Ασκήσεις στις Βασικές έννοιες (Μέρος 2)

Πιθανότητες και Στατιστική

Τμήμα Πληροφορικής, Πανεπιστήμιο Πειραιώς

2023-2024

Βασικοί ορισμοί και βασικοί τύποι

Η έννοια της πιθανότητας

- Ορισμός κλασικής πιθανότητας (De Moivre 1711, Laplace 1812)
  Αν A είναι ενδεχόμενο του δειγματικού χώρου Ω, τότε

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{πλήθος ευνοϊκών για το } A \text{ αποτελεσμάτων}}{\text{πλήθος δυνατών αποτελεσμάτων}}$$

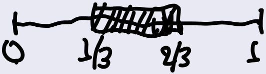
- Στατιστικός ορισμός πιθανότητας (όριο σχετικής συχνότητας)
  Αν $N_n(A)$ είναι το πλήθος πραγματοποιήσεων του ενδεχομένου A του δειγματικού χώρου Ω, μετά από n δοκιμές, τότε

$$P(A) = \lim_{n\to\infty} \frac{N_n(A)}{n}$$

  Πολλές επαναλήψεις, μεγάλα μεγέθη $\Rightarrow$ σύγκλιση σε μια τιμή (πιθανότητα).

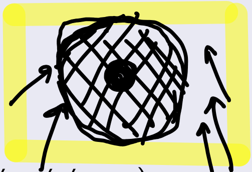
Ορισμός - Πιθανότητα σε υπεραριθμήσιμους χώρους.

Έστω Ω ένας υπεραριθμήσιμος δειγματικός χώρος οριζόμενος από μια περιοχή του (μοναδιάστατου ή διδιάστατου ή τρισδιάστατου) χώρου στην οποία οποιεσδήποτε στοιχειώδεις περιοχές είναι εξίσου πιθανές. Η πιθανότητα ενός ενδεχομένου A οριζόμενου από μια περιοχή του δειγματικού χώρου Ω δίδεται από την σχέση



$$P(A) = \frac{\mu(A)}{\mu(\Omega)}$$

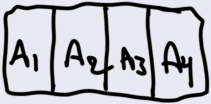
1



όπου $\mu(A)$ και $\mu(\Omega)$ είναι το μέτρο (μήκος, ή εμβαδόν, ή όγκος) των περιοχών A και Ω αντίστοιχα.

Αξιωματικός ορισμός πιθανότητας (Kolmogorov 1930)

Έστω συνάρτηση $P : \mathcal{P}(\Omega) \to \mathbb{R}$. Η P ονομάζεται **συνάρτηση πιθανότητας** ή **μέτρο πιθανότητας** αν και μόνο αν



- $P(A) \geq 0$, για κάθε $A \subseteq \Omega$.
- $P(\Omega) = 1$.
- $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, για κάθε (αριθμήσιμη) οικογένεια $(A_i)_{i \in \mathbb{N}^*}$ ξένων ανά δύο υποσυνόλων του Ω.

Η τριάδα $(\Omega, \mathcal{P}(\Omega), P)$ ονομάζεται **χώρος πιθανότητας**, ή **μοντέλο πιθανότητας**.

Ορισμός - Δεσμευμένη πιθανότητα

Η **δεσμευμένη πιθανότητα του ενδεχομένου** A **δεδομένου του ενδεχομένου** $B \neq \varnothing$ συμβολίζεται με $P(A|B)$ και ορίζεται ως εξής:

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

Βασικοί ορισμοί και βασικοί τύποι

Ορισμός - Ανεξάρτητα ενδεχόμενα

- Τα ενδεχόμενα $A, B \subseteq \Omega$ ονομάζονται **ανεξάρτητα** αν και μόνο αν
$$P(A \cap B) = P(A)P(B).$$
Ισοδύναμα, $P(A|B) = P(A)$ ή $P(B|A) = P(B)$.
- Τα ενδεχόμενα $A_1, A_2, \ldots, A_n \subseteq \Omega$ ονομάζονται (πλήρως) ανεξάρτητα αν και μόνο αν για κάθε υποσύνολο $\{i_1, i_2, \ldots, i_k\}$ του $[n]$, όπου $2 \leq k \leq n$, ισχύει ότι
$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}).$$

Ορισμός - Ανεξαρτησία υπό συνθήκη.

Έστω $\varnothing \neq C \subseteq \Omega$. Τα $A, B \subseteq \Omega$ ονομάζονται **υπό συνθήκη** ή **υπό δέσμευση ανεξάρτητα δεδομένου του** C αν και μόνο αν
$$P(A \cap B|C) = P(A|C)P(B|C).$$

Βασικοί ορισμοί και βασικοί τύποι

Ιδιότητες αξιωματικού ορισμού πιθανότητας

Αν $A, B, C \subseteq \Omega$, τότε

1. $P(\overline{A}) = 1 - P(A)$
2. $P(\varnothing) = 0$
3. $P(A) \leq 1$
4. $A \subseteq B \Rightarrow P(A) \leq P(B)$
5. $P(A \setminus B) = P(A \cap \overline{B}) = P(A) - P(A \cap B)$
6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
7. $P(A \cup B \cup C) =$ $P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)-P(B\cap C)+P(A\cap B\cap C)$

Ιδιότητες δεσμευμένης πιθανότητας

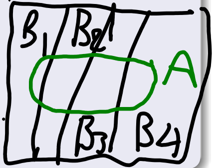
Αν $A, B, C, D \subseteq \Omega$, με $D \neq \varnothing$ τότε

1. $P(\overline{A}|D) = 1 - P(A|D)$
2. $P(\varnothing|D) = 0$
3. $P(A|D) \leq 1$
4. $P(A \setminus B|D) = P(A \cap \overline{B}|D) = P(A|D) - P(A \cap B|D)$
5. $A \subseteq B \Rightarrow P(A|D) \leq P(B|D)$
6. $P(A \cup B|D) = P(A|D) + P(B|D) - P(A \cap B|D)$
7. $P(A \cup B \cup C|D) = P(A|D) + P(B|D) + P(C|D) - P(A \cap B|D) - P(A \cap C|D) - P(B \cap C|D) + P(A \cap B \cap C|D)$

Τύπος ολικής πιθανότητας

Αν τα (μη κενά) σύνολα της οικογένειας $(B_i)_{i\in[n]}$, $n \in \mathbb{N}^*$, αποτελούν διαμέριση του Ω, τότε για κάθε $A \subseteq \Omega$ ισχύει

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



Τύπος του Bayes (Τύπος Bayes)

- Για κάθε $A, B \subseteq \Omega$, με $A, B \neq \varnothing$, ισχύει ότι

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

- Αν η οικογένεια $(A_i)_{i\in[n]}$, $n \in \mathbb{N}^*$, αποτελεί διαμέριση του Ω, τότε για κάθε μη κενό ενδεχόμενο $B \subseteq \Omega$ και για κάθε $i \in [n]$ ισχύει

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}.$$

Βασικοί ορισμοί και τύποι

Στην περίπτωση που τα ενδεχόμενα ενός δειγματικού χώρου δεν είναι κατ΄ ανάγκη ανεξάρτητα, ο παρακάτω τύπος είναι πολύ χρήσιμος.

Κανόνας του πολλαπλασιασμού

Έστω $A_1, A_2, \ldots, A_n \subseteq \Omega$ με $P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$. Τότε

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

Απόδειξη.

Ισχύει ότι

$$\begin{aligned}
&P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \\
&= P(\cancel{A_1})\frac{P(A_1 \cancel{\cap A_2})}{\cancel{P(A_1)}}\,\frac{P(A_1 \cap \cancel{A_2 \cap A_3})}{P(\cancel{A_1 \cap A_2})}\cdots\frac{P(A_1 \cap A_2 \cap \cdots \cap A_n)}{P(A_1 \cap A_2 \cancel{\cap \cdots \cap} A_{n-1})} \\
&= P(A_1 \cap A_2 \cap \cdots A_n).
\end{aligned}$$

$\square$

Παρ)ση: Το πλεονεκτημα του κανονα του πολ/σμου ειναι οτι μπορουμε να επιλεξουμε εμεις την σειρα που θα μπουν τα $A_1, A_2, \dots, A_n$

$$P(A_1 \cap A_2 \cap A_3 \cap A_4)$$

$$= P(A_3) \cdot P(A_4 | A_3) \cdot P(A_1 | A_3 \cap A_4) \cdot P(A_2 | A_1 \cap A_3 \cap A_4)$$

Άσκηση (Μανιώδεις παίχτες)

Ο Chevalier de Mere έπαιζε τόσο συχνά ζάρια που εμπειρικά γνώριζε ότι αν ρίξει τρία ζάρια οι ενδείξεις τους αθροίζουν συχνότερα στο 11 παρά στο 12, εν τούτοις προσπαθώντας να υπολογίσει τις αντίστοιχες πιθανότητες θεωρούσε ως δειγματικό χώρο τις μη διατεταγμένες τριάδες, όπου στο ενδεχόμενο A: τα τρία ζάρια αθροίζουν στο 11 αντιστοιχούν οι 6 τριάδες:

$$\{6,4,1\} \quad \{6,3,2\} \quad \{5,5,1\} \quad \{5,4,2\} \quad \{5,3,3\} \quad \{4,4,3\}$$

ενώ στο ενδεχόμενο B: τα τρια ζάρια αθροίζουν στο 12 αντιστοιχούν οι 6 τριάδες:

$$\{6,5,1\} \quad \{6,4,2\} \quad \{6,3,3\} \quad \{5,5,2\} \quad \{5,4,3\} \quad \{4,4,4\}$$

οπότε σύμφωνα με τον Chevalier de Mere οι πιθανότητες θα έπρεπε να είναι ίσες. Γι αυτό απευθύνθηκε στον Pascal για να του δώσει την λύση. Τι θα του απαντούσατε αν απευθυνόταν σε σας;

Λύση

Ο δειγματικός χώρος Ω του πειράματος αποτελείται από τις **διατεταγμένες τριάδες των ενδείξεων των τριών ζαριών**, δηλαδή $\Omega = \{(x, y, z) : 1 \leq x, y, z \leq 6\}$, οπότε $|\Omega| = 6 \cdot 6 \cdot 6 = 6^3$.
Έστω A το ενδεχόμενο τα τρία ζάρια να αθροίζουν στο 11 και B το ενδεχόμενο τα τρία ζάρια να αθροίζουν στο 12.
Το ενδεχόμενο A αποτελείται από τις 27 παρακάτω ισοπίθανες τριάδες:

(6, 4, 1)	(6, 3, 2)	(6, 2, 3)	(6, 1, 4)	(5, 5, 1)	(5, 4, 2)	(5, 3, 3)	(5, 2, 4)	(5, 1, 5)
(4, 6, 1)	(4, 5, 2)	(4, 4, 3)	(4, 3, 4)	(4, 2, 5)	(4, 1, 6)	(3, 6, 2)	(3, 5, 3)	(3, 3, 5)
(3, 4, 4)	(3, 2, 6)	(2, 6, 3)	(2, 5, 4)	(2, 4, 5)	(2, 3, 6)	(1, 6, 4)	(1, 5, 5)	(1, 4, 6)

Επομένως, $P(A) = \dfrac{27}{6^3}$.

Ασκήσεις

Λύση (συνέχεια)

Το ενδεχόμενο B αποτελείται από τις 25 παρακάτω ισοπίθανες τριάδες:

$(6,5,1)$	$(6,4,2)$	$(6,3,3)$	$(6,2,4)$	$(6,1,5)$	$(5,6,1)$	$(5,5,2)$	$(5,4,3)$	$(5,3,4)$
$(5,2,5)$	$(5,1,6)$	$(4,6,2)$	$(4,5,3)$	$(4,4,4)$	$(4,3,5)$	$(4,2,6)$	$(3,6,3)$	$(3,5,4)$
$(3,4,5)$	$(3,3,6)$	$(2,6,4)$	$(2,5,5)$	$(2,4,6)$	$(1,6,5)$	$(1,5,6)$		

Επομένως, $P(B) = \dfrac{25}{6^3}$.

Επομένως, είναι πιο πιθανό να φέρουμε 11 από ότι 12.

Παρατήρηση: Αν θεωρήσουμε ότι ο δειγματικός χώρος έχει μη διατεταγμένες τριάδες τότε τα στοιχειώδη ενδεχόμενα δεν είναι ισοπίθανα, πχ η 3-άδα $\{5,5,1\}$ μπορεί να προκύψει με 3 τρόπους, ενώ η τριάδα $\{6,4,1\}$ με 6 τρόπους.

Άσκηση (Εξάρες)

i) Να βρεθεί ο ελάχιστος αριθμός ρίψεων δύο ζαριών ώστε η πιθανότητα να έρθουν ‘‘εξάρες’’ σ’ αυτές τις ρίψεις να είναι μεγαλύτερη ή ίση του $1/2$·

ii) Αν διπλασιάσουμε τον αριθμό των ρίψεων του προηγούμενου ερωτήματος, θα διπλασιαστεί η πιθανότητα να έρθουν ‘‘εξάρες’’;

Λύση

(*i*) Έστω ότι ρίχνουμε τα δύο ζάρια n φορές και A_n το ενδεχόμενο μια τουλάχιστον φορά στις n ρίψεις να φέρουμε ‘‘εξάρες’’. Ψάχνουμε το ελάχιστο n ώστε $P(A_n) \geq \frac{1}{2}$. Θα υπολογίσουμε το ενδεχόμενο $\overline{A_n}$: Καμία φορά στις n ρίψεις να μην φέρουμε ‘‘εξάρες’’. Ισχύει ότι

$$P(\overline{A_n}) = \underbrace{\frac{35}{36} \cdot \frac{35}{36} \cdots \frac{35}{36}}_{n \text{ φορές}} = \left(\frac{35}{36}\right)^n$$

Λύση (συνέχεια)

Επομένως,

$$P(A_n) \geq \frac{1}{2} \Leftrightarrow P(\overline{A_n}) \leq \frac{1}{2} \Leftrightarrow \left(\frac{35}{36}\right)^n \leq \frac{1}{2} \Leftrightarrow 2 \leq \left(\frac{36}{35}\right)^n$$
$$\Leftrightarrow \ln 2 \leq n \ln \frac{36}{35} \Leftrightarrow n \geq \frac{\ln 2}{\ln 36 - \ln 35} = 24.605.$$

Άρα, απαιτούνται τουλάχιστον 25 ρίψεις των δύο ζαριών για να φέρουμε ‘‘εξάρες’’ με πιθανότητα τουλάχιστον $\frac{1}{2}$.
(ii) Όχι, η πιθανότητα να έρθουν ‘‘εξάρες’’ δεν είναι γραμμική συνάρτηση του αριθμού n των ρίψεων. Συγκεκριμένα, με $2 \cdot 25 = 50$ ρίψεις, η αντίστοιχη πιθανότητα είναι

$$P(A_{50}) = 1 - P(\overline{A_{50}}) = 1 - \left(\frac{35}{36}\right)^{50} = 0.755 = 75.5\%$$

Ασκήσεις

περιέχουν την λέξη buy

Άσκηση (Ανεπιθύμητη αλληλογραφία)

Ένας χρήστης έχει παρατηρήσει ότι, από τα email που λαμβάνει καθημερινά, το 60% είναι γραμμένα στα Αγγλικά και το υπόλοιπο 40% στα Ελληνικά. Επίσης, έχει παρατηρήσει ότι, από τα email στα Αγγλικά, το 90% είναι ανεπιθύμητα (spam) και, από τα email στα Ελληνικά, το 30% είναι ανεπιθύμητα.

α) Να βρεθεί η πιθανότητα ένα email που διαβάζει ο χρήστης να είναι spam.

β) Να βρεθεί η πιθανότητα ένα spam email που διαβάζει ο χρήστης να είναι γραμμένο στα Ελληνικά.

α) $P(S)$

English Greek

E = Το ενδεχομενο ενα email να ειναι στα Αγγλικά

G = — " — Ελληνικα

S = — " — να ειναι spam

Γνωριζουμε $P(E) = 0.6$ $P(G) = 0.4$

$P(S|E) = 0.9$ $P(S|G) = 0.3$

Απο δω προκυπτουν και τα εξης: $P(\bar{S}|E) = 1 - P(S|E) = 0.1$

$P(\bar{S}|G) = 1 - P(S|G) = 0.7$

α) Ψάχνουμε την πιθανοτητα $P(S)$

Τα E, G διαμεριζουν τον δειγματικό χώρο,
απο τον τυπο της ολικης πιθανότητας

$$P(S) = P(E) \cdot P(S|E) + P(G) \cdot P(S|G)$$

$$= 0.6 \cdot 0.9 + 0.4 \cdot 0.3 = 0.66$$

β) Ψάχνουμε την πιθανότητα $P(G|S)$

Θα χρησιμοποιώντας τον τύπο του Bayes έχουμε

$$P(G|S) = \frac{P(S|G) \cdot P(G)}{P(S)}$$

$$= \frac{0.3 \cdot 0.4}{0.66} = 0.1818$$

από το α) ερώτημα

Άσκηση (Οικιακή βοηθός) [SOS Θεμα εξετασεων]

Μια οικιακή βοηθός ζήτησε από τους προηγούμενους εργοδότες της μια συστατική επιστολή για μια καινούργια δουλειά. Εκτιμά ότι έχει 80% πιθανότητα να προσληφθεί αν η συστατική είναι καλή, 40% πιθανότητα αν η επιστολή είναι μέτρια, και 10% πιθανότητα αν η επιστολή είναι κακή. Επιπλέον, εκτιμά ότι οι πιθανότητες να είναι η συστατική καλή, μέτρια ή κακή είναι 0.7, 0.2 και 0.1 αντίστοιχα.

α) Πόσο σίγουρη είναι ότι θα προσληφθεί στην νέα δουλειά;

β) Με δεδομένο ότι προσλαμβάνεται, πόσο πιθανό είναι να έλαβε καλή, μέτρια, ή κακή συστατική επιστολή;

γ) Με δεδομένο ότι δεν προσλαμβάνεται, πόσο πιθανό είναι να έλαβε καλή, μέτρια ή κακή συστατική επιστολή;

Λύση R_1: Το ενδεχόμενο να πάρει καλή συστατική
R_2: —"— μέτρια —"—
R_3: —"— κακή —"—
A: —"— να προσληφθεί

$$\boxed{\begin{array}{ll} P(R_1)=0.7 & P(A|R_1)=0.8 \\ P(R_2)=0.2 & P(A|R_2)=0.4 \\ P(R_3)=0.1 & P(A|R_3)=0.1 \end{array}}$$

α) Ψάχνουμε την πιθανότητα $P(A)$

Τα R_1, R_2, R_3 διαμερίζουν τον δειγματικό χώρο
άρα από τον τύπο της ολικής πιθανότητας

$$P(A) = P(R_1)\cdot P(A|R_1) + P(R_2)\cdot P(A|R_2) + P(R_3)\cdot P(A|R_3)$$
$$= 0.7\cdot 0.8 + 0.2\cdot 0.4 + 0.1\cdot 0.1 = 0.65$$

Άρα, με βάση τα παραπάνω η πιθανότητα να προσληφθεί
είναι 65 %.

γ) Δεδομένου ότι δεν προσλαμβάνεται πόσο πιθανό είναι να έλαβε καλή, μέτρια ή κακή επιστολή;

Θα χρησιμοποιηθεί ο τύπος του Bayes

$$P(R_1 \mid \bar{A}) = \frac{P(\bar{A} \mid R_1) \cdot P(R_1)}{P(\bar{A})} = \frac{(1-0.8)\cdot 0.7}{1-0.65} = 0.4$$

$$P(R_2 \mid \bar{A}) = \frac{P(\bar{A} \mid R_2) \cdot P(R_2)}{P(\bar{A})} = \frac{(1-0.4)\cdot 0.2}{0.35} = 0.3428$$

$$P(R_3 \mid \bar{A}) = \frac{P(\bar{A} \mid R_3) \cdot P(R_3)}{P(\bar{A})} = \frac{(1-0.1)\cdot 0.1}{0.35} = 0.2571$$

Άρα, το πιο πιθανό σενάριο είναι ότι έλαβε ή μέτρια ή κακή συστατική επιστολή.

β) $P(R_1|A) = \dfrac{P(A|R_1)\cdot P(R_1)}{P(A)} = \ldots$

$P(R_2|A) = \dfrac{P(A|R_2)\cdot P(R_2)}{P(A)} = \ldots$

$P(R_3|A) = \dfrac{P(A|R_3)\cdot P(R_3)}{P(A)} = \ldots$

Άσκηση (Ποδοσφαιρόφιλοι)

Σε ένα μεγάλο δείγμα ποδοσφαιρόφιλων διαπιστώθηκε ότι 20% υποστηρίζουν την Α.Ε.Κ., 35% τον Π.Α.Ο. και 45% τον Ο.Σ.Φ.Π. Από τους συγκεριμένους φιλάθλους της Α.Ε.Κ. υπάρχει ποσοστό 4% που υποστηρίζει ως δεύτερη προτίμηση τον Πανιώνιο, το 2% των φιλάθλων του Π.Α.Ο. υποστηρίζει ως δεύτερη ομάδα επίσης τον Πανιώνιο και το 3% των φιλάθλων του Ο.Σ.Φ.Π. υποστηρίζει ως δεύτερη επιλογή τον Πανιώνιο.

Επιλέγουμε τυχαία ένα φίλαθλο:

α) Να προσδιορισθεί η πιθανότητα να υποστηρίζει τον Πανιώνιο.

β) Να βρεθεί ποια είναι η πιο πιθανή ομάδα που υποστηρίζει δεδομένου ότι δηλώνει ότι δεν είναι φίλαθλος του Πανιώνιου.

Λύση:

A = Το ενδεχόμενο κάποιος φίλαθλος να είναι AEK

Π = —〃— ΠΑΟ

Ο = —〃— ΟΣΦΠ

Β = —〃— Πανιώνιος

$$\boxed{\begin{aligned} P(A) &= 0.2 \\ P(\Pi) &= 0.35 \\ P(O) &= 0.45 \end{aligned}} \qquad \boxed{\begin{aligned} P(B|A) &= 0.04 \\ P(B|\Pi) &= 0.02 \\ P(B|O) &= 0.03 \end{aligned}}$$

α) Ψάχνουμε την πιθανότητα $P(B)$

Τα Α, Π, Ο διαμερίζουν τον δειγματικό χώρο άρα από τον τύπο της ολικής πιθανότητας έχουμε

$$P(B) = P(A)\cdot P(B|A) + P(\Pi)\cdot P(B|\Pi) + P(O)\cdot P(B|O)$$
$$= 0.2\cdot 0.04 + 0.35\cdot 0.02 + 0.45\cdot 0.03$$
$$= 0.0285 \simeq 3\%$$

β) Ψάχνουμε να βρούμε ποια από τις πιθανότητες $P(A|\bar{B})$, $P(\Pi|\bar{B})$, $P(O|\bar{B})$ είναι η μεγαλύτερη

Χρησιμοποιώντας τον τύπο του Bayes έχουμε

$$P(A|\bar{B}) = \frac{P(\bar{B}|A)\cdot P(A)}{P(\bar{B})} = \frac{(1-0.04)\cdot 0.2}{1-0.0285} = \frac{0.96\cdot 0.2}{0.9715} = 0.197$$

$$P(\Pi|\bar{B}) = \frac{P(\bar{B}|\Pi)\cdot P(\Pi)}{P(\bar{B})} = \frac{(1-0.02)\cdot 0.35}{0.9715} = 0.3530$$

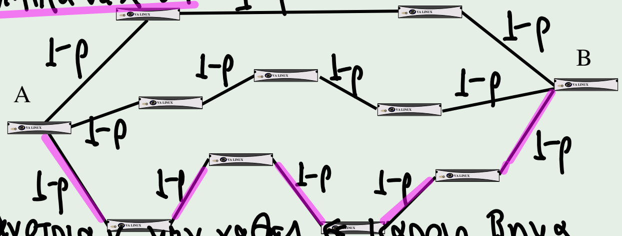
$$P(O|\bar{B}) = \frac{P(\bar{B}|O)\cdot P(O)}{P(\bar{B})} = \frac{(1-0.03)\cdot 0.45}{0.9715} = 0.4493$$

Άρα, το πιθανότερο είναι να είναι ΟΣΦΠ

Άσκηση (Δρομολόγηση πακέτων)

Ανάμεσα σε δύο υπολογιστές A και B ενός δικτύου υπάρχουν 3 διαφορετικά μονοπάτια δρομολόγησης πακέτων με μήκη 3, 4, 5 αντίστοιχα, όπως φαίνεται στο επόμενο σχήμα:

p: Η πιθανότητα να χαθεί



1-p: Η πιθανότητα να μην χαθεί σε κάποιο βήμα

Για την επικοινωνία μεταξύ τους, ο υπολογιστής A επιλέγει να χρησιμοποιήσει ένα από τα 3 μονοπάτια με πιθανότητα $1/2$, $1/3$ και $1/6$ αντίστοιχα.

Σε κάθε δεσμό που συνδέει δύο κόμβους των μονοπατιών υπάρχει πιθανότητα p να χαθεί το πακέτο που μεταφέρεται.

Να υπολογισθεί, συναρτήσει του p, η πιθανότητα ένα πακέτο που αποστέλλεται από τον υπολογιστή A να φτάσει στον υπολογιστή B.

Λύση:

P_1: Το ενδεχόμενο το πακέτο να σταλεί μέσω του 1ου μονοπατιού (με μήκος 3)

P_2: — // — 2ου μονοπατιού (με μήκος 4)

P_3: — // — 3ου μονοπατιού (με μήκος 5)

S: Το ενδεχόμενο το πακέτο να μην χαθεί και να φτάσει στο B

Γνωρίζουμε ότι $P(P_1)=\frac{1}{2}$, $P(P_2)=\frac{1}{3}$, $P(P_3)=\frac{1}{6}$

Για να μην χαθει ενα πακετο που διασχιζει ενα μονοπατι με k βηματα, πρεπει να μην χαθει σε κανενα απο τα k βηματα (πρεπει να τα διασχισει ολα τα k βηματα).

Η πιθανοτητα να συμβει αυτο ειναι $(1-p)^k$

αρα

$$\boxed{\begin{aligned} P(S \mid P_1) &= (1-p)^3 \\ P(S \mid P_2) &= (1-p)^4 \\ P(S \mid P_3) &= (1-p)^5 \end{aligned}}$$

Επειδη, τα P_1, P_2, P_3 αποτελουν διαμεριση του δειγματικου χωρου, απο τον τυπο της ολικης πιθανοτητας εχουμε

$$P(S) = P(P_1) \cdot P(S \mid P_1) + P(P_2) \cdot P(S \mid P_2) + P(P_3) \cdot P(S \mid P_3)$$

$$= \frac{1}{2}(1-p)^3 + \frac{1}{3}(1-p)^4 + \frac{1}{6}(1-p)^5$$

Ασκηση (Χαμενο εικοσαευρο)

Κάποιος βαζει στην τσεπη μας 5 χαρτονομισματα που εχουν αξια 100 ευρω. Οπως περπαταμε κατα λαθος κανουμε για κινηση και μας πεφτει ενα απο αυτα στον δρομο και το χανουμε. Ποια ειναι η πιθανοτητα να μας επεσε 20 ευρω;

Λυση Αρχικα θα βρουμε ολα τα πιθανα σεναρια για τα οποια 5 χαρτονομισματα εχουν αξια 100 ευρω

$$\boxed{\begin{aligned} &\Sigma 1: \quad 5 \quad 20 \text{ ευρω} \\ &\Sigma 2: \quad 50 + 20 + 10 + 10 + 10 \\ &\Sigma 3: \quad 50 + 20 + 20 + 5 + 5 \end{aligned}} \Bigg\} \text{ Εξισου πιθανα και τα 3 σεναρια}$$

Εστω A: Το ενδεχομενο να χαθηκε 20 ευρω

Τα $\Sigma_1, \Sigma_2, \Sigma_3$ διαμερίζουν τον δειγματικο χωρο

Αρα, απο τον τυπο της ολικής πιθανότητας εχουμε οτι

$$P(A) = P(\Sigma_1) \cdot P(A|\Sigma_1) + P(\Sigma_2) \cdot P(A|\Sigma_2) + P(\Sigma_3) \cdot P(A|\Sigma_3)$$

$$= \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot \frac{1}{5} + \frac{1}{3} \cdot \frac{2}{5} = 0.5333$$

Αρα, η πιθανοτητα να χαθηκε 20 ευρω ειναι 53%

Άσκηση (Κατηγοριοποίηση δεδομένων)

Για μια ομάδα ατόμων έχουμε συλλέξει τα παρακάτω δεδομένα

age	income	student	credit rating	buys computer
≤ 30	high	no	fair	no
≤ 30	high	no	excellent	no
31..40	high	no	fair	yes
> 40	medium	no	fair	yes
> 40	low	yes	fair	yes
> 40	low	yes	excellent	no
31..40	low	yes	excellent	yes
≤ 30	medium	no	fair	no
≤ 30	low	yes	fair	yes
> 40	medium	yes	fair	yes
≤ 30	medium	yes	excellent	yes
31..40	medium	no	excellent	yes
31..40	high	yes	fair	yes
> 40	medium	no	excellent	no

14 εγγραφές

#yes = 9

no = 5

Με βάση τα παραπάνω δεδομένα που θα κατηγοριοποιούσαμε ένα άτομο με τα παρακάτω χαρακτηριστικά

age	income	student	credit rating	buys computer
> 40	high	no	excellent	?

A_1 = Το ενδεχόμενο ένα άτομο να έχει ηλικία >40
I_1 = Το ενδεχόμενο ένα άτομο να έχει high income
S_1 = Το ενδεχόμενο ένα άτομο να μην σπουδάζει
C_1 = Το ενδεχόμενο ένα άτομο να έχει excellent credit rating
B = Το ενδεχόμενο ένα άτομο να αγοράσει ένα προϊόν

Ιδέα:
Για να κατηγοριοποιήσουμε το συγκεκριμένο άτομο θα συγκρίναμε τις παρακάτω πιθανότητες

$$P(B \mid A_1 \cap I_1 \cap S_1 \cap C_1) = 1$$

$$P(\bar{B} \mid A_1 \cap I_1 \cap S_1 \cap C_1) = 2$$

και όποια έχει την μεγαλύτερη τιμή θα θεωρήσουμε ότι ισχύει το B ή το B ανάλογα

Χρησιμοποιώντας τον τύπο του Bayes έχουμε

$$① = \frac{P(A_1 \cap I_1 \cap S_1 \cap C_1 \mid B) \cdot P(B)}{P(A_1 \cap I_1 \cap S_1 \cap C_1)}$$

$$② = \frac{P(A_1 \cap I_1 \cap S_1 \cap C_1 \mid \overline{B}) \cdot P(\overline{B})}{P(A_1 \cap I_1 \cap S_1 \cap C_1)}$$

Παρατηρούμε ότι τα δύο κλάσματα έχουν κοινό παρονομαστή, οπότε για να αποφασίσω αρκεί να συγκρίνω τους αριθμητές.

Υποθέτουμε* ότι τα ενδεχόμενα A_1, I_1, S_1, C_1 είναι ανεξάρτητα δεδομένου του B (και του $\overline{B}$)

[*Μας ενδιαφέρει να μην υπάρχουν ισχυρές εξαρτήσεις ανάμεσα τους]

Αυτές οι ιδέες ονομάζονται naive bayes classification

Με βαση τα παραπανω:
Για τον αριθμητη του 1 εχουμε:

$$P(A_1|B) \cdot P(I_1|B) \cdot P(S_1|B) \cdot P(C_1|B) \cdot P(B) =$$

$$\frac{3}{9} \cdot \frac{2}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{9}{14} = 0.005 \rightarrow$$ (ΔΕΝ ΕΙΝΑΙ ΠΙΘΑΝΟΤΗΤΑ)

Για τον αριθμητη του 2 εχουμε:

$$P(A_1|\overline{B}) \cdot P(I_1|\overline{B}) \cdot P(S_1|\overline{B}) \cdot P(C_1|\overline{B}) \cdot P(\overline{B}) =$$

$$\frac{2}{5} \cdot \frac{2}{5} \cdot \frac{4}{5} \cdot \frac{3}{5} \cdot \frac{5}{14} = 0.027$$

Επειδη 2 > 1 θεωρουμε οτι ειναι πιο πιθανο να ισχυει το $\overline{B}$ (δηλαδη θα θεωρησουμε οτι το ατομο δω θα αγορασει το προϊον)

Άσκηση (Έλεγχος ποιότητας)

Ένας ελεγκτής ποιότητας εξετάζει μια παρτίδα με 100 τεμάχια ενός προϊόντος επιλέγοντας 5 τεμάχια (χωρίς επανατοποθέτηση). Αν κανένα από τα τεμάχια δεν είναι ελαττωματικό, τότε η παρτίδα γίνεται αποδεκτή, αλλιώς υποβάλλεται σε περαιτέρω ελέγχους. Να βρεθεί η πιθανότητα μια παρτίδα που περιέχει 14 ελαττωματικά τεμάχια να γίνει αποδεκτή.

Λύση

Έστω A_i το ενδεχόμενο το i-στό τεμάχιο που ελέγχεται να μην είναι ελαττωματικό, $i = 1, 2, 3, 4, 5$.
Μια παρτίδα γίνεται δεκτή αν πραγματοποιηθούν και τα 5 ενδεχόμενα, επομένως μας ενδιαφέρει η πιθανότητα $P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5)$.
Από τον κανόνα του πολλαπλασιασμού έχουμε ότι

$$\boxed{P(A_1 \cap \cdots \cap A_5)} = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)P(A_4|A_1 \cap A_2 \cap A_3) \cdot P(A_5|A_1 \cap A_2 \cap A_3 \cap A_4).$$

Λύση (συνέχεια)

Επίσης,

$P(A_1) = \frac{86}{100}$, $P(A_2|A_1) = \frac{85}{99}$, $P(A_3|A_1 \cap A_2) = \frac{84}{98}$,

$P(A_4|A_1 \cap A_2 \cap A_3) = \frac{83}{97}$, $P(A_5|A_1 \cap A_2 \cap A_3 \cap A_4) = \frac{82}{96}$.

Επομένως,

$$P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) = \frac{86 \cdot 85 \cdot 84 \cdot 83 \cdot 82}{100 \cdot 99 \cdot 98 \cdot 97 \cdot 96} = 0.462577.$$

Αν είχαμε επανατοποθέτηση των τεμαχίων, το πρόβλημα θα λυνόταν πιο εύκολα:

$P(A_1) = P(A_2|A_1) = P(A_3|A_1 \cap A_2) = P(A_4|A_1 \cap A_2 \cap A_3) = P(A_5|A_1 \cap A_2 \cap A_3 \cap A_4) = \frac{86}{100}$.

Επομένως,

$$P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) = \left(\frac{86}{100}\right)^5 = 0.470427.$$

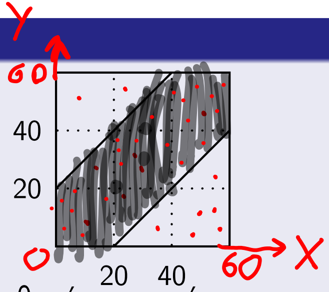
Συνάντηση στο Δημαρχείο

Δύο φίλοι Χ και Υ συμφωνούν να συναντηθούν στην πλατεία Κοραή κάποια στιγμή στο διάστημα μεταξύ 5:00 μ.μ. και 6:00 μ.μ. Ο καθένας τους φτάνει κάποια τυχαία στιγμή, περιμένει για 20 λεπτά, και αν δεν έρθει ο άλλος φεύγει. Ποια η πιθανότητα να συναντηθούν;

Λύση

Αν $x \in [0, 60]$ είναι η χρονική στιγμή άφιξης του Χ στο διάστημα 5:00 μέχρι 6:00 και $y \in [0, 60]$ είναι η χρονική στιγμή άφιξης του Υ, τότε μπορούμε να θεωρήσουμε ως δειγματικό χώρο το σύνολο των σημείων του τετραγώνου $[0, 60] \times [0, 60]$.
Οι δύο φίλοι θα συναντηθούν ανν $|x - y| \leq 20$, επομένως το ενδεχόμενο A οι δύο φίλοι να συναντηθούν αντιστοιχεί στα σημεία του χωρίου με $|x - y| \leq 20$, το οποίο περιγράφεται στο σχήμα:

Λύση (συνέχεια).



$\boxed{3600}$

Επομένως, η ζητούμενη πιθανότητα ισούται με το πηλίκο των εμβαδόν των δύο χωρίων:

$$P(A) = \frac{60^2 - 2\frac{1}{2}40^2}{60^2} = 1 - \left(\frac{2}{3}\right)^2 = 1 - \frac{4}{9} = \frac{5}{9} \simeq 0.55.$$

Άσκηση (Επέτειος γάμου)

Ο Πάνος και η Ελένη σχεδιάζουν χωριστά και μυστικά ο ένας από τον άλλον να πάρουν δώρο για την επέτειό τους. Σκοπεύουν να πληρώσουν με κάρτα από τον κοινό τους λογαριασμό για έκτακτες αγορές, που έχει υπόλοιπο 500 ευρώ. Αν ο καθένας σχεδιάζει να πάρει δώρα οποιασδήποτε αξίας από 100 μέχρι 300 ευρώ, να υπολογισθεί η πιθανότητα

i) να μην φτάσουν τα χρήματα στον ένα από τους δύο,

ii) να μείνουν τουλάχιστον 100 ευρώ για δείπνο σε εστιατόριο.

Λύση

Θέτουμε X, Y τα ποσά που θα ξοδεύσει ο Πάνος και η Μαρία αντίστοιχα. Το ζεύγος (X, Y) μπορεί να θεωρηθεί ως τυχαίο σημείο στο τετράγωνο $[100, 300] \times [100, 300]$.

i) Για μην φθάσουν τα χρήματα, θα πρέπει $X + Y > 500$, το οποίο συμβαίνει για όλα τα σημεία $(X, Y) \in [300, 300]$ που βρίσκονται πάνω από την ευθεία $X + Y = 500$, οπότε

$$P(X + Y > 500) = \frac{100 \cdot 100/2}{200^2} = \frac{1}{8} = 0.125.$$

ii) Για να μείνουν τουλάχιστον 100 ευρώ για δείπνο, θα πρέπει $X + Y \leq 400$, το οποίο συμβαίνει για όλα τα σημεία $(X, Y) \in [300, 300]$ που βρίσκονται όχι πάνω από την ευθεία $X + Y = 400$.

$$P(X+Y \leq 400) = 1 - P(X+Y > 400) = 1 - \frac{200 \cdot 200/2}{200^2} = 1 - \frac{1}{2} = 0.5.$$

Άσκηση (Ελικόπτερα)

Δέκα ελικόπτερα χρησιμοποιούνται για την αναζήτηση ενός αγνοούμενου ορειβάτη, ο οποίος μπορεί να βρίσκεται σε μία από δύο πιθανές περιοχές με πιθανότητες 0.7 και 0.3 αντίστοιχα. Κάθε ελικόπτερο μπορεί να χρησιμοποιηθεί σε μία μόνο περιοχή και έχει πιθανότητα 0.2 να βρει τον αγνοούμενο, αν αυτός βρίσκεται στην περιοχή που ερευνά.
Να βρεθεί πόσα ελικόπτερα πρέπει να σταλούν σε κάθε περιοχή, ώστε η πιθανότητα να βρεθεί ο αγνοούμενος να είναι η μέγιστη δυνατή.
Ποια είναι η πιθανότητα να βρεθεί ο αγνοούμενος σε αυτή την περίπτωση;
(Υπόδειξη: Υποθέστε ότι k ελικόπτερα κατανέμονται στην πρώτη περιοχή και $10 - k$ ελικόπτερα στην δεύτερη περιοχή.)

Λύση

Θέτουμε A το ενδεχόμενο να βρεθεί ο ορειβάτης από κάποιο ελικόπτερο, B_1 το ενδεχόμενο ο ορειβάτης να βρίσκεται στην πρώτη περιοχή και B_2 το ενδεχόμενο να βρίσκεται στην δεύτερη περιοχή.
Γνωρίζουμε ότι $P(B_1) = 0.7$ και $P(B_2) = 0.3$.
Αν κατανέμονται k ελικόπτερα στην πρώτη περιοχή και $10 - k$ ελικόπτερα στην δεύτερη, τότε

$$P(A|B_1) = 1 - P(\overline{A}|B_1) = 1 - 0.8^k \quad \text{και}$$
$$P(A|B_2) = 1 - P(\overline{A}|B_2) = 1 - 0.8^{10-k}.$$

Επομένως, από τον τύπο της ολικής πιθανότητας, έχουμε

$$\begin{aligned} P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) \\ &= (1 - 0.8^k) \cdot 0.7 + (1 - 0.8^{10-k}) \cdot 0.3 \\ &= 1 - 0.7 \cdot 0.8^k - 0.3 \cdot 0.8^{10-k} \end{aligned}$$

Λύση (συνέχεια)

Οι τιμές της $P(A)$ για $k = 0, 1, 2, \ldots, 10$ δίνονται στον επόμενο πίνακα:

k	0	1	2	3	4	5
$P(A)$	0.267	0.399	0.501	0.578	0.634	0.672
k	6	7	8	9	10	
$P(A)$	0.693	0.699	0.690	0.666	0.624	

από όπου προκύπτει ότι η $P(A)$ μεγιστοποιείται όταν 7 ελικόπτερα κατανέμονται στην πρώτη περιοχή (και 3 στη δεύτερη) και η αντίστοιχη πιθανότητα να βρεθεί ο ορειβάτης είναι $0.699 \approx 70\%$.

Άσκηση (Σπασμένο τηλέφωνο)

Δύο φίλοι κάθονται σε μια σειρά ανθρώπων, ο πρώτος στην αρχή της σειράς και ο δεύτερος στο τέλος της. Ανάμεσά τους κάθονται n άτομα $L_1, L_2, \ldots, L_n$, τα οποία έχουν την τάση να λένε ψέμματα. Αν ακούσουν ΝΑΙ τότε με πιθανότητα $p < 1$ μεταφέρουν στον επόμενο ΟΧΙ και με πιθανότητα $1 - p$ μεταφέρουν ΝΑΙ, και αν ακούσουν ΟΧΙ τότε με πιθανότητα p μεταφέρουν ΝΑΙ και με πιθανότητα $1 - p$ μεταφέρουν ΟΧΙ. Να βρεθεί η πιθανότητα p_n ο L_n να πει στον δεύτερο φίλο ΝΑΙ δεδομένου ότι ο πρώτος φίλος είπε στον L_1 ΝΑΙ.

Λύση

Θέτουμε p_i την πιθανότητα ο L_i να πει ΝΑΙ (ανεξαρτήτως τι άκουσε), οπότε $p_1 = 1 - p$ και, για κάθε $i = 2, 3, \ldots, n$,

$$p_i = p_{i-1}(1 - p) + (1 - p_{i-1})p = (1 - 2p)p_{i-1} + p.$$

Λύση (συνέχεια)

Επομένως, η ακολουθία (p_i) ικανοποιεί την γραμμική αναγωγική σχέση

$$p_i - (1-2p)p_{i-1} = p \text{ για } i \geq 2, \qquad p_1 = 1-p.$$

Από την χαρακτηριστική εξίσωση $x - (1-2p) = 0$, προκύπτει ότι η λύση της ομογενούς είναι η $q_i = c(1-2p)^i$. Επιπλέον, επειδή p είναι σταθερά που δεν εξαρτάται από το i, η μερική λύση είναι μια σταθερά A. Αντικαθιστώντας στην αναγωγική σχέση της p_i, έχουμε ότι

$$A - (1-2p)A = p \Leftrightarrow A = 1/2.$$

Επομένως, η γενική λύση είναι $p_i = q_i + A = c(1-2p)^i + \frac{1}{2}$. Για $i = 1$, έχουμε ότι $c(1-2p) + \frac{1}{2} = 1 - p \Leftrightarrow c = \frac{1}{2}$, άρα τελικά ισχύει ότι $\boxed{p_i = \frac{1}{2}((1-2p)^i + 1).}$

Λύση (συνέχεια)

Επομένως, η ζητούμενη πιθανότητα ισούται με

$$p_n = \frac{1}{2}\left(\left(1-2p\right)^n + 1\right)\right).$$

Καθώς $n \to \infty$ η πιθανότητα $p_n \to 1/2$, δηλαδή καθώς αυξάνει το n χάνεται η πληροφορία που μεταφέρουν οι ενδιάμεσοι αφού στο τέλος όλα είναι ισοπίθανα.

Άσκηση (Παράδοξα με μπάλες)

Έχουμε n δοχεία, καθένα εκ των οποίων περιέχει a λευκές και b μαύρες μπάλες. Μια μπάλα επιλέγεται τυχαία από το πρώτο δοχείο και μεταφέρεται στο δεύτερο, στην συνέχεια μια μπάλα επιλέγεται τυχαία από το δεύτερο δοχείο και μεταφέρεται στο τρίτο, κ.ο.κ. Στο τέλος, μια μπάλα επιλέγεται τυχαία από το τελευταίο δοχείο. Να βρεθεί η πιθανότητα η μπάλα που επιλέχθηκε να είναι λευκή.

Λύση

Θέτουμε A_1 το ενδεχόμενο να μεταφέρθηκε λευκή μπάλα από το πρώτο στο δεύτερο δοχείο και A_2 το ενδεχόμενο να επιλεγεί μια λευκή μπάλα από το δεύτερο δοχείο μετά την πρώτη μεταφορά. Ισχύει ότι

$$P(A_1) = \frac{a}{a+b}, \qquad P(\overline{A_1}) = \frac{b}{a+b} \quad \text{και}$$
$$P(A_2|A_1) = \frac{a+1}{a+b+1}, \qquad P(A_2|\overline{A_1}) = \frac{a}{a+b+1}$$

Λύση (συνέχεια)

Αφού τα A_1, $\overline{A_1}$ διαμερίζουν τον δειγματικό χώρο, από τον τύπο της ολικής πιθανότητας έχουμε ότι

$$\begin{aligned}
P(A_2) &= P(A_2|A_1)P(A_1) + P(A_2|\overline{A_1})P(\overline{A_1}) \\
&= \frac{a+1}{a+b+1} \cdot \frac{a}{a+b} + \frac{a}{a+b+1} \cdot \frac{b}{a+b} \\
&= \frac{a(a+1+b)}{(a+b+1)(a+b)} = \frac{a}{a+b}
\end{aligned}$$

Άρα, **η πιθανότητα να επιλεγεί μια λευκή μπάλα από το δεύτερο δοχείο μετά από την μεταφορά είναι η ίδια όπως και πριν την μεταφορά**. Συνεπώς, και η αντίστοιχη πιθανότητα να επιλεγεί μια λευκή μπάλα από το τρίτο δοχείο είναι επίσης ίδια με πριν, ομοίως για το τέταρτο κ.ο.κ., όπως και για το τελευταίο. Άρα, η πιθανότητα να επιλεγεί μια λευκή μπάλα από το n-οστό δοχείο ισούται με $a/(a+b)$.

Άσκηση (Επαλήθευση πολυωνυμικών ταυτοτήτων)

Έστω ότι θέλουμε να ελέγξουμε αν ισχύει η ισότητα

$$(x+1)(x+2)(x+3)(x+4) = x^4 + 10x^3 + 35x^2 + 50x + 25$$

Υπάρχουν δύο προσεγγίσεις που μπορούμε να ακολουθήσουμε:

1. Να κάνουμε όλους τους πολλαπλασιασμούς στην αριστερή πλευρά της ισότητας, να ομαδοποίησουμε ως προς τις δυνάμεις του x, και να ελέγξουμε αν το αποτέλεσμα ισούται με την δεξιά πλευρά.
2. Να θέσουμε $x = 0$ και στις δύο πλευρές ώστε να πάρουμε $24 = 26$ και να συμπεράνουμε ότι η ισότητα δεν ισχύει.

Άσκηση (Επαλήθευση πολυωνυμικών ταυτοτήτων - συνέχεια)

Γενικότερα, έστω ότι θέλουμε να επαληθεύσουμε την πολυωνυμική ταυτότητα

$$F(x) = G(x)$$

όπου τα πολυώνυμα F, G δίδονται σε διαφορετικές μορφές:

$$F(x) = \prod_{i=1}^{d}(x - a_i) \text{ και } G(x) = \sum_{i=1}^{d} c_i x^i$$

Η κλασική προσέγγιση είναι να πολλαπλασιάσουμε τους όρους του $F(x)$, να ομαδοποιήσουμε τις δυνάμεις του x και να επαληθεύσουμε την ταυτότητα. Το κόστος αυτής της προσέγγισης είναι $O(d^2)$ πολλαπλασιασμοί των συντελεστών: αυτό σημαίνει ότι ο αριθμός των πολλαπλασιασμών αυξάνει όπως η συνάρτηση Cd^2 όπου το C είναι μια θετική σταθερά. Υπάρχει και ένας άλλος τρόπος ο οποίος είναι υπολογιστικά φθηνότερος, αλλά δεν δίνει πάντα την σωστή απάντηση!

Ασκήσεις

Άσκηση (Επαλήθευση πολυωνυμικών ταυτοτήτων - συνέχεια)

Ένας τυχαιοποιημένος (randomized) αλγόριθμος για το πρόβλημα επαλήθευσης των πολυωνυμικών ταυτοτήτων:
Είσοδος: Τα πολυώνυμα F, G βαθμού d και ο αριθμός n των επαναλήψεων.
Έξοδος: Η απάντηση ΝΑΙ αν τα πολυώνυμα είναι ίσα και ΟΧΙ αν δεν είναι ίσα.

- Σε κάθε επανάληψη διαλέγουμε με ίση πιθανότητα ένα τυχαίο αριθμο r από το σύνολο $\{1, 2, \ldots, 100d\}$
    - Υπολογίζουμε τα $F(r)$, $G(r)$ (κόστος υπολογισμού $O(d)$ το καθένα)
    - Αν $F(r) \neq G(r)$, τότε τα δύο πολυώνυμα δεν είναι ίσα, και σταματάμε.
      Αν $F(r) = G(r)$, συνεχίζουμε μέχρι να συμπληρωθούν οι n επαναλήψεις.
- Αν συμπληρωθούν οι n επαναλήψεις, θεωρούμε ότι τα πολυώνυμα είναι ίσα.

Ο αλγόριθμος αυτός ονομάζεται τυχαιοποιημένος διότι κατά την εκτέλεση του χρησιμοποιεί τυχαίους αριθμούς.

Άσκηση (Επαλήθευση πολυωνυμικών ταυτοτήτων - συνέχεια)

Αν ο αλγόριθμος συμπεράνει ότι τα δύο πολυώνυμα δεν είναι ίσα, αυτό είναι πάντα σωστό.
Αν ο αλγόριθμος συμπεράνει ότι τα δύο πολυώνυμα είναι ίσα, αυτό δεν ισχύει πάντα.
Να υπολογισθεί η πιθανότητα ο αλγόριθμος να συμπεράνει ότι δύο πολύωνυμα είναι ίσα, δεδομένου ότι δεν είναι ισα.

Λύση

Αν τα πολυώνυμα F, G δεν είναι ίσα, ο αλγόριθμος συμπεραίνει εσφαλμένα ότι είναι ίσα όταν $F(r) - G(r) = 0$. Η διαφορά $F(x) - G(x)$ είναι ένα πολυώνυμο βαθμού d, άρα έχει το πολύ d πραγματικές ρίζες. Επομένως, σε κάθε επανάληψη η πιθανότητα το r που επιλέγεται από το διάστημα $\{1, 2, \ldots, 100d\}$ να είναι μια τις d ρίζες είναι το πολύ $d/100d = 1/100$. Άρα, σε κάθε επανάληψη η πιθανότητα σφάλματος είναι το πολύ $1/100$. Συνολικά, σε n επαναλήψεις η πιθανότητα σφάλματος γίνεται μικρότερη ή ίση από $1/100^n$.

Άσκηση (Καρδιακή πάθηση)

Διαθέτουμε μια συλλογή ιατρικών δεδομένων στην οποία για κάθε άτομο έχουμε καταγράψει τα παρακάτω γνωρίσματα:

- Exercise (Yes, No)
- Diet (Healthy, Unhealthy)
- Chest pain (Yes, No)
- Blood pressure (Yes, No)
- Heart disease (Yes, No)

BBN's
Bayesian Belief Networks

Οι τιμές των γνωρίσματων Exercise και Diet επηρρεάζουν τις τιμές του γνωρίσματος Heart disease (π.χ. η ανθυγεινή διατροφή και η απουσία άσκησης μπορεί να δημιουργήσουν προβλήματα στην καρδιά). Υπάρχει μεταξύ τους μια **σχέση αιτίου – αποτελέσματος**.
Από την άλλη, οι τιμές του γνωρίσματος Heart disease επηρρεάζουν τις τιμές των γνωρισμάτων Blood pressure και Heart disease (π.χ. προβλήματα στην καρδιά μπορεί να έχουν ως συνέπεια υψηλή πίεση και πόνο στο στήθος). Υπάρχει μεταξύ τους μια **σχέση αιτίου – αποτελέσματος**.

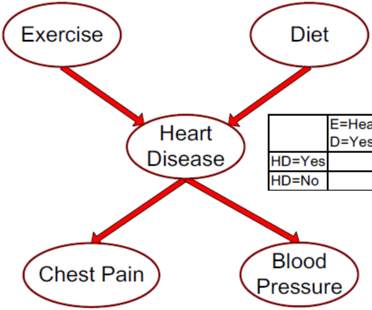
Ασκήσεις

Άσκηση (Καρδιακή πάθηση - συνέχεια)

Οι προηγούμενες εξαρτήσεις μπορούν να απεικονισθούν με την βοήθεια ενός προσανατολισμένου γραφήματος. Παράλληλα, με βάση αυτά τα δεδομένα, έχουμε υπολογίσει τις παρακάτω (απλές και δεσμευμένες) πιθανότητες:

Exercise=Yes	0.7
Exercise=No	0.3

Diet=Healthy	0.25
Diet=Unhealthy	0.75



	E=Healthy D=Yes	E=Healthy D=No	E=Unhealthy D=Yes	E=Unhealthy D=No
HD=Yes	0.25	0.45	0.55	0.75
HD=No	0.75	0.55	0.45	0.25

	HD=Yes	HD=No
CP=Yes	0.8	0.01
CP=No	0.2	0.99

	HD=Yes	HD=No
BP=High	0.85	0.2
BP=Low	0.15	0.8

Άσκηση (Καρδιακή πάθηση - συνέχεια)

Ένας άνθρωπος έχει τα παρακάτω γνωρίσματα

Exercise	Diet	Chest pain	Blood Pressure
Yes	Unhealthy	No	Yes

Πόσο πιθανό είναι να έχει καρδιακή πάθηση ο άνθρωπος αυτός;

Άσκηση (Άλογα)

Στους ετήσιους πανευρωπαϊκούς αγώνες ιπποδρομίας συμμετέχουν άλογα από όλη την Ευρώπη τα οποία δεν έχουν αγωνιστεί ποτέ μεταξύ τους. Κάποιος επίδοξος τζογαδόρος μελέτησε τις τελευταίες 200 κούρσες των 3 φαβορί και συνόψισε τα αποτελέσματά τους (κατάταξη στην 1η, 2η, 3η θέση ή στις υπόλοιπες) και τις αντίστοιχες πιθανότητες στον επόμενο πίνακα:

Κατάταξη	Αστραπή	Άνεμος	Φάντασμα
1η	0.15	0.30	0.20
2η	0.10	0.05	0.30
3η	0.70	0.25	0.30
Άλλη	0.05	0.40	0.20

Με βάση τα παραπάνω δεδομένα ποιό άλογο θεωρείτε ότι είναι το πιο προβλέψιμο;