ΚΕΦΑΛΑΙΟ 4

# ΣΧΕΣΕΙΣ ΕΝΔΕΧΟΜΕΝΩΝ - ΤΥΧΑΙΕΣ ΜΕΤΑΒΛΗΤΕΣ - ΚΑΤΑΝΟΜΗ ΠΙΘΑΝΟΤΗΤΑΣ

## ΔΕΣΜΕΥΜΕΝΗ ΠΙΘΑΝΟΤΗΤΑ

Σε πολλά προβλήματα, το γεγονός ότι κάποιο ενδεχόμενο B έχει συμβεί είναι πολύ πιθανό να επηρεάσει την πιθανότητα να συμβεί κάποιο άλλο γεγονός A. Αν για παράδειγμα, ρίξουμε δύο ζάρια και ορίσουμε τα ενδεχόμενα A={άθροισμα 5} B={ένα από τα δύο είναι 3} τότε P(A)=4/36. Εάν όμως ξέρουμε ήδη ότι το ένα ζάρι είναι 3 η πιθανότητα του A γίνεται 2/11. Αυτό συμβαίνει γιατί στην περίπτωση αυτή ο αρχικός δειγματικός χώρος των 36 σημείων ελαττώνεται στο δειγματικό χώρο

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

Όταν οι δειγματικοί χώροι έχουν ισοπίθανα απλά ενδεχόμενα, όπως το προηγούμενο παράδειγμα, ο καθορισμός πιθανοτήτων ενδεχομένων δοθέντος ότι κάποιο άλλο ενδεχόμενο έχει συμβεί είναι εύκολος. Εάν όμως τα απλά ενδεχόμενα δεν είναι ισοπίθανα ο καθορισμός τέτοιων πιθανοτήτων δεν είναι πάντα εύκολος. Ο ορισμός που ακολουθεί θα διευκολύνει τον υπολογισμό πιθανοτήτων της μορφής αυτής.

**Ορισμός:** Έστω S ο δειγματικός χώρος ενός πειράματος και B ένα ενδεχόμενο με P(B)>0. Η πιθανότητα να συμβεί ένα ενδεχόμενο A δοθέντος ότι το B έχει ήδη συμβεί λέγεται *δεσμευμένη πιθανότητα (condiotinal probability) του A δοθέντος του B,* συμβολίζεται με P(A|B) και ορίζεται ως εξής:

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

Χρησιμοποιώντας τον ορισμό στο παράδειγμα μας έχουμε

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{2/36}{11/36} = \frac{2}{11}$$

Ο ορισμός της δεσμευμένης πιθανότητας είναι χρήσιμος και για τον υπολογισμό πιθανοτήτων τομής ενδεχομένων.

**Πολλαπλασιαστικός κανόνας**

$P(A \cap B) = P(A|B)P(B) \qquad (P(B) > 0)$

ή, εναλλακτικά

$P(A \cap B) = P(B|A)P(A) \qquad (P(A) > 0).$

Ο πολλαπλασιαστικός κανόνας μπορεί να επεκταθεί σε n ενδεχόμενα.

**Θεώρημα:** Θεωρούμε τα ενδεχόμενα $A_1, A_2, ..., A_n$ με την ιδιότητα $P\left(\bigcap_{i=1}^{n-1} A_i\right) > 0$. Τότε

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1 \cap A_2 \cap A_n)$$
$$= P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) ... P\left(A_n | \bigcap_{i=1}^{n-1} A_i\right)$$

**Απόδειξη:** Το θεώρημα ισχύει για n = 2. Έστω ότι ισχύει για n = k. Θα αποδείξουμε ότι ισχύει και για n = k + 1. Έχουμε

$$P[A_1 \cap A_2 \cap ... \cap A_{k+1}] = P\left[\left(\bigcap_{i=1}^{k} A_i\right) \bigcap A_{k+1}\right] = P\left[A_{k+1} \middle| \bigcap_{i=1}^{k} A_i\right] P\left(\bigcap_{i=1}^{k} A_i\right) =$$
$$= P\left[A_{k+1} \middle| \bigcap_{i=1}^{k} A_i\right] x P\left[A_k \middle| \bigcap_{i=1}^{k-1} A_i\right] x...x \ P[A_2 | A_2] P(A_1)$$

Από την μορφή του πολλαπλασιαστικού τύπου, είναι φανερό ότι κατά κάποιο τρόπο θεωρούμε σαν δεδομένο ότι πρώτα συμβαίνει το $A_1$ και κατόπιν κατά σειρά τα $A_2$, $A_3$, …, $A_n$.

Ο ορισμός της δεσμευμένης πιθανότητας ικανοποιεί τα τρία αξιώματα του Kolmogorov, όπως φαίνεται από το θεώρημα που ακολουθεί.

**Θεώρημα:** Αν $P(B) > 0$, τότε

1) $P(A|B) \geq 0$
2) $P(B|B) = 1$
3) Αν τα ενδεχόμενα $A_i$ , i=1,2,... είναι αμοιβαία ξένα μεταξύ τους, τότε

$$P\left(\bigcap_{i=1}^{\infty} A_i \mid B\right) = \sum_{i=1}^{\infty} P(A_i \mid B)$$

**Απόδειξη:**

1) $P(A/B)=\dfrac{P(A\cap B)}{P(B)}\geq 0$

2) $P(B/B) = \dfrac{P(B)}{P(B)} = 1$

3) $$P\left[\bigcup_{i=1}^{\infty} A_i | B\right] = \frac{P\left[\left(\bigcup_{i=1}^{\infty} A_i\right)\bigcap B\right]}{P(B)} = \frac{P\left[\bigcup_{i=1}^{\infty}\left(A_i \bigcap B\right)\right]}{P(B)} =$$

$$= \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)} = \sum_{i=1}^{\infty} P(A_i/B)$$

**Παράδειγμα:** Σύνθετα ανταλλακτικά συναρμολογούνται σε μια βιομηχανία η οποία χρησιμοποιεί δύο διαφορετικά συστήματα συναρμολόγησης A και A΄. Το σύστημα A χρησιμοποιεί παλαιότερα μηχανήματα από το σύστημα A΄ με αποτέλεσμα να είναι πιο αργό και μικρότερης εμπιστοσύνης. Ας υποθέσουμε ότι κάποια μέρα το σύστημα A συναρμολόγησε 8 ανταλλακτικά από τα οποία 2 είναι ελαττωματικά (E) και 6 μη ελαττωματικά (E΄), ενώ το A΄ συναρμολόγησε 1 ελαττωματικό και 9 μη ελαττωματικά.
Δηλαδή, ισχύει ο εξής συνοπτικός πίνακας:

| | E | E΄ |
|---|---|---|
| A | 2 | 6 |
| A΄ | 1 | 9 |

Χωρίς να έχει τις παραπάνω πληροφορίες περί ελαττωματικών ένας διευθυντής διαλέγει στην τύχη 1 από 18 ανταλλακτικά για επίδειξη. Τότε P(το ανταλλακτικό προέρχεται από το σύστημα

A)=8/18=0.44. Εάν όμως το ανταλλακτικό βρεθεί να είναι ελαττωματικό P(A|E)=2/3 ή ισοδύναμα

$$P(A \mid E) = \frac{P(AE)}{P(E)} = \frac{2/18}{3/18} = 2/3$$

**Παράδειγμα:** Μια δέσμη αποτελείται από 12 αντικείμενα 4 από τα οποία είναι ελαττωματικά. Διαλέγουμε στην τύχη 3 αντικείμενα από την δέσμη, χωρίς επανάθεση και θέλουμε να βρούμε την πιθανότητα να είναι και τα τρία ελαττωματικά.

**Λύση:** Έστω $A_i$ ={το i αντικείμενο είναι ελαττωματικό}, i=1,2,3.
Μας ενδιαφέρει η $P(A_1 \cap A_2 \cap A_3)$.
Είναι

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid A_1 \cap A_2) \\ &= (4/12)\,(3/11)\,(2/10) = 1/55 \end{aligned}$$

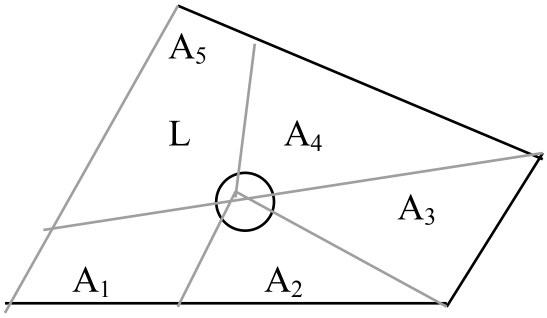
**Ορισμός:** Μια *διαμέριση* ενός συνόλου S είναι μια πεπερασμένη συλλογή $A_1, A_2, \ldots, A_n$ από υποσύνολα του S που ικανοποιούν τις ακόλουθες δύο συνθήκες:
1) $S = A_1 \cup A_2 \cup \ldots \cup A_n$
2) $A_i \cap A_j = \varnothing$, $i,j = 1,2,\ldots,n \quad i \neq j$

**Θεώρημα Ολικής Πιθανότητας** *(theorem of total probability)*: Έστω ότι $A_1, A_2, \ldots, A_n$ είναι μια διαμέριση του δειγματικού χώρου S τέτοια ώστε $P(A_i) \neq 0$, $i=1,2,\ldots,n$. Τότε για κάθε ενδεχόμενο E έχουμε

$$P(E) = \sum_{i=1}^{n} P(A_i) P(E \mid A_i)$$

(από το πολλαπλασιαστικό θεώρημα).

Το θεώρημα ολικής πιθανότητας χρησιμεύει στον υπολογισμό πιθανοτήτων ενδεχομένων. Όπως είναι προφανές από το θεώρημα, η πιθανότητα ενός ενδεχομένου μπορεί να υπολογισθεί μέσω των δεσμευμένων πιθανοτήτων του ενδεχομένου σε σχέση με τα στοιχεία κάποιας διαμέρισης θεωρώντας τον σταθμικό μέσο αυτών των πιθανοτήτων με βάση τις πιθανότητες των στοιχείων της διαμέρισης.




$$P(L) = \sum_i P(L \mid A_i)P(A_i)$$

**Παράδειγμα:** Τρία κουτιά περιέχουν στοιχεία μερικά από τα οποία είναι ελαττωματικά. Η αναλογία φαίνεται στον παρακάτω πίνακα.

| | Αριθμός στοιχείων | Αριθμός ελαττωματικών |
|---|---|---|
| κουτί 1 | 10 | 4 |
| κουτί 2 | 6 | 1 |
| κουτί 3 | 8 | 3 |

Διαλέγουμε ένα κουτί στην τύχη και στην συνέχεια διαλέγουμε ένα στοιχείο στην τύχη από το κουτί αυτό. Να βρεθεί η πιθανότητα το στοιχείο να είναι ελαττωματικό.

**Λύση:** Έστω $A_i$ = {το στοιχείο προέρχεται από το κουτί i} i=1,2,3.
E = {το στοιχείο είναι ελαττωματικό}
Από το θεώρημα ολικής πιθανότητα έχουμε
$P(E) = P(E|A_1)P(A_1) + P(E|A_2)P(A_2) + P(E|A_3)P(A_3) =$

$= (4/10)(1/3) + (1/6)(1/3) + (3/8)(1/3) = 113/360$

Μια άλλη εφαρμογή της δεσμευμένης πιθανότητας είναι το θεώρημα του Bayes (Άγγλος κληρικός του 18ου αιώνα). Όπως θα δούμε στην συνέχεια το θεώρημα του Bayes είναι το αρχικό σημείο μιας ολόκληρης στατιστικής φιλοσοφίας γνωστής ως *Μπεϋζιανή Στατιστική (Bayesian Statistics).*

**Θεώρημα:** Έστω $A_1, A_2, ..., A_n$ μια διαμέριση του δειγματικού χώρου S με $P(A_i)>0$ για κάθε $i=1,2,...,n$. Τότε, για κάθε ενδεχόμενο E με $P(E)>0$ έχουμε ότι

$$P(A_k \mid E) = \frac{P(E \mid A_k)P(A_k)}{\sum_{i=1}^{n} P(E \mid A_i)P(A_i)}$$

ή ισοδύναμα

$$P(A_k \mid E) = \frac{P(E \mid A_k)P(A_k)}{P(E)}$$

**Απόδειξη:** $E=(A_1 \cap E) \cup (A_2 \cap E) \cup ... \cup (A_n \cap E)$.
Προφανώς $A_1 \cap E, A_2 \cap E, ... A_n \cap E$ είναι μια διαμέριση του E.
Επομένως,

$$P(A_k \mid E) = \frac{P(A_k \cap E)}{P(E)} = \frac{P(A_k)P(E \mid A_k)}{\sum_{i=1}^{n} P(A_i \mid E)P(E)}$$

**Σημείωση:** Το νόημα του παραπάνω θεωρήματος ίσως να προκαλεί κάποια σύγχυση. Στα παραδείγματα που έχουμε συναντήσει μέχρι τώρα οι πιθανότητες $P(A_i)$ των στοιχείων $A_i$ της διαμέρισης ήταν γνωστές και μας ενδιέφεραν οι δεσμευμένες πιθανότητες σε σχέση με τα στοιχεία της διαμέρισης. Υπάρχουν όμως προβλήματα όταν είναι άγνωστες οι πιθανότητες των στοιχείων της διαμέρισης και ο σκοπός του πειράματος είναι να προσεγγίσουμε, ή να βρούμε, τις ακριβείς τιμές των πιθανοτήτων αυτών.

Πριν να γίνει το πείραμα είναι δυνατόν να έχουμε κάποια υποκειμενική αντίληψη, η εκτίμηση, για τις πιθανότητες $P(A_k)$ *(εκ των προτέρων ή a-priori πιθανότητες)*. Αφού κάνουμε το πείραμα μια ή περισσότερες φορές και παρατηρήσουμε τα αποτελέσματα του πειράματος, ίσως αποφασίσουμε να αναθεωρήσουμε τις απόψεις μας για τις πιθανότητες αυτές. Στην περίπτωση αυτή κατασκευάζουμε (μέσω του θεωρήματος του Bayes) τις *εκ των υστέρων (a-posteriori) πιθανότητες.*

**Παράδειγμα:** Στην αρχή του έτους, διατυπώθηκαν τρεις οικονομικές θεωρίες για την πιθανή εξέλιξη της Ελληνικής Οικονομίας. Όταν διατυπώθηκαν και οι τρεις θεωρίες φαίνονταν ισοπίθανες. Στο τέλος του έτους εξετάσθηκε η πραγματική κατάσταση της οικονομίας με αναφορά στις τρεις θεωρίες. Η ανάλυση κατέληξε στο συμπέρασμα ότι αν η πρώτη θεωρία ήταν αληθινή, η οικονομία θα είχε πιθανότητα 0.6 να καταλήξει στην παρούσα κατάσταση. Οι αντίστοιχες πιθανότητες για την δεύτερη και την τρίτη πρόβλεψη είναι 0.4 και 0.2. Να υπολογισθεί η πιθανότητα με την οποία η παρούσα κατάσταση της οικονομίας μπορεί να θεωρηθεί αποτέλεσμα της θεωρίας i, i=1,2,3.

**Λύση:** Έστω

$$A_i = \{\text{η θεωρία i είναι σωστή}\}, \quad i=1,2,3$$

Έχουμε

$$P(A_1) = P(A_2) = P(A_3) = 1/3$$

Έστω

$$E = \{\text{η οικονομία βρίσκεται στην παρούσα κατάσταση}\}$$

Τότε με βάση το θεώρημα του Bayes έχουμε

$$P(A_1 \mid E) = \frac{P(E \mid A_1)P(A_1)}{P(E \mid A_1)P(A_1) + P(E \mid A_2)P(A_2) + P(E \mid A_3)P(A_3)}$$

$$= \frac{(1/3)(6/10)}{(1/3)(6/10) + (1/3)(4/10) + (1/3)(2/10)} = \frac{6}{12} = \frac{1}{2}$$

Με όμοιο τρόπο, βρίσκουμε

$$P(A_2 \mid E) = \frac{(1/3)(4/10)}{4/10} = \frac{1}{3}$$

$$P(A_3 \mid E) = \frac{(1/3)(2/10)}{4/10} = \frac{1}{6}$$

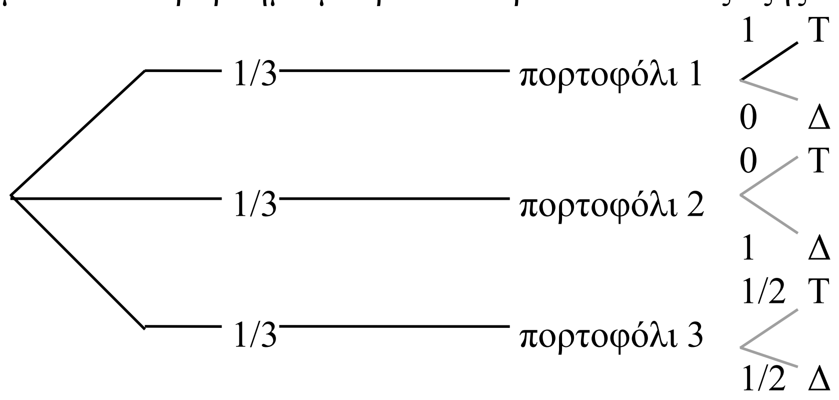
**Παράδειγμα.** ***(Παράδοξο του Russel):*** Έχουμε τρία όμοια πορτοφόλια καθένα από τα οποία περιέχει 2 νομίσματα. Το ένα περιέχει δύο τάλληρα, το άλλο δύο δεκάρικα και το τρίτο ένα τάλληρο και ένα δεκάρικο. Διαλέγουμε ένα πορτοφόλι στην τύχη και βγάζουμε το ένα από τα δύο νομίσματα που βρίσκεται να είναι τάλληρο. Να βρεθεί η πιθανότητα το άλλο νόμισμα στο πορτοφόλι που διαλέξαμε να είναι επίσης τάλληρο. (Δηλαδή η πιθανότητα να έχουμε διαλέξει το πρώτο πορτοφόλι).

**Λύση:** Έστω Ε={το νόμισμα που διαλέξαμε ήταν τάλληρο}
και έστω $A_i$ ={το νόμισμα το πήραμε από το i πορτοφόλι}, i=1,2,3.
Τότε

$$P(A_1 \mid E) = \frac{P(E \mid A_1)P(A_1)}{P(E \mid A_1)P(A_1) + P(E \mid A_2)P(A_2) + P(E \mid A_3)P(A_3)}$$

$$= \frac{1 * (1/3)}{(1 * 1/3) + (0 * 1/3) + (1/2 * 1/3)} = \frac{2}{3}$$

**Σημείωση:** Με μια επιπόλαιη ματιά ίσως κάποιος παρασυρθεί να απαντήσει ότι αφού το νόμισμα ήταν τάλληρο θα πρέπει να προερχόταν ή από το πρώτο ή από το τρίτο πορτοφόλι και επομένως η πιθανότητα να ήταν από το πρώτο είναι 1/2. Η λογική εξήγηση του γεγονότος ότι η σωστή απάντηση είναι 2/3 και όχι 1/2 βρίσκεται στο ότι 2 από τα 3 τάλληρα που έχουν τα τρία πορτοφόλια βρίσκονταν στο πρώτο πορτοφόλι.
Σχηματικά το πρόβλημα μπορεί να παρουσιασθεί ως εξής:




**Παράδειγμα:** Ένα άγνωστο άτομο με συγκεκριμένα χαρακτηριστικά (χρώμα μαλλιών, ομάδα αίματος κ.λ.π.) διέπραξε ένα αδίκημα. Είναι γνωστό ότι η πιθανότητα να έχει κάποιο άτομο στον πληθυσμό αυτά τα χαρακτηριστικά είναι p. Ένας ύποπτος έχει συλληφθεί από την αστυνομία και βρέθηκε να έχει όλα τα παραπάνω χαρακτηριστικά.

Παρ' ότι η αστυνομία δεν έχει αποδείξεις για την ενοχή του, το γεγονός ότι έχει όλα τα χαρακτηριστικά του δράστη αυξάνει την πιθανότητα να είναι ένοχος. Υποθέτουμε ότι πριν η αστυνομία διαπιστώσει ότι ο ύποπτος είχε τα χαρακτηριστικά του δράστη, η πιθανότητα ότι ο ύποπτος ήταν και ο δράστης ήταν 50-50 (0.5).
Χρησιμοποιώντας την πρόσθετη πληροφορία ότι ο ύποπτος έχει όλα τα χαρακτηριστικά του δράστη μπορούμε με το θεώρημα του Bayes, να αναθεωρήσουμε την αρχική μας πιθανότητα ότι ο ύποπτος είναι ένοχος ως εξής: Έστω Ε το ενδεχόμενο "ένοχος", Α το ενδεχόμενο "αθώος" και C το ενδεχόμενο ότι κάποιος έχει τα χαρακτηριστικά του δράστη. Ξέρουμε ότι P(C|E)=1 και υποθέσαμε ότι P(C)=p.
Επομένως

$$P(E \mid C) = \frac{P(C \mid E)P(E)}{P(C)} = \frac{1(1/2)}{p} = \frac{1}{2p} > \frac{1}{2}$$

**Σημείωση:** Ένας από τους τρόπους χρησιμοποίησης του θεωρήματος του Bayes που προκαλεί πολλές αντιθέσεις είναι στην περίπτωση που προσπαθεί κανείς να υποστηρίξει μια σχέση αιτίου-αιτιατού με βάση το παραπάνω θεώρημα και να καθορίσει την αιτία από το αποτέλεσμα.

Στην ιατρική διάγνωση για παράδειγμα, κάποιο αποτέλεσμα Ε (ορισμένα συμπτώματα) παρατηρούνται σ' έναν ασθενή. Ο γιατρός πρέπει να καθορίσει την αιτία Α που προκάλεσε τα συμπτώματα αυτά. Συνήθως, ο γιατρός έχει δεδομένα (πληροφορίες) για την πιθανότητα P(E|A) δηλαδή πόσο πιθανά είναι τα συγκεκριμένα συμπτώματα δεδομένου ότι ο ασθενής έχει μια συγκεκριμένη ασθένεια. Αυτό που ο γιατρός θέλει είναι η πιθανότητα P(A|E), δηλαδή η πιθανότητα ότι ο ασθενής έχει την συγκεκριμένη ασθένεια Α που προκαλεί τα συμπτώματα Ε.

**Παράδειγμα:** Ένα παιδί με εξανθήματα και πυρετό επισκέπτεται τον γιατρό. Ο γιατρός θεωρεί σαν πιθανές αιτίες των συμπτωμάτων αυτών (Σ) την ανεμοβλογιά (Α) που έχει επιδημία την εποχή αυτή, την οστρακιά (Π) και ίσως κάποια τρίτη άγνωστη αρρώστεια (Β). Έστω ότι τα παραπάνω συμπτώματα εμφανίζονται πάντα σε παιδιά που έχουν προσβληθεί από οστρακιά, σχεδόν πάντα (85% των περιπτώσεων) σε παιδιά που υποφέρουν από ανεμοβλογιά και μόνο στο 33% των παιδιών που υποφέρουν από κάποια άλλη ασθένεια. Δηλαδή

$$P(\Sigma|A)=0.85,\ P(\Sigma|\Pi)=1,\ P(\Sigma|B)=0.33$$

Ο γιατρός γνωρίζει ότι οι συχνότητες των τριών παραπάνω ασθενειών την εποχή αυτή είναι: P(A)=0.15 P(Π)=0.005 και P(B)=0.075. Με βάση τα παραπάνω στοιχεία οι πιθανότητες να έχει το παιδί τις τρεις παραπάνω ασθένειες υπολογίζονται ως εξής:

$$P(A\mid\Sigma)=\frac{P(\Sigma\mid A)P(A)}{P(\Sigma\mid A)P(A)+P(\Sigma\mid\Pi)P(\Pi)+P(\Sigma\mid B)P(B)}$$

$$=\frac{(0.85)(0.150)}{(0.85)(0.150)+(1)(0.005)+(0.33)(0.075)}=0.81$$

Με όμοιο τρόπο, βρίσκουμε ότι

$$P(\Pi\mid\Sigma)=\frac{(1)(0.005)}{0.157}=0.032$$

και

$$P(B\mid\Sigma)=\frac{(0.33)(0.075)}{0.157}=0.157$$

**Σημείωση:** Είναι χρήσιμο να παρατηρήσει κανείς ότι τα ενδεχόμενα Α, Π και Β δεν αποτελούν διαμέριση του δειγματικού χώρου. Πράγματι P(A)+P(Π)+P(B)=0.23 και όχι 1. Αυτό που έχει συμβεί είναι ότι αγνοήσαμε το ενδεχόμενο Υ (το παιδί είναι υγιές) που έχει πιθανότητα P(Y)=0.77. Αυτό όμως δεν επηρεάζει τους υπολογισμούς μας μια και μπορούμε να υποθέσουμε ότι P(Σ|Y)=0.

## ΑΝΕΞΑΡΤΗΤΑ ΚΑΙ ΑΣΥΜΒΑΤΑ ΕΝΔΕΧΟΜΕΝΑ

Ο ορισμός της δεσμευμένης πιθανότητας μας επιτρέπει να αναθεωρήσουμε την πιθανότητα P(A) ενός ενδεχομένου Α που αρχικά έχουμε δώσει στο Α όταν μας δοθεί η πληροφορία ότι κάποιο άλλο ενδεχόμενο Β έχει συμβεί. Η νέα πιθανότητα είναι η P(A|B).

Συμβαίνει πολλές φορές να έχουμε $P(A|B) \neq P(A)$, γεγονός που σημαίνει ότι η πληροφορία ότι το B έχει συμβεί έχει σαν αποτέλεσμα να μεταβληθεί η πιθανότητα να συμβεί το A. Σε άλλες όμως περιπτώσεις αυτό δεν συμβαίνει. Για τις τελευταίες αυτές περιπτώσεις δίνουμε τον εξής ορισμό.

**Ορισμός:** Τα ενδεχόμενα A και B λέγονται *ανεξάρτητα (independent)* (ή ακριβέστερα *στοχαστικά ανεξάρτητα ή στατιστικά ανεξάρτητα ή ανεξάρτητα κατά πιθανότητα*), αν

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

Σε μερικές περιπτώσεις, χρησιμοποιείται η ορολογία "A ανεξάρτητο του B" αν $P(A|B)=P(A)$. Ομοίως "B ανεξάρτητο του A" αν $P(B|A)=P(B)$. Στην ορολογία αυτή υποτίθεται ότι $P(B)>0$ στην πρώτη περίπτωση και $P(A)>0$ στην δεύτερη περίπτωση. Είναι προφανές (λόγω της πολλαπλασιαστικής αρχής) ότι αν το A είναι ανεξάρτητο του B τότε και το B είναι ανεξάρτητο του A και αντίστροφα (οπότε τα A και B είναι ανεξάρτητα) με την προϋπόθεση ότι $P(A)>0$ και $P(B)>0$. Με τον ορισμό της ανεξαρτησίας όμως που δώσαμε οι συνθήκες $P(A)>0$ και $P(B)>0$ δεν είναι απαραίτητες.

**Σημείωση:** Ο ορισμός της ανεξαρτησίας είναι ισοδύναμος με το ότι τα ενδεχόμενα A και B είναι ανεξάρτητα αν το γεγονός ότι το ένα έχει συμβεί δεν επηρεάζει την πιθανότητα να συμβεί το άλλο.

**Σημείωση:** Από τον ορισμό, προκύπτει ότι κάθε γεγονός A είναι ανεξάρτητο από το αδύνατο γεγονός $\varnothing$ και από το βέβαιο γεγονός S.

**Παράδειγμα:** Έστω ότι βγάζουμε ένα χαρτί από μια τράπουλα και έστω

$$A = \{\text{το χαρτί είναι άσσος}\}$$
$$B = \{\text{το χαρτί είναι σπαθί}\}$$

Έχουμε

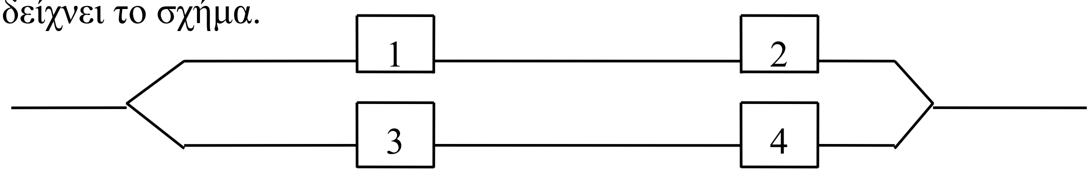
$$P(A) = 4/52 = 1/13, \quad P(B) = 13/52 = 1/4$$

Εξ' άλλου

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

Επομένως A και B είναι ανεξάρτητα.

**Παράδειγμα:** Ένα σύστημα αποτελείται από 4 εξαρτήματα όπως δείχνει το σχήμα.




Το σύστημα λειτουργεί αν ή το υποσύστημα 1-2 λειτουργεί ή το υποσύστημα 3-4 λειτουργεί, (μια και τα δύο συστήματα συνδέονται παράλληλα). Προφανώς ένα υποσύστημα δουλεύει αν και τα δύο εξαρτήματά του δουλεύουν (μια και τα δύο εξαρτήματα σε κάθε υποσύστημα συνδέονται στην σειρά). Κάθε εξάρτημα δουλεύει ή χαλάει ανεξάρτητα από τα άλλα και κάθε ένα δουλεύει με πιθανότητα 0.9. Να βρεθεί η πιθανότητα να λειτουργεί ολόκληρο το σύστημα. (Η πιθανότητα αυτή ονομάζεται *συντελεστής αξιοπιστίας του συστήματος (system reliability coefficient)*).

**Λύση:** Έστω $A_i$ (i=1,2,3,4) το ενδεχόμενο ότι το i εξάρτημα δουλεύει. Τα $A_i$ είναι αμοιβαία ανεξάρτητα. Το ενδεχόμενο ότι το 1-2 υποσύστημα δουλεύει είναι το $A_1 \cap A_2$. Ομοίως $A_3 \cap A_4$ είναι το ενδεχόμενο ότι το υποσύστημα 3-4 δουλεύει. Η πιθανότητα που ζητάμε είναι

$P[(A_1 \cap A_2) \cup (A_3 \cap A_4)]=$

$$= P(A_1 \cap A_2)+P(A_3 \cap A_4)-P[(A_1 \cap A_2) \cap (A_3 \cap A_4)] =$$

$$= P(A_1)\, P(A_2) + P(A_3)\, P(A_4) - P(A_1)\, P(A_2)\, P(A_3)\, P(A_4) =$$

$$= (0.9)\,(0.9) + (0.9)\,(0.9) - (0.9)^4 = 0.9639$$

**Παράδειγμα (συνέχεια):** Έστω x = $P(A_i)$, i=1,2,3,4. Ποιά είναι η τιμή του x που θα δώσει συντελεστή αξιοπιστίας του συστήματος 0.99;

**Λύση:** Θέλουμε P (το σύστημα λειτουργεί) = 0.99 $\Leftrightarrow$

$\Leftrightarrow x^2 + x^2 - x^4 = 0.99$

$\Leftrightarrow \psi^2 - 2\psi + (0.99) = 0$ (με $\psi = x^2$)

$\Leftrightarrow \psi = 0.9 \Rightarrow x \cong 0.95$

**Θεώρημα:** Αν A και B είναι ανεξάρτητα τότε:

1) $A$ , $B'$ είναι ανεξάρτητα.
2) $A'$, $B$ είναι ανεξάρτητα.
3) $A'$, $B'$ είναι ανεξάρτητα.

**Απόδειξη:** 1) Έχουμε ότι $P(A \cap B) = P(A)\,P(B)$. Προφανώς το B και $B'$ αποτελούν μία διαμέριση του S. Επομένως και τα $A \cap B$, $A \cap B'$ αποτελούν μια διαμέριση του A. Τότε

$$P(A) = P(A \cap B) + P(A \cap B')$$

δηλαδή,

$$P(A \cap B') = P(A) - P(A)P(B) = P(A) - (1 - P(B)) = P(A)\,P(B')$$

Άρα τα A και $B'$ είναι ανεξάρτητα.
Με όμοιο τρόπο αποδεικνύονται οι προτάσεις (2) και (3).

**Παράδειγμα λανθασμένης χρήσης της έννοιας της ανεξαρτησίας ενδεχομένων.** ***(Η δίκη Collins στις ΗΠΑ):*** Το 1968, έγινε στο Λος Άντζελες η δίκη για μία ληστεία που είχε γίνει το 1964 από ένα ζευγάρι. Σύμφωνα με το κατηγορητήριο, ο άνδρας ήταν μαύρος με γένεια και μουστάκι και η γυναίκα λευκή με ξανθά μαλλιά και αλογουρά. Στην ληστεία, χρησιμοποίησαν ένα κίτρινο αυτοκίνητο. Ένα ζευγάρι που είχε όλα αυτά τα χαρακτηριστικά συνελήφθη και καταδικάστηκε. Η καταδίκη τους βασίσθηκε στην εξής επιχειρηματολογία που ανέλυσε ένας καθηγητής μαθηματικών ενός τοπικού κολλεγίου:

Από την σύνθεση του πληθυσμού του Λος Άντζελες προκύπτει ότι οι πιθανότητες για κάθε ένα από τα παρακάτω χαρακτηριστικά είναι:

P(ζευγάρι από διαφορετικές φυλές σε αυτοκίνητο) = 1/1000
P(κίτρινο αυτοκίνητο) = 1/10
P(μαύρος άνδρας με γένεια ) = 1/10
P(άνδρας με μουστάκι) = 1/4
P(ξανθή κοπέλλα) = 1/3
P(κοπέλλα με αλογουρά) = 1/10

Με την υποθέση ότι τα παραπάνω χαρακτηριστικά είναι ανεξάρτητα (κάτι που δεν αποδείχθηκε, αλλά ούτε αμφισβητήθηκε στην δίκη) και πολλαπλασιάζοντας τις αντίστοιχες πιθανότητες, βρέθηκε ότι η πιθανότητα να έχει ένα ζευγάρι όλα τα παραπάνω χαρακτηριστικά είναι 1/120000000. Είναι δηλαδή πολύ σπάνιο να βρεθεί ζευγάρι που να έχει όλα τα παραπάνω χαρακτηριστικά, δεδομένου ότι την εποχή αυτή ζούσαν στην περιοχή περίπου 12000000 ζευγάρια. Αφού λοιπόν βρέθηκε ένα ζευγάρι με όλα τα χαρακτηριστικά, το ζευγάρι αυτό θα πρέπει να είναι και οι ληστές. Με την επιχειρηματολογία αυτή, το ζευγάρι καταδικάσθηκε. Αργότερα όμως έκαναν έφεση και χρησιμοποιώντας μια άλλη επιχειρηματολογία βασισμένη στις πιθανότητες, αθωώθηκαν.

Το πρόβλημα δημιουργήθηκε από την άκριτη χρήση του πολλαπλασιαστικού κανόνα. Πριν χρησιμοποιηθεί ο κανόνας αυτός θα πρέπει να ελεγχθεί η ανεξαρτησία ή, εναλλακτικά, να χρησιμοποιηθεί η δεσμευμένη πιθανότητα.

Υπάρχει βέβαια και μια άλλη αντίρρηση στην επιχειρηματολογία του Εισαγγελέα. Οι υπολογισμοί πιθανοτήτων με την χρήση κανόνων όπως ο πολλαπλασιαστικός, αναπτύχθηκαν για την αντιμετώπιση τυχερών παιχνιδιών όπου η βασική διαδικασία που διέπει το παιχνίδι μπορεί να επαναληφθεί με ανεξάρτητες επαναλήψεις και κάτω από τις ίδιες συνθήκες. Ο Εισαγγελέας, στο συγκεκριμένο πρόβλημα, προσπάθησε να εφαρμόσει αυτή την θεωρία σε ένα μοναδικό φαινόμενο. Κάτι που ή συνέβη ή δεν συνέβη στις 18 Ιουνίου 1964 στις 11:30 π.μ.. Τί σημαίνει η πιθανότητα στο συγκεκριμένο πλαίσιο; Ήταν στην ευθύνη του Εισαγγελέα να απαντήσει στην ερώτηση αυτή και να αποδείξει ότι η θεωρία εφαρμόζεται στην συγκεκριμένη περίπτωση. Ο Εισαγγελέας υπολόγισε τις "πιθανότητες" για δύο "ενδεχόμενα" πηδώντας από το ένα στο άλλο και αντίστροφα. Το πρώτο ενδεχόμενο ήταν ότι οι κατηγορούμενοι ήταν ένοχοι. Το δεύτερο ενδεχόμενο ήταν ότι δεν υπήρχε άλλο ζευγάρι στο Λος Άντζελες την εποχή εκείνη με τα ίδια χαρακτηριστικά. Με την κλασσική στατιστική προσέγγιση (της σχετικής συχνότητας), η έννοια της πιθανότητας δεν μπορεί να εφαρμοστεί τόσο καλά. Ακόμα και κάποιος που ακολουθεί την Μπεϋζιανή προσέγγιση θα είχε κάποιες δυσκολίες εδώ δεδομένου ότι δεν υπάρχει κάποιο γενικό μοντέλο πιθανοτήτων που να συνδέει τα δεδομένα με την υπόθεση ενοχής ή αθωότητας.

Υπήρχαν άλλα ζευγάρια στο Λος Άντζελες που να ικανοποιούν τα ίδια χαρακτηριστικά; Καταρχήν η ερώτηση αυτή μοιάζει να είναι ένα στατιστικό ερώτημα που θα μπορούσε να απαντηθεί με την λήψη ενός δείγματος. Παρ' όλα αυτά, κάποιοι υπολογισμοί μπορούν να αποδείξουν ότι δειγματοληψία από τα ζευγάρια που ζούσαν στο Λος Άντζελες δεν θα έλυνε το πρόβλημα με κάποιο αξιόπιστο βαθμό εμπιστοσύνης (για την σχέση δείγματος με πληθυσμό, βλέπε μεθόδους δειγματοληψίας). Για την σωστή απάντηση στο ερώτημα αυτό θα έπρεπε να γίνει μία πλήρης απογραφή.

**Παράδειγμα.** ***(Νόμος των Hardy-Weinberg):*** Θεωρούμε ένα απλό γονίδιο το οποίο μπορεί να βρίσκεται σε μια από τις εξής δύο καταστάσεις. Επικρατούσα κατάσταση, έστω Α και υπολειπόμενη κατάσταση, έστω α. Κάθε άτομο σε ένα πληθυσμό είναι φορέας δύο τέτοιων γονιδίων. Επομένως, οι δυνατοί συνδυασμοί γονιδίων είναι:

ΑΑ (επικρατών ομόζυγος)
Αα (ετερόζυγος)
αα (υπολειπόμενος ομόζυγος)

Έστω ότι στην πρώτη γενεά ενός πληθυσμού τα ποσοστά των ατόμων που έχουν αυτά τα τρία είδη γονιδίων είναι $p_1$, $p_2$, και $p_3$ αντίστοιχα (για αρσενικά και θηλυκά). Σε ένα τυχαίο ζευγάρωμα, το αρσενικό και το θηλυκό επιλέγονται τυχαία και κάθε ένα συνεισφέρει ανεξάρτητα από το άλλο ένα από τα γονίδια στον απόγονο. Να βρεθεί η γενεά στην οποία επιτυγχάνεται η κατάσταση ισορροπίας της κατανομής των γονοτύπων.

**Λύση:** Η πιθανότητα ότι το αρσενικό συνεισφέρει στον απόγονο της πρώτης γενεάς ένα γονίδιο της μορφής Α είναι:

$P$ = P(μεταφέρεται το Α) = P(A|AA)P(AA)+P(A|Aα)P(Aα)

= P(AA)+P(A|Aα)P(Aα) = $p_1+p_2/2$

Ομοίως

q = P(μεταφέρεται το α) = P(α|Aα)P(Aα)+P(α|αα)P(αα) = $(1/2)p_2+p_3$

(ή ισοδύναμα: P(μεταφέρεται το α) = 1 - p = $(1/2)p_2 + p_3$ μια και $p_1+p_2+p_3 = 1$).

Είναι προφανές ότι οι πιθανότητες αυτές είναι ίδιες για τα γονίδια που συνεισφέρουν τα αρσενικά και τα θηλυκά. Επομένως, τα ποσοστά των ατόμων της δεύτερης γενεάς που είναι φορείς γονιδίων της μορφής ΑΑ, Αα και αα, έστω $p_1', p_2', p_3'$ αντίστοιχα είναι:

$p_1'$ = P(AA στην δεύτερη γενεά)

= P(αρσενικό της πρώτης γενεάς συνεισφέρει ένα Α $\cap$θηλυκό της πρώτης γενεάς συνεισφέρει ένα Α)

= [P(άτομο της πρώτης γενεάς συνεισφέρει ένα Α)]$^2$ = $p^2$

Ομοίως

$p_2'$ = P(Aα στην δεύτερη γενεά)

= P[{αρσενικό της πρώτης γενεάς συνεισφέρει ένα Α $\cap$ θηλυκό της πρώτης γενεάς συνεισφέρει ένα α} $\cup$ {αρσενικό της πρώτης γενεάς συνεισφέρει ένα α $\cap$ θηλυκό της πρώτης γενεάς συνεισφέρει ένα Α}] = 2p(1-p) = 2pq

Τέλος,

$p_3'$ = P(αα στην δεύτερη γενεά) = $q^2 = (1-p)^2$

Επαναλαμβάνοντας την ίδια λογική για τους απογόνους της δεύτερης γενεάς (δηλαδή για την τρίτη γενεά) θα έχουμε ότι:

$p^*$ = P(μεταφέρεται το Α από ένα άτομο της δεύτερης γενεάς)

= P(A|AA στην δεύτερη γενεά) P(AA στην δεύτερη γενεά)

+ P(A|Aα στην δεύτερη γενεά) P(Aα στην δεύτερη γενεά)

= $p_1' + (1/2)\ p_2' = p^2 + p(1-p) = p$

Δηλαδή, η πιθανότητα $p^*$ είναι ίση με την p και επομένως το ποσοστό ατόμων της τρίτης γενεάς με γονίδια της μορφής AA είναι

$$p_1'' = P(\text{AA στην τρίτη γενεά}) = p^2$$

Με όμοιο τρόπο, βρίσκουμε

$$p_2'' = P(\text{Aα στην τρίτη γενεά}) = 2p(1-p)$$

και

$$p_3'' = P(\text{αα στην τρίτη γενεά}) = (1-p)^2$$

Επομένως, τα γονότυπα της τρίτης γενεάς (και όλων των επομένων γενεών) θα έχουν την ίδια αναλογία όπως και τα γονότυπα της δεύτερης γενεάς. Δηλαδή, η κατάσταση ισορροπίας της κατανομής των γονοτύπων επιτυγχάνεται μετά από μία γενεά. Ο νόμος αυτός αποδείχθηκε το 1908 από τον μαθηματικό G. H. Hardy.

**Ορισμός:** Τα ενδεχόμενα $A_1, A_2, \ldots A_n$, $n \geq 1$ είναι *ανεξάρτητα* αν ο πολλαπλασιαστικός τύπος ισχύει για κάθε συνδυασμό δύο ή περισσοτέρων από αυτά. Αν δηλαδή

$$P(A_{i_1}A_{i_2}...A_{i_k}) = P(A_{i_1})P(A_{i_2})...P(A_{i_k}), \qquad 1 \leq i_1 < i_2 \ldots < i_k \leq n$$

Για παράδειγμα, τα ενδεχόμενα A, B, Γ είναι ανεξάρτητα αν

P(AB)=P(A)P(B), P(AΓ)=P(A)P(Γ), P(BΓ)=P(B)P(Γ) και

P(ABΓ)=P(A)P(B)P(Γ).

Γενικά αν έχουμε n ενδεχόμενα θα πρέπει να εξετάσουμε $\binom{n}{2}+\binom{n}{3}+\ldots+\binom{n}{n} = 2^n - n + 1$ περιπτώσεις.

**Ορισμός:** Τα ενδεχόμενα $A_1, A_2, \ldots A_n$ θα λέγονται *ανεξάρτητα κατά ζεύγη* αν $P(A_iA_j)=P(A_i)P(A_j)$, $i, j=1, 2, \ldots, n$, $i \neq j$.

**Σημείωση:** Είναι δυνατόν να έχουμε ανεξαρτησία κατά ζεύγη χωρίς να έχουμε ανεξαρτησία.

**Παράδειγμα:** Στρίβουμε δύο νομίσματα και θεωρούμε τα ενδεχόμενα

A=(το πρώτο νόμισμα είναι K)

B=(το δεύτερο νόμισμα είναι K)

Γ=(ακριβώς ένα K στο πείραμα)

Έχουμε

$$S=(KK, K\Gamma, \Gamma K, \Gamma\Gamma)$$
$$P(A)=P(B)=P(\Gamma)=1/2$$

Επίσης, επειδή

$$AB=\{KK\}, A\Gamma=\{K\Gamma\}, B\Gamma=\{\Gamma K\}$$

Θα είναι

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

Δηλαδή, τα A, B, Γ είναι ανεξάρτητα κατά ζεύγη.

Όμως $AB\Gamma=\varnothing \Rightarrow P(AB\Gamma)=0$

ενώ $P(A)P(B)P(\Gamma)=1/8 \neq P(AB\Gamma)$

Επομένως, τα A, B, Γ είναι εξαρτημένα.

**Ορισμός:** *(Ανεξάρτητες δοκιμές).* Έστω ότι ένα πείραμα Ε αποτελείται από μια ακολουθία n δοκιμών $\Delta_1$, $\Delta_2$, …, $\Delta_n$. Οι δοκιμές είναι *ανεξάρτητες,* αν το αποτέλεσμα κάθε μιας από αυτές δεν επηρεάζει τις πιθανότητες των ενδεχομένων στις άλλες δοκιμές. Επιπλέον, αν το ενδεχόμενο $A_i$ αντιστοιχεί στην i δοκιμή και $P(A_i)$ είναι η πιθανότητα του ενδεχομένου αυτού (i=1, 2, …, n), τότε

$$P(A_1 A_2 A_3 ... A_n) = P(A_1)P(A_2)P(A_3)...P(A_n)$$

**Ορισμός:** Τα ενδεχόμενα A και B λέγονται *ασυμβίβαστα* ή *ασύμβατα* ή *ξένα μεταξύ τους* (*mutually exclusive* ή *disjoint events*), αν η πραγματοποίηση του ενός δεν επιτρέπει την πραγματοποίηση του άλλου, αν δηλαδή

$$P(A\cap B)=0$$

**Σημείωση:** Προφανώς, στην περίπτωση των ξένων μεταξύ τους ενδεχομένων, ισχύει ότι $P(A\cup B) = P(A)+P(B)$.

Για παράδειγμα, στο ρίξιμο ενός ζαριού, τα ενδεχόμενα A={αποτέλεσμα "άσσος"} και B={αποτέλεσμα "έξι"} είναι ασύμβατα. ($P(A)=1/6$, $P(B)=1/6$, $P(A\cap B)=0$ και $P(A\cup B)=P(A)+P(B)=1/6+1/6=2/6$).

## Σχέση Ξένων Μεταξύ τους και Ανεξαρτήτων Ενδεχομένων

Τα *ασυμβίβαστα*, ή αλλιώς *ξένα μεταξύ τους, ενδεχόμενα* (*mutually exclusive*) και τα ανεξάρτητα ενδεχόμενα (*independent*) είναι δύο έννοιες που αναφέρονται σε ζευγάρια ενδεχομένων και εκφράζουν κάποιας μορφής σχέση μεταξύ τους. Η σχέση αυτή όμως είναι τελείως διαφορετική στην μια περίπτωση από την άλλη.

*Ξένα μεταξύ τους* είναι δύο ενδεχόμενα που η πραγματοποίηση του ενός δεν επιτρέπει την πραγματοποίηση του άλλου.

*Ανεξάρτητα* είναι δύο ενδεχόμενα η πραγματοποίηση του ενός εκ των οποίων δεν μεταβάλλει την πιθανότητα πραγματοποίησης του άλλου.

Με τα ξένα μεταξύ τους ενδεχόμενα σχετίζεται ο προσθετικός ή αθροιστικός κανόνας για τις πιθανότητες. Ο κανόνας αυτός αναφέρεται στην πιθανότητα ότι τουλάχιστον ένα από δύο πράγματα μπορούν να συμβούν.

Με τα ανεξάρτητα ενδεχόμενα συνδέεται ο πολλαπλασιαστικός κανόνας. Ο κανόνας αυτός χρησιμοποιείται για να προσδιορίσει την πιθανότητα ότι δύο ενδεχόμενα θα συμβούν ταυτόχρονα.

Επομένως, το πρώτο βήμα για να αποφασίσει κανείς εάν θα χρησιμοποιήσει τον πολλαπλασιαστικό ή τον αθροιστικό κανόνα είναι να απαντήσει στην ερώτηση: Με ενδιαφέρει η πιθανότητα P(A ή B), η πιθανότητα P(A και B) ή κάτι τελείως διαφορετικό;

Θα πρέπει να προσέχουμε ιδιαίτερα ότι, άθροιση πιθανοτήτων δύο ενδεχομένων απαιτεί ότι τα ενδεχόμενα αυτά είναι αμοιβαία ξένα μεταξύ τους.

Πολλαπλασιασμός μη δεσμευμένων πιθανοτήτων δύο ενδεχομένων απαιτεί ότι τα ενδεχόμενα αυτά είναι ανεξάρτητα. (Για εξαρτημένα ενδεχόμενα, ο πολλαπλασιαστικός κανόνας χρησιμοποιεί δεσμευμένες πιθανότητες).

Τα δύο παραδείγματα που ακολουθούν είναι χαρακτηριστικά κακής εφαρμογής των εννοιών των ανεξαρτήτων ενδεχομένων και των ξένων μεταξύ τους ενδεχομένων.

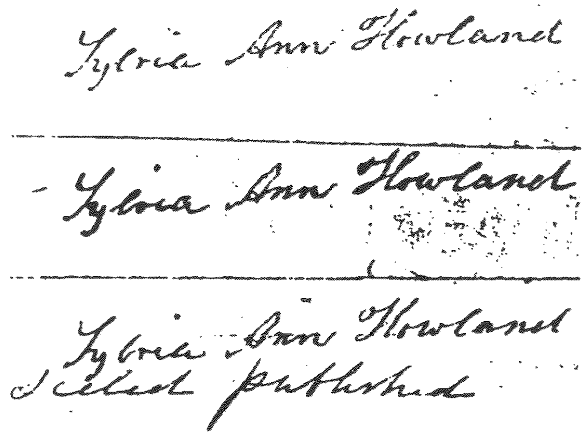
**Παράδειγμα λανθασμένης χρήσης του πολλαπλασιαστικού κανόνα των πιθανοτήτων για ανεξάρτητα ενδεχόμενα.** ***(Η διαθήκη Howland):*** Μια από τις πρώτες εφαρμογές της θεωρίας των πιθανοτήτων και των στατιστικών ενδείξεων στο δικαστικό σύστημα των Ηνωμένων Πολιτείων έγινε το 1967, όταν ο *Benjamin Peirce*, καθηγητής των μαθηματικών στο Πανεπιστήμιο του Harvard, κατέθεσε ως μάρτυρας στο δικαστήριο στην δίκη αμφισβήτησης της ομοιότητας της υπογραφής μιας διαθήκης και της υπογραφής που είχε μπει σε μια πρόσθετη παράγραφο της διαθήκης αυτής. *("Paul Meier & Sandy Zabell: Benjamin Peirce & the Howland Will". Journal of the American Statistical Association September 1980, 497-506).* Ο *Peirce*, χρησιμοποιώντας τον πολλαπλασιαστικό κανόνα των πιθανότητων, υποστήριξε ότι οι δύο υπογραφές είχαν γίνει από το ίδιο άτομο. Το συμπέρασμα αυτό το στήριξε σε μια λεπτομερή σύγκριση των κατακόρυφων γραμμών σε 42 άλλες υπογραφές του αποθανόντος για τις οποίες δεν είχε υπάρξει αμφισβήτηση. Σύμφωνα με τον πολλαπλασιαστικό κανόνα, ο *Peirce* εκτίμησε ότι η πιθανότητα για όμοια κατακόρυφη γραμμή είναι 0.2. Στην συγκεκριμένη διαθήκη, όλες οι 30 κατακόρυφες γραμμές στην πρόσθετη υπογραφή ταίριαζαν με τις κατακόρυφες γραμμές στην τελική υπογραφή της διαθήκης. Χρησιμοποιώντας τον πολλαπλασιαστικό κανόνα, υπολόγισε οτι η πιθανότητα να υπάρχουν 30 τέτοιες συμπτώσεις είναι $(0.2)^{30}$ και εξ αυτού συμπέρανε ότι ενδεχόμενο με τόσο μικρή πιθανότητα είναι μάλλον αδύνατον. Πρέπει να σημειώσουμε ότι η χρήση του πολλαπλασιαστικού κανόνα υποθέτει έμμεσα ανεξαρτησία σε όλες τις 30 κατακόρυφες γραμμές. Στην μαρτυρία του, δεν υπάρχει αναφορά για το κατά πόσο αυτή η κρίσιμη υπόθεση είναι εύλογη ή βάσιμη. Επίσης, δεν έλαβε υπόψη του το γεγονός ότι η εκτίμηση 0.2 για την πιθανότητα βασίστηκε σε 42 υπογραφές που μπήκαν σε διαφορετικές χρονικές στιγμές, ενώ η υπογραφή της διαθήκης και η υπογραφή στην πρόσθετη παράγραφο της διαθήκης γράφτηκαν, σύμφωνα με τις μαρτυρίες, την ίδια μέρα. Δεδομένου, όμως, ότι την εποχή εκείνη, η γνώση των πιθανότητων δεν ήταν διαδεδομένη και ταυτόχρονα, λόγω του μεγάλου σεβασμού που υπήρχε για την ακαδημαϊκή αξιοπιστία του *Peirce*, ο δικηγόρος της αντίθετης πλευράς δεν αμφισβήτησε τους ισχυρισμούς αυτούς.

**Η αυθεντική υπογραφή (αρ. 1) και
οι δύο αμφισβητούμενες υπογραφές (αρ. 10 και 15)**

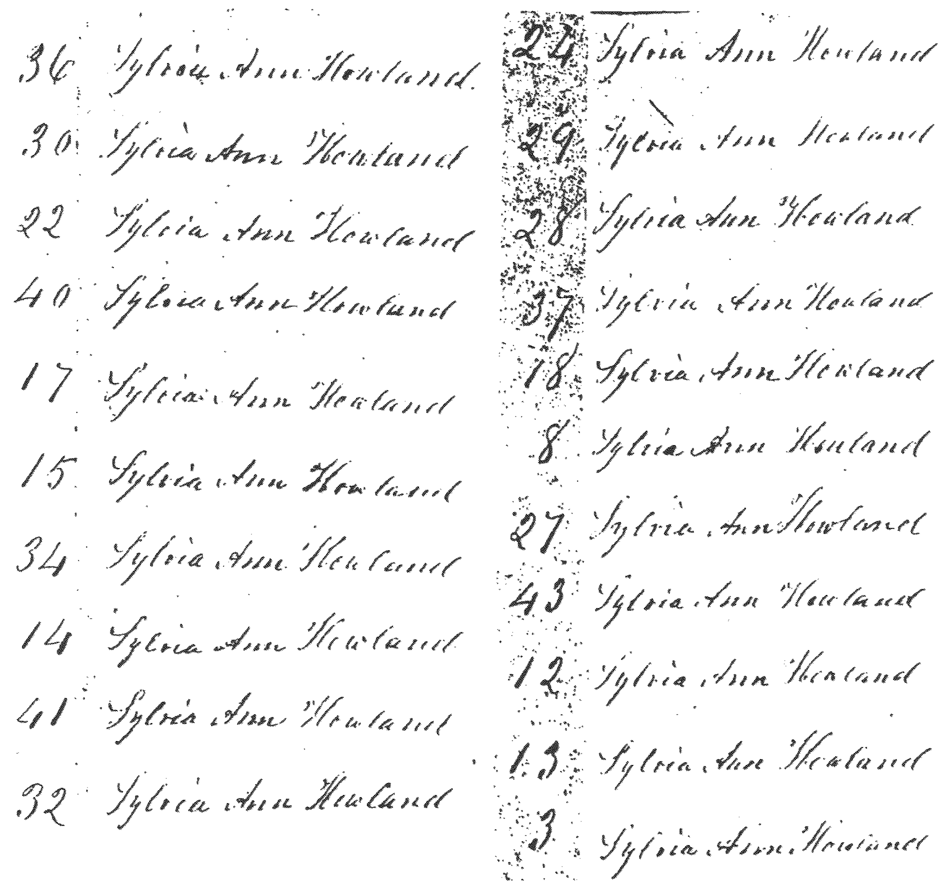
1

10

15



**Μερικές από τις 42 υπογραφές που χρησιμοποιήθηκαν για σύγκριση**



**Παράδειγμα λανθασμένης χρήσης του αθροιστικού κανόνα των πιθανοτήτων.** ***(Το παράδοξο του Chevalier De Méré):*** Τον 17ο αιώνα, οι Γάλλοι παίκτες τυχερών παιχνιδιών στοιχημάτιζαν πολλές φορές στο ενδεχόμενο: Όταν ένα ζάρι ριχτεί 4 φορές, ποιά είναι η πιθανότητα να εμφανισθεί τουλάχιστον ένας άσσος. Ένα άλλο τυχερό παιχνίδι που έπαιζαν ήταν στην περίπτωση όπου ένα ζευγάρι από ζάρια ριχνόταν 24 φορές και ενδιέφερε η πιθανότητα να εμφανισθούν άσσοι τουλάχιστον μία φορά.

Ο Chevalier De Méré, ένας Γάλλος ευγενής της περιόδου εκείνης, πίστευε ότι τα δύο παραπάνω ενδεχόμενα είναι ισοπίθανα. Η λογική του για το πρώτο παιχνίδι ήταν η εξής:

- Σε ένα ρίξιμο ζαριού, η πιθανότητα άσσου έναι 1/6.
- Σε τέσσερις ρίψεις του ζαριού, η πιθανότητα τουλάχιστον ενός άσσου είναι 4 φορές το 1/6=2/3.

Για το δεύτερο παιχνίδι, χρησιμοποίησε το εξής επιχείρημα:

- Όταν ρίξει κανείς ένα ζευγάρι ζάρια, η πιθανότητα άσσων είναι 1/36.
- Επομένως, σε 24 ρίψεις ενός ζευγαριού ζαριών, η πιθανότητα να πάρει κανείς τουλάχιστον ένα ζευγάρι άσσων είναι 24 φορές το 1/36=2/3.

Με την παραπάνω επιχειρηματολογία, οι πιθανότητες για τα δύο αυτά τυχερά παιχνίδια ήταν ίδιες, δηλαδή 2/3. Η εμπειρία, όμως, είχε δείξει ότι το πρώτο ενδεχόμενο είναι περισσότερο πιθανό να εμφανισθεί από ότι το δεύτερο. Η αντίφαση αυτή έγινε γνωστή ως *το παράδοξο του Chevalier De Méré,* και οφειλόταν στην λανθασμένη χρήση της έννοιας των αμοιβαία ξένα μεταξύ τους ενδεχομένων.

Ο De Méré ρώτησε τον φιλόσοφο Blaise Pascal για το πρόβλημα αυτό και ο Pascal το έλυσε με την βοήθεια ενός φίλου του του Piere de Fermat. (Ο Fermat ήταν δικαστής και ταυτόχρονα μέλος του κοινοβουλίου που είναι γνωστός σήμερα για την έρευνα στα μαθηματικά που έκανε αργά το βράδυ μετά τις άλλες ασχολίες του). Ο Fermat αντελήφθη ότι ο De Méré προσέθετε πιθανότητες για ενδεχόμενα τα οποία δεν ήταν ξένα μεταξύ τους. Στην πραγματικότητα, εάν κανείς προχωρούσε την επιχειρηματολογία του De Méré λίγο περισσότερο, θα μπορούσε να καταλήξει στο συμπέρασμα ότι η πιθανότητα να έχει κανείς ως αποτέλεσμα άσσο σε 6 ρίψεις ενός ζαριού είναι 6/6 ή, ισοδύναμα, 100%, κάτι που θα έπρεπε να είναι λάθος. Το ερώτημα που τίθεται είναι πώς να υπολογίσει κανείς σωστά τις πιθανότητες αυτές.

Σε τέσσερις ρίψεις ενός ζαριού, υπάρχουν $6^4 = 1.296$ δυνατά αποτελέσματα. Σε 24 ρίψεις ενός ζεύγους ζαριών υπάρχουν, $36^{24} \approx 2.2 \times 10^{37}$ αποτελέσματα. Ο υπολογισμός όμως των ευνοϊκών ενδεχομένων σε κάθε μια από τις δύο περιπτώσεις είναι αρκετά δύσκολος.

Ας δούμε όμως τον συλλογισμό του De Méré.

Στο πρώτο από τα παιχνίδια, αν Α, Β, Γ και Δ είναι τα ενδεχόμενα άσσου στην πρώτη, δεύτερη, τρίτη και τέταρτη δοκιμή αντίστοιχα, μας ενδιαφέρει η

$$P(A \cup B \cup \Gamma \cup \Delta),$$

δηλαδή η πιθανότητα άσσου σε μια τουλάχιστον από τις τέσσερις δοκιμές.

Ο De Méré, θεώρησε έμμεσα ότι τα ενδεχόμενα Α, Β, Γ, Δ είναι ξένα μεταξύ τους και κατέληξε ότι

$$P(A \cup B \cup \Gamma \cup \Delta) = P(A) + P(B) + P(\Gamma) + P(\Delta) =$$
$$= 1/6 + 1/6 + 1/6 + 1/6 = 4/6 = 2/3$$

Με τον ίδιο τρόπο μπορεί κανείς να περιγράψει την επιχειρηματολογία του De Méré στο δεύτερο πρόβλημα.

Ο τρόπος και οι σκέψεις που ο Pascal και ο Fermat χρησιμοποίησαν για να επιλύσουν σωστά το πρόβλημα δεν έχει καταγραφεί στην ιστορία. Ιστορικά στοιχεία όμως για την αλληλογραφία του Pascal και του Fermat μπορεί κανείς να βρει στο βιβλίο του David, F.N. *Games, Gods and Gambling*. Griffin, 1962. Ο Pascal και ο Fermat έλυσαν το πρόβλημα με ένα έμμεσο μαθηματικό συλλογισμό.

Θα μπορούσε όμως κανείς να φαντασθεί τον διάλογο του Pascal και του Fermat ως εξής:

*Pascal.* Ας κοιτάξουμε πρώτα το πρώτο παιχνίδι.

*Fermat*. Η πιθανότητα να κερδίσει κανείς είναι δύσκολο να υπολογισθεί, γι' αυτό, ας δουλέψουμε με την πιθανότητα του αντίθετου ενδεχομένου, δηλαδή του ενδεχομένου να χάσει. Τότε,

η πιθανότητα να κερδίσει είναι 100% - την πιθανότητα να χάσει.

*Pascal*. Συμφωνώ. Ο παίκτης θα χάσει όταν σε καμιά από τις 4 ζαριές δεν έλθει άσσος. Πώς όμως υπολογίζουμε αυτήν την πιθανότητα;
*Fermat*. Δεν είναι δύσκολο. Ας αρχίσουμε με το ρίξιμο ενός ζαριού. Ποιά είναι η πιθανότητα ότι στο πείραμα αυτό το αποτέλεσμα δεν θα είναι άσσος;
*Pascal*. Θα πρέπει το αποτέλεσμα να είναι ή 2 ή 3 ή 4 ή 5 ή 6. Επομένως, η πιθανότητα είναι 5/6.
*Fermat*. Ωραία. Και ποιά είναι η πιθανότητα ότι στις δύο πρώτες ζαριές το αποτέλεσμα δεν θα είναι άσσος;
*Pascal*. Μπορούμε να χρησιμοποιήσουμε τον πολλαπλασιαστικό κανόνα. Η πιθανότητα ότι και στην πρώτη ζαριά και στην δεύτερη ζαριά δεν θα υπάρχει άσσος είναι $5/6 \times 5/6 = (5/6)^2$. Αυτό, γιατί οι ζαριές είναι ανεξάρτητες μεταξύ τους, έτσι δεν είναι;
*Fermat*. Τί θα γίνει με τρεις ζαριές;
*Pascal*. Προφανώς, θα είναι $5/6 \times 5/6 \times 5/6 = (5/6)^3$.
*Fermat*. Εντάξει. Τί γίνεται με τέσσερις ζαριές;
*Pascal*. Θα πρέπει να είναι $(5/6)^4$.
*Fermat*. Ναι. Και αυτό είναι περίπου 0.482 ή 48,2%.
*Pascal*. Επομένως, η πιθανότητα να χάσει κανείς αυτό το παιχνίδι είναι 48,2% και

η πιθανότητα νίκης είναι 100% - πιθανότητα να χάσει κανείς = 100% - 48,2% = 51,8%.

*Fermat*. Ωραία. Αυτό δίνει την λύση στο πρώτο παιχνίδι. Η πιθανότητα δηλαδή λύσης είναι λίγο περισσότερο από 50%. Ας δούμε τώρα το δεύτερο παιχνίδι.
*Pascal*. Σε μια ζαριά με δύο ζάρια, υπάρχει 1/36 πιθανότητα άσσων και 35/36 πιθανότητα να μην έχουμε άσσους. Επομένως με, τον πολλαπλασιαστικό κανόνα, σε 24 ζαριές δύο ζαριών, η πιθανότητα να μην έχουμε άσσους θα πρέπει να είναι

$$(35/36)^{24}$$

*Fermat*. Ωραία. Αυτό είναι 50.9% και αυτό αποτελεί την πιθανότητα να χάσει κανείς. Επομένως,

πιθανότητα κέρδους = 100% - πιθανότητα να χάσει κανείς = 100% - 50.9% = 49.1%.

*Pascal*. Πράγματι και αυτό είναι λίγο λιγότερο από 50%. Γι' αυτό, παρατηρούμε λιγότερες φορές νίκης στο δεύτερο παιχνίδι από ό,τι στο πρώτο. Χρειάζεται, όμως, να επαναλάβει κανείς πολλές φορές το παιχνίδι για να διαπιστώσει την διαφορά.

**Σημείωση 1:** Η σωστή απάντηση με την χρήση του αξιώματος του Kolmogorov θα μπορούσε να βρεθεί αν ο De Méré είχε λάβει υπόψη του το γεγονός ότι τα ενδεχόμενα Α, Β, Γ, Δ δεν είναι ξένα μεταξύ τους. Επομένως,

$P(A \cup B \cup \Gamma \cup \Delta) = P(A) + P(B) + P(\Gamma) + P(\Delta) - P(AB) - P(A\Gamma) - P(A\Delta) - P(B\Gamma) - P(B\Delta) - P(\Gamma\Delta) - P(AB\Gamma) - P(AB\Delta) - P(B\Gamma\Delta) - P(AB\Gamma\Delta) =$
$= 4/6 - 6/36 - 2(1/6)^3 - (1/6)^4 = 2/3 - 1/6 - 1/108 - 1/1296 \approx 0.49$

**Σημείωση 2:** Από τον αριθμό των δυνατών ενδεχομένων που υπολογίσαμε προηγουμένως για τα δύο αυτά παιχνίδια, είναι φανερό ότι, παρότι θα μπορούσε κανείς να χρησιμοποιήσει τον κλασσικό ορισμό πιθανότητας του Pascal ως πηλίκου του αριθμού των ευνοϊκών διά του συνολικού αριθμού ενδεχομένων, είναι μάλλον εξαιρετικά δύσκολο να προσδιορισθεί ο αριθμός των ευνοϊκών ενδεχομένων.

**Σημείωση 3:** Το παράδειγμα αυτό αποτελεί κλασσική περίπτωση χρησιμοποίησης μιας στρατηγικής υπολογισμού των πιθανοτήτων. Αν είναι δύσκολο να υπολογίσει κανείς την πιθανότητα ενός ενδεχομένου, βρίσκει την πιθανότητα του συμπληρωματικού του ενδεχομένου και στην συνέχεια την αφαιρεί από την μονάδα.

## ΤΥΧΑΙΕΣ ΜΕΤΑΒΛΗΤΕΣ

Οι δειγματικοί χώροι που συναντήσαμε μέχρι τώρα στα παραδείγματά μας μπορούν να καταταγούν σε δυο κατηγορίες:

1) Ποιοτικοί ή περιγραφικοί
2) Ποσοτικοί ή αριθμητικοί.

Παράδειγμα της πρώτης κατηγορίας είναι το φύλο ενός παιδιού (S={A,K}), ενώ της δεύτερης είναι ο αριθμός των K σε τρία στριψίματα ενός νομίσματος. Ο υπολογισμός πιθανοτήτων ενδεχομένων ή συνδυασμού ενδεχομένων είναι βέβαια ευκολότερος στην δεύτερη περίπτωση όπου χρησιμοποιούνται αριθμοί.

Για τον ευκολότερο υπολογισμό πιθανοτήτων σε δειγματικό χώρο οποιασδήποτε μορφής θα ήταν επιθυμητό να ορισθεί ένας κανόνας (ή μια συνάρτηση) που να αντιστοιχεί ένα πραγματικό αριθμό k σε κάθε στοιχείο A του δειγματικού χώρου S. Η συνάρτηση αυτή ονομάζεται *τυχαία μεταβλητή (random variable).* Η χρήση της τυχαίας μεταβλητής διευκολύνει επίσης την κατάσταση όταν ενδιαφερόμαστε να υπολογίσουμε πιθανότητες συνάρτησης ενδεχομένων.

**Ορισμός:** Δοθέντος ενός δειγματικού χώρου S και ενός συνόλου B υποσυνόλων του S, ορίζουμε ως *τυχαία μεταβλητή (τ.μ.) (random variable)* μια μονοσήμαντη συνάρτηση με πεδίο ορισμού το S και τιμές στην ευθεία των πραγματικών αριθμών R

$$X : S \to R$$

τέτοια ώστε το σύνολο $X^{-1}$ (I) να είναι ένα ενδεχόμενο για κάθε διάστημα $I \subseteq R$.

Η τυχαία μεταβλητή δηλαδή δεν είναι τίποτα άλλο από μια πραγματική συνάρτηση που ορίζεται στα στοιχεία του δειγματικού χώρου. Είναι τυχαία με την έννοια ότι η τιμή της εξαρτάται από το τυχαίο αποτέλεσμα ενός πειράματος που καθορίζει ένα στοιχείο του δειγματικού χώρου (πεδίου ορισμού της τυχαίας μεταβλητής).

**Παράδειγμα:** Έστω ότι στρίβουμε ένα αμερόληπτο νόμισμα δύο φορές και ενδιαφερόμαστε για τον αριθμό των K που θα εμφανισθούν.

S = {KK, KΓ, ΓK, ΓΓ}

Έστω X ο αριθμός των K που παρατηρήθηκαν.

$$X: \quad \begin{array}{l} KK \rightarrow X(KK) = 2 \\ K\Gamma \rightarrow X(K\Gamma) = 1 \\ \Gamma K \rightarrow X(\Gamma K) = 1 \\ \Gamma\Gamma \rightarrow X(\Gamma\Gamma) = 0 \end{array}$$

Το ενδεχόμενο

A = {τουλάχιστο ένα K στις δύο δοκιμές)

μπορεί να εκφρασθεί μέσω της τυχαίας μεταβλητής X σαν $\{X \geq 1\}$. Επομένως $P(A) = P(X \geq 1)$.

Οι πιθανότητες επάγονται στο πεδίο τιμών της τυχαίας μεταβλητής μέσω των πιθανοτήτων που έχουν ορισθεί στον δειγματικό χώρο. Στο παράδειγμά μας το πεδίο τιμών της X είναι το $\{0,1,2\}$ και οι πιθανότητες που αντιστοιχούν σε αυτές μέσω του S είναι 1/4, 1/2, και 1/4 αντίστοιχα.

Είναι προφανές ότι σε ένα δειγματικό χώρο είναι δυνατό να ορισθούν ταυτόχρονα περισσότερες από μια τυχαίες μεταβλητές.

**Παράδειγμα:** Έστω οτι παίρνουμε 3 χαρτιά από μια τράπουλα.

$$S = \{(1,2,3), (1,2,4), \ldots\}$$

(με όλους τους συνδυασμούς χαρτιών και χρωμάτων).

Μπορούμε να ορίσουμε

X = αριθμός μαύρων χαρτιών

Y = αριθμός σπαθιών

Z = αριθμός χαρτιών με αριθμό μεγαλύτερο του 5.

κ.λ.π.

## Διακριτές Τυχαίες Μεταβλητές

**Ορισμός:** Μια τυχαία μεταβλητή $X:S \rightarrow R$ λέγεται *διακριτή (discrete)* (ή απαριθμητή ή ασυνεχής) αν ισχύει μία από τις παρακάτω συνθήκες:

1) S είναι ένα πεπερασμένο σύνολο.
2) X(S) είναι ένα πεπερασμένο σύνολο (έστω και αν S δεν είναι πεπερασμένο)
3) X(S) είναι αριθμήσιμο σύνολο.

**Παράδειγμα:** Για κάθε ενδεχόμενο A του S, μπορούμε να ορίσουμε μία τυχαία μεταβλητή $I_A$ που λέγεται τυχαία μεταβλητή-δείκτης του A ή τυχαία μεταβλητή του Bernoulli για το A ως εξής:

$$I_A(x) = \begin{cases} 1 \text{ αν } & x \in A \\ 0 \text{ αν } & x \notin A \end{cases} \qquad \forall \quad x \in S$$

Όταν το $I_A$ παίρνει την τιμή 1 αυτό σημαίνει ότι το ενδεχόμενο A έχει συμβεί, δηλαδή το πείραμα κατέληξε σε ένα στοιχειώδες ενδεχόμενο $x \in A$.

**Σημείωση:** Αργότερα θα έχουμε την ευκαιρία να μιλήσουμε για συναρτήσεις τυχαίων μεταβλητών. Αν $X:S \to R$ είναι μία τυχαία μεταβλητή και $f:R \to R$ είναι μία πραγματική συνάρτηση ορισμενή στο σύνολο των πραγματικών αριθμών, μπορούμε να ορίσουμε μία τυχαία μεταβλητή ως εξής: Για κάθε $x \in S$ έστω $f(X)(x)=f(X(x))$. Δηλαδή, πρώτα εφαρμόζουμε την τυχαία μεταβλητή X στο σημείο x στη συνεχεία δε εφαρμόζουμε την συνάρτηση f στον αριθμό αυτό παίρνοντας τον πραγματικό αριθμό $f(X(x))$. Για παράδειγμα, αν $f(x)=x^2$ μπορούμε να ορίσουμε την τυχαία μεταβλητή $X^2$ σαν $X^2(x)=(X(x))^2$.

## ΚΑΤΑΝΟΜΗ ΠΙΘΑΝΟΤΗΤΑΣ

Έστω $\{X=\alpha\} = \{x \in S : X(x) = \alpha\}$

**Ορισμός:** Έστω X μία διακριτή τυχαία μεταβλητή ορισμένη στο χώρο πιθανοτήτων S. Η συνάρτηση

$$p_x = P(x) \equiv P_X(x) = P(\{X=x\})$$

με πεδίο ορισμού τις τιμές της X και πεδίο τιμών τις πιθανότητες των τιμών αυτών λέγεται *συνάρτηση πιθανότητας* της X ή αλλιώς *κατανομή πιθανότητας (probability distribution)* της X.

Η συνάρτηση πιθανότητας δίνει την πιθανότητα με την οποία η τυχαία μεταβλητή X παίρνει την τιμή $x \in R$.

Από εδώ και στο εξής θα γράφουμε $P(X=x)$ και θα εννοούμε $P(\{X=x\})$.

Στο παράδειγμα όπου X είναι ο αριθμός των K που παρατηρήθηκαν όταν στρίβουμε ένα νόμισμα δυο φορές, έχουμε ότι η συνάρτηση πιθανότητας της X είναι

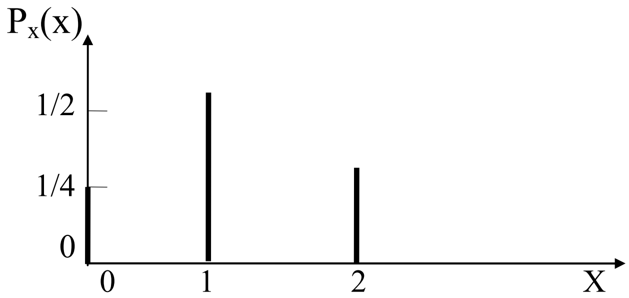
$$P_X(x) = \begin{cases} 1/4 & x = 0 \\ 1/2 & x = 1 \\ 1/4 & x = 2 \end{cases}$$

Απο τον παραπάνω ορισμό και τα αξιώματα των πιθανοτήτων, είναι φανερό ότι η συνάρτηση πιθανότητας $P_X(x)$ έχει τις εξής ιδιότητες:

1) $P_X(x) \geq 0$ , για κάθε $x \in R$
2) $\Sigma P_X(x) = 1$
3) $P(X \in A) = \sum_{x \in A} P_X(x)$ όπου A υποσύνολο του πεδίου τιμών της X.

Οι ιδιότητες (1) και (2) δίνουν τις αναγκαίες και ικανές συνθήκες που πρέπει να πληροί μία συνάρτηση για να είναι συνάρτηση πιθανότητας.

Πολλές φορές είναι χρήσιμο και σκόπιμο να έχουμε μία παραστατική μορφή της συνάρτησης πιθανότητας. Έτσι η γραφική παράσταση της συνάρτησης πιθανότητας του παραδείγματος είναι




**Παράδειγμα:** Διαλέγουμε ένα δείγμα από 3 αντικείμενα από ένα κουτί που περιέχει 12 αντικείμενα 3 από τα οποία είναι ελαττωματικά. Έστω X η τυχαία μεταβλητή που αναφέρεται στον αριθμό των ελαττωματικών αντικειμένων στο δείγμα. Το X μπορεί να πάρει τις τιμές 0,1,2, και 3 με πιθανότητες.

$$P(X=0)=\frac{\binom{3}{0}\binom{9}{3}}{\binom{12}{3}}=\frac{84}{220}$$

$$P(X=1)=\frac{\binom{3}{1}\binom{9}{2}}{\binom{12}{3}}=\frac{108}{220}$$

$$P(X=2)=\frac{\binom{3}{2}\binom{9}{1}}{\binom{12}{3}}=\frac{27}{220}$$

$$P(X=3)=\frac{\binom{3}{3}\binom{9}{0}}{\binom{12}{3}}=\frac{11}{220}.$$

## ΣΥΝΕΧΕΙΣ ΤΥΧΑΙΕΣ ΜΕΤΑΒΛΗΤΕΣ

**Ορισμός:** Μια τυχαία μεταβλητή X λέγεται (απόλυτα) *συνεχής (continuous)* αν υπάρχει μία μη αρνητική συνάρτηση f(x) που να ικανοποιεί την εξίσωση

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

για μια τυχαία τιμή του x , $-\infty < x < +\infty$.

Η συνάρτηση f(x) λέγεται *συνάρτηση πυκνότητας πιθανότητας (probability density function).*

Προφανώς αν X είναι (απόλυτα) συνεχής, $P(\alpha<X<\beta)= \int_{\alpha}^{\beta} f(t)dt$ .

Συνεχείς τυχαίες μεταβλητές συναντώνται συχνά, ειδικά σε πειράματα στα οποία παίρνονται μετρήσεις. Ας υποθέσουμε, για παράδειγμα, ότι 200 άνθρωποι μετρούν το ίδιο αντικείμενο του οποίου το μήκος είναι περίπου 25 εκατοστά. Στις περιπτώσεις αυτές, συνήθως, η απάντηση στρογγυλεύεται σε κάποιο δεκαδικό ψηφίο. Στο παράδειγμά μας, αν οι μετρήσεις στρογγυλευθούν στο πλησιέστερο εκατοστό ο δειγματικός χώρος θα μπορούσε να είναι το πεπερασμένο σύνολο {22, 23, 24, 25, 26, 27, 28}.

Αν οι μετρήσεις στρογγυλευθούν στο πλησιέστερο εκατοστό του εκατοστού ο δειγματικός χώρος θα μπορούσε να είναι πάλι ένα πεπερασμένο σύνολο της μορφής {22.00, 22.01, 22.02, ... 27.99, 28}.

Από το άλλο μέρος, πριν την στρογγυλοποίηση, οποιαδήποτε τιμή στο διάστημα [27, 28] είναι δυνατόν να προκύψει ως αποτέλεσμα μιας μέτρησης. Επομένως το διάστημα [22, 28] θα μπορούσε να θεωρηθεί ως δειγματικός χώρος του πειράματος αυτού με την έννοια ότι κάθε ένα από τα άπειρα στοιχεία του συνόλου αυτού είναι πιθανό ενδεχόμενο του πειράματος.

Είναι φυσικό ότι στην περίπτωση αυτή η $P(X=\alpha)$ όπου $\alpha \in X(S)$ είναι ίση με το μηδέν.

**Ιδιότητες της f(x):** Από τον ορισμό της f(x) προκύπτουν εύκολα οι εξής ιδιότητες:

1) $f(x) \geq 0$ σχεδόν παντού
2) $\int_{-\infty}^{+\infty} f(x)dx = 1$ .

(Σχεδόν παντού σημαίνει για όλα τα σημεία του R εκτός από ένα αριθμήσιμο, το πολύ, πλήθος σημείων του R).

## ΣΥΝΑΡΤΗΣΗ ΚΑΤΑΝΟΜΗΣ

**Ορισμός:** Έστω X μία τυχαία μεταβλητή. Η συνάρτηση $F:R \to R$ που ορίζεται ως

$$F_X(\alpha) = P(X\leq\alpha) \equiv P(s:X(s) \leq \alpha) \quad \forall\ \alpha\in R$$

λέγεται *συνάρτηση κατανομής (probability function)* ή *αθροιστική συνάρτηση κατανομής* της τυχαίας μεταβλητής X. Γνώση της $F_X(\alpha)$ καθορίζει τις πιθανότητες όλων των γεγονότων στο R.

**Σημείωση:** Στη συνέχεια θα χρησιμοποιούμε το συμβολισμό F(α) (χωρίς δηλαδή τον δείκτη X) όταν αυτό δεν δημιουργεί ασάφεια.

**Παράδειγμα:** Στο παράδειγμα όπου X είναι ο αριθμός των K που παρατηρήθηκαν όταν στρίβουμε ένα νόμισμα δυο φορές, έχουμε ότι η συνάρτηση κατανομής είναι

$$F(\alpha)=\begin{cases} 0 & \alpha<0 \\ 1/4 & 0\le\alpha<1 \\ 3/4 & 1\le\alpha<2 \\ 1 & 2\le\alpha \end{cases}$$

**Ιδιότητες:**

1) Η F(α) είναι μη φθίνουσα συνάρτηση του α. (Αν δηλαδή $\alpha_1 < \alpha_2 \Rightarrow F(\alpha_1) \le F(\alpha_2)$).

2) Η F(α) είναι συνεχής από τα δεξιά $\forall\ \alpha \in R$.
(Δηλαδη $\lim\limits_{\alpha_n \to a+0} F(\alpha_n) = F(\alpha)$).

3) Η μέγιστη τίμη της F(α) είναι το 1 και η ελάχιστη το 0.
(Δηλαδή $\lim\limits_{a\to+\infty} F(\alpha) = 1$ και $\lim\limits_{a\to-\infty} F(\alpha) = 0$).

Αν η τυχαία μεταβλητή X είναι διακριτή τότε η συνάρτηση $F(\cdot)$ είναι βαθμωτή με F(α) - F(α-1) = P(X=α).
Οι παρακάτω ιδιότητες της συνάρτησης $F_X(\cdot)$, όπου X διακριτή, είναι προφανείς.

1) $P(\alpha < X \le b) = F(b) - F(\alpha)$
2) $P(\alpha \le X \le b) = F(b) - F(\alpha-1)$
3) $P(\alpha < X < b) = F(b-1) - F(\alpha)$
4) $P(\alpha \le X < b) = F(b-1) - F(\alpha-1)$.

**Γραφική Παράσταση της Συνάρτησης F(·)**

Για το παράδειγμα όπου X είναι ο αριθμός των K που παρατηρήθηκαν όταν στρίβουμε ένα νόμισμα δυο φορές, έχουμε: