Στατιστική Ι

Γιώργος Τσιρογιάννης

Τμήμα Διοίκησης Επιχειρήσεων Αγροτικών Προϊόντων και Τροφίμων,

Πανεπιστήμιο Πατρών

Ακαδημαϊκό έτος: 2020-2021

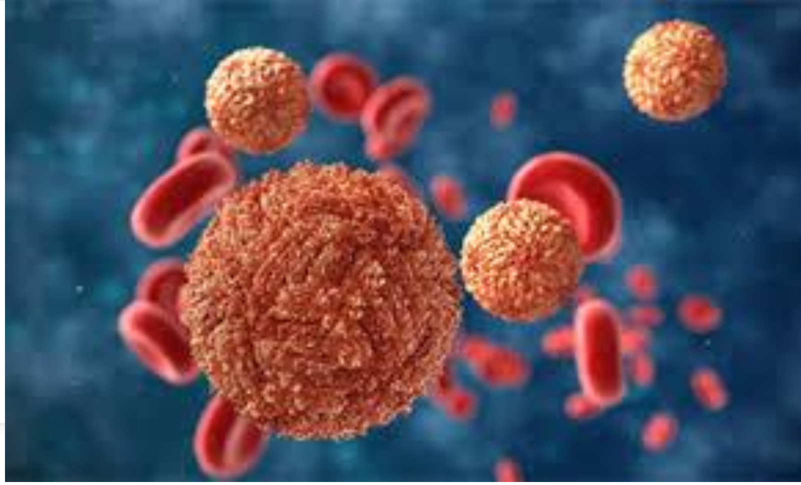
Διάλεξη 4η

Θεώρημα του Bayes

Ανεξαρτησία ενδεχομένων

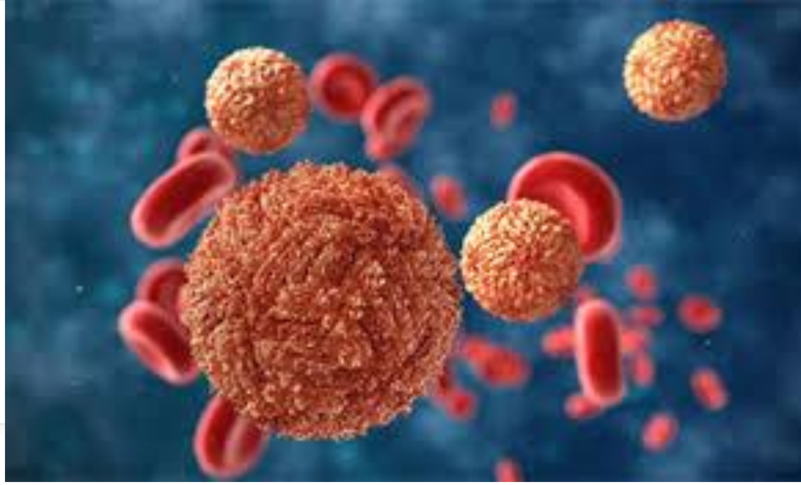
4.4 και 4.5

Παράδειγμα



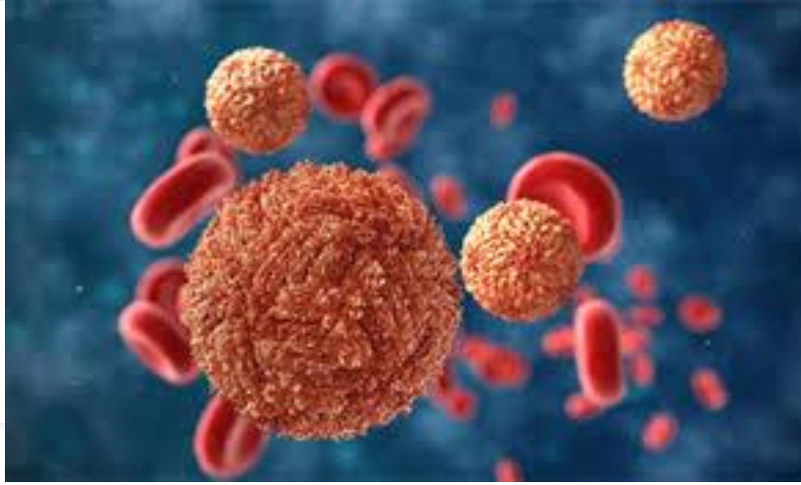
- Ποσοστό φορέων του ιού: 0.75%
- Αν κάποιος υποβληθεί σε τεστ για τον ιό:
  - Μετά την ανακοίνωση του αποτελέσματος η πιθανότητα μεταβάλλεται γιατί το τεστ δεν είναι αλάνθαστο
    - Αν το άτομο είναι πράγματι φορέας του ιού, τότε το τέστ κάνει διάγνωση με 92% ακρίβεια
    - Αν το άτομο δεν είναι φορέας του ιού, τότε το τέστ κάνει διάγνωση με 96% ακρίβεια

Παράδειγμα (συν.)



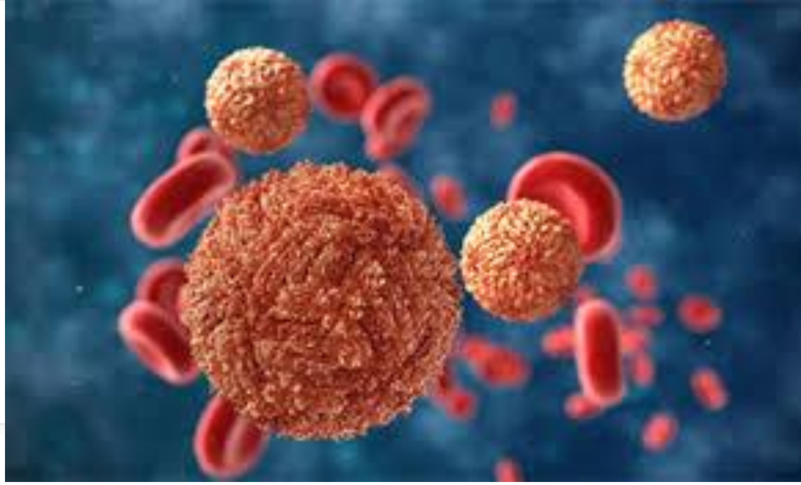
- Ενδεχόμενα:
  - B: ο κάτοικος είναι φορέας του ιού
  - Θ: το αποτέλεσμα του τεστ είναι θετικό
- Πιθανότητες από τα δεδομένα:
  - $P(B)=0.0075$
- Πως η πιθανότητα αλλάζει όταν γνωρίζουμε το αποτέλεσμα του τεστ για έναν κάτοικο;

Παράδειγμα (συν.)



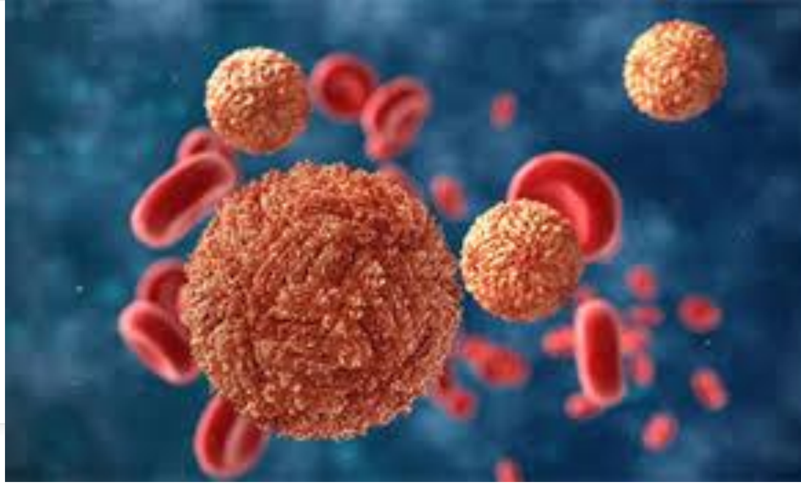
- **Ερώτηση:** ποια είναι η πιθανότητα ο συγκεκριμένος κάτοικος να είναι φορέας (ενδεχόμενο Β), αν έχει πάρει τα αποτελέσματα του τέστ (δηλ. το ενδεχόμενο Θ έχει συμβεί);
- Αν το τέστ ήταν «τέλειο/ιδανικό» η πιθανότητα θα ήταν προφανώς 1.
- Για τα μη-τελεια τεστ: ζητάμε να υπολογίσουμε την **δεσμευμένη πιθανότητα** του Β (conditional probability) δοθέντος του Θ και συμβολίζεται $P(B|\Theta)$
- Τι εννοούμε όταν γράφουμε $P(B|\Theta')$

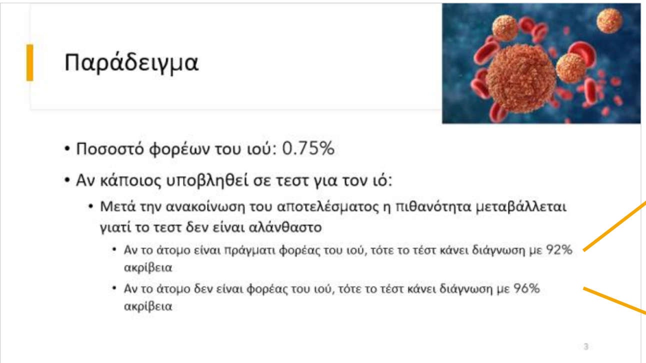
Παράδειγμα (συν.)



- Η πιθανότητα P(B) είναι εκείνη που γνωρίζουμε πριν μάθουμε το αποτέλεσμα του τεστ και καλείται εκ των προτέρων πιθανότητα (a priori probability)
- Η πιθανότητα P(B|Θ) είναι εκείνη που προσαρμόζουμε εφόσον μάθουμε το αποτέλεσμα του τεστ και καλείται εκ των υστέρων πιθανότητα (posterior probability)

Παράδειγμα (συν.)





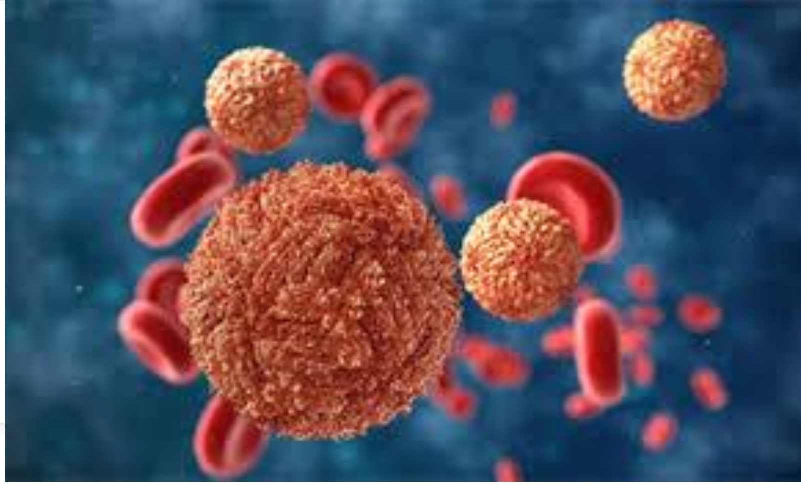
$P(\Theta|B)=0.92$

$P(\Theta'|B')=0.96$

Ενδιαφέρον να υπολογίσουμε

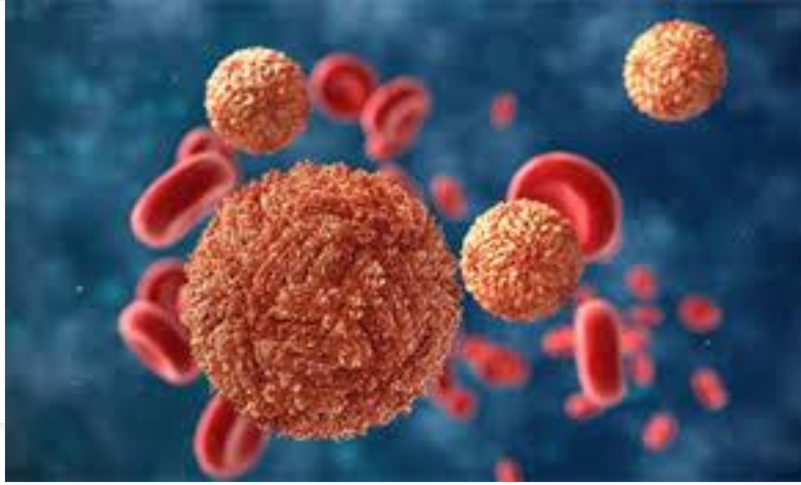
$P(B|\Theta)$, $P(B|\Theta')$ και $P(\Theta)$

Παράδειγμα (συν)

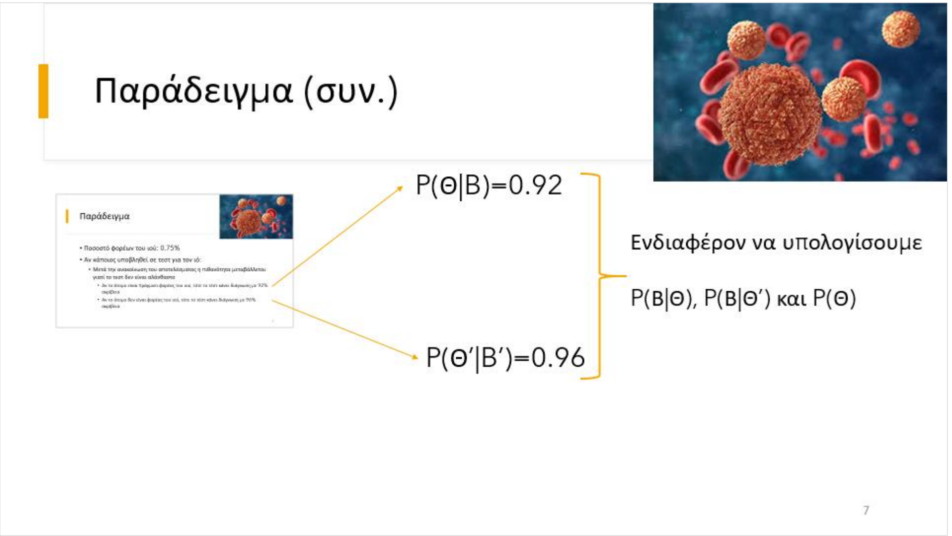


- Ποια είναι η πιθανότητα κάποιος να υποβληθεί στο τεστ και να «βγει» θετικό;

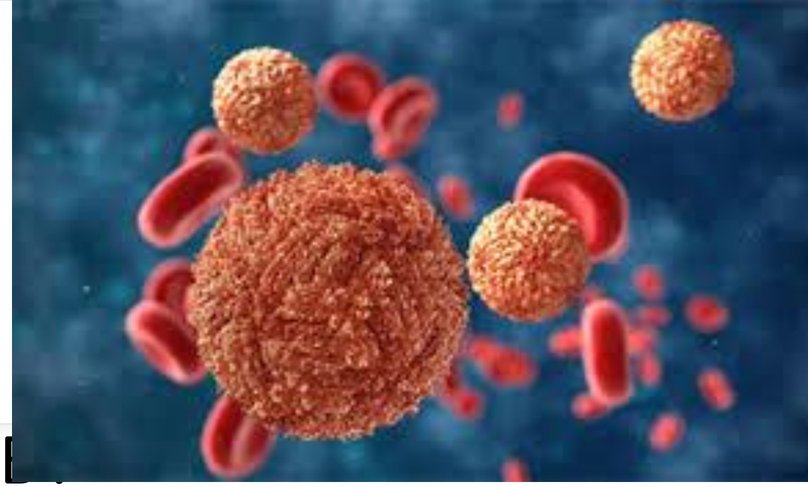
Παράδειγμα (συν)



- Ενδεχόμενα:
  - B: ο κάτοικος είναι φορέας του ιού
  - Θ: το αποτέλεσμα του τεστ είναι θετικό
- Δεδομένα:
  - $P(B)=0.0075$, $P(B')=1-P(B)=0.9925$
  - $P(\Theta|B)=0.92$, $P(\Theta'|B')=0.96$
- Ζητούμενο: η πιθανότητα $P(\Theta)$

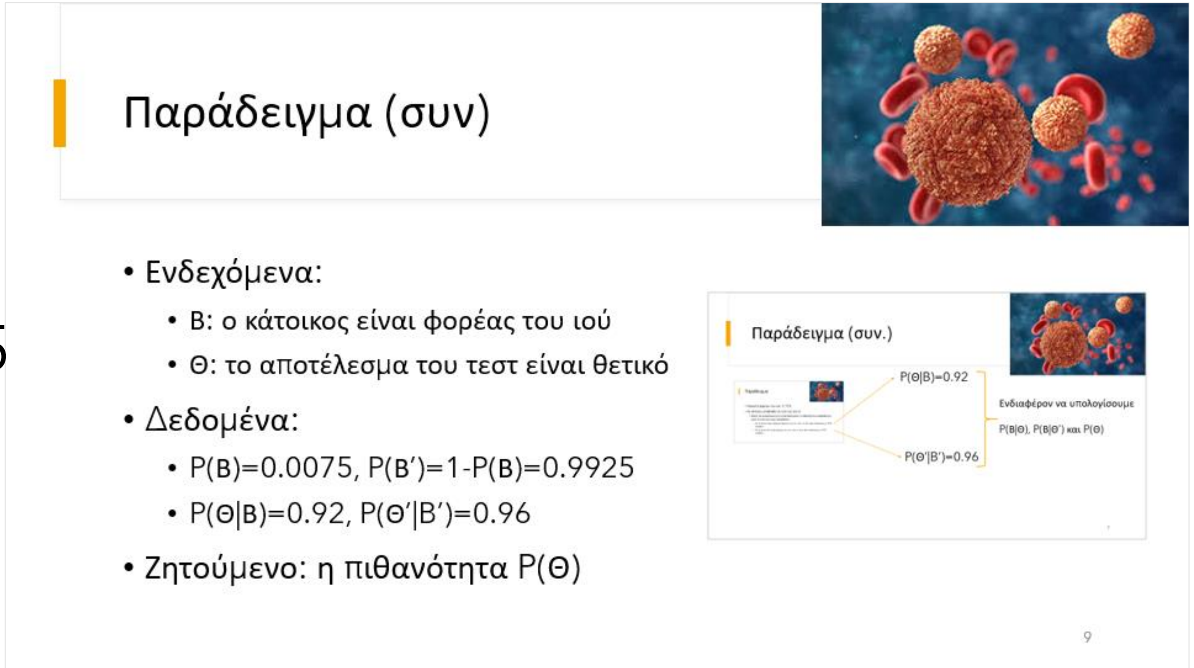


Παράδειγμα (συν.)

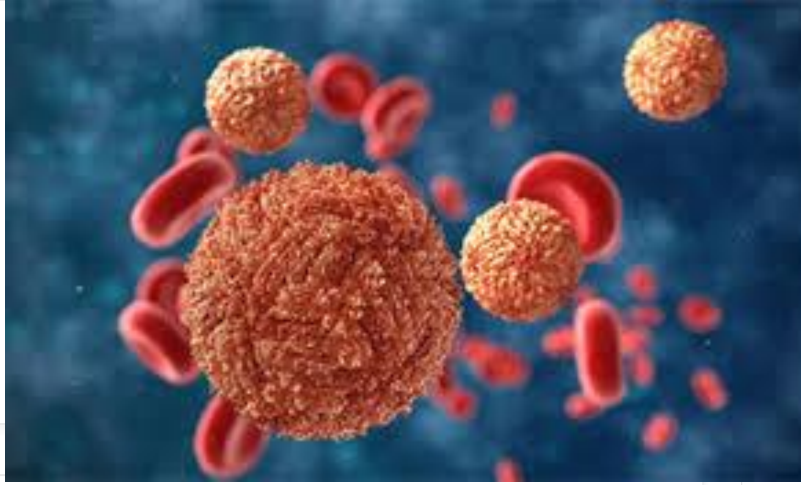


- Θεωρούμε την διαμέριση του ΔΧ Ω βάσει των B, B
  - Προφανώς είναι διαμέριση γιατί $B \cup B' = \Omega$ και $B \cap B' = \emptyset$
- Από το θεώρημα της ολική πιθανότητας έχουμε:
  - $P(\Theta) = P(\Theta|B)\,P(B) + P(\Theta|B')\,P(B')$

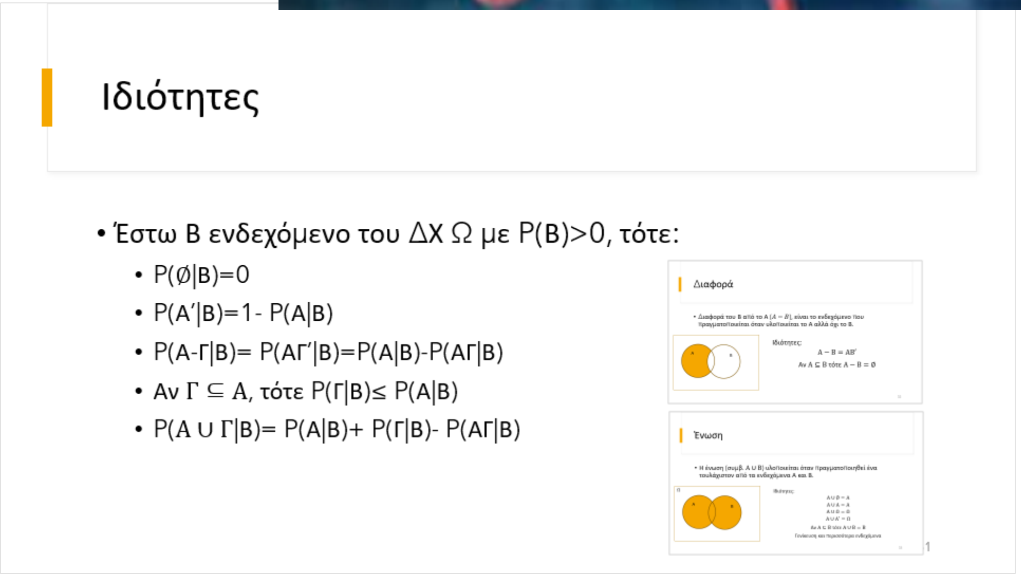
$$\underbrace{P(\Theta|B)}_{0.92}\quad \underbrace{P(B)}_{0.0075}\quad \underbrace{P(\Theta|B')}_{?}\quad \underbrace{P(B')}_{0.9925}$$



Παράδειγμα (συν.)

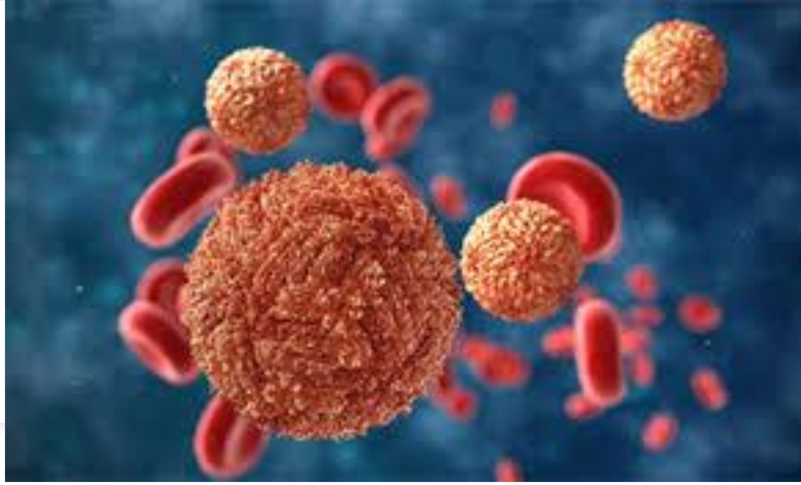


- $P(\Theta|B')=1-P(\Theta'|B')=1-0.96=0.04$



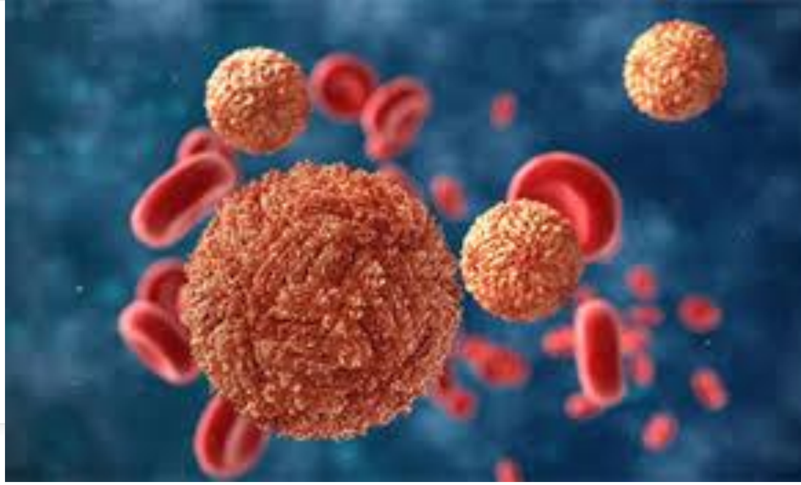
- $P(\Theta) = P(\Theta|B)\ P(B) + P(\Theta|B')\ P(B')=$ $0.92*0.0075+0.04*0.9925=0.0466$

Παράδειγμα (συν)



- Ποια είναι η πιθανότητα κάποιος κάτοικος να υποβληθεί στο τεστ και όταν να «βγει» θετικό, να είναι πράγματι φορέας του ιού;

Παράδειγμα (συν)



Τι γνωρίζουμε/έχουμε υπολογίσει έως τώρα

- Ενδεχόμενα:
  - Β: ο κάτοικος είναι φορέας του ιού
  - Θ: το αποτέλεσμα του τεστ είναι θετικό

- Πιθανότητες από τα δεδομένα:

  $P(B)=0.0075$

  $P(\Theta|B)=0.92$

  $P(\Theta'|B')=0.96$

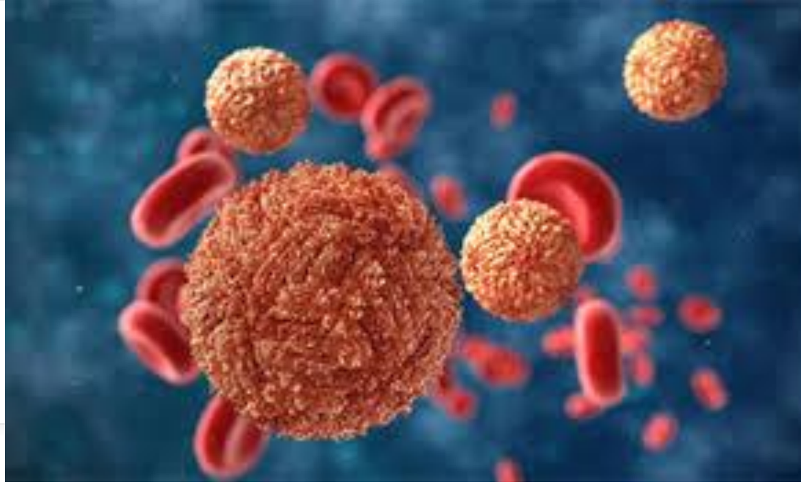
Έχουμε υπολογίσει

- Πιθανότητα κάποιος να υποβληθεί στο τεστ και να «βγει» θετικό

  $P(\Theta)=0.0466$

Ποια είναι η πιθανότητα κάποιος κάτοικος να υποβληθεί στο τεστ και όταν να «βγει» θετικό, να είναι πράγματι φορέας του ιού;

Παράδειγμα (συν)



Από τον ορισμό της δεσμευμένης πιθανότητας

Δεσμευμένη πιθανότητα (ορισμός)

- Έστω ενδεχόμενα Α και Β του ΔΧ Ω, και P(B)>0, τότε η δεσμευμένη πιθανότητα του Α δοθέντος του Β (conditional probability) ισούται με

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

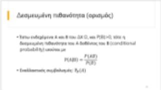
- Εναλλακτικός συμβολισμός: $\mathrm{P}_B(A)$

34

και το πολλαπλασιαστικό τύπο

Πολλαπλασιαστικός τύπος

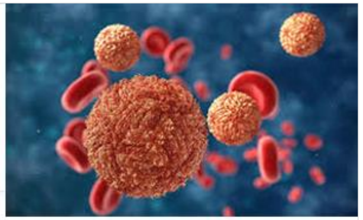
- Έστω ενδεχόμενα Α και Β του ΔΧ Ω, και P(B)>0, τότε ισχύει:
  - P(AB)=P(B)*P(A|B)
- Έστω ενδεχόμενα Α και Β του ΔΧ Ω, και P(A)>0, τότε ισχύει:
  - P(AB)=P(A)*P(B|A)



35

$$\mathrm{P(B|\Theta)} = \frac{P(\mathrm{B\Theta})}{P(\Theta)} = \frac{P(\mathrm{\Theta IB})\mathrm{P(B)}}{P(\Theta)} = \frac{0.92 * 0.0075}{0.0466} = 0.1480$$

Παράδειγμα (συν)



Τι γνωρίζουμε/έχουμε υπολογίσει έως τώρα

- Ενδεχόμενα:
  - Β: ο κάτοικος είναι φορέας του ιού
  - Θ: το αποτέλεσμα του τεστ είναι θετικό
- Πιθανότητες από τα δεδομένα:
  - P(B)=0.0075
  - P(Θ|B)=0.92
  - P(Θ'|B')=0.96

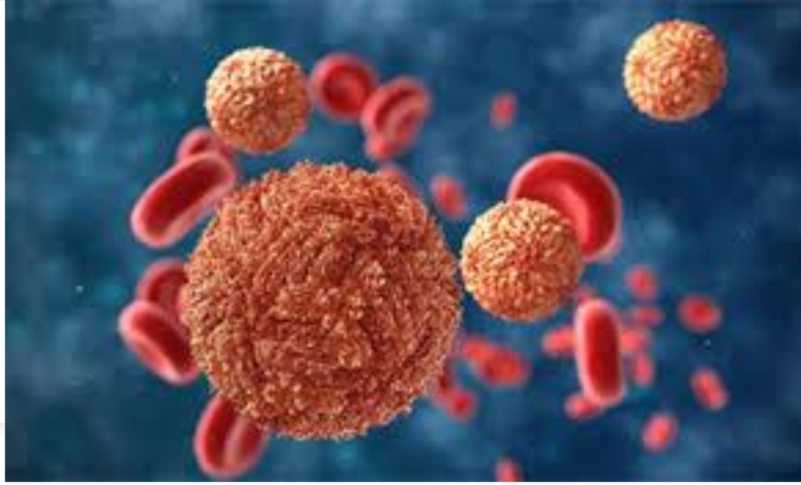
Έχουμε υπολογίσει

- Πιθανότητα κάποιος να υποβληθεί στο τεστ και να «βγει» θετικό
  - P(Θ) =0.0466

Ποια είναι η πιθανότητα κάποιος κάτοικος να υποβληθεί στο τεστ και όταν να «βγει» θετικό, να είναι πράγματι φορέας του ιού;

13

Παράδειγμα (συν)

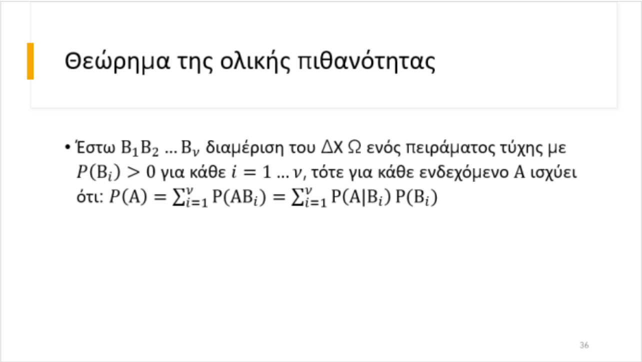


- Ποια είναι η πιθανότητα κάποιος κάτοικος να υποβληθεί στο τεστ και όταν να «βγει» θετικό, να είναι πράγματι φορέας του ιού;
- Απάντηση: $\mathrm{P(B|\Theta)} = 0.1480$
- Τι σημαίνει ο αριθμός;
- Για έναν τυχαία επιλεγμένο κάτοικο που έκανε το τέστ για τον ιό, και το αποτέλεσμα βγήκε θετικό, η πιθανότητα ο συγκεκριμένος κάτοικος να έχει πραγματικά τον ιό είναι 14.8%!!!
- Η πιθανότητα ο συγκεκριμένος κάτοικος να μην φέρει τον ιό, παρόλο που έχει πάρει θετικό αποτέλεσμα στο τέστ είναι 85,2%!!!

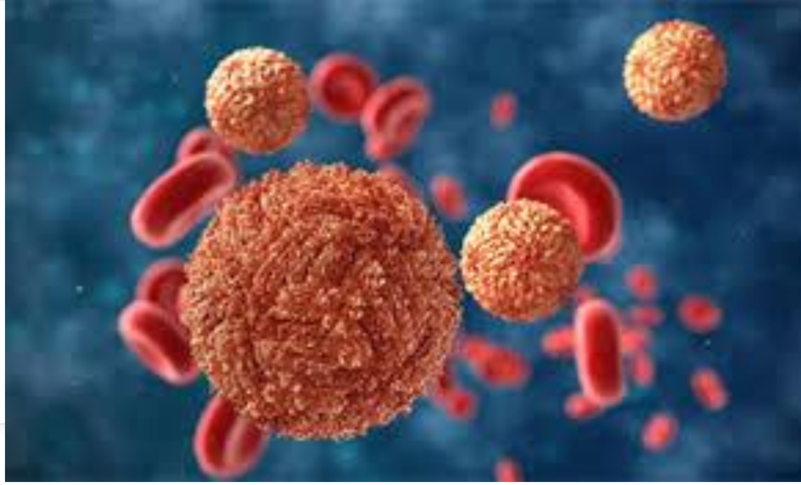
Θεώρημα του Bayes

- Έστω $\mathrm{B}_1 \mathrm{B}_2 \ldots \mathrm{B}_\nu$ διαμέριση του ΔΧ Ω με $P(B_i) > 0$ για όλα $i = 1,2, \ldots, n$ τότε για κάθε ενδεχόμενο A του Ω με $P(A) > 0$ ισχύει

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^{\nu} P(A|B_i)P(B_i)}$$



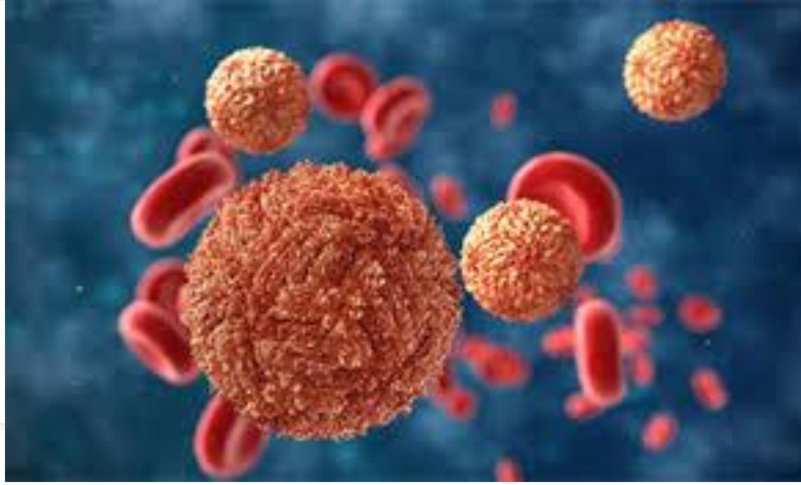
Παράδειγμα (συν)



- Ποια είναι η πιθανότητα κάποιος κάτοικος να υποβληθεί στο τεστ και όταν να «βγει» θετικό, να είναι πράγματι φορέας του ιού;
- Απάντηση: $\mathrm{P}(\mathrm{B}|\Theta) = 0.1480$
- Τι σημαίνει ο αριθμός;
- Για έναν τυχαία επιλεγμένο κάτοικο που έκανε το τέστ για τον ιό, και το αποτέλεσμα βγήκε θετικό, η πιθανότητα ο συγκεκριμένος κάτοικος να έχει πραγματικά τον ιό είναι 14.8%!!!
- Η πιθανότητα ο συγκεκριμένος κάτοικος να μην φέρει τον ιό, παρόλο που έχει πάρει θετικό αποτέλεσμα στο τέστ είναι 85,2%!!!

?

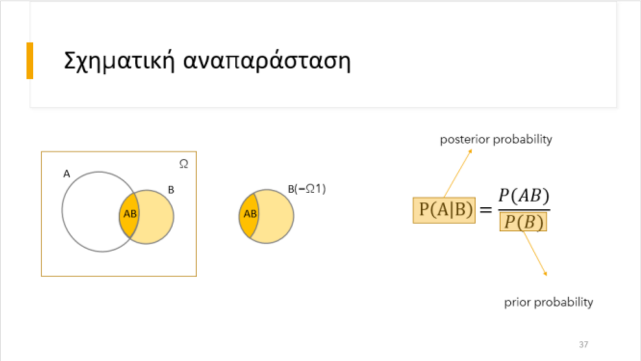
Παράδειγμα (συν)

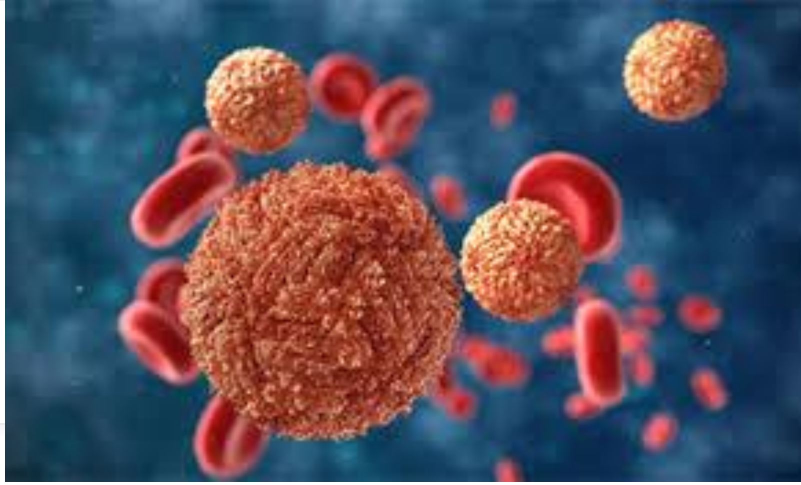


Δεδομένα:

- Τα διαγνωστικό τεστ είναι πολύ αξιόπιστο. Το τεστ που διενεργείται για το αν ένας κάτοικος φέρει τον ιό:
  - κάνει σωστή διάγνωση με πιθανότητα 0.92 για τους φορείς (δηλ. αν επιλέξουμε 100 κατοίκους που φέρουν τον ιό και τους υποβάλουμε στο τεστ, για τους 92 θα βγει θετικό).
  - κάνει σωστή διάγνωση με πιθανότητα 0.96 για τους μη-φορείς (δηλ. αν επιλέξουμε 100 κατοίκους που δεν φέρουν τον ιό και τους υποβάλουμε στο τεστ, για τους 96 θα βγει αρνητικό).
    - Συνεπώς: Υπάρχει πιθανότητα 0.04 ή 4 στους 100 μη-φορείς που το τεστ τους χαρακτηρίζει ως φορείς.
- Όταν κάποιος κάτοικος υποβληθεί στο τεστ και το αποτέλεσμα βγει θετικό, τότε υπάρχει 0.1480 πιθανότητα να είναι φορέας (δηλαδή από τα 100 θετικά τεστ, μόνο 15 κάτοικοι είναι φορείς)
  - Συνεπώς, όταν κάποιος κάτοικος υποβληθεί στο τεστ και το αποτέλεσμα βγει θετικό, τότε κατά 85% πιθανότητα δεν είναι φορέας!!!

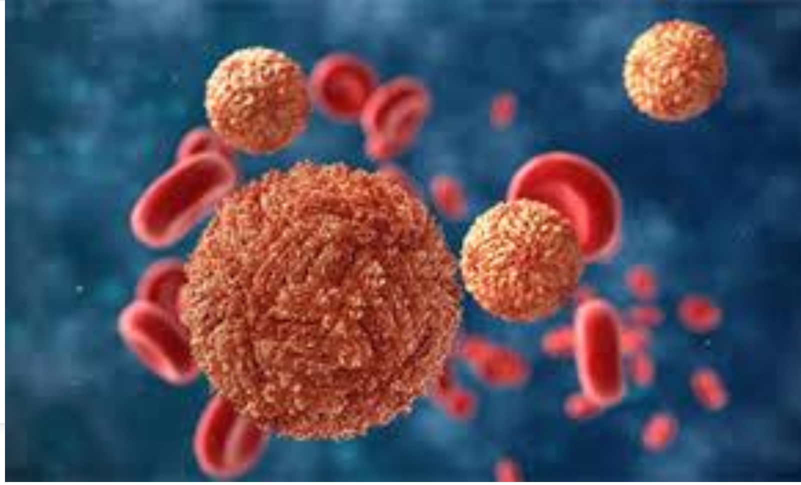
Παράδειγμα (συν)





- Γιατί;
- Υποθέτουμε μια πόλη 10000 κατοίκων:
  - Η ασθένεια είναι σπάνια (δηλ. P(B)=0.0075) οπότε 75 κάτοικοι φέρουν τον ιό και οι 9925 είναι μη φορείς
  - Από τους 75 φορεις το τεστ θα εντοπίσει του 0.92*75=69, 6 φορείς δεν θα εντοπισθούν
  - Από τους 9925 μη-φορείς, οι 9528 θα βγουν αρνητικοί, ενώ 9925- 9528=397 θα βγουν θετικοί.
  - Συνεπώς στα 10000 τεστ, 397+69=466 θα βγουν θετικά.
  - Το πηλίκο 69/466=0.1480 είναι η πιθανότητα ένας κάτοικος με θετικό αποτέλεσμα στο τεστ να είναι φορέας.
  - Αυτό συμβαίνει γιατί η πιθανότητα 0.1480 κυριαρχείται από τους 397 μη φορείς που το τέστ εντοπίζει ως θετικούς.

Παράδειγμα (συν)



- Το «παράξενο» αποτέλεσμα οφείλεται στην σπανιότητα της ασθένειας στον γενικό πληθυσμό (δηλαδή 75 στους 10000)

	P(B)			
	0.0075	0.01	0.02	0.05
P(B|Θ)	0.1480	0.1885	0.3194	0.5476

- **Μια διαφορετική οπτική γωνία:** Στην πραγματικότητα η πιθανότητα 14.8% είναι 20x φορές μεγαλύτερο από την αρχικό ποσοστό 0.75% στο γενικό πληθυσμό!!!

Ανεξάρτητα ενδεχόμενα

Ανεξάρτητα ενδεχόμενα

- Στα παραδείγματα δεσμευμένης πιθανότητας, είδαμε ότι η πιθανότητες P(A) και P(A|B) ήταν γενικά διαφορετικές:
  - $P(A) < P(A|B)$
  - $P(A) > P(A|B)$
  - Δηλαδή: η γνώση ότι εμφανίστηκε το ενδεχόμενο Β, μεταβάλει την πιθανότητα/«γνώση» μας για το Α
- Επίσης όταν τα ενδεχόμενα Α, Β είναι ξένα μεταξύ τους, τότε
  - $P(A|B)=0$ και $P(B|A)=0$
  - Δηλαδή: η γνώση ότι εμφανίστηκε το ενδεχόμενο Β, δεν μεταβάλει την πιθανότητα/«γνώση» μας για το Α (ότι είναι αδύνατο)
- Όταν το Α είναι υποσύνολο του Β, η γνώση ότι συνέβη το Β μας δίνει απόλυτη γνώση για το Α
  - $P(A|B)=1$

Ανεξάρτητα ενδεχόμενα

- Τι συμβαίνει στην περίπτωση $P(A|B)= P(A|B')=P(A)$;
- Ερμηνεία:
  - Η γνώση ότι το ενδεχόμενο Β συνέβη ή δεν-συνέβη δεν προσφέρει καμία νέα γνώση σχετικά με το Α.
  - Η πραγματοποίηση ή μη του ενδεχομένου Β, δεν έχει καμία επίδραση στην πραγματοποίηση του Α.

Ανεξάρτητα ενδεχόμενα

- $P(A|B) = P(A) \leftrightarrow \frac{P(AB)}{P(B)} = P(A) \leftrightarrow P(AB) = P(A)P(B) \leftrightarrow$
$$\frac{P(AB)}{P(A)} = P(B) \leftrightarrow P(B|A) = P(B)$$
- Αν η πραγματοποίηση του B δεν έχει καμία επίδραση στο A, τότε το A δεν έχει καμία επίδραση στο B (και αντιστρόφως).

Ανεξάρτητα ενδεχόμενα (ορισμός)

- Δύο ενδεχόμενα A και B του ΔΧ Ω λέγονται ανεξάρτητα (independent events) αν ισχύει $P(AB)=P(A)P(B)$
- Αν ισχύει $P(AB) = P(A)P(B)$ τότε τα ενδεχόμενα A και είναι ανεξάρτητα.

Χρήσιμες προτάσεις

- Αν A,B είναι ανεξάρτητα ενδεχόμενα τότε είναι ανεξάρτητα και τα:
  - A' και B'
  - A και B'
  - A' και B
- Αν $P(A|B)=P(A|B')$ τότε $P(A|B)=P(A|B')=P(A)$
- Το βέβαιο Ω και αδύνατο ενδεχόμενο ∅ είναι ανεξάρτητο με κάθε ενδεχόμενο A

Παράδειγμα

- Ρίχνουμε ένα αμερόληπτο νόμισμα 2 φορές
- Μας ενδιαφέρουν τα ενδεχόμενα:
  - Η ένδειξή στην δεύτερη ρίψη είναι «κεφαλή»
  - Η ένδειξή στην πρώτη ρίψη είναι «γράμματα»

Παράδειγμα (συν)

- Ο δειγματικός χώρος είναι Ω={κκ, γγ, γκ, κγ}
- A={κκ, γκ} και B={γγ, γκ}
- P(A)=2/4=0.5 και P(B)=2/4=0.5
- P(AB)=P({γκ})=1/4=0.25
- Οπότε P(A)P(B)=0.5*0.5=0.25=P(AB), δηλαδή ανεξάρτητα ενδεχόμενα

Ανεξάρτητα ενδεχόμενα (ορισμός ν>2)

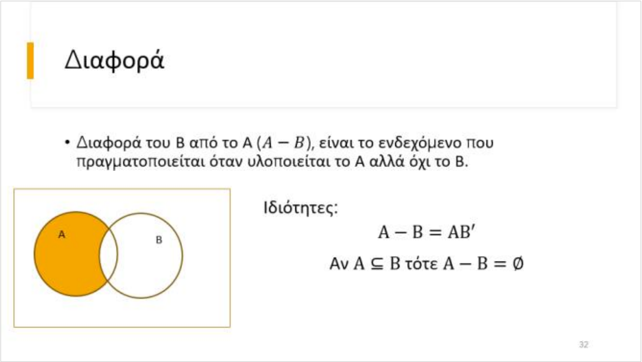
- Θα λέμε ότι ν ενδεχόμενα $A_1 A_2 \dots A_\nu$ ενός ΔΧ Ω είναι ανεξάρτητα ή τελείως ανεξάρτητα αν για οποιοδήποτε σύνολο k δεικτών $i_1 i_2 \dots i_k$ από το σύνολο {1,2,…,ν} ισχύει

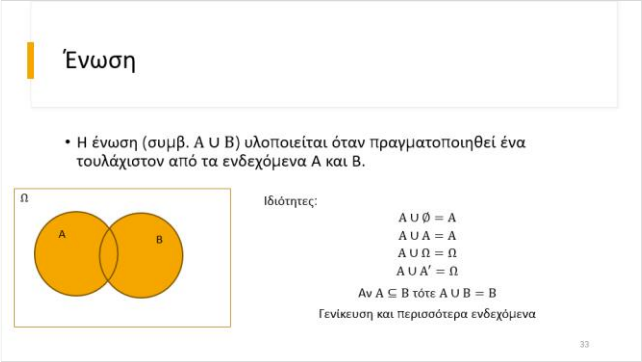
$$P(A_1 A_2 \dots A_k) = P(A_1)P(A_2) \dots P(A_k)$$

Backup

Ιδιότητες

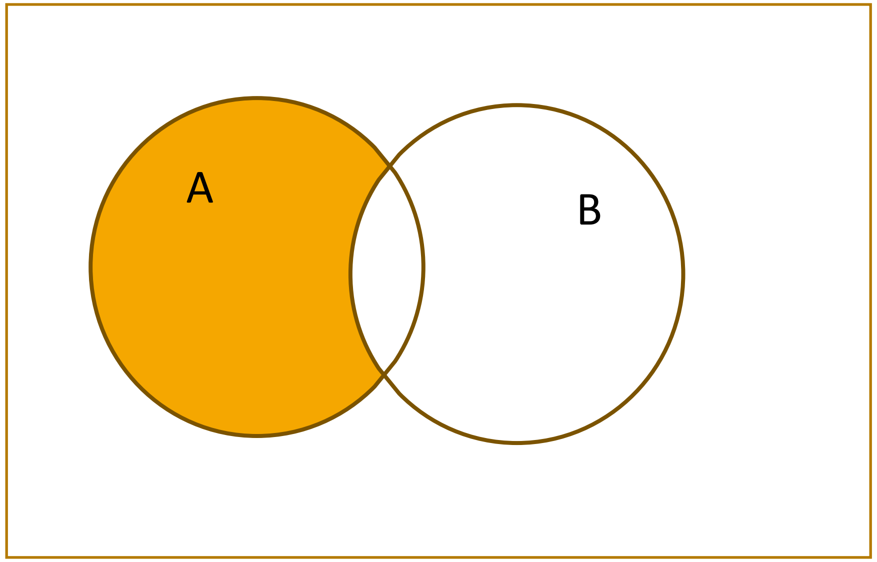
- Έστω Β ενδεχόμενο του ΔΧ Ω με P(Β)>0, τότε:
  - P(Ø|B)=0
  - P(A′|B)=1- P(A|B)
  - P(A-Γ|B)= P(AΓ′|B)=P(A|B)-P(AΓ|B)
  - Αν $\Gamma \subseteq A$, τότε P(Γ|B)≤ P(A|B)
  - P($A \cup \Gamma$|B)= P(A|B)+ P(Γ|B)- P(AΓ|B)





Διαφορά

- Διαφορά του Β από το Α $(A - B)$, είναι το ενδεχόμενο που πραγματοποιείται όταν υλοποιείται το Α αλλά όχι το Β.



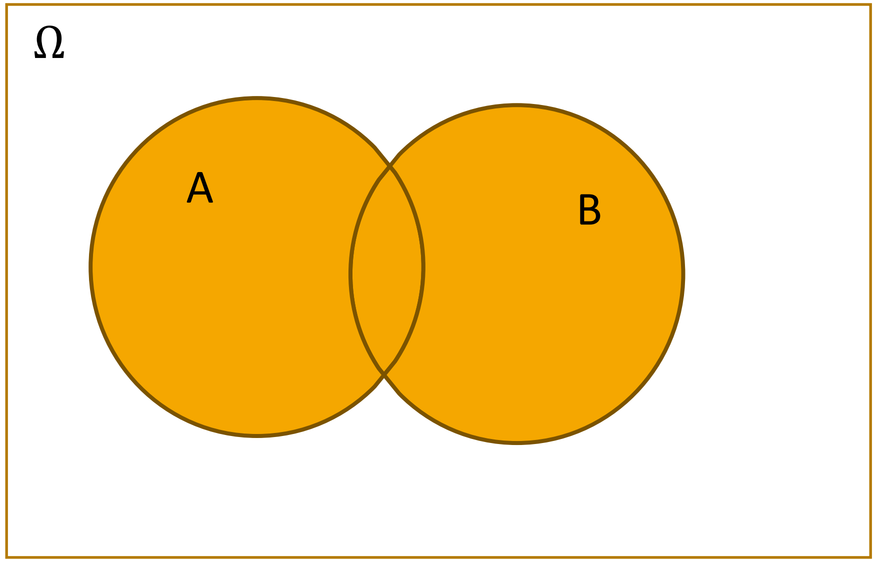
Ιδιότητες:

$$\mathrm{A} - \mathrm{B} = \mathrm{AB}'$$

$$\text{Αν } \mathrm{A} \subseteq \mathrm{B} \text{ τότε } \mathrm{A} - \mathrm{B} = \emptyset$$

Ένωση

- Η ένωση (συμβ. $A \cup B$) υλοποιείται όταν πραγματοποιηθεί ένα τουλάχιστον από τα ενδεχόμενα Α και Β.



Ιδιότητες:

$$A \cup \emptyset = A$$
$$A \cup A = A$$
$$A \cup \Omega = \Omega$$
$$A \cup A' = \Omega$$

Αν $A \subseteq B$ τότε $A \cup B = B$

Γενίκευση και περισσότερα ενδεχόμενα

Δεσμευμένη πιθανότητα (ορισμός)

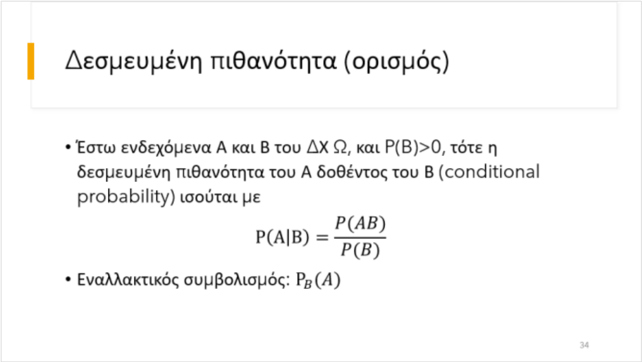
- Έστω ενδεχόμενα A και B του ΔΧ Ω, και P(B)>0, τότε η δεσμευμένη πιθανότητα του A δοθέντος του B (conditional probability) ισούται με

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

- Εναλλακτικός συμβολισμός: $\mathrm{P}_B(A)$

Πολλαπλασιαστικός τύπος

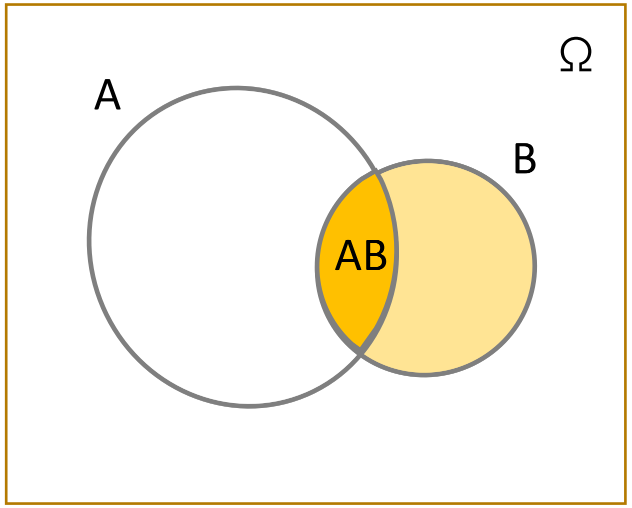
- Έστω ενδεχόμενα A και B του ΔΧ Ω, και P(B)>0, τότε ισχύει:
  - $P(AB)=P(B)*P(A|B)$
- Έστω ενδεχόμενα A και B του ΔΧ Ω, και P(A)>0, τότε ισχύει:
  - $P(AB)=P(A)*P(B|A)$

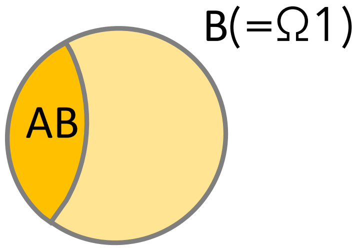


Θεώρημα της ολικής πιθανότητας

- Έστω $\mathrm{B}_1\mathrm{B}_2 \ldots \mathrm{B}_\nu$ διαμέριση του ΔΧ Ω ενός πειράματος τύχης με $P(\mathrm{B}_i) > 0$ για κάθε $i = 1 \ldots \nu$, τότε για κάθε ενδεχόμενο A ισχύει ότι: $P(\mathrm{A}) = \sum_{i=1}^{\nu} \mathrm{P}(\mathrm{AB}_i) = \sum_{i=1}^{\nu} \mathrm{P}(\mathrm{A}|\mathrm{B}_i)\,\mathrm{P}(\mathrm{B}_i)$

Σχηματική αναπαράσταση





posterior probability

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

prior probability