστό ότι το 80% των παρόμοιων ερευνών έδειξαν υψηλή ζήτηση όταν η ζήτηση ήταν υψηλή ενώ, όταν η ζήτηση ήταν μέτρια και χαμηλή, τα αντίστοιχα ποσοστά ήταν 60% και 15%. Με βάση τα δεδομένα αυτά, να ξαναϋπολογίσετε τις πιθανότητες για τα τρία επίπεδα ζήτησης.

19. Η επιχείρηση Α αγοράζει από ορισμένο προμηθευτή κυλίνδρους σε μεγάλες παρτίδες. Η αναλογία των ελαττωματικών είναι συνήθως ίση με 0.001. Μερικές φορές πάντως η αναλογία ανεβαίνει στο 0.02 επειδή παρουσιάζονται ορισμένα προβλήματα στην παραγωγική διαδικασία των κυλίνδρων. Έστω B_1 και B_2 τα ενδεχόμενα η αναλογία των ελαττωματικών να ισούται με 0.001 και 0.02 αντίστοιχα. Από προηγούμενη πείρα είναι γνωστό ότι $P(B_1) = 0.95$ και $P(B_2) = 0.05$. Αν από ορισμένη παρτίδα πάρουμε με τυχαία δειγματοληψία έναν κύλινδρο και είναι ελαττωματικός ποια είναι η πιθανότητα του B_1;

20. Η Δημόσια Επιχείρηση Πετρελαίου πρέπει να αποφασίσει αν θα κάνει ή όχι γεωτρήσεις σε ορισμένο μέρος της Ηπείρου. Σύμφωνα με την κρίση των υπεύθυνων της εταιρείας, η οποία βασίζεται σε προηγούμενη εμπειρία και σε μια αρχική ανάλυση των χαρακτηριστικών της περιοχής, για τα τρία ενδεχόμενα-καταστάσεις της φύσης Ξερή τρύπα (Θ_1), Μέτρια ποσότητα (Θ_2), Μεγάλη ποσότητα (Θ_3), δίνονται οι ακόλουθες a priori πιθανότητες: $P(\Theta_1) = 0.60$, $P(\Theta_2) = 0.25$, $P(\Theta_3) = 0.15$. Οι εταιρείες εξόρυξης πετρελαίου χρησιμοποιούν ένα σεισμικό τεστ το οποίο είναι θετικό στο 20% των περιπτώσεων όταν υπάρχει μεγάλο κοίτασμα, στο 55% των περιπτώσεων όταν υπάρχει μέτριο και στο 80% όταν υπάρχει ξερή τρύπα. Αν το τεστ που έκανε η ΔΕΠ βγήκε θετικό να ξαναϋπολογιστούν οι πιθανότητες για τις καταστάσεις της φύσης.

21. Οι 260 εργαζόμενοι της επιχείρησης Α ομαδοποιήθηκαν στον ακόλουθο πίνακα ανάλογα με το φύλο και την ηλικία τους.

	18-22	22-25	25-30	30-40	40-55	55-65	Σύνολο
Άνδρες	12	17	34	46	33	28	170
Γυναίκες	11	22	28	19	8	2	90
Σύνολο	23	39	62	65	41	30	260

Η επιχείρηση αποφάσισε να κάνει απολύσεις και ζητείται:

i) Αν η πρώτη απόλυση γίνει με κλήρωση, ποια είναι η πιθανότητα να αφορά γυναίκα ηλικίας 18–22 ετών;

ii) Αν η επιχείρηση αποφάσισε να μην απολύσει άνδρες και μια απόλυση γίνει πάλι με κλήρωση, ποια είναι η πιθανότητα να απολύσει γυναίκα ηλικίας 18–22 ετών;

iii) Αν αποφάσισε να απολύσει άτομο ηλικίας 18–22 ετών, ποια είναι η πιθανότητα να είναι γυναίκα;

22. Υπηρεσία του Ευρωπαϊκού Κοινοβουλίου κατηγορήθηκε για σεξιστικές διακρίσεις σε βάρος των γυναικών, διότι προσέλαβε για ορισμένη διευθυντική θέση περισσότερους άνδρες από γυναίκες. Οι υπεύθυνοι των προσλήψεων δικαιολόγησαν τους περισσότερους άνδρες με το μεγαλύτερο αριθμό ανδρών υποψηφίων. Στον ακόλουθο πίνακα ταξινομήθηκαν όσοι υπέβαλαν για την εργασία αυτή ανάλογα με το φύλο τους και το αν τελικά προσλήφθηκαν ή όχι. Συμφωνείτε ότι οι προσλήψεις έγιναν ανεξάρτητα από το φύλο; Να δώσετε τις πληροφορίες του πίνακα με τη μορφή ενός δέντρου πιθανότητας.

Πρόσληψη / Υποψήφιοι	Ναι (Ν)	Όχι (Ο)	Άθροισμα
Άνδρες (Α)	231	99	330
Γυναίκες (Γ)	63	27	90
Άθροισμα	294	126	420

23. Δύο αγόρια και δύο κορίτσια με τα ίδια τυπικά προσόντα είναι υποψήφια για δύο θέσεις στην επιχείρησή σας και για να μην κατηγορηθείτε ότι κάνετε διακρίσεις με βάση το φύλο θέλετε να προκρίνετε δύο άτομα με διαδοχική κλήρωση (που σημαίνει χωρίς επανάθεση). Ποια είναι η πιθανότητα να μην πάρουμε δύο άτομα του ίδιου φύλου; Να γίνει η δενδροειδής παράσταση του δειγματικού χώρου αυτού του πειράματος.

24. Σας έδωσαν δύο ολόιδια κουτιά και 5 άσπρες και 5 μαύρες μπάλες και συμφωνήσατε με τον καθηγητή σας ότι θα τραβήξει μια μπάλα από ένα κουτί και αν είναι μαύρη θα περάσετε το μάθημα, διαφορετικά κόβεστε. Πώς πρέπει να οργανώσετε τις μπάλες σας στα δύο κουτιά ώστε να μεγιστοποιήσετε την πιθανότητα να περάσετε το μάθημα;

25. Αν για ένα παιδί που γεννιέται η πιθανότητα να είναι αγόρι ισούται με την πιθανότητα να είναι κορίτσι και το φύλο σε διαδοχικές γεννήσεις είναι ανεξάρτητο, ποια είναι η πιθανότητα για ένα ζευγάρι με 8 παιδιά να έχει:

i) 4 κορίτσια και 4 αγόρια;

ii) τα 4 πρώτα κορίτσια και τα 4 επόμενα αγόρια;

iii) και τα 8 του ίδιου φύλου;

26. Σε ένα κλειστό κουτί υπάρχουν 2 μπάλες άσπρες (Α) και δύο μαύρες (Μ) ίδιου ακριβώς μεγέθους και ζητείται:

i) Αν παίρνουμε μπάλες χωρίς επανάθεση μέχρις ότου πάρουμε μια άσπρη, να προσδιοριστεί ο δειγματικός χώρος του πειράματος καθώς και οι πιθανότητες που αντιστοιχούν σε όλα τα σημεία του,

ii) Αν πάρουμε δύο μπάλες, με επανάθεση, ποια είναι η πιθανότητα να πάρουμε μια Α και μια Μ αν η σειρά εμφάνισης δεν έχει σημασία;

Να γίνουν οι αντίστοιχες δενδροειδείς παραστάσεις.

27. Αν το ποσοστό φοροφυγάδων στους επαγγελματίες του κλάδου ΑΛΦΑ ισούται με 0.45 ποιά είναι η πιθανότητα σε τυχαίο δείγμα 4 επαγγελματιών στους οποίους θα γίνει εξονυχιστικός έλεγχος να είναι όλοι φοροφυγάδες; Ποιά είναι η πιθανότητα να είναι ένας τουλάχιστον φοροφυγάς;

28. Για τη διάγνωση μιας σπάνιας μορφής καρκίνου του αίματος χρησιμοποιείται ένα τεστ το οποίο έχει ευαισθησία ίση με 0.80 και ειδικότητα ίση με 0.90. Γνωστός σας έκανε το τεστ το οποίο βγήκε θετικό και απελπισμένος σχολιάζει το πόσο άτυχος είναι αφού η πιθανότητα να προσβληθεί κάποιος από αυτήν την ασθένεια ισούται με 1%. Ποιά είναι η πιθανότητα να νοσεί πραγματικά;