Θεωρία Πιθανοτήτων και Στατιστική

Διδάσκων: Επαμεινώνδας Διαμαντόπουλος
Επικοινωνία: epdiaman@ee.duth.gr

Περιεχόμενα 2ου μαθήματος

- Διατάξεις (ασκήσεις).
- Συνδυασμοί.
- Δεσμευμένη Πιθανότητα.
- Δεσμευμένη πιθανότητα και ανεξάρτητα ενδεχόμενα.
- Τύπος του Bayes.
- Ευαισθησία, ειδικότητα και άλλα μέτρα αξιολόγησης δίτιμης διαδικασίας πρόβλεψης.
- Ο κανόνας του Bayes σε όρους ευαισθησίας και ειδικότητας.
- Καμπύλη λειτουργικού χαρακτηριστικού δέκτη (ROC curve).

Γνωστικοί στόχοι 2ου μαθήματος

Στο τέλος αυτού του μαθήματος, ο φοιτητής πρέπει να είναι σε θέση :

- Να διαχωρίζει τις έννοιες διάταξη και συνδυασμό και να είναι σε θέση να υπολογίζει το πλήθος διατάξεων και συνδυασμών σε απλά προβλήματα.
- Να υπολογίζει δεσμευμένες πιθανότητες με χρήση του ορισμού.
- Να μπορεί να εφαρμόζει τον νόμο της ολικής πιθανότητας.
- Να μπορεί να εφαρμόζει τον κανόνα του Bayes.
- Να γνωρίζει βασικά μέτρα αξιολόγησης δίτιμης διαδικασίας πρόβλεψης.
- Να κατανοεί τις πληροφορίες που περιέχει μία καμπύλη ROC.

Κώδικας R – Διατάξεις P(5, 3) (χωρίς επανάθεση)

```
library(gtools)

res <- permutations(n=5, r=3, v=1:5)
print ("Διατάξεις P(5, 3) (χωρίς επανάθεση)")
print (res)

print ("Πλήθος διατάξεων P(5, 3) (χωρίς επανάθεση)")
print (nrow(res))
```

$\{α, β, γ, δ\}$

$\underline{4}\ \underline{3}\ \underline{2}\ \underline{1} \to 4! = 24$

$\underline{4} \cdot \underline{3} \to 12$

άλυτη

Κώδικας R – Διατάξεις P(5, 3) (με επανάθεση)

```
library(gtools)

res1 <- permutations(n=5, r=3, v=1:5, repeats.allowed=T)
print ("Διατάξεις (με επανάθεση)")
print (res1)

print ("Πλήθος διατάξεων (με επανάθεση)")
print (nrow(res1))
```

{α, β, γ, δ, ε}

$\underline{5}\ \underline{5}\ \underline{5}$ 5^3

Κώδικας R – Συνδυασμοί C(5, 2)

```
library(gtools)

res <- permutations(n=5, r=2, v=1:5)
print ("Διατάξεις P(5, 2) (χωρίς επανάθεση)")
print (res)
print ("Συνδυασμοί C(5, 2)")
combinations(5, 2)

res <- permutations(n=5, r=3, v=1:5)
print ("Διατάξεις P(5, 3) (χωρίς επανάθεση)")
print (res)
print ("Συνδυασμοί C(5, 3)")
combinations(5, 3)
```

$\{\alpha, \beta, \gamma, \delta, \varepsilon\}$

$\underline{5} \cdot \underline{4} = 20$

$\underline{\alpha}\ \underline{\beta}$

$\underline{\beta}\ \underline{\alpha}$

$\underline{\alpha}\ \underline{\gamma}$

$\underline{\gamma}\ \underline{\alpha}$

$\underline{\varepsilon}\ \underline{\alpha}$

$\frac{20}{2} = 10$

$\frac{5 \cdot 4 \cdot 3!}{2! \cdot 3!} = \binom{5}{2}$

Διατάξεις (ασκήσεις)

$P(n,k) = n\cdot(n-1)\cdots(n-k+1)$

(γ)(i) Συνολικές: $P(7,5)$. Ευνοϊκές BA + AB, $P(5,3)$

Ασκήσεις στις διατάξεις

{A, B, Γ, Δ, E, Z, H}

1. Για τη δημιουργία “λέξεων” μπορούμε να χρησιμοποιήσουμε τα Α, Β, Γ, Δ, Ε, Ζ, Η.
(α) Πόσες διαφορετικές “λέξεις” με 3 γράμματα, *χωρίς επανάληψη*, μπορούν να δημιουργηθούν;
(β) Ποια είναι η πιθανότητα μία τέτοια “λέξη” 3 γραμμάτων:
  (i) Να μην περιέχει φωνήεν.
  (ii) Να περιέχει το Α.
  (iii) Να περιέχει τουλάχιστον ένα από τα Α ή Β. (′) : κανένα από τα Α, Β.
(γ) Ποια είναι η πιθανότητα μία τέτοια “λέξη” 5 γραμμάτων:
  (i) Να περιέχει τα Α, Β διαδοχικά. AB + BA
  (ii) Να περιέχει και το Α και το Β.

AB _ _ _ $P(5,3)\cdot 4\cdot 2$

Υπόδειξη
(β) ~~(ii) και~~ (iii): Υπολογίστε την πιθανότητα του συμπληρωματικού γεγονότος.

β (iii) Ευνοϊκές: $P(5,3) = 5\cdot 4\cdot 3 = 60$

$P(\{\text{κανένα } A,B\}) = \frac{60}{210}$

$P(\{\text{τουλ. ένα } A,B\}) = \frac{150}{210}$.

(α) $P(7,3) = 7\cdot 6\cdot 5 = 210$.

(β)(i) Ευνοϊκές: $P(4,3) = 4\cdot 3\cdot 2 = 24$. $P(\{\text{όχι φωνήεν}\}) = \frac{24}{210}$.

(ii) A _ _ / _ A _ / _ _ A Ευνοϊκές: $3\cdot P(6,2) = 3\cdot 6\cdot 5 = 90$, $P(\{\text{περιέχει } A\}) = \frac{90}{210}$.

άλυτη

$$\frac{n!}{s_1!\cdot s_2!\cdots s_k!}$$

Ασκήσεις στις διατάξεις

3. (i) Να βρεθεί το πλήθος των “λέξεων” που μπορούμε να γράψουμε με όλα τα γράμματα της λέξης “ΦΟΙΤΗΤΗΣ”.
(ii) Πόσες από αυτές ξεκινάνε με “ΗΗ”;
(iii) Να βρεθεί η πιθανότητα μία λέξη να μην ξεκινάει με “ΗΗ”.

(i) 8 γράμματα ⟹ P(8,8) = 8! μεταθέσεις. Όμως έχουμε 2 "Η", άρα

8 7 6 5 4 3 2 1

Το πλήθος των λέξεων είναι $\frac{8!}{2!} = \frac{8\cdot7\cdot6\cdot5\cdot4\cdot3\cdot\cancel{2}\cdot1}{\cancel{2}} = 20.160$.

_ _ H_1 _ _ H_2 _

_ _ H_2 _ _ H_1 _

Επίσης έχουμε 2 "Τ", άρα τελικά:

Πλήθος: $\frac{8!}{2!\cdot2!} = \frac{20.160}{2} = 10.080$.

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

(ii) H H _ _ _ _ _ _ : $\frac{6!}{2!} = \frac{6\cdot5\cdot4\cdot3\cdot\cancel{2}}{\cancel{2}} = 360$. P("όχι ΗΗ στην αρχή") = $1 - \frac{360}{10.080}$.

άλυτη

Ασκήσεις στις διατάξεις

4. Δέκα φοιτητές τοποθετούνται στη σειρά. Ποια είναι η πιθανότητα, οι δύο νεότεροι να είναι δίπλα – δίπλα;

Υπόδειξη: Υπολογίστε το πλήθος των δυνατών τοποθετήσεων των 2 νεότερων στις 10 θέσεις και μετά για κάθε μία από αυτές τις τοποθετήσεις, υπολογίστε το πλήθος των συνδυασμών για τους υπόλοιπους φοιτητές.

άλυτη

Ασκήσεις στις διατάξεις

5. Έχουμε 3 βιβλία, 1 με μπλε, 1 με κόκκινο και 1 με πράσινό εξώφυλλο και μία βιβλιοθήκη με 4 διαφορετικά ράφια.

(α) Σε κάθε ράφι τοποθετούνται όσα βιβλία θέλουμε. Να βρεθεί το πλήθος των διαφορετικών τοποθετήσεων των 3 βιβλίων στα 4 ράφια.

(β) Σε κάθε ράφι μπορεί να τοποθετηθεί ακριβώς ένα βιβλίο. Να βρεθεί το πλήθος των διαφορετικών τοποθετήσεων των 3 βιβλίων στα 4 ράφια.

άλυτη

Ασκήσεις στις διατάξεις

6. (α) Να βρεθεί το πλήθος των τρόπων με τους οποίους μπορούν να μείνουν 5 άτομα σε 5 μονόκλινα δωμάτια ενός ξενοδοχείου.
(β) Να βρεθεί το πλήθος των τρόπων με τους οποίους μπορούν να μείνουν 3 άτομα σε 5 μονόκλινα δωμάτια ενός ξενοδοχείου.

(α) $P(5,5)=5!$: $5\cdot4\cdot3\cdot2\cdot1=120$.

(β) $5\cdot4\cdot3$: $P(5,3)=60$

άλυτη

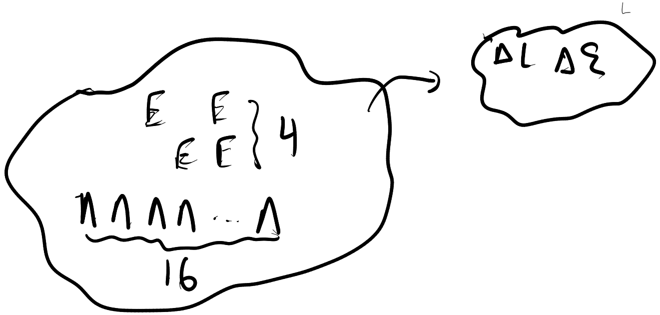
Συνδυασμοί

Ασκήσεις στους συνδυασμούς

1. Ένα δέμα με 20 σκληρούς δίσκους παρελήφθη από κατάστημα ηλεκτρονικών υπολογιστών. Τέσσερις από τους δίσκους βρέθηκε να είναι ελαττωματικοί. Ένα δείγμα 2 δίσκων επιλέγεται τυχαία.

(α) Με πόσους τρόπους μπορεί να ληφθεί το δείγμα των 2 δίσκων;

(β) Ποια είναι η πιθανότητα κανένας από τους δίσκους να μην είναι ελαττωματικός;



(α) $C(20,2) = \binom{20}{2} = \dfrac{20!}{2!\,(20-2)!} = \dfrac{20!}{2\cdot 18!} = \dfrac{20\cdot 19}{2} = 190$

(β) Ευνοϊκές: $C(16,2) = \binom{16}{2} = \dfrac{16!}{2!\cdot 14!} = \dfrac{16\cdot 15}{2} = 120$

$P(\text{"κανένας ελατ."}) = \dfrac{120}{190}$.

άλυτη

Ασκήσεις στους συνδυασμούς

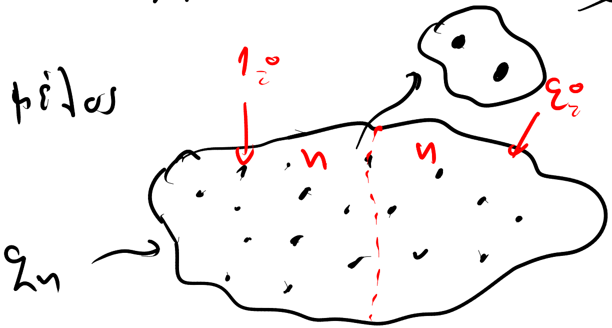
2. Να αποδείξετε την ταυτότητα $\binom{2n}{2} = 2\binom{n}{2} + n^2, \ n \geq 2.$
(α) αλγεβρικά
(β) με επιχειρήματα συνδυαστικής.

(α) $$\binom{2n}{2} = \frac{(2n)!}{2!\cdot(2n-2)!} = \frac{\cancel{2}n\cdot(2n-1)\cdot\cancel{(2n-2)!}}{\cancel{2}\cdot\cancel{(2n-2)!}} = n\cdot(2n-1) = 2n^2-n.$$

$$2\binom{n}{2}+n^2 = \cancel{2}\cdot\frac{n!}{\cancel{2!}\cdot(n-2)!}+n^2 = \frac{n\cdot(n-1)\cdot\cancel{(n-2)!}}{\cancel{(n-2)!}}+n^2 = n^2-n+n^2 = 2n^2-n.$$

(β) $C(2n,2)$: 1ο μέλος

$\binom{2n}{2}$



$n\cdot n$

Περιπτώσεις: 1 από 1ο + 1 από 2ο : n^2

2 από 1ο + 0 από 2ο : $C(n,2)=\binom{n}{2}$

0 από 1ο + 2 από 2ο : $C(n,2)=\binom{n}{2}$

άλυτη

Ασκήσεις στους συνδυασμούς

3. (α) Να αποδείξετε ότι $\binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{n} = 2^n$.

Υπόδειξη: Χρησιμοποιείστε το διώνυμο του Νεύτωνα

(β) Αποδείξτε ότι το πλήθος των υποσυνόλων ενός συνόλου με n στοιχεία είναι 2^n.
Σημείωση: το κενό θεωρείται και αυτό ένα υποσύνολο.

άλυτη

Ασκήσεις στους συνδυασμούς

4. Στο παιχνίδι Bridge ο κάθε παίκτης παίρνει από 13 χαρτιά.
(α) Πόσες δυνατές 13άδες υπάρχουν;

Π1 : 13άδα

(β) Ποια είναι η πιθανότητα σε δύο συνεχόμενα παιχνίδια ένας παίκτης να πάρει ακριβώς την ίδια 13άδα;

(α) 52 φύλλα. Επιλέγω 13. $C(52, 13) = \binom{52}{13} = \frac{52!}{13! \cdot 39!} = 635.013.559.600$

Συνολικές 13άδες

(β) Ευνοϊκές: 1 $\quad P(\text{"ίδια 13άδα"}) = \frac{1}{635.013.559.600}$.

άλυτη

Άσκηση αντιστοίχισης

5. Να αντιστοιχίσετε τους τύπους με τα προβλήματα των οποίων αποτελούν λύση.

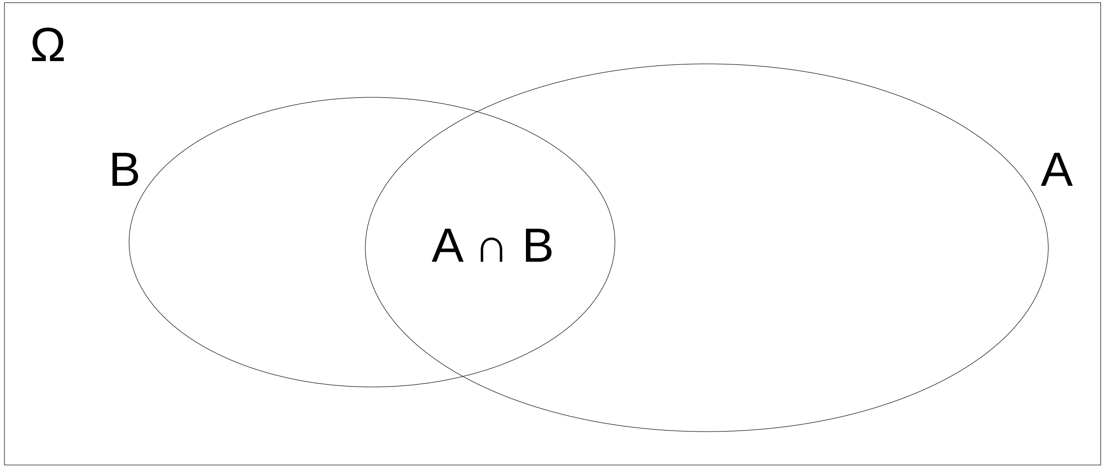
3^5	Πλήθος τοποθετήσεων 3 διαφορετικών σφαιρών σε 5 διαφορετικούς κάδους.
	Πλήθος λέξεων με 3 διαφορετικά γράμματα από το σύνολο {α, β, γ, δ, ε}
5^3	Πλήθος λέξεων με 3 γράμματα (όχι κατ' ανάγκη διαφορετικά) από το σύνολο {α, β, γ, δ, ε}
	Πλήθος διατάξεων από 5 ίδιες κόκκινες μπάλες και 3 ίδιες μπλε μπάλες.
P(5, 3)	Πλήθος δυνατών υποσυνόλων του A, όπου N(A) = 5·3.
	Πλήθος συναρτήσεων από το A στο B, όπου N(A) = 5 και N(B) = 3.
$2^{5\cdot 3}$	Πλήθος απεικονίσεων από το A στο B, όπου N(A) = 5 και N(B) = 3.

Δεσμευμένη Πιθανότητα

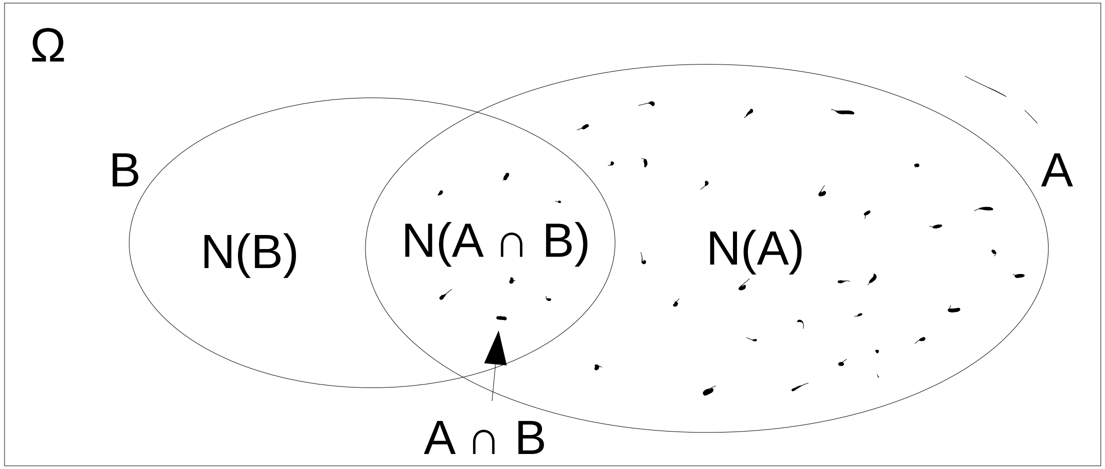
Ορισμός

Η πιθανότητα να συμβεί το ενδεχόμενο Β *γνωρίζοντας ότι έχει συμβεί* το ενδεχόμενο Α συμβολίζεται με **P(B | A)** και ονομάζεται **δεσμευμένη πιθανότητα του Β δεδομένου του Α.**

Από τον ορισμό της, η δεσμευμένη πιθανότητα P(B | A) δηλώνει την **πιθανότητα να συμβεί το Β αν το σύνολο Α θεωρηθεί ως ο δειγματοχώρος όλων των εφικτών ενδεχομένων.**



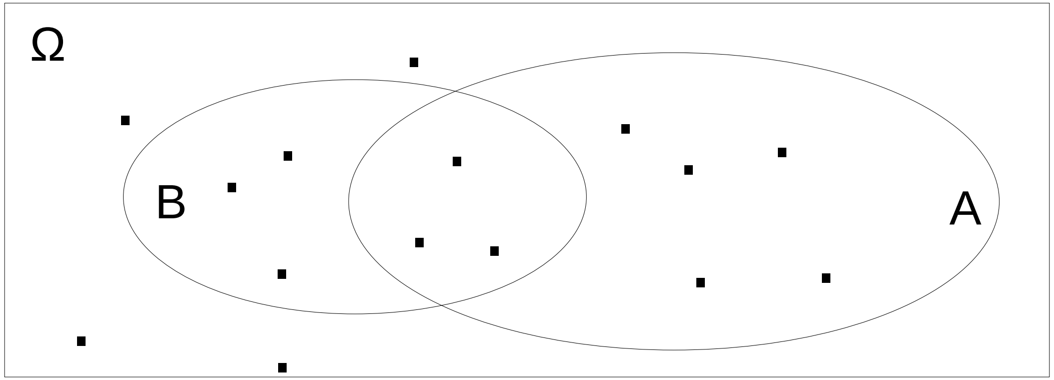
Δεσμευμένη Πιθανότητα



Από τον ορισμό της δεσμευμένης πιθανότητας, υπολογίζουμε:

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

Δεσμευμένη Πιθανότητα



Παράδειγμα
Για το δειγματοχώρο του σχήματος, η δεσμευμένη πιθανότητα του B δεδομένου του A, είναι $P(B \mid A) = \frac{N(B \cap A)}{N(A)} = \frac{3}{8}$ ενώ η δεσμευμένη πιθανότητα του A δεδομένου του B, είναι $P(A \mid B) = \frac{N(A \cap B)}{N(B)} = \frac{3}{6} = 0{,}5.$

Δεσμευμένη Πιθανότητα

Συνοψίζοντας,

– Δεσμευμένη πιθανότητα P(B | A) ονομάζεται η πιθανότητα να συμβεί το ενδεχόμενο B *γνωρίζοντας ότι έχει συμβεί* το ενδεχόμενο A.

– Η δεσμευμένη πιθανότητα υπολογίζεται ως:

$$\mathrm{P}\left(\mathrm{B}|\mathrm{A}\right)=\frac{\mathrm{P}\left(\mathrm{B}\cap\mathrm{A}\right)}{\mathrm{P}\left(\mathrm{A}\right)},$$

από την οποία προκύπτει ισοδύναμα, P(B ∩ A) = P(A) · P(B | A).

– Φανερά,

$$\mathbf{P(B \mid A) \neq P(A \mid B).}$$

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

Δεσμευμένη Πιθανότητα

Από τη σχέση $P(B \cap A) = P(A) \cdot P(B \mid A)$, με αναδρομικό συλλογισμό, βρίσκουμε πως για οποιαδήποτε ενδεχόμενα $A_1, A_2, \ldots, A_n$ ισχύει

- $P(A_2 \cap A_1) = P(A_1) \cdot P(A_2 \mid A_1)$,
- $P(A_3 \cap A_2 \cap A_1) = P(A_2 \cap A_1) \cdot P(A_3 \mid A_2 \cap A_1) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_2 \cap A_1)$,

… και γενικότερα

- $P(\cap_{n \in N} A_n) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_2 \cap A_1) \cdot \ldots \cdot P(A_n \mid A_{n-1} \cap \ldots \cap A_2 \cap A_1)$

Δεσμευμένη Πιθανότητα

Από τη σχέση P(B ∩ A) = P(B | A) · P(A) συνάγουμε ότι αν Γ είναι ένα ενδεχόμενο με μη μηδενική πιθανότητα, τότε

$$P(B \cap A \mid \Gamma) = P(B \mid A \cap \Gamma) \cdot P(A \mid \Gamma).$$

Η σχέση P(B ∩ A | Γ) = P(B | A ∩ Γ) · P(A | Γ), περιγράφει την παρατήρηση πως οποιοδήποτε αποτέλεσμα πιθανότητας ισχύει για την άνευ όρων πιθανότητα, παραμένει αληθές εάν όλα εξαρτηθούν από κάποιο γεγονός.

Απόδειξη

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

Σημείωση
Η απόδειξη μπορεί να γίνει και από ένα διάγραμμα Venn.

Ασκήσεις στη δεσμευμένη πιθανότητα

1. Ρίχνουμε ένα ζάρι και ορίζουμε Α = {το αποτέλεσμα είναι άρτιος αριθμός} και Β = {το αποτέλεσμα είναι μεγαλύτερο ή ίσο του 5}. Να βρεθεί η P(Β | Α).

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

$\Omega = \{1,2,3,4,5,6\}$ $\quad A\cap B = \{6\}$

$A = \{2,4,6\}$, $B = \{5,6\}$

$P(A) = \frac{3}{6}$, $P(B\cap A) = \frac{1}{6}$. $\quad P(B|A) = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}$.

άλυτη

Ασκήσεις στη δεσμευμένη πιθανότητα

2. Ρίχνουμε ένα ζάρι και ορίζουμε Α = {άρτιος αριθμός}, Β = {περιττός αριθμός} και Γ = {4 ή 6}. Να βρεθούν οι P(B | A) και P(A | Γ).

άλυτη

Ασκήσεις στη δεσμευμένη πιθανότητα

3. Σε τρεις ρίψεις νομίσματος βγήκε δύο φορές “Γράμματα”. Βρείτε την δεσμευμένη πιθανότητα να βγήκε στην δεύτερη ρίψη “Γράμματα”.

$A = \{$ "έχουν βγει 2 Γ" $\}$ $\quad B = \{$ "στη 2η ρίψη είναι Γ" $\}$. $\quad P(B|A) = \dfrac{P(B\cap A)}{P(A)}$

$\underline{\Omega} = \{ KKK, KKΓ, KΓK, ΓKK, \underbrace{\overset{B}{KΓΓ}}_{A}, \underbrace{ΓKΓ}_{A}, \underbrace{\overset{B}{ΓΓK}}_{A}, ΓΓΓ \}$

$P(A) = \dfrac{3}{8}$. $\quad P(B\cap A) = \dfrac{2}{8}$

$P(B|A) = \dfrac{\frac{2}{8}}{\frac{3}{8}} = \dfrac{2}{3}$.

άλυτη

Ασκήσεις στη δεσμευμένη πιθανότητα

4. Σε 7 κάρτες είναι γραμμένα τα γράμματα « Λ, Λ, Ο, Ο, Ο, Τ, Τ». Αφού τις ανακατέψουμε διαλέγουμε τυχαία πέντε και τις βάζουμε στη σειρά. Βρείτε την πιθανότητα να σχηματισθεί η λέξη ΛΟΤΤΟ.

Υπόδειξη: $P(\cap_{n \in N} A_n) = P(A_1) \cdot P(A_2 \,|\, A_1) \cdot P(A_3 \,|\, A_2 A_1) \cdot \ldots \cdot P(A_n \,|\, A_{n-1} \ldots A_2 A_1)$

άλυτη

Ασκήσεις στη δεσμευμένη πιθανότητα

5. Το 51% των κατοίκων μιας συγκεκριμένης περιοχής είναι άνδρες (και το 49% γυναίκες). Από προηγούμενη έρευνα γνωρίζουμε ότι το 4,2% των κατοίκων αυτής της περιοχής πάσχει από αχρωματοψία. Επίσης από την ίδια έρευνα γνωρίζουμε ότι 4% των κατοίκων της περιοχής αυτής είναι άνδρες που πάσχουν από αχρωματοψία.

(α) Θεωρούμε τα ενδεχόμενα A = {το άτομο πάσχει από αχρωματοψία} και B = {το άτομο είναι άνδρας}. Να βρείτε τις πιθανότητες P(A), P(B), P(A ∩ B).

(β) Αν επιλέξουμε τυχαία ένα άτομο από αυτή την περιοχή και είναι άνδρας, ποια είναι η πιθανότητα να πάσχει από αχρωματοψία.

(α) $P(A) = 0{,}042$, $P(B) = 0{,}51$

$P(A \cap B) = 0{,}04$

(β) $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{0{,}04}{0{,}51} = \dfrac{4}{51}$

Ασκήσεις στη δεσμευμένη πιθανότητα

5. Το 51% των κατοίκων μιας συγκεκριμένης περιοχής είναι άνδρες (και το 49% γυναίκες). Από προηγούμενη έρευνα γνωρίζουμε ότι το 4,2% των κατοίκων αυτής της περιοχής πάσχει από αχρωματοψία. Επίσης από την ίδια έρευνα γνωρίζουμε ότι 4% των κατοίκων της περιοχής αυτής είναι άνδρες που πάσχουν από αχρωματοψία.

(α) Θεωρούμε τα ενδεχόμενα A = {το άτομο πάσχει από αχρωματοψία} και
B = {το άτομο είναι άνδρας}. Να βρείτε τις πιθανότητες P(A), P(B), P(A ∩ B).

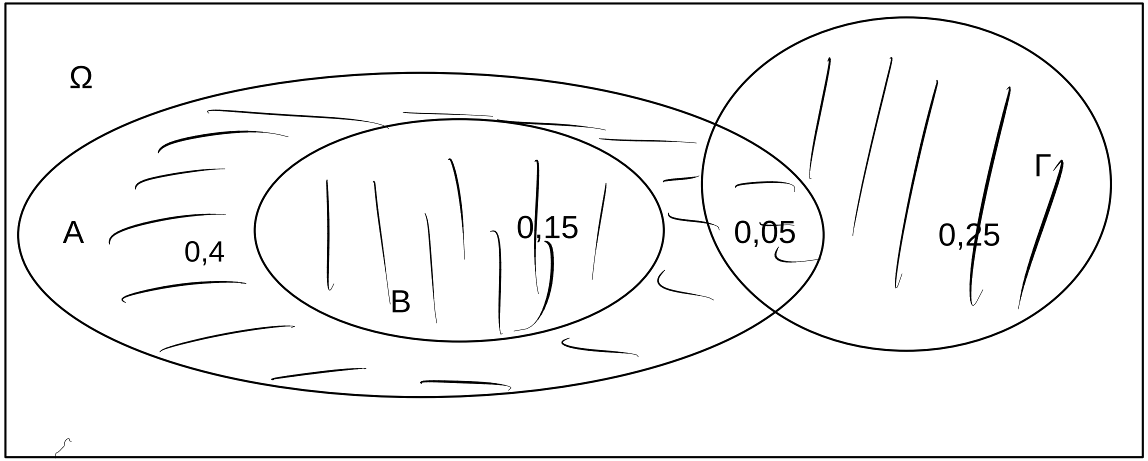
(β) Αν επιλέξουμε τυχαία ένα άτομο από αυτή την περιοχή και είναι άνδρας, ποια είναι η πιθανότητα να πάσχει από αχρωματοψία.

Λύση
(α) $P(A) = 0,042$, $P(B) = 0,51$, $P(A \cap B) = 0,04$.

(β) $P(A \mid B) = P(A \cap B) / P(B) = 0,04 / 0,51 = 0,078 = 7,8\%$.

Ασκήσεις στη δεσμευμένη πιθανότητα

6. Για τα ενδεχόμενα Α, Β, Γ του διαγράμματος, γνωρίζουμε ότι Ρ(Α – (ΒUΓ)) = 0,4, Ρ(Γ – Α) = 0,25, Ρ(Β) = 0,15, Ρ(Α ∩ Γ) = 0,05.



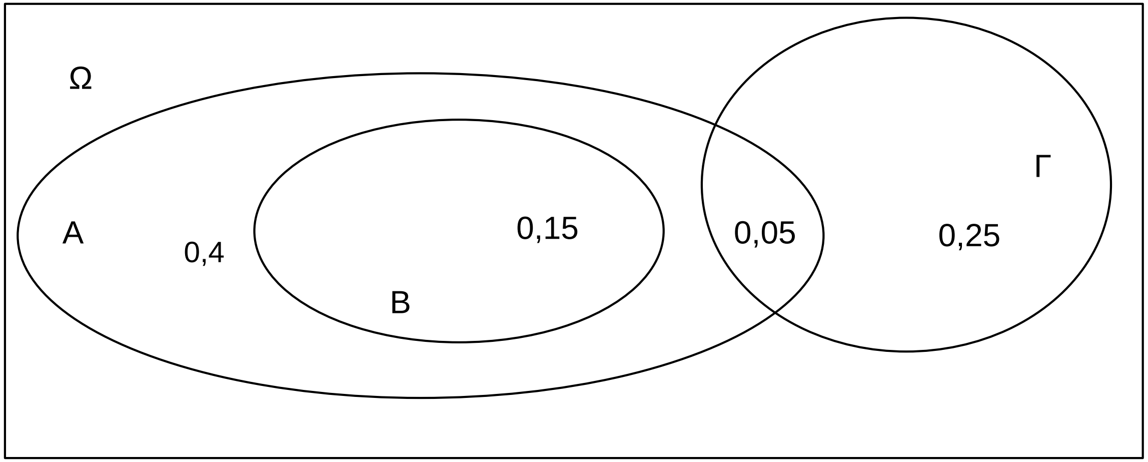
Να βρεθούν οι πιθανότητες

$P(A) = 0,4 + 0,15 + 0,05 = 0,6$

(α) Ρ(Α), Ρ(Γ), Ρ((Α U Γ)’)

(β) Ρ(Α | Β), Ρ(Β | Α), Ρ(Γ | Α), Ρ(Α | Γ).

Ασκήσεις στη δεσμευμένη πιθανότητα



Λύση

(α) $P(A) = 0,4 + 0,15 + 0,05 = 0,6$

$P(\Gamma) = 0,25 + 0,05 = 0,3$

$P((A \cup \Gamma)') = 1 – P(A \cup \Gamma)$

$= 1 – 0,85$

$= 0,15.$

(β) $P(A \mid B) = P(A \cap B) / P(B) = 0,15 / 0,15 = 1.$

$P(B \mid A) = P(B \cap A) / P(A) = 0,15 / 0,6 = 0,25.$

$P(\Gamma \mid A) = P(\Gamma \cap A) / P(A) = 0,05 / 0,6 = 0,083.$

$P(A \mid \Gamma) = P(A \cap \Gamma) / P(\Gamma) = 0,05 / 0,3 = 0,167.$

Δεσμευμένη Πιθανότητα και Ανεξαρτησία

Γνωρίζουμε ότι δύο γεγονότα ονομάζονται **ανεξάρτητα** αν

$$\mathbf{P(B \cap A) = P(B) \cdot P(A).}$$

Η ανεξαρτησία δύο γεγονότων εκφράζεται και με όρους δεσμευμένης πιθανότητας όπως δείχνει η παρακάτω πρόταση.

Πρόταση
Τα γεγονότα A και B είναι **ανεξάρτητα** αν και μόνο αν $\mathbf{P(B \mid A) = P(B)}$.

Απόδειξη
Πράγματι, από τη σχέση ορισμού $P(B \cap A) = P(A) \cdot P(B \mid A)$ παίρνουμε
$P(B) \cdot P(A) = P(A) \cdot P(B \mid A)$ ή $P(B \mid A) = P(B)$.

Ανεξάρτητα vs Ξένα ενδεχόμενα (updated)

Ανεξάρτητα ενδεχόμενα

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

$P(A \cup B) = P(A) + P(B) – P(A \cdot B)$

A, B ανεξάρτητα → A’, B’ ανεξάρτητα.

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

Ξένα ενδεχόμενα

$P(A \cap B) = 0$

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

A, B ξένα → A’, B’ όχι ξένα.

$P(B \mid A) = 0 \neq P(B)$

Ανεξαρτησία δεδομένου τρίτου γεγονότος

Ορισμός
Τα γεγονότα Α και Β λέγονται **ανεξάρτητα, δεδομένου ενός άλλου γεγονότος Γ**, αν

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

Αν τα Α, Β είναι ανεξάρτητα μεταξύ τους τότε θα είναι και ανεξάρτητα δεδομένου οποιουδήποτε άλλου μη κενού γεγονότος Γ.

Το αντίθετο όμως δεν ισχύει, δηλαδή τα Α, Β μπορεί να είναι ανεξάρτητα δεδομένου κάποιου τρίτου γεγονότος Γ χωρίς να είναι ανεξάρτητα μεταξύ τους.

$P(B \mid A \cap \Gamma) = P(B \mid \Gamma) \not\Rightarrow P(B \mid A) = P(B)$

Αν τα Α, Β είναι ανεξάρτητα δεδομένου του Γ, δεν είναι οπωσδήποτε ανεξάρτητα μεταξύ τους.

Παράδειγμα
Έστω $\Omega = \{ (X_1, X_2, X_3), X_1, X_2, X_3 = 0, 1, 2$ ή $3\}$. Παρατηρούμε (πως;) ότι $N(\Omega) = 64$.

Θεωρούμε τα γεγονότα $A = \{X_1 = 2\}$, $B = \{X_1 + X_2 + X_3 = 3\}$.

Υπολογίζουμε (πως;) ότι

$$P(A) = 16 / 64,\ P(B) = 10 / 64,\ P(A \cap B) = 2 / 64,\ P(B \mid A) = 2 / 16).$$

Άρα, $P(B \mid A) = 2 / 16 \neq 10 / 64 = P(B)$, δηλαδή, **τα Α και Β δεν είναι ανεξάρτητα.**

Σημείωση
Τους αριθμούς X_1, X_2, X_3, που λαμβάνουν με τυχαίο τρόπο τις τιμές 0, 1, 2, 3 στο μέλλον θα τους ονομάσουμε διακριτές τυχαίες μεταβλητές.

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

Έστω τώρα το γεγονός $\Gamma = \{X_1 + X_2 = 2\}$.

Είναι $A \cap \Gamma = \{X_1 = 2 \text{ και } X_1 + X_2 = 2\} = \{X_1 = 2 \text{ και } X_2 = 0\}$, από όπου εύκολα βρίσκουμε

- $B \mid A \cap \Gamma = \{X_1 + X_2 + X_3 = 3 \mid X_1 = 2 \text{ και } X_2 = 0\}$ και
- $B \mid \Gamma = \{X_1 + X_2 + X_3 = 3 \mid X_1 + X_2 = 2\}$,

Προκύπτει (πως;) ότι

$P(B \mid A \cap \Gamma) = P(B \mid \Gamma) = 1 / 64$, συνεπώς $P(B \mid A \cap \Gamma) = P(B \mid \Gamma)$, δηλαδή

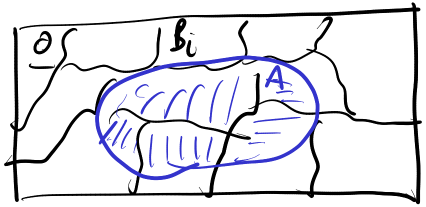
τα Α και Β είναι ανεξάρτητα δεδομένου του Γ.

Νόμος της Ολικής Πιθανότητας (updated)

Νόμος της Ολικής Πιθανότητας

Έστω B_i, i = 1, 2, … είναι μία διαμέριση του δειγματοχώρου Ω, που αποτελείται από ξένα μεταξύ τους σύνολα, δηλαδή

$$\bigcup_{n=1}^{\infty} B_n = \Omega \text{ και } B_i \cap B_j = \emptyset \,,\ i \neq j.$$



Αν $A \subseteq \Omega$, ένα οποιοδήποτε ενδεχόμενο, τότε σύμφωνα με το Νόμο της Ολικής Πιθανότητας:

$$P(A) = \sum_{n=1}^{\infty} P(A \cap B_n)$$

Σε όρους δεσμευμένης πιθανότητας, ο παραπάνω τύπος παίρνει τη μορφή

$$\boxed{P(A) = \sum_{n=1}^{\infty} P(A \mid B_n) \cdot P(B_n)}$$

Σημείωση: Ο Νόμος της Ολικής Πιθανότητας ισχύει και στην περίπτωση κατά την οποία $A \subset \bigcup_{n=1}^{\infty} B_n \subset \Omega$.

Ασκήσεις στο Νόμο της Ολικής Πιθανότητας

1. Οι ηλεκτρικοί λαμπτήρες προωθούνται στην αγορά συσκευασμένοι σε χαρτοκιβώτια των 25 λαμπτήρων. Ας υποθέσουμε ότι από ένα χαρτοκιβώτιο που περιέχει 3 ελαττωματικούς λαμπτήρες εξάγουμε 2 λαμπτήρες.
Να υπολογισθούν οι πιθανότητες των ενδεχομένων Α και Β εξαγωγής ελαττωματικού λαμπτήρα κατά την πρώτη και δεύτερη εξαγωγή αντίστοιχα, αν η εξαγωγή γίνει
(α) με επανάθεση του λαμπτήρα στο χαρτοκιβώτιο
(β) χωρίς επανάθεση.

(α) 3 Ε, 22 Λ, 2

1η εξαγωγή: $P(A) = \frac{3}{25}$, 2η εξαγωγή: $P(B) = \frac{3}{25}$

Ν.Ο.Π.

$$P(B) = P(B|E_1)\cdot P(E_1) + P(B|E_1')\cdot P(E_1') = \frac{2}{24}\cdot\frac{3}{25} + \frac{3}{24}\cdot\frac{22}{25}$$

(β) 1η εξαγωγή: $P(A) = \frac{3}{25}$.

$A = E_1 = \{$ο 1ος είναι Ε$\}$

$$B = (B \cap E_1) \cup (B \cap E_1')$$

$$= \frac{72}{24\cdot 25} = 0{,}12 = 12\%.$$

Ω E_1 E_1'

Ασκήσεις στο Νόμο της Ολικής Πιθανότητας

Λύση
A = {εξαγωγή ελαττωματικού λαμπτήρα κατά την πρώτη εξαγωγή}
B = {εξαγωγή ελαττωματικού λαμπτήρα κατά την δεύτερη εξαγωγή}

$$P(A) = 3 / 25 = 0,12, \; P(A') = 22 / 25 = 0,88.$$

Υπολογίζουμε τις δεσμευμένες πιθανότητες για τις δύο περιπτώσεις
(α) επανάθεση του 1ου λαμπτήρα στο χαρτοκιβώτιο: $P(B \mid A) = P(B \mid A') = 3 / 25 = 0,12$.
(β) χωρίς επανάθεση του 1ου λαμπτήρα: $P(B \mid A) = 2 / 24 = 0,083$, $P(B \mid A') = 3 / 24 = 0,125$.

Από το νόμο της ολικής πιθανότητας υπολογίζουμε:
(α) Με επανάθεση:

$$P(B) = P(B \mid A) \cdot P(A) + P(B \mid A') \cdot P(A') = 0,12 \cdot 0,12 + 0,12 \cdot 0,88 = 0,12.$$

(β) Χωρίς επανάθεση:

$$P(B) = P(B \mid A) \cdot P(A) + P(B \mid A') \cdot P(A') = 0,083 \cdot 0,12 + 0,125 \cdot 0,88 = 0,21.$$

Ασκήσεις στο Νόμο της Ολικής Πιθανότητας

2. Σωματίδιο περνάει ανάμεσα από τρεις καταμετρητές, όπου μπορεί να πέσει σε καθέναν από αυτούς με πιθανότητες 0.3, 0.2 και 0.4 αντίστοιχα. Εάν πέσει στον πρώτο καταμετρητή, τότε καταγράφεται με πιθανότητα 0.6, εάν πέσει στον δεύτερο καταγράφεται με πιθανότητα 0.5 και στον τρίτο με πιθανότητα 0.55. Βρείτε την πιθανότητα καταγραφής του σωματιδίου.

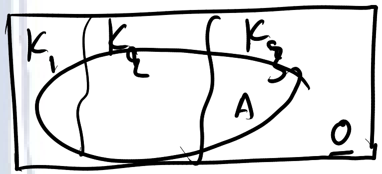
$P(K_1)=0{,}3 \quad P(K_2)=0{,}2 \quad P(K_3)=0{,}4$

$A=\{$το σωματίδιο καταγράφεται$\}$

$P(A|K_1)=0{,}6\ ,\ P(A|K_2)=0{,}5\ ,\ P(A|K_3)=0{,}55\ .$

$$P(A)=P(A|K_1)\cdot P(K_1)+P(A|K_2)P(K_2)+P(A|K_3)\cdot P(K_3)=0{,}6\cdot 0{,}3+0{,}5\cdot 0{,}2+0{,}55\cdot 0{,}4$$

$$=0{,}18+0{,}1+0{,}22=0{,}5=50\%.$$



άλυτη

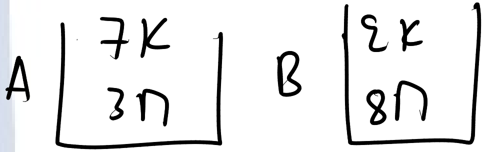
$$P(\Pi) = P((\Pi \cap A) \cup (\Pi \cap B))$$

ή

Ασκήσεις στο Νόμο της Ολικής Πιθανότητας

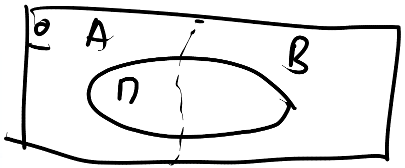
3. Έχουμε μπροστά μας δύο σάκους τον Α και τον Β. Ο Α έχει 7 κόκκινους και 3 πράσινους βόλους, ενώ ο Β έχει 2 κόκκινους και 8 πράσινους βόλους. Διαλέγουμε στην τύχη ένα σάκο και επιλέγουμε τυχαία ένα βόλο. Να βρεθεί η πιθανότητα αυτός να είναι πράσινος.

Ν.Ο.Π.



$$P(\Pi) = P(\Pi|A)\cdot P(A) + P(\Pi|B)\cdot P(B) =$$

$$= \frac{3}{10}\cdot 0,5 + \frac{8}{10}\cdot 0,5 = 0,15 + 0,4 = 0,55 = 55\%$$



άλυτη

Ασκήσεις στο Νόμο της Ολικής Πιθανότητας

4. Ένα κατάστημα προμηθεύεται τα προϊόντα που εμπορεύεται από δύο εταιρείες, την Α και την Β. Η εταιρεία Α προμηθεύει το 80% των προϊόντων και από αυτά μόνο το 1% αποδεικνύεται ελαττωματικό. Η εταιρεία Β προμηθεύει το υπόλοιπο 20% των προϊόντων και το 3% αυτών αποδεικνύεται ότι είναι ελαττωματικά.
Εάν ένας πελάτης αγοράσει από το κατάστημα τυχαία ένα προϊόν, ποια είναι η πιθανότητα αυτό να είναι ελαττωματικό;

Υπόδειξη
Η επιλογή γίνεται σε δύο φάσεις. Στην πρώτη επιλέγεται τυχαία η εταιρεία και στη δεύτερη το προϊόν.

$E = \{$ ελαττωματικό προϊόν $\}$. $P(A) = 0{,}8$. $P(E|A) = 0{,}01$.

$P(B) = 0{,}2$. $P(E|B) = 0{,}03$

Ν.Ο.Π. $P(E) = P(E|A) \cdot P(A) + P(E|B) \cdot P(B) = 0{,}01 \cdot 0{,}8 + 0{,}03 \cdot 0{,}2 = 0{,}008 + 0{,}006 = 0{,}014 =$

$= 1{,}4\%$.

άλυτη

Συμπλήρωμα και δεσμευμένη πιθανότητα

Παρατήρηση
Γνωρίζουμε ότι $P(A') = 1 – P(A)$. Μία χρήσιμη παρατήρηση είναι πως το ίδιο ισχύει όταν το A δεσμευτεί ως προς ένα άλλο ενδεχόμενο B:

$$\mathbf{P(A' \mid B) = 1 – P(A \mid B).}$$

Η απόδειξη είναι άμεση από τον ορισμό της δεσμευμένης πιθανότητας:

$$P(A' \mid B) = P(A' \cap B) / P(B) = P(B – A) / P(B) = [P(B) – P(A \cap B)] / P(B) = 1 – P(A \mid B).$$

Σημείωση
Η παραπάνω ισότητα δεν ισχύει όταν το συμπλήρωμα είναι στο B, δηλαδή η σχέση $P(A \mid B') = 1 – P(A \mid B)$, <u>δεν είναι σωστή</u>.

Θεώρημα (ή Τύπος) Bayes

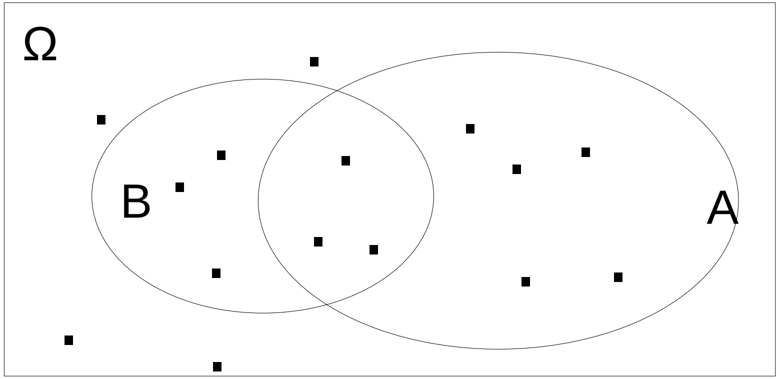
Θεώρημα (ή Τύπος) Bayes

Θεώρημα ή Κανόνας του Bayes
Για κάθε ενδεχόμενα A, B με P(A), P(B) $\neq$ 0, ισχύει $P(B \mid A) = \dfrac{P(A \mid B) \cdot P(B)}{P(A)}$.

Απόδειξη
Η απόδειξη είναι άμεση: $P(B \mid A) = \dfrac{P(B \cap A)}{P(A)} = \dfrac{P(A \mid B) \cdot P(B)}{P(A)}$.

Παράδειγμα
Για τα ενδεχόμενα A και B του διαγράμματος, είναι P(A) = 8 / 15, P(B) = 6 / 15, P(A | B) = 3 / 6, άρα

$$P(B \mid A) = \frac{P(A \mid B) \cdot P(B)}{P(A)} = \frac{\frac{3}{6} \cdot \frac{6}{15}}{\frac{8}{15}} = \frac{3}{8}.$$



Σημείωση: Επαληθεύστε τον υπολογισμό από το διάγραμμα!

Θεώρημα (ή Τύπος) Bayes

Αν B_i, i = 1, 2, … μία διαμέριση του Ω ξένων μεταξύ τους ενδεχομένων, τότε για κάθε ενδεχόμενο A είναι

$$P(A) = \sum_{n=1}^{\infty} P(A \mid B_n) \cdot P(B_n)$$ Ν.Ο.Π.

Στην περίπτωση αυτή, ο κανόνας του Bayes γράφεται ως:

$$P(B \mid A) = \frac{P(A \mid B) \cdot P(B)}{P(A)} = \frac{P(A \mid B) \cdot P(B)}{\sum_{n=1}^{\infty} P(A \mid B_n) \cdot P(B_n)}.$$

Ιδιαίτερα, καθώς για κάθε B είναι Ω = B U B’, είναι

$$P(B \mid A) = \frac{P(A \mid B) \cdot P(B)}{P(A \mid B) \cdot P(B) + P(A \mid B') \cdot P(B')}.$$

Ασκήσεις στον τύπο του Bayes

1. Σε ένα νησί φτάνουν καθημερινά πλοία, που αναχωρούν από Πειραιά και Ραφήνα και σε ποσοστά 60% και 40% αντιστοίχως. Το 10% των πλοίων από Πειραιά και το 5% των πλοίων από Ραφήνα φθάνουν με καθυστέρηση στο νησί. Αν μια μέρα επιλέξουμε τυχαία ένα πλοίο που φτάνει στο νησί, να βρεθούν οι πιθανότητες:
i) Να φτάνει με καθυστέρηση
ii) Αν φτάνει με καθυστέρηση, να έρχεται από Πειραιά.

Λύση
$\Omega = \{B_1 =$ “Φτάνει από Πειραιά”, $B_2 =$ “Φτάνει από Ραφήνα”$\}$ και $A =$ “Φτάνει με καθυστέρηση”.

(i) Από το νόμο της ολικής πιθανότητας υπολογίζουμε:

$$P(A) = P(A \mid B_1) \cdot P(B_1) + P(A \mid B_2) \cdot P(B_2) = 0{,}1 \cdot 0{,}6 + 0{,}05 \cdot 0{,}4 = 0{,}06 + 0{,}02 = 0{,}08.$$

(ii) Από τον κανόνα του Bayes, υπολογίζουμε:

$$P(B_1 \mid A) = \frac{P(A \mid B_1) \cdot P(B_1)}{P(A)} = \frac{0{,}1 \cdot 0{,}6}{0{,}08} = \frac{3}{4} = 0{,}75.$$

Ασκήσεις στον τύπο του Bayes

2. Σε ένα εργοστάσιο το 60% των εργαζομένων είναι άνδρες και το 40% είναι γυναίκες. Από τους άνδρες καπνίζει το 50% και από τις γυναίκες το 30%.

Αν επιλέξουμε τυχαία ένα άτομο που καπνίζει, ποια η πιθανότητα να είναι γυναίκα;

Λύση
Ω = {Α = "Άνδρας", Γ = "Γυναίκα"} και Κ = "Καπνίζει".

Είναι Ρ(Α) = 0,6, Ρ(Γ) = 0,4, Ρ(Κ | Α) = 0,5, Ρ(Κ | Γ) = 0,3. Αναζητούμε την πιθανότητα Ρ(Γ | Κ).

Από το νόμο της ολικής πιθανότητας, υπολογίζουμε:

$$P(K) = P(K \mid A)\cdot P(A) + P(K \mid \Gamma)\cdot P(\Gamma) = 0{,}5\cdot 0{,}6 + 0{,}3\cdot 0{,}4 = 0{,}3 + 0{,}12 = 0{,}42.$$

Από τον κανόνα του Bayes, υπολογίζουμε:

$$P(\Gamma \mid K) = \frac{P(K \mid \Gamma)\cdot P(\Gamma)}{P(K)} = \frac{0{,}3\cdot 0{,}4}{0{,}42} = \frac{6}{21}.$$

Ασκήσεις στον τύπο του Bayes

3. Σε δείγματα ενός υλικού βρέθηκε το μετάλλευμα A σε ποσοστό 40% και το μετάλλευμα B σε ποσοστό 55%. Είναι γνωστό ότι η πιθανότητα να υπάρχει το B δεδομένης της ύπαρξης του A είναι 80%. Να βρεθεί η πιθανότητα, σε ένα δείγμα από το υλικό:

α) Να υπάρχει ένα τουλάχιστον από τα A και B.

β) Να μην υπάρχει κανένα από τα A, B.

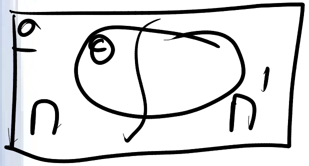
Λύση

(α) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$= P(A) + P(B) - P(B \mid A) \cdot P(A)$

$= 0{,}4 + 0{,}55 - 0{,}8 \cdot 0{,}4 = 0{,}63.$

(β) $P(A' \cap B') = P((A \cup B)') = 1 - P(A \cup B) = 1 - 0{,}63 = 0{,}37.$



Ασκήσεις στον τύπο του Bayes

4. Το 1‰ ενός πληθυσμού πάσχει από μια σοβαρή ασθένεια. Ένα καινούργιο τεστ διάγνωσης της ασθένειας έχει πιθανότητα θετικού σφάλματος (θετικό τεστ, ενώ το άτομο είναι υγιές) 1% και πιθανότητα αρνητικού σφάλματος (αρνητικό τεστ, ενώ το άτομο πάσχει από την ασθένεια) 5%.

Για ένα τυχαίο άτομο από τον πληθυσμό αυτό το τεστ είναι θετικό. Να βρείτε την πιθανότητα το άτομο να πάσχει πράγματι από την ασθένεια αυτή.

$P(\Pi) = 0{,}001$ $\quad$ $\Pi = \{$ το άτομο πάσχει $\}$. $\Theta = \{$ τεστ θετικό $\}$

$P(\Theta|\Pi') = 0{,}01$ $\quad$ $P(\Pi|\Theta) = \dfrac{P(\Theta|\Pi)\cdot P(\Pi)}{P(\Theta)} = \dfrac{0{,}95\cdot 0{,}001}{0{,}01094} = 0{,}087 = 8{,}7\%$

$P(\Theta'|\Pi) = 0{,}05$

Ν.Ο.Π : $P(\Theta) = P(\Theta|\Pi)\cdot P(\Pi) + P(\Theta|\Pi')\cdot P(\Pi') = 0{,}95\cdot 0{,}001 + 0{,}01\cdot 0{,}999$

$= 0{,}01094$

$P(\Theta|\Pi) = 1 - P(\Theta'|\Pi) = 1 - 0{,}05 = 0{,}95$

άλυτη

Ασκήσεις στον τύπο του Bayes

5. Γνωρίζουμε ότι 5% των εγκύων γυναικών που παρακολουθούνται από μία κλινική παρουσιάζουν βακτηριουρία. Επίσης είναι γνωστό ότι 30% των εγκύων γυναικών που παρουσιάζουν βακτηριουρία και 1% των εγκύων γυναικών που δεν παρουσιάζουν βακτηριουρία, πάσχουν από πυελονεφρίτιδα.

Να υπολογισθούν οι πιθανότητες όπως μία έγκυος γυναίκα που παρακολουθείται στην κλινική αυτή και προσέρχεται για προγραμματισμένη εξέταση
(α) παρουσιάσει βακτηριουρία και να πάσχει από πυελονεφρίτιδα,
(β) πάσχει από πυελονεφρίτιδα και
(γ) παρουσιάζει βακτηριουρία δεδομένου ότι πάσχει από πυελονεφρίτιδα.

$P(B) = 0{,}05$. $P(\Pi \mid B) = 0{,}3$. $P(\Pi \mid B') = 0{,}01$.

(α) $P(\Pi \cap B) = P(\Pi \mid B) \cdot P(B) = 0{,}3 \cdot 0{,}05 = 0{,}015 = 1{,}5\%$

(β) $P(\Pi) = P(\Pi \mid B) \cdot P(B) + P(\Pi \mid B') \cdot P(B') = 0{,}3 \cdot 0{,}05 + 0{,}01 \cdot 0{,}95 = 0{,}015 + 0{,}0095 =$
$= 0{,}0245 = 2{,}45\%$

(γ) $P(B \mid \Pi) = \dfrac{P(\Pi \mid B) \cdot P(B)}{P(\Pi)} = \dfrac{0{,}3 \cdot 0{,}05}{0{,}0245} = 0{,}184$
$= 18{,}4\%$.

άλυτη

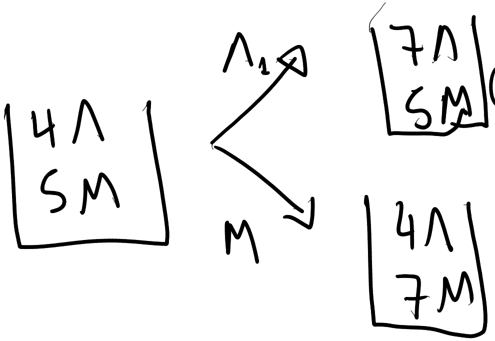
$\Lambda_1 = \{$η $1^η$ μπάλλα είναι $\Lambda\}$, $\Lambda_2 = \{$η $2^η$ μπάλλα είναι $\Lambda\}$

Ασκήσεις στον τύπο του Bayes

6. Μία κάλπη περιέχει 4 λευκές και 5 μαύρες σφαίρες. Παίρνουμε μία σφαίρα και αν είναι λευκή την ξαναβάζουμε με άλλες τρεις λευκές ενώ αν είναι μαύρη την ξαναβάζουμε με άλλες δύο μαύρες.

εξάγουμε

α) Ποια η πιθανότητα ώστε να ~~είναι~~ λευκή με την δεύτερη λήψη;
β) Αν η δεύτερη λήψη βγάλει λευκή σφαίρα ποια η πιθανότητα ώστε η πρώτη να ήταν λευκή;



(α) $P(\Lambda_2) = P(\Lambda_2|\Lambda_1)\cdot P(\Lambda_1) + P(\Lambda_2|\Lambda_1')\cdot P(\Lambda_1') =$

$$= \frac{7}{12}\cdot\frac{4}{9} + \frac{4}{11}\cdot\frac{5}{9} = \frac{28}{108} + \frac{20}{99} = 0{,}461 = 46{,}1\%.$$

(β) $$P(\Lambda_1|\Lambda_2) = \frac{P(\Lambda_2|\Lambda_1)\cdot P(\Lambda_1)}{P(\Lambda_2)} = \frac{\frac{7}{12}\cdot\frac{4}{9}}{0{,}461} = \frac{0{,}259}{0{,}461} = 0{,}562 = 56{,}2\%.$$

Ασκήσεις στον τύπο του Bayes

7. Υποθέτουμε ότι ο επιπολασμός μίας ασθένειας σε ένα δεδομένο πληθυσμό είναι 5%. Γνωρίζουμε ότι το 80% από εκείνους που έχουν την ασθένεια εμφανίζουν ένα ορισμένο εργαστηριακό εύρημα ενώ μόνο 10% από τους μη ασθενείς παρουσιάζουν το ίδιο εύρημα.

Ποια είναι η πιθανότητα ένα τυχαίο άτομο του πληθυσμού που εμφανίζει το συγκεκριμένο εύρημα να έχει πράγματι την ασθένεια;

$A = \{\text{το άτομο ασθενεί}\}$, $P(A) = 0,05$.

$E = \{\text{εύρημα}\}$. $P(E|A) = 0,8$. $P(E|A') = 0,1$.

$$P(A|E) = \frac{P(E|A)\cdot P(A)}{P(E)} = \frac{P(E|A)\cdot P(A)}{P(E|A)P(A) + P(E|A')P(A')} = \frac{0,8\cdot 0,05}{0,8\cdot 0,05 + 0,1\cdot 0,95} =$$

$$= 0,296 = 29,6\%.$$

άλυτη

$S = \{\text{spam email}\}$ $A = \{\text{ανίχνευση}\}$

Ασκήσεις στον τύπο του Bayes

8. Υπολογίζεται ότι το 50% των email είναι spam email. Ένα λογισμικό για το φιλτράρισμα αυτών των ανεπιθύμητων μηνυμάτων ηλεκτρονικού ταχυδρομείου ισχυρίζεται ότι μπορεί να ανιχνεύσει το 99% των ανεπιθύμητων μηνυμάτων ηλεκτρονικού ταχυδρομείου και η πιθανότητα για ψευδώς θετικά (ένα μη spam email που ανιχνεύεται ως spam) είναι 5%.

Αν ένα email $\underbrace{\text{εντοπιστεί ως ανεπιθύμητο}}^{S}$, τότε ποια είναι η πιθανότητα να είναι στην πραγματικότητα ένα μη ανεπιθύμητο email;

$P(S) = 0{,}5$

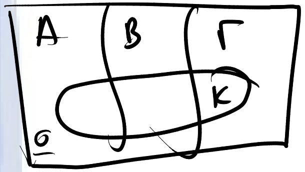
$P(A|S) = 0{,}99$

$P(A|S') = 0{,}05$

$$P(S'|A) = \frac{P(A|S')\cdot P(S')}{P(A)} = \frac{0{,}05\cdot 0{,}5}{0{,}99\cdot 0{,}5 + 0{,}05\cdot 0{,}5} = 0{,}048 = 4{,}8\%.$$

$$P(A) = P(A|S)\cdot P(S) + P(A|S')\cdot P(S') = 0{,}99\cdot 0{,}5 + 0{,}05\cdot 0{,}5$$

άλυτη



Ασκήσεις στον τύπο του Bayes

9. Η εταιρεία Α προμηθεύει το 40% των υπολογιστών που πωλούνται και καθυστερεί στις παραδόσεις στο 5% των περιπτώσεων. Η εταιρεία Β προμηθεύει το 30% των υπολογιστών που πωλούνται και καθυστερεί στο 3% των παραδόσεων. Η εταιρεία Γ προμηθεύει ακόμα άλλο 30% και καθυστερεί στο 2,5% των παραδόσεων.

Αν ένας υπολογιστής παραδοθεί με καθυστέρηση τότε ποια είναι η πιθανότητα να προέρχεται από την εταιρεία Α;

$P(A)=0{,}4 \quad P(K|A)=0{,}05 \quad P(B)=0{,}3 \quad P(K|B)=0{,}03 \quad P(\Gamma)=0{,}3 \quad P(K|\Gamma)=0{,}025$

$$P(A|K)=\frac{P(K|A)\,P(A)}{P(K|A)P(A)+P(K|B)P(B)+P(K|\Gamma)\cdot P(\Gamma)}=\frac{0{,}05\cdot 0{,}4}{0{,}05\cdot 0{,}4+0{,}03\cdot 0{,}3+0{,}025\cdot 0{,}3}=0{,}181$$

άλυτη

Ασκήσεις στον τύπο του Bayes

10. Έχουμε 3 νομίσματα, δύο αμερόληπτα, με πιθανότητα 0,5 για Κορώνα ή Γράμματα ενώ το τρίτο δεν είναι αμερόληπτο δείχνοντας Κορώνα με πιθανότητα 0,75.

Κάποιος επιλέγει τυχαία ένα από τα νομίσματα, το πετάει 3 φορές και έρχεται 3 φορές Κορώνα. Ποια είναι η πιθανότητα να είναι αυτό το μη αμερόληπτο νόμισμα;

άλυτη

Ασκήσεις στον τύπο του Bayes

11. Μία ασφαλιστική εταιρεία κατατάσσει τους ανθρώπους σε μία από τις τρεις κατηγορίες: καλοί πληρωτές, μεσαίοι πληρωτές και κακοί πληρωτές. Οι καταγραφές τους δηλώνουν ότι τα άτομα καλών, μεσαίων και κακών πληρωμών έχουν πιθανότητες να εμπλακούν σε τροχαίο ατύχημα μέσα σ’ ένα χρόνο 0,05, 0,15 και 0,30 αντίστοιχα.
(α) Αν 20% του πληθυσμού είναι καλοί πληρωτές, 50% είναι μεσαίοι πληρωτές και 30% είναι κακοί πληρωτές, τι ποσοστό ανθρώπων έχουν ατύχημα μέσα σ’ ένα σταθερό χρόνο;

(β) Αν ένας κάτοχος ασφαλιστικού συμβολαίου δεν είχε ατύχημα μέσα σ’ ένα χρόνο, ποια είναι η πιθανότητα ότι είναι καλός πληρωτής;

άλυτη

Κανόνας Bayes για 3 ενδεχόμενα

Έστω Α, Β, Γ τρία ενδεχόμενα ενός δειγματοχώρου Ω. Τότε:

$$P\left(A \mid B \cap \Gamma\right) = \frac{P\left(B \mid A \cap \Gamma\right) \cdot P\left(A \mid \Gamma\right)}{P\left(B \mid \Gamma\right)}.$$

Αξιολόγηση δίτιμης διαδικασίας πρόβλεψης

Ευαισθησία, ειδικότητα, PPV, NPV, PAC

Για κάθε διαδικασία που προβλέπει αν υπάρχει ένα χαρακτηριστικό σε ένα υποκείμενο (Όχι / Ναι, Αρνητικό / Θετικό, Παρών / Απών), έχει καθιερωθεί ένα σύνολο δεικτών που αξιολογούν την προβλεπτική της ικανότητα. Οι κυριότεροι δείκτες είναι:

- Το **ποσοστό ορθής ταξινόμησης** ή **ακρίβεια** (percentage accuracy in classification – **PAC** ή **accuracy**)
- Η **ευαισθησία** (**sensitivity** ή **TPR**) του τεστ (αληθώς ταξινομημένα ως θετικά - true positives).
- Η **προσδιοριστικότητα** ή **ειδικότητα** (**specificity** ή **TNR**) του τεστ (αληθώς ταξινομημένα ως αρνητικά - true negatives).
- Η **θετική διαγνωστική του αξία** (positive predictive value – **PPV**) (το ποσοστό των ταξινομημένων ως θετικά προς το σύνολο των θετικών προβλέψεων του μοντέλου).
- Η **αρνητική διαγνωστική του αξία** (negative predictive value - **NPV**) (το ποσοστό των ταξινομημένων ως αρνητικά προς το σύνολο των αρνητικών συμβάντων).

Ευαισθησία, ειδικότητα, PPV, NPV, PAC

Ο παρακάτω πίνακας συμπτώσεων περιγράφει εμπειρικά δεδομένα μίας διαδικασία πρόβλεψης.

		Πρόβλεψη (Prediction)		
		Όχι	Ναι	Σύνολο
Πραγματική κατάσταση (Observation)	Όχι	α	β	O_1 (= α + β)
	Ναι	γ	δ	O_2 (= γ + δ)
	Σύνολο	P_1 (= α + γ)	P_2 (= β + δ)	Σ (= α + β + γ + δ)

Βάσει του παραπάνω πίνακα είναι:

Ποσοστό ορθής ταξινόμησης (percentage accuracy in classification) PAC = (α + δ) / Σ.

Ευαισθησία (sensitivity) = δ / O_2. (TPR: αληθώς ταξινομημένα ως θετικά - true positives).

Ειδικότητα (specificity) = α / O_1. (TNR: αληθώς ταξινομημένα ως αρνητικά - true negatives).

Θετική διαγνωστική του αξία (positive predictive value) PPV = δ / P_2.

(PPV: ποσοστό των πραγματικά θετικών προς το σύνολο των θετικών προβλέψεων του μοντέλου)

Αρνητική διαγνωστική του αξία (negative predictive value) NPV = α / P_1.

(NPV: ποσοστό των πραγματικά αρνητικών προς το σύνολο των αρνητικών προβλέψεων του μοντέλου)

Ευαισθησία, ειδικότητα, PPV, NPV, PAC

Παράδειγμα 1
Σε πρόσφατη έρευνα έγινε σύγκριση της αξιοπιστίας του rapid test ID – NOW με το μοριακό (RT-PCR). Σε αυτήν, συμμετείχαν 974 ασθενείς, 785 συμπτωματικοί και 189 ασυμπτωματικοί ασθενείς. Από τους 785 συμπτωματικούς, οι 21 ήταν θετικοί τόσο στο rapid test όσο και με RT-PCR, και 2 μόνο με RT-PCR. Και οι 189 ασυμπτωματικοί ασθενείς βρέθηκαν αρνητικοί και στα δύο τεστ.

Βάσει του πίνακα είναι:

- Ποσοστό ορθής ταξινόμησης
  PAC = 972 / 974 = 99,8%.
- Ευαισθησία = 21 / 23 = 91,3%
- Προσδιοριστικότητα = 951 / 951 = 100%
- PPV = 21 / 21 = 100%
- NPV = 951 / 953 = 99,8%.

Πίνακας: Σύγκριση rapid test ID-NOW και RT-PCR				
		Rapid Test ID-NOW		
		Αρνητικό*	Θετικό	Σύνολο
	Αρνητικό	$\alpha = 951$	$\beta = 0$	$O_1 = 951$
RT-PCR	Θετικό	$\gamma = 2$	$\delta = 21$	$O_2 = 23$
	Σύνολο	$P_1 = 953$	$P_2 = 21$	$\Sigma = 974$
(*) 9 από τους 951 είχαν άκυρο rapid test				

Πηγή δεδομένων: eLife 2021;10:e65726 doi: 10.7554/eLife.65726

Ευαισθησία, ειδικότητα, PPV, NPV, PAC

Άσκηση

Σε πρόσφατη έρευνα, μεταξύ 1.465 ατόμων που ελέγχθηκαν, η RT-PCR ήταν θετική σε 141 άτομα. Η δοκιμή αντιγόνου Roche/SD Biosensor rapid test ήταν θετική σε 94 ασθενείς και αρνητική σε 1.368 άτομα, ενώ δεν ήταν έγκυρη σε 3 ασθενείς. Η συνολική ευαισθησία του rapid test ήταν 65,3% ενώ η προσδιοριστικότητα του ήταν 99,9%.

(α) Να συμπληρωθεί ο πίνακας

(β) Να υπολογιστούν:

(i) Το ποσοστό ορθής ταξινόμησης (PAC)
(ii) Η θετική διαγνωστική αξία (PPV)
(iii) Η αρνητική διαγνωστική αξία (NPV).

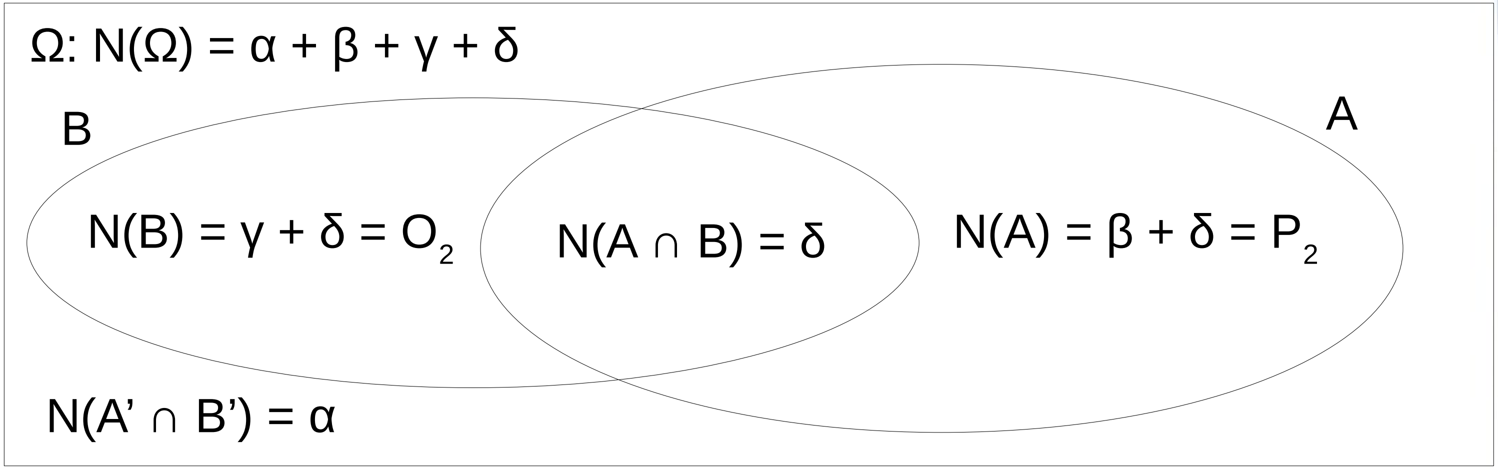
Πίνακας: Σύγκριση rapid test Roche/SD και RT-PCR				
		Rapid Test Roche/SD		
		Αρνητικό*	Θετικό	Σύνολο
	Αρνητικό	α =	β =	O_1 =
RT-PCR	Θετικό	γ =	δ =	O_2 =
	Σύνολο	P_1 =	P_2 =	Σ =
(*) Αρνητικά και άκυρα μαζί				

Πηγή δεδομένων:Jegerlehner, S., Suter-Riniker, F., Jent, P., Bittel, P., & Nagler, M. (2021). Diagnostic accuracy of a SARS-CoV-2 rapid antigen test in real-life clinical settings. International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases, 109, 118–122. https://doi.org/10.1016/j.ijid.2021.07.010

Ευαισθησία, ειδικότητα, PPV, NPV, PAC ως δεσμευμένες πιθανότητες

		Πρόβλεψη (Prediction)		
		Όχι	Ναι	Σύνολο
Πραγματική κατάσταση (Observation)	Όχι	α	β	O_1 (= α + β)
	Ναι	γ	δ	O_2 (= γ + δ)
	Σύνολο	P_1 (= α + γ)	P_2 (= β + δ)	Σ (= α + β + γ + δ)

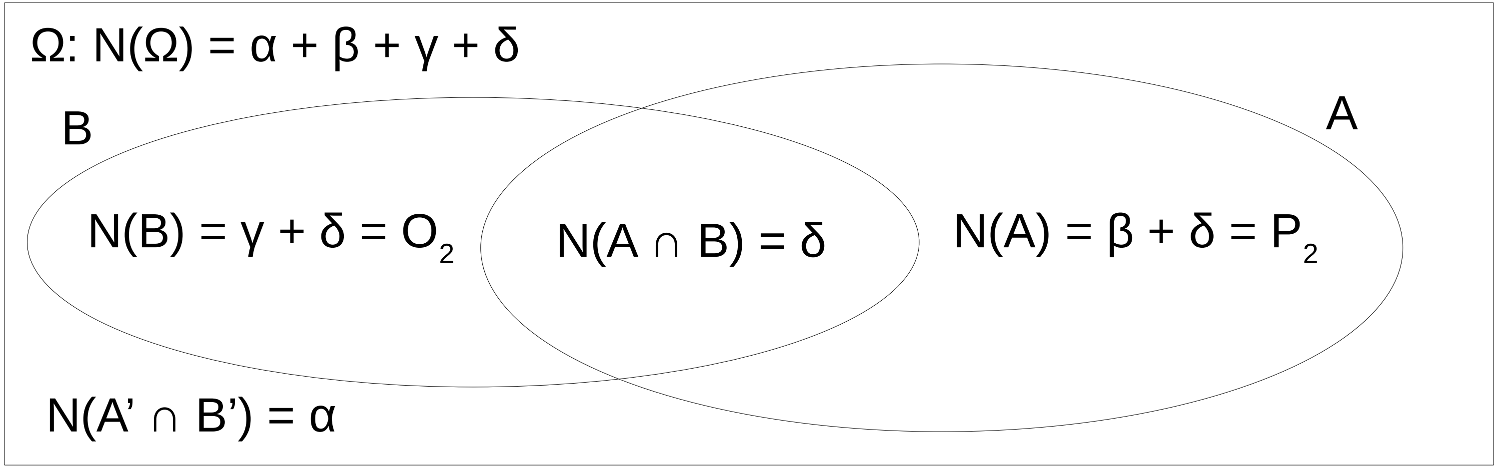
Ο πίνακας πρόβλεψης είναι δυνατό να αναπαρασταθεί ως ένα διάγραμμα Venn.



A = {Πρόβλεψη θετική} και B = {Παρατήρηση θετική}

(π.χ. A = {το rapid test είναι θετικό} και B = { το άτομο πάσχει από COVID-19})

Ευαισθησία, ειδικότητα, PPV, NPV, PAC ως δεσμευμένες πιθανότητες



A = {Πρόβλεψη θετική} και B = {Παρατήρηση θετική} (π.χ. A = {το rapid test είναι θετικό} και B = { το άτομο πάσχει από COVID-19})

Είναι $N(B) = O_2$, $N(B') = O_1$, $N(A) = P_2$, $N(A') = P_1$, $N(A \cap B) = \delta$, $N(A' \cap B') = \alpha$, συνεπώς:

- **Ευαισθησία** $= \delta / O_2 = N(A \cap B) / N(B) = P(A \mid B)$
- **Ειδικότητα** $= \alpha / O_1 = N(A' \cap B') / N(B') = P(A' \mid B')$
- **Θετική διαγνωστική του αξία** PPV $= \delta / P_2 = N(A \cap B) / N(A) = P(B \mid A)$
- **Αρνητική διαγνωστική του αξία** NPV $= \alpha / P_1 = N(A' \cap B') / N(A') = P(B' \mid A')$

Παράδειγμα

Θεωρούμε την ασθένεια του καρκίνου του πνεύμονα και το κάπνισμα ως σύμπτωμα του. Υποθέτουμε ότι το 90% των ατόμων με καρκίνο του πνεύμονα και το 30% των ατόμων χωρίς καρκίνο του πνεύμονα είναι καπνιστές, ποια είναι η ευαισθησία και η ειδικότητα του καπνίσματος ως σύμπτωμα που οδηγεί στον καρκίνο του πνεύμονα;

$A = \{\text{καρκίνος}\}$, $K = \{\text{κάπνισμα}\}$

$P(K|A) = 0,9$ Ευαισθησία $= P(K|A) = 0,9 = 90\%$.

$P(K|A') = 0,3$ Ειδικότητα $= P(K'|A') = 1 - 0,3 = 0,7 = 70\%$.

Άσκηση

Σε μια πόλη με 1.000 κατοίκους το 10% νοσεί από τη νόσο Χ. Ένα νέο τεστ για τη νόσο Χ, έχει ευαισθησία 80% και ειδικότητα 70%. Πόσα άτομα με τη νόσο Χ δεν θα ανιχνευθούν από αυτό το τεστ;

άλυτη

Παράδειγμα

Θεωρούμε την ασθένεια του καρκίνου του πνεύμονα και το κάπνισμα ως σύμπτωμα του. Υποθέτουμε ότι το 90% των ατόμων με καρκίνο του πνεύμονα και το 30% των ατόμων χωρίς καρκίνο του πνεύμονα είναι καπνιστές, ποια είναι η ευαισθησία και η ειδικότητα του καπνίσματος ως σύμπτωμα που οδηγεί στον καρκίνο του πνεύμονα;

Απάντηση

A = {το άτομο καπνίζει}, B = {το άτομο ασθενεί από καρκίνο του πνεύμονα}

Είναι:

- Ευαισθησία = P(A | B) = 0,9
- Ειδικότητα = P (A’ | B’) = 1 – P (A | B’) = 0,7

Σύνοψη των τύπων

Φανερά, τα μέτρα αξιολόγησης PAC, ευαισθησία, ειδικότητα, PPV, NPV ισχύουν για κάθε δίτιμη διαδικασία αξιολόγησης, δηλαδή για κάθε “σύμπτωμα” Α που μπορεί να συσχετίζεται με κάποια “ασθένεια” Β.

Συμπερασματικά, οι σχέσεις που καταγράψαμε είναι οι εξής:

- P(A | B) = ευαισθησία = TPR (True Positive Rate)
- P(A’ | B) = 1 – P(A | B) = 1 – ευαισθησία = FNR (False Negative Rate)
- P (A’ | B’) = ειδικότητα = TNR (True Negative Rate)
- P(A | B’) = 1 – ειδικότητα = FPR (False Positive Rate)
- P(B | A) = PPV (Positive Predictive Value)
- P(B’ | A’) = NPV (Negative Predictive Value)

Ο κανόνας του Bayes σε όρους ευαισθησίας και ειδικότητας

A = {το άτομο έχει το σύμπτωμα} και B = {το άτομο έχει την ασθένεια}

Από τις σχέσεις που καταγράψαμε, ο κανόνας του Bayes,

$$P(B \mid A) = \frac{P(A \mid B) \cdot P(B)}{P(A \mid B) \cdot P(B) + P(A \mid B') \cdot P(B')},$$

παίρνει τη μορφή

$$PPV = \frac{\text{ευαισθησία} \cdot x}{\text{ευαισθησία} \cdot x + (1 - \text{ειδικότητα}) \cdot (1 - x)},$$

όπου x = P(B), η πιθανότητα το άτομο να έχει την ασθένεια. Επιπλέον,

$$NPV = P(B' \mid A') = \frac{P(A' \mid B') \cdot P(B')}{P(A' \mid B') \cdot P(B') + P(A' \mid B) \cdot P(B)} = \frac{\text{ειδικότητα} \cdot (1 - x)}{\text{ειδικότητα} \cdot (1 - x) + (1 - \text{ευαισθησία}) \cdot x}.$$

Ο κανόνας του Bayes σε όρους ευαισθησίας και ειδικότητας

A = {το άτομο έχει το σύμπτωμα} και B = {το άτομο έχει την ασθένεια}

Εναλλακτικά, καθώς

- ευαισθησία = TPR (true positive rate), 1 – ευαισθησία = FNR (false negative rate)
- ειδικότητα = TNR (true negative rate), 1 – ειδικότητα = FPR (false positive rate),

μπορούμε να ξαναγράψουμε τους τύπους ως:

$$\mathrm{PPV} = \frac{\mathrm{TPR} \cdot x}{\mathrm{TPR} \cdot x + \mathrm{FPR} \cdot (1 - x)}, \qquad \mathrm{NPV} = \frac{\mathrm{TNR} \cdot (1 - x)}{\mathrm{TNR} \cdot (1 - x) + \mathrm{FNR} \cdot x},$$

όπου x = P(B), η πιθανότητα το άτομο να έχει την ασθένεια.

Ο κανόνας του Bayes σε όρους ευαισθησίας και ειδικότητας

Παράδειγμα 1

Γνωρίζουμε ότι το 20% του ενήλικου πληθυσμού είναι υπερτασικό. Ένα νέο αυτόματο μηχάνημα αρτηριακής πίεσης, ταξινομεί ως υπερτασικούς το 84% των ατόμων που πράγματι πάσχουν από υπέρταση και το 23% του πληθυσμού με φυσιολογική πίεση.

Ποια είναι η θετική προγνωστική αξία (PPV) και ποια η αρνητική προγνωστική αξία (NPV) του μηχανήματος;

Λύση

Γνωρίζουμε ότι, x = P(ασθένεια) = 0,2, ευαισθησία = P(σύμπτωμα | ασθένεια) = 0,84 και ειδικότητα = P (χωρίς σύμπτωμα | καμία ασθένεια) = 1 – 0,23 = 0,77. Υπολογίζουμε,

$$\text{PPV} = \frac{\text{ευαισθησία} \cdot \text{x}}{\text{ευαισθησία} \cdot \text{x} + \left(1 - \text{ειδικότητα}\right) \cdot \left(1 - \text{x}\right)} = \frac{0{,}84 \cdot 0{,}2}{0{,}84 \cdot 0{,}2 + \left(1 - 0{,}77\right) \cdot \left(1 - 0{,}2\right)} = 0{,}48.$$

$$\text{NPV} = \frac{\text{ειδικότητα} \cdot \left(1 - \text{x}\right)}{\text{ειδικότητα} \cdot \left(1 - \text{x}\right) + \left(1 - \text{ευαισθησία}\right) \cdot \text{x}} = \frac{0{,}77 \cdot 0{,}8}{0{,}77 \cdot 0{,}8 + \left(1 - 0{,}84\right) \cdot 0{,}2} = 0{,}95.$$

Ο κανόνας του Bayes σε όρους ευαισθησίας και ειδικότητας

Παράδειγμα 2

Ένας κλινικός ιατρός θέλει να χρησιμοποιήσει ένα διαγνωστικό τεστ για να εκτιμήσει την παρουσία ή την απουσία πνευμονικής εμβολής (ΠΕ) σε κάποιον που εμφανίζει κάποια σχετικά συμπτώματα (ο επιπολασμός στη συγκεκριμένη πληθυσμιακή ομάδα είναι γνωστό ότι είναι 50%).

Ο κλινικός ιατρός μπορεί να επιλέξει ανάμεσα σε δύο διαφορετικές εξετάσεις

(α) D-dimer με 97% ευαισθησία και 41% ειδικότητα ή

(β) υπερηχογράφημα συμπίεσης (CUS) με 49% ευαισθησία και 96% ειδικότητα.

Για τον αποκλεισμό της πνευμονικής εμβολής, η αρνητική προγνωστική αξία είναι αυτή που μας ενδιαφέρει. Ποιο από τα δύο τεστ είναι πιο χρήσιμο για αυτόν τον σκοπό;

Ο κανόνας του Bayes σε όρους ευαισθησίας και ειδικότητας

Λύση

(α) D-dimer με 97% ευαισθησία και 41% ειδικότητα

$$\text{NPV} = \frac{\text{ειδικότητα} \cdot (1 - \text{x})}{\text{ειδικότητα} \cdot (1 - \text{x}) + (1 - \text{ευαισθησία}) \cdot \text{x}} = \frac{0{,}41 \cdot (1 - 0{,}5)}{0{,}41 \cdot (1 - 0{,}5) + (1 - 0{,}97) \cdot 0{,}5} = 0{,}93 = 93\,\%.$$

(β) υπερηχογράφημα συμπίεσης (CUS) με 49% ευαισθησία και 96% ειδικότητα.

$$\text{NPV} = \frac{\text{ειδικότητα} \cdot (1 - \text{x})}{\text{ειδικότητα} \cdot (1 - \text{x}) + (1 - \text{ευαισθησία}) \cdot \text{x}} = \frac{0{,}96 \cdot (1 - 0{,}5)}{0{,}96 \cdot (1 - 0{,}5) + (1 - 0{,}49) \cdot 0{,}5} = 0{,}65 = 65\,\%.$$

Συμπέρασμα

Το D-dimer είναι περισσότερο κατάλληλο για τον αποκλεισμό της πνευμονικής εμβολής.

Πηγή: https://doi.org/10.1016/j.jclinepi.2020.12.021

Ο κανόνας του Bayes σε όρους ευαισθησίας και ειδικότητας

Παράδειγμα

Καταγράφηκε για 20 φοιτητές η επιτυχία τους στις εξετάσεις και ο χρόνος προετοιμασίας τους. Στα δεδομένα προσαρμόστηκε λογιστικό μοντέλο πρόβλεψης το οποίο αντιστοίχισε σε κάθε έναν φοιτητή μία πιθανότητα επιτυχίας. Να δείξετε ότι το όριο αποδοχής (cut – off score) α = 0,7, προβλέπει σωστά το 75% των περιπτώσεων και εξασφαλίζει ευαισθησία 60% και ειδικότητα 90%.

Φοιτητής	1	2	3	4	5	6	7	8	9	10
Ώρες μελέτης	0,5	0,75	1	1,25	1,5	1,75	1,75	2	2,25	2,5
Επιτυχία	0	0	0	0	0	0	1	0	1	0
Πρόβλεψη επιτυχίας	0,035	0,05	0,071	0,1	0,139	0,191	0,191	0,256	0,334	0,422
Φοιτητής	11	12	13	14	15	16	17	18	19	20
Ώρες μελέτης	2,75	3	3,25	3,5	4	4,25	4,5	4,75	5	5,5
Επιτυχία	1	0	1	0	1	1	1	1	1	1
Πρόβλεψη επιτυχίας	0,515	0,607	0,693	0,766	0,874	0,91	0,937	0,956	0,969	0,985

Ο κανόνας του Bayes σε όρους ευαισθησίας και ειδικότητας

Λύση
Ταξινομούμε τους φοιτητές κατά σειρά επιτυχίας.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0,035	0,05	0,071	0,1	0,139	0,191	0,191	0,256	0,334	0,422	0,515	0,607	0,693	0,766	0,874	0,91	0,937	0,956	0,969	0,985

Πίνακας: Πρόβλεψη επιτυχίας με α = 0,7				
		Πρόβλεψη		
		0	1	Σύνολο
Επιτυχία	0	9	1	10
	1	4	6	10
	Σύνολο	13	7	20

Από τον πίνακα υπολογίζουμε:

PAC = (9 + 6) / 20 = 15 / 20 = 75%.

Ευαισθησία = 6 / 10 = 0,6 = 60%

Ειδικότητα = 9 / 10 = 0,9 = 90%

Ασκήσεις

1. Ο παρακάτω πίνακας δείχνει τα αποτελέσματα από την εξέταση της διαγνωστικής ακρίβειας ενός νέου γρήγορου τεστ για HIV σε 100.000 άτομα, σε σύγκριση με το πρότυπο τεστ αναφοράς. Οι σειρές του πίνακα αντιπροσωπεύουν το αποτέλεσμα της εξέτασης και οι στήλες την πραγματική κατάσταση της νόσου.

	Τεστ +	Τεστ -	Σύνολο
HIV+	378	2	380
HIV-	397	98.823	99.220
Σύνολο	775	98.825	100.000

Να υπολογιστούν τα PAC, ευαισθησία, ειδικότητα, NPV, PPV.

Ασκήσεις

Λύση

	Τεστ +	Τεστ -	Σύνολο
HIV+	378	2	380
HIV-	397	98.823	99.220
Σύνολο	775	98.825	100.000

- Ποσοστό ορθής ταξινόμησης PAC = (378 + 98.823) / 100.000 = 99,2%.
- Ευαισθησία = 378 / 380 = 99,5%
- Ειδικότητα = 98.823 / 99.220 = 99,6%
- PPV = 378 / 775 = 48,8%
- NPV = 98.823 / 98.825 = 99,9%.

Καμπύλη λειτουργικού χαρακτηριστικού δέκτη

Receiver Operating Characteristic (ROC) - curve

Καμπύλη λειτουργικού χαρακτηριστικού δέκτη (ROC curve)

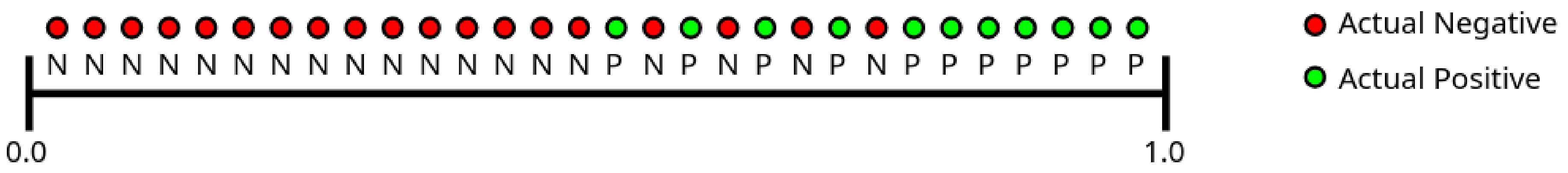
Η ονομασία “Καμπύλη ROC” χρονολογείται από τον Δεύτερο Παγκόσμιο Πόλεμο και τους ηλεκτρολόγους μηχανικούς που υποστήριζαν τη λειτουργία των ραντάρ.

Όταν η ενίσχυση του σήματος στο ραντάρ ήταν μηδενική, τότε δεν ανιχνευόταν κανένα εχθρικό αεροπλάνο. Με την αύξηση της ενίσχυσης, ανιχνευόταν όλο και περισσότερα αντικείμενα, αυξάνοντας ωστόσο και την ποσότητα του θορύβου η οποία οδηγούσε σε εσφαλμένα θετικά αποτελέσματα (False Positive), δηλαδή αναγνώριση αντικειμένων που μπορεί να φαίνονταν σαν αεροπλάνα αλλά δεν ήταν, όπως σύννεφα βροχής ή κοπάδια πουλιών. Από κάποια τιμή, η περαιτέρω αύξηση της ενίσχυσης ήταν αντιπαραγωγική, καθώς ο θόρυβος (False Positive) αρχίζει να υπερτερεί των σημάτων (True Positive).

Η καμπύλη ROC επινοήθηκε ως ένα απλό στη δημιουργία του διάγραμμα που θα βοηθούσε το χειριστή να αποφασίσει ποιο επίπεδο ενίσχυσης θα προσέφερε ικανοποιητική ευαισθησία (True Positive Rate, TPR) και σχετικά μεγάλη ειδικότητα ή ισοδύναμα μικρή πιθανότητα εσφαλμένης ανίχνευσης (False Positive Rate FPR = 1 – ειδικότητα).

Πηγή: Streiner, D. L., & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristics curves. Canadian journal of psychiatry. Revue canadienne de psychiatrie, 52(2), 121–128. https://doi.org/10.1177/070674370705200210

Καμπύλη ROC



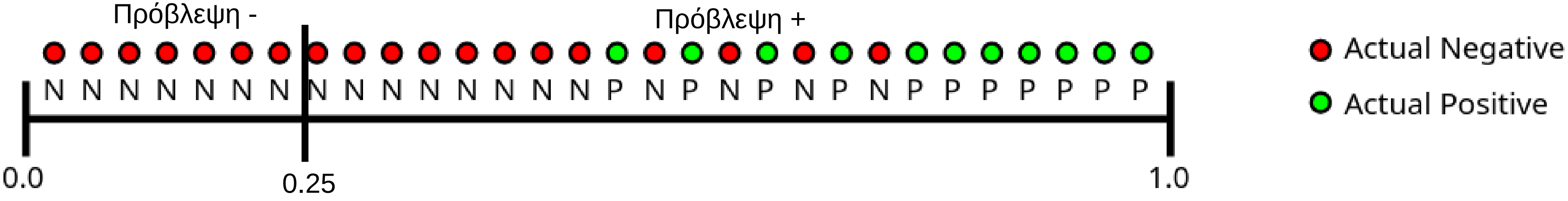
Στην προσπάθεια προσαρμογής ενός μοντέλου πρόβλεψης spam email, έγινε χρήση δείγματος 30 email τα οποία **ήταν spam**, ή **δεν ήταν**.

Μοντέλο πρόβλεψης αξιοποίησε άλλα χαρακτηριστικά των email (αποστολέας, πλήθος παραληπτών, περιεχόμενο κλπ) και απέδωσε μία πιθανότητα μεταξύ 0 και 1 σε κάθε ένα από αυτά να είναι spam. Τα 30 email τοποθετήθηκαν στην σειρά ως προς την εκτιμώμενη πιθανότητα. Το ερώτημα που τώρα πρέπει να απαντηθεί είναι:

Πάνω από ποια τιμή εκτιμώμενης πιθανότητας, το μοντέλο έχει ικανοποιητική ευαισθησία και αρκετά μεγάλη ειδικότητα;

Πηγή: https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall

Καμπύλη ROC



Για παράδειγμα, στα παραπάνω δεδομένα, επιλέγοντας ως όριο την τιμή πιθανότητας 0,25, ο πίνακας συμπτώσεων της πρόβλεψης με την πραγματική κατάσταση των email είναι:

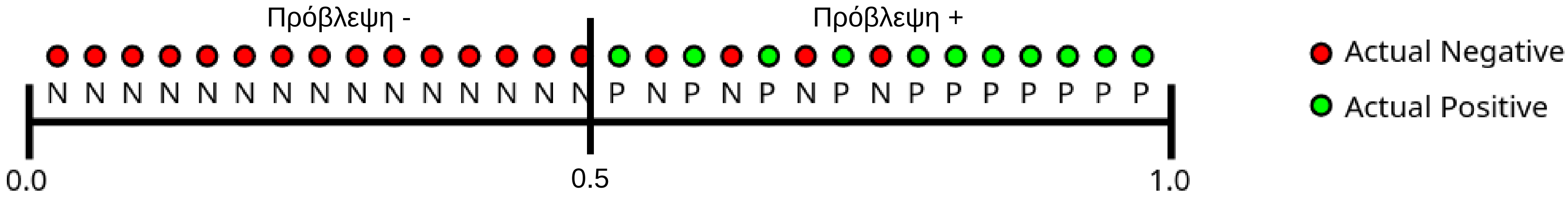
Πίνακας: Όριο αποδοχής 0,25				
		Πρόβλεψη		
		Όχι	Ναι	Σύνολο
Το email είναι spam	Όχι	α = 7	β = 12	$O_1 = 19$
	Ναι	γ = 0	δ = 11	$O_2 = 11$
	Σύνολο	$P_1 = 7$	$P_2 = 23$	Σ = 30

Από τον πίνακα υπολογίζουμε, πως στην επιλογή ορίου (cut – off score) 0,25 αντιστοιχεί:

Ευαισθησία (sensitivity) = $\delta / O_2 = 11 / 11 = 100\%$

Ειδικότητα (specificity) = $\alpha / O_1 = 7 / 19 = 36{,}8\%$

Καμπύλη ROC



Επιλέγοντας ως όριο θετικότητας την πιθανότητα 0,5, ο πίνακας συμπτώσεων της πρόβλεψης του μοντέλου με την πραγματική κατάσταση των email είναι:

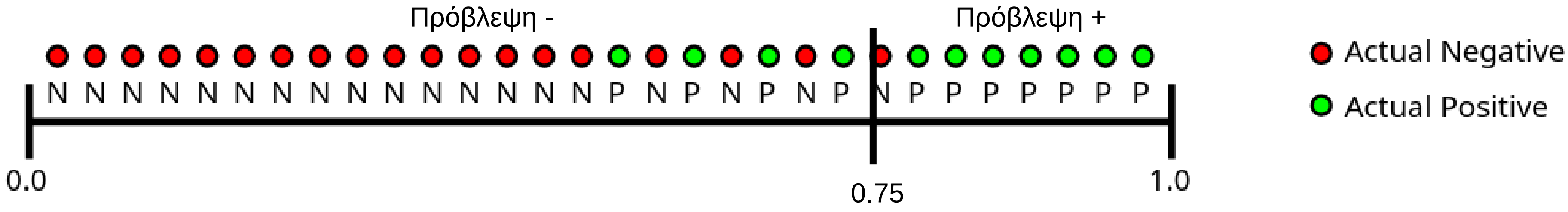
Πίνακας: Όριο αποδοχής 0,5				
		Πρόβλεψη		
		Όχι	Ναι	Σύνολο
Το email είναι spam	Όχι	α = 15	β = 4	O_1 = 19
	Ναι	γ = 0	δ = 11	O_2 = 11
	Σύνολο	P_1 = 15	P_2 = 15	Σ = 30

Από τον πίνακα υπολογίζουμε, πως στην επιλογή ορίου (cut – off score) 0,5 αντιστοιχεί:

Ευαισθησία (sensitivity) = δ / O_2 = 11 / 11 = 100%

Ειδικότητα (specificity) = α / O_1 = 15 / 19 = 78,9%

Καμπύλη ROC

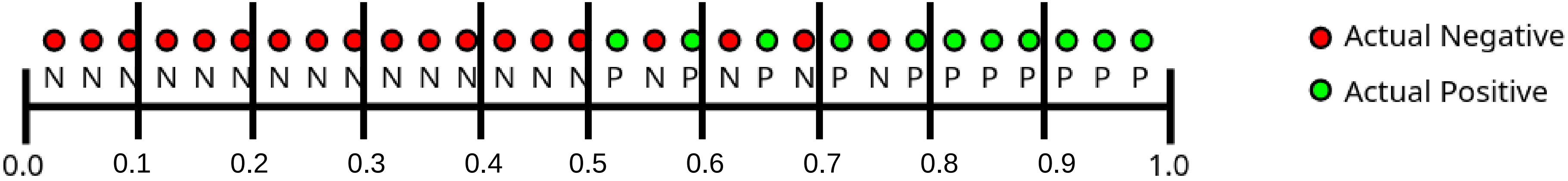


Επιλέγοντας ως όριο θετικότητας την πιθανότητα 0,75, ο πίνακας συμπτώσεων της πρόβλεψης του μοντέλου με την πραγματική κατάσταση των email είναι:

Πίνακας: Όριο αποδοχής 0,75				
		Πρόβλεψη		
		Όχι	Ναι	Σύνολο
Το email είναι spam	Όχι	α = 18	β = 1	O_1 = 19
	Ναι	γ = 4	δ = 7	O_2 = 11
	Σύνολο	P_1 = 22	P_2 = 8	Σ = 30

Από τον πίνακα υπολογίζουμε, πως στην επιλογή ορίου (cut – off score) 0,5 αντιστοιχεί:
Ευαισθησία (sensitivity) = δ / O_2 = 7 / 11 = 63,6%
Ειδικότητα (specificity) = α / O_1 = 18 / 19 = 94,7%

Καμπύλη ROC



Η διαδικασία που ακολουθήσαμε για cut-off score α = 0,25, 0,5 και 0,75 είναι δυνατόν να επαναληφθεί για κάθε τιμή μεταξύ 0 και 1. Στον πίνακα παρουσιάζονται οι τιμές για α = 0,1, 0,2, …, 1.

α	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
1 – ειδικότητα (FPR)	100%	84,2%	68,4%	52,6%	36,8%	21,1%	15,8%	5,3%	0,0%	0,0%	0,0%
ευαισθησία (TPR)	100%	100%	100%	100%	100%	100%	81,8%	72,7%	54,5%	27,3%	0,0%

α	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
1 – ειδικότητα (FPR)	100%	84,2%	68,4%	52,6%	36,8%	21,1%	15,8%	5,3%	0,0%	0,0%	0,0%
ευαισθησία (TPR)	100%	100%	100%	100%	100%	100%	81,8%	72,7%	54,5%	27,3%	0,0%

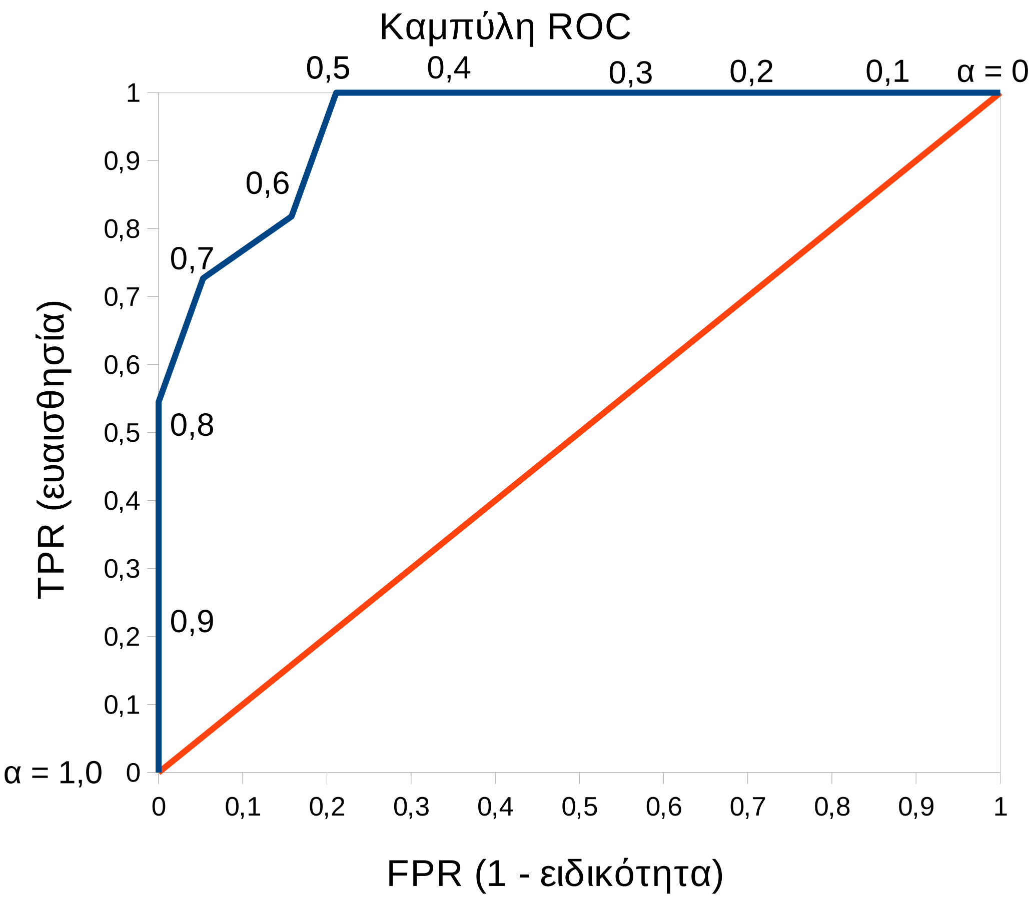
Τα ζεύγη των τιμών

(FPR, TPR) = (1 – ειδικότητα, ευαισθησία),

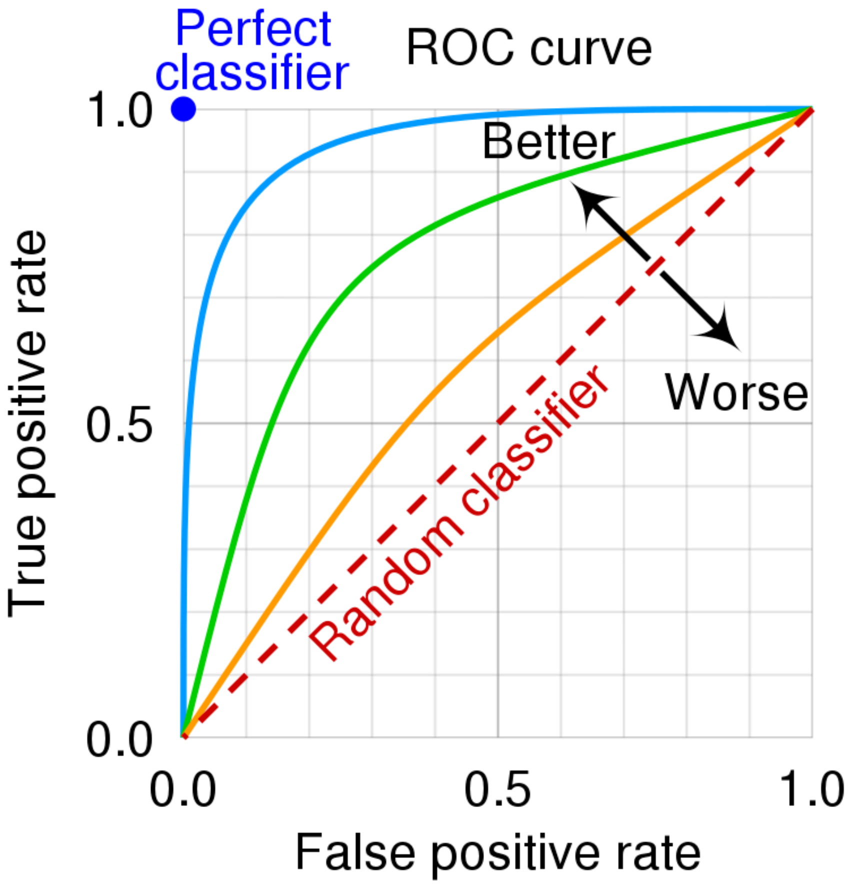
αν αναπαρασταθούν σε διαγραμματική μορφή, συνιστούν την καμπύλη λειτουργικού χαρακτηριστικού δέκτη (καμπύλη ROC). Ο ερευνητής μπορεί να επιλέξει το όριο α με το οποίο επιτυγχάνεται αξιοπρεπής ευαισθησία και ειδικότητα στο μοντέλο πρόγνωσης που έχει αναπτύξει.

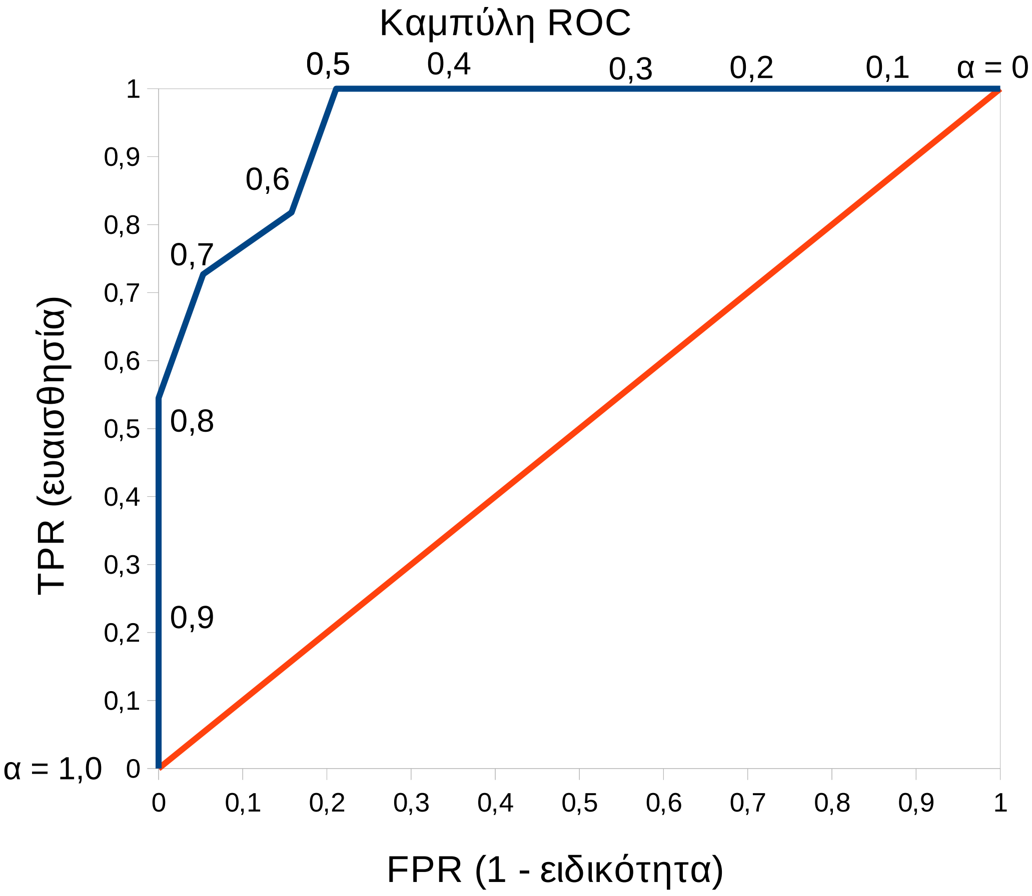
Η καμπύλη ROC, συνήθως απεικονίζεται μαζί με την ευθεία y = x, η οποία αντιστοιχεί στην τυχαία ταξινόμηση, όπου FPR = TPR. Ως εκ τούτου, επιθυμία είναι η επιλογή cut-off score α που να εξασφαλίζει ζεύγος (FPR, TPR) όσο το δυνατόν πιο μακριά από αυτή.

Στην συγκεκριμένη περίπτωση η επιλογή α = 0,5 ή α = 0,6 αντιστοιχεί σε μοντέλο με ικανοποιητική επίδοση στην πρόβλεψη των spam email.



Καμπύλη λειτουργικού χαρακτηριστικού δέκτη (ROC curve)





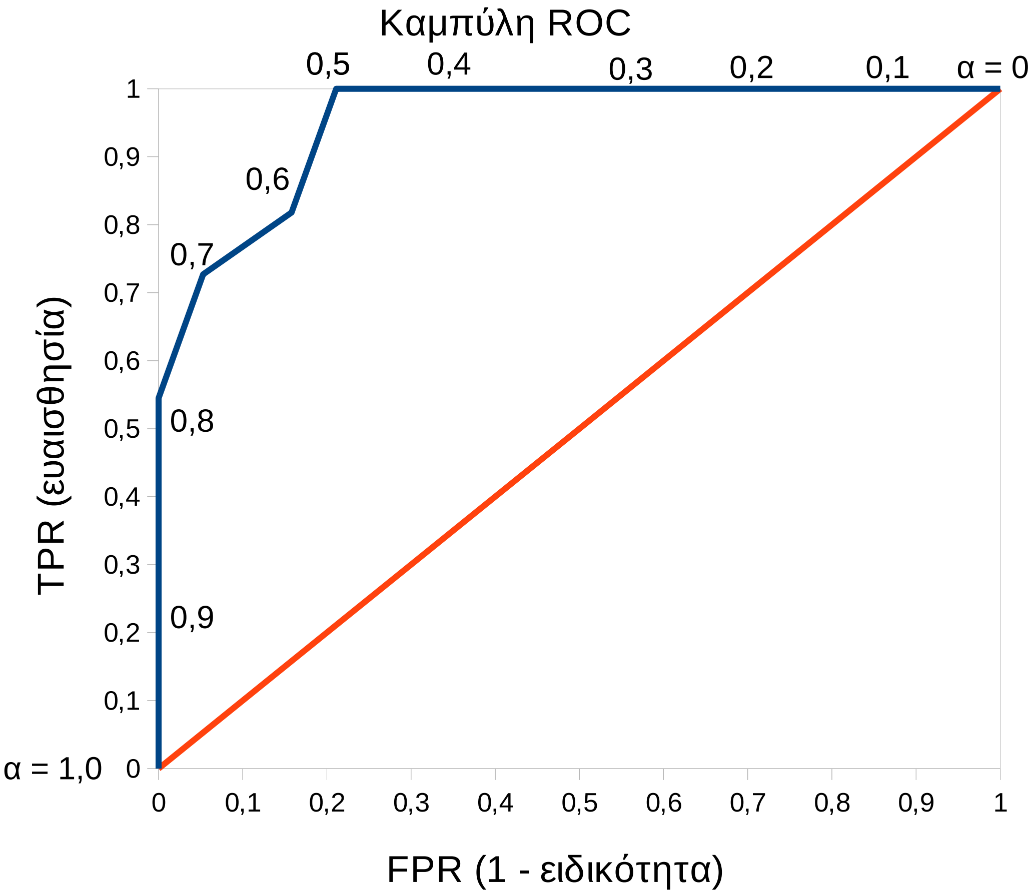
Καμπύλη λειτουργικού χαρακτηριστικού δέκτη (ROC curve)

Το εμβαδόν της περιοχής κάτω από την ROC καμπύλη (Area Under Curve – AUC) είναι ένα βασικό μέτρο αξιολόγησης της μεθόδου ταξινόμησης.

Αν AUC < 0,5 τότε η επίδοση του μοντέλου είναι χειρότερη από την τυχαία επιλογή και πρέπει να απορριφθεί εξ' ολοκλήρου.

Όσο πιο κοντά στο 1 βρίσκεται το AUC τόσο καλύτερη είναι η διαδικασία ταξινόμησης στο σύνολό της.

Επιπλέον, το AUC επιτρέπει τη σύγκριση δύο ή περισσότερων διαφορετικών μοντέλων που στοχεύουν στην ταξινόμηση των παρατηρήσεων.



Καμπύλη λειτουργικού χαρακτηριστικού δέκτη (ROC curve)

Παράδειγμα 2
Ένας τρόπος διάγνωσης του υποθυρεοειδισμού, είναι το επίπεδο της ορμόνης T4. Για τον έλεγχο της αξιοπιστίας της μέτρησης αυτής ελέγχθηκαν 125 ασθενείς από τους οποίους οι 32 ήταν βέβαιο ότι έπασχαν από υποθυρεοειδισμό και βρέθηκαν τα αποτελέσματα του πίνακα.

(α) Να συμπληρωθούν πίνακες αξιολόγησης της διαγνωστικής ικανότητας της ορμόνης T4 στα επίπεδα, των 5,0, 7,0 και 9,0 microliter/dl.

(β) Να συμπληρωθεί ο πίνακας και να γίνει η καμπύλη ROC

	Υποθυρεοειδισμός		
Ορμόνη T4	Ναι	Όχι	Σύνολο
< 5,0	18	1	19
5,1 – 7,0	7	17	24
7,1 – 9,0	4	36	40
> 9,0	3	39	42
Σύνολο	32	93	125

Όριο διάγνωσης T4	5,0	7,0	9,0
Ευαισθησία			
Ειδικότητα			
1 – Ειδικότητα			

Πηγή δεδομένων: Goldstein, B. J., & Mushlin, A. I. (1987). Use of a single thyroxine test to evaluate ambulatory medical patients for suspected hypothyroidism. Journal of general internal medicine, 2(1), 20–24. https://doi.org/10.1007/BF02596245

Καμπύλη λειτουργικού χαρακτηριστικού δέκτη (ROC curve)

Όριο 5,0	Υποθυρεοειδισμός		
Ορμόνη Τ4	Ναι	Όχι	Σύνολο
< 5,0	18	1	19
≥ 5,0	14	92	104
Σύνολο	32	93	125

Όριο 5,0: Ευαισθ. = 18/19, Ειδ. = 92/104

Όριο 7,0: Ευαισθ. = 25/43, Ειδ. = 75/82

Όριο 9,0: Ευαισθ. = 29/83, Ειδ. = 39/42

Όριο 7,0	Υποθυρεοειδισμός		
Ορμόνη Τ4	Ναι	Όχι	Σύνολο
< 7,0	25	18	43
≥ 7,0	7	75	82
Σύνολο	32	93	125

Όριο 9,0	Υποθυρεοειδισμός		
Ορμόνη Τ4	Ναι	Όχι	Σύνολο
< 9,0	29	54	83
≥ 9,0	3	39	42
Σύνολο	32	93	125

Πηγή δεδομένων: Goldstein, B. J., & Mushlin, A. I. (1987). Use of a single thyroxine test to evaluate ambulatory medical patients for suspected hypothyroidism. Journal of general internal medicine, 2(1), 20–24. https://doi.org/10.1007/BF02596245

Καμπύλη λειτουργικού χαρακτηριστικού δέκτη (ROC curve)

2. Όριο 5,0: Ευαισθ. = 18/19 = 0,947, Ειδ. = 92/104

Όριο 7,0: Ευαισθ. = 25/43, Ειδ. = 75/82

Όριο 9,0: Ευαισθ. = 29/83, Ειδ. = 39/42

Όριο T4 για διάγνωση	5,0	7,0	9,0
Ευαισθησία	0,947	0,581	0,349
Ειδικότητα	0,885	0,915	0,929
1 – Ειδικότητα	0,115	0,085	0,071

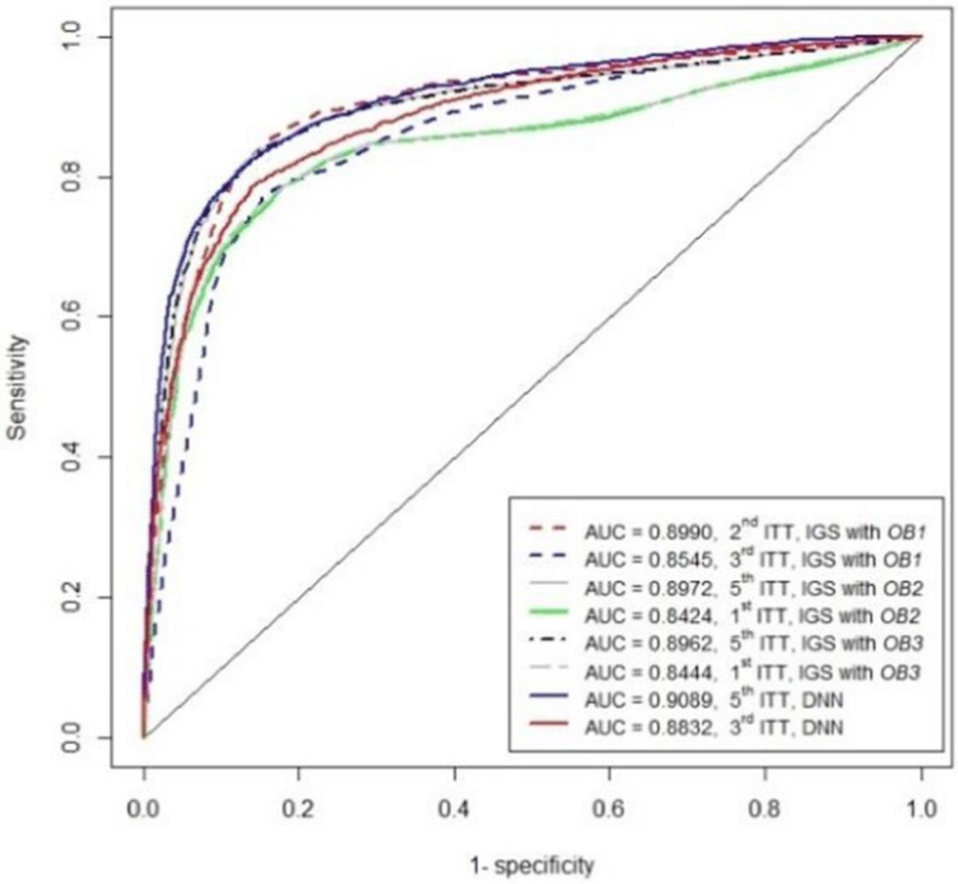
Πηγή δεδομένων: Goldstein, B. J., & Mushlin, A. I. (1987). Use of a single thyroxine test to evaluate ambulatory medical patients for suspected hypothyroidism. Journal of general internal medicine, 2(1), 20–24. https://doi.org/10.1007/BF02596245

Καμπύλη λειτουργικού χαρακτηριστικού δέκτη (ROC curve)

Παράδειγμα 3
Σε πρόσφατη έρευνα, δεδομένα από εθνικό σύστημα υγείας της Ταϊβάν χρησιμοποιήθηκαν για την ανάπτυξη ενός συστήματος υποστήριξης κλινικών αποφάσεων με στόχο την πρόβλεψη του οξέος εμφράγματος του μυοκαρδίου.

Αναπτύχθηκαν 8 διαφορετικά μοντέλα πρόβλεψης, τα οποία αξιοποίησαν 75 διαφορετικά χαρακτηριστικά των ασθενών.

Τα 8 διαφορετικά μοντέλα πρόβλεψης συγκρίθηκαν ως προς την ποιότητα της πρόβλεψης που προσφέρουν με ένα πολλαπλό διάγραμμα ROC.



Πηγή: Wu, Fu-Hsing & Lai, Huey-Jen & Lin, Hsuan-Hung & Chan, Po-Chou & Tseng, Chien-Ming & Chang, Kun-Min & Chen, Yung-Fu & Lin, Chih-Sheng. (2022). Predictive models for detecting patients more likely to develop acute myocardial infarctions. The Journal of Supercomputing. 78. 1-29. 10.1007/s11227-021-03916-z.