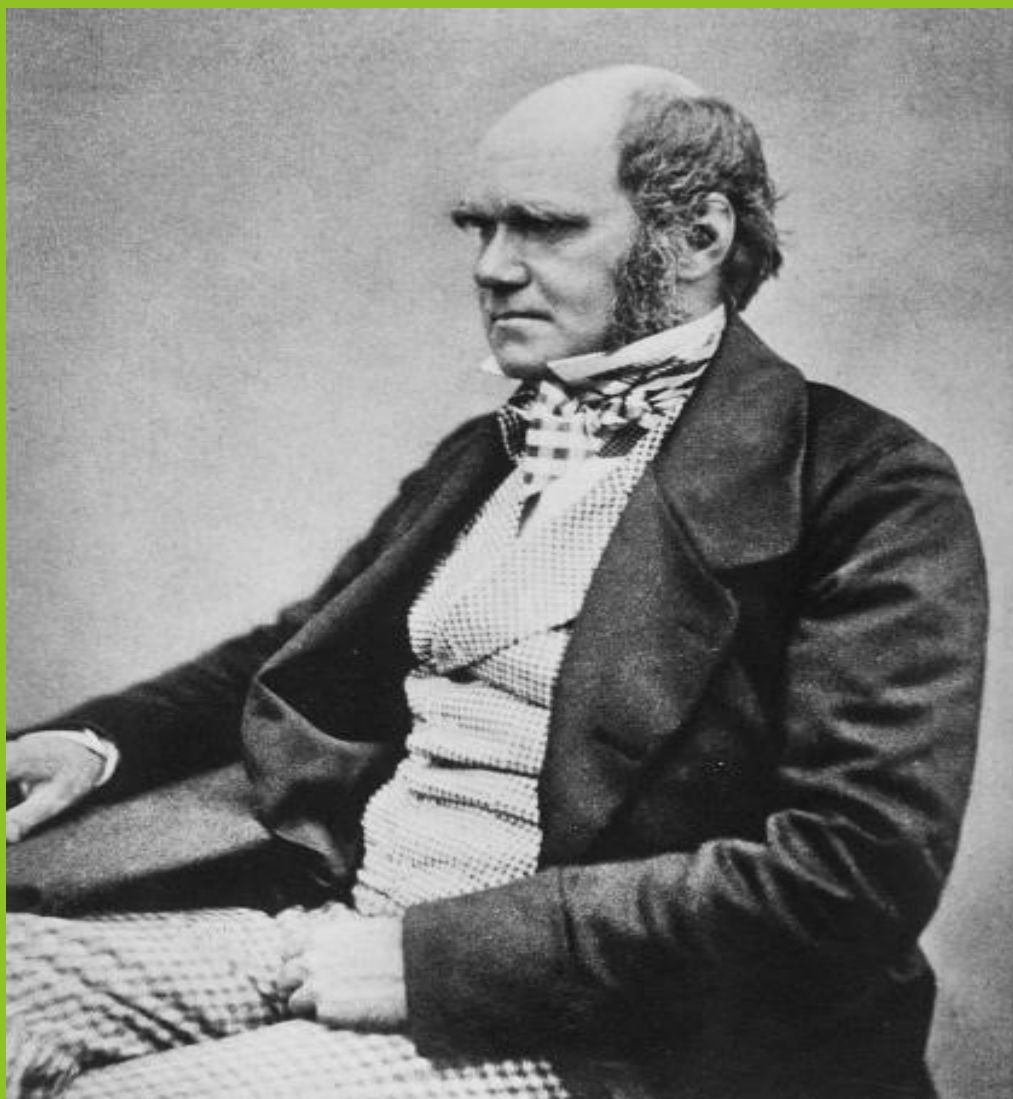


# Συστήματα Επιχειρηματικής Ευφυΐας

Γενετικοί αλγόριθμοι (GA) : Από τον Δαρβίνο (1859) στον J. Holland (1975).

(Ένα ταξίδι στον υπέροχο κόσμο της επιλογής, της διασταύρωσης και της μετάλλαξης).



Charles Darwin 1809 - 1882



John Henry Holland 1929 - 2015

# Δεύτερη Διάλεξη - Περιεχόμενα (1)

- ▶ Έννοιες της Θεωρίας Εξέλιξης των ειδών
- ▶ Στοιχειώδεις έννοιες Βιολογίας
- ▶ Η γέννηση των Γενετικών αλγορίθμων
- ▶ Ο κανόνας της Φυσικής Επιλογής
- ▶ Γενική μορφή Γενετικών αλγορίθμων
- ▶ Τα 5 απαραίτητα συστατικά του Γενετικού αλγορίθμου
- ▶ Κωδικοποίηση λύσεων
- ▶ Συνάρτηση καταλληλότητας
- ▶ Επιλογή γονέων
- ▶ Αλγόριθμοι επιλογής

## Δεύτερη Διάλεξη - Περιεχόμενα (2)

- ▶ Διασταύρωση (Crossover ή mating)
- ▶ Μετάλλαξη (mutation)
- ▶ Ανανέωση πληθυσμού
- ▶ Σύγκλιση
- ▶ Πλεονεκτήματα των Γενετικών αλγορίθμων
- ▶ Περιορισμοί των Γενετικών αλγορίθμων

# Έννοιες της Θεωρίας Εξέλιξης των ειδών (1)

- ▶ Η Θεωρία Εξέλιξης των ειδών (Evolution of Species) διατυπώθηκε από τον Δαρβίνο (1859)
- ▶ Βασίζεται σε δύο βασικές υποθέσεις:
  - ❑ α) ότι όλα τα είδη προήλθαν, μέσω τροποποιήσεων, από κοινά προγονικά είδη και
  - ❑ β) ότι οι τροποποιήσεις αυτές οφείλονται στη φυσική επιλογή που δρα πάνω στις διαφορές που υπάρχουν μεταξύ των ατόμων ενός είδους.
- ▶ Επίσης, υιοθετεί την υπόθεση ότι τα άτομα ενός πληθυσμού προσαρμόζονται στο περιβάλλον με διαφορετικό τρόπο και ότι η προσαρμογή δημιουργεί νέα κληροδοτούμενα χαρακτηριστικά στους απογόνους.
- ▶ Ακόμη, τα καλά προσαρμοσμένα άτομα (άρα τα ισχυρά-εύρωστα) έχουν μεγαλύτερη πιθανότητα επιβίωσης και αναπαραγωγής έναντι των ασθενικών.

Η πρωτότυπη δημοσίευση του Δαρβίνου υπάρχει στο link:

<http://darwin-online.org.uk/content/frameset?itemID=F373&viewtype=text&pageseq=1>

## Έννοιες της Θεωρίας Εξέλιξης των ειδών (2)

- ▶ Ο Δαρβίνος αναγνώρισε ίχνη εξελικτικών ιδεών στο έργο *Περί Φυσικής Ακροάσεως* του **Αριστοτέλη**. Επίσης, ο **Αναξίμανδρος** (610-546 π.Χ) και ο **Εμπεδοκλής** (492-432 π.Χ) υποστήριξαν ότι τα ζώα μεταλλάσσονται και αφανίζονται.
- ▶ Η Θεωρία του Δαρβίνου δίνει λύση στο πρόβλημα της δυσκολίας της ταξινόμησης των ειδών (λόγω της συνάφειάς τους) και εξηγεί την ποικιλομορφία της φύσης.
- ▶ Αν και το έργο του Δαρβίνου ανέτρεπε τις κατεστημένες θρησκευτικές αντιλήψεις της εποχής του, μέσα σε 20 χρόνια έγινε ευρύτατα αποδεκτό τόσο από την Ακαδημαϊκή κοινότητα όσο και από τον απλό κόσμο, χάρη στην εκλαϊκευμένη παρουσίασή του.
- ▶ Αποτελεί τον πυρήνα της σύγχρονης αντίληψης σχετικά με τις ιδιότητες και τη συμπεριφορά των διάφορων πληθυσμών (με την ευρύτερη έννοια).

# Στοιχειώδεις έννοιες Βιολογίας (1)

- ▶ **Χρωμοσώματα (chromosomes)** : είναι πολύπλοκα οργανικά μόρια τα οποία κωδικοποιούν τη δομή και τα χαρακτηριστικά τους.
  - ❑ Οι προσπάθειες των ατόμων να προσαρμοστούν στο περιβάλλον τους, έχει σαν αποτέλεσμα την αλλαγή των χαρακτηριστικών των ατόμων. Όλες αυτές οι αλλαγές έχουν άμεση σχέση με αντίστοιχες αλλαγές στα χρωμοσώματα των ατόμων.
- ▶ **Γονίδια (genes)** : είναι τα μικρότερα δομικά μέρη από τα οποία αποτελούνται τα χρωμοσώματα.
  - ❑ Τα γονίδια «φέρουν» την γενετική πληροφορία που χαρακτηρίζει το κάθε άτομο.
  - ❑ Η δημιουργία ενός νέου ατόμου προϋποθέτει μια διαδικασία αποκωδικοποίησης της γενετικής πληροφορίας.



## Στοιχειώδεις έννοιες Βιολογίας (2)

- ▶ Γονότυπος (genotype) : είναι το σύνολο της κωδικοποιημένης γενετικής πληροφορίας των γονιδίων.
- ▶ Φαινότυπος (phenotype) : η κωδικοποιημένη γενετική πληροφορία (γονότυπος) προσδίδει κάποια ορατά χαρακτηριστικά στα άτομα. Η ορατή εικόνα των ατόμων καλείται φαινότυπος.
- ▶ Αναπαραγωγή (reproduction) : είναι η διαδικασία παραγωγής νέου ατόμου από δύο γονείς.
  - Το νέο άτομο κληρονομεί γενετική πληροφορία από την μητέρα και τον πατέρα εξ ίσου. Άρα, κάθε χρωμόσωμα περιέχει μισά γονίδια από την μητέρα και μισά από τον πατέρα.
  - Άρα, λογικά, ο φαινότυπος του ατόμου θα ομοιάζει με τους φαινοτύπους των γονέων. (βλ. μετάλλαξη)



# Στοιχειώδεις έννοιες Βιολογίας (3)

- ▶ **Αλληλόμορφα γονίδια (alleles).** Δύο γονίδια (ένα από κάθε γονέα) που είναι υπεύθυνα για τον καθορισμό της τιμής του ίδιου χαρακτηριστικού του απογόνου (δηλαδή, που διεκδικούν την ίδια θέση σε ένα χρωμόσωμα) λέγονται *αλληλόμορφα (alleles)*.
- ▶ **Κυρίαρχο ή επικρατές (dominant) γονίδιο :** λέγεται αυτό που, μεταξύ δύο αλληλόμορφων, τελικά θα επικρατήσει και θα ευθύνεται για την τιμή του γονιδίου του απογόνου και συνεπώς για τον φαινότυπο.
- ▶ **υπολειπόμενο (recessive) γονίδιο :** λέγεται αυτό που τελικά δεν επικρατεί.
  - Τόσο το κυρίαρχο όσο και το υπολειπόμενο γονίδιο κληροδοτούνται στις επόμενες γενεές.

# Στοιχειώδεις έννοιες Βιολογίας (4)

► **Μετάλλαξη (mutation)** : κατά την διαδικασία αντιγραφής των γονιδίων των γονέων στο νέο απόγονο, συνήθως συμβαίνουν λάθη. Έτσι, το άτομο δεν κληρονομεί την ακριβή κωδικοποιημένη πληροφορία των χρωμοσωμάτων των γονέων. Το ίδιο φαινόμενο μπορεί να είναι αποτέλεσμα εξωγενών παραγόντων.

- Ένα μεταλλαγμένο χρωμόσωμα περιέχει διαφορετική πληροφορία σε επίπεδο γονιδίου από αυτήν που θα έπρεπε να είχε αν δεν συνέβαινε η μετάλλαξη.
- Η μετάλλαξη μπορεί να αποβεί καταστροφική (π.χ. το κάπνισμα προκαλεί μετάλλαξη στα γονίδια με αποτέλεσμα την εμφάνιση του καρκίνου).

## Στοιχειώδεις έννοιες Βιολογίας (5)

- Όμως, το λάθος το οποίο οφείλεται στη μετάλλαξη μερικές φορές αποβαίνει σε όφελος του ατόμου, καθώς είναι δυνατόν να βελτιώσει τον γονότυπο και άρα τον φαινότυπο.
- Αν δεν υπήρχε το φαινόμενο της μετάλλαξης τότε όλοι οι απόγονοι θα ήταν πανομοιότυποι των γονέων τους. Άρα, δεν θα υπήρχε ούτε ποικιλομορφία, ούτε εξέλιξη.

# Η γέννηση των Γενετικών αλγορίθμων

► Εμπνευστής : John Holland (1975)

- ❑ Θεώρησε ότι κάποιες έννοιες της Θεωρίας Εξέλιξης μπορούν να προσομοιωθούν από ΗΥ, ως αλγόριθμοι, προκειμένου να επιλυθούν δύσκολα προβλήματα.
- ❑ Χρησιμοποίησε σειρές από bits για να αναπαραστήσει τις υποψήφιες λύσεις (χρωμοσώματα).

# Ο κανόνας της Φυσικής Επιλογής (1)

- ▶ Ο Holland στήριξε την πρότασή του στον κανόνα της Φυσικής επιλογής, σύμφωνα με την οποία:
  - Οι οργανισμοί που δεν μπορούν να προσαρμοστούν στο περιβάλλον (λύσεις που δεν ικανοποιούν κάποια κριτήρια) εκλείπουν.
  - Οι απόγονοι (λύσεις επόμενης γενεάς) διαφοροποιούνται από τους προγόνους (λύσεις προηγούμενης γενεάς), ενώ συνήθως υπερσχύουν αυτοί που συγκεντρώνουν τα καλύτερα χαρακτηριστικά των προγόνων (δηλαδή οι καλύτερες λύσεις) και πολλαπλασιάζονται.

## Ο κανόνας της Φυσικής Επιλογής (2)

- ❑ Με μικρή πιθανότητα συμβαίνουν μεταλλάξεις, οι οποίες, τις περισσότερες φορές, οδηγούν τους απογόνους στον θάνατο (απόρριψη λύσεων). Είναι όμως πιθανόν να δημιουργήσουν έναν απόγονο (λύση) πιο εύρωστο από τους προγόνους του.
- ❑ Τα άτομα ενός πληθυσμού έχουν την ικανότητα να προσαρμόζονται (και να επιβιώνουν) σε ένα μεταβαλλόμενο περιβάλλον, με την προϋπόθεση ότι το περιβάλλον δεν έχει μεγάλο ρυθμό μεταβολής.
- ▶ Χαρακτηριστικό παράδειγμα της εφαρμογής του κανόνα της Φυσικής επιλογής αποτελεί ένας πληθυσμός από λαγούς (με διαφορετικό βαθμό ευφυίας και γρηγοράδας), ο οποίος απειλείται από αλεπούδες. Οι γρηγορότεροι και εξυπνότεροι θα επιβιώσουν και θα πολλαπλασιαστούν. Όμως θα επιβιώσουν και κάποιοι λιγότερο έξυπνοι και πιο αργοί, από τύχη.
- ❑ Ευτυχώς, εξελίσσονται και οι αλεπούδες, διότι αλλιώς θα κατακλυζόμασταν από λαγούς.

# Γενική μορφή Γενετικών αλγορίθμων

- ▶ 1. Δημιούργησε με τυχαίο τρόπο έναν αρχικό πληθυσμό  $\Pi$  από  $N$  υποψήφια λύσεις, οι οποίες δεν θα είναι αναγκαστικά εφικτές (feasible).
- ▶ 2. Αξιολόγησε κάθε λύση του πληθυσμού  $\Pi$ , δηλαδή υπολόγισε πόσο κοντά βρίσκεται στην επιθυμητή λύση, χρησιμοποιώντας μια συνάρτηση καταλληλότητας (fitness function).
- ▶ 3. Δημιούργησε  $N/2$  ζευγάρια από λύσεις του πληθυσμού  $\Pi$ , επιλεγμένες όχι απαραίτητα μια φορά, δίνοντας προτεραιότητα στις πιο κατάλληλες.
- ▶ 4. Κάθε ζευγάρι διασταυρώνεται και παράγει δύο λύσεις-απογόνους (offsprings).
- ▶ 5. Αντικατάστησε τον αρχικό πληθυσμό  $\Pi$  από το σύνολο των απογόνων  $\Pi'$ , που συνήθως είναι βελτιωμένος.
- ▶ 6. Αν τα κριτήρια τερματισμού δεν έχουν ικανοποιηθεί επανάλαβε από το βήμα 2, αλλιώς τερμάτισε.



# Τα 5 απαραίτητα συστατικά του Γενετικού αλγορίθμου

- ▶ Σε κάθε πρόβλημα στο οποίο πρόκειται να εφαρμοστεί ο Γενετικός αλγόριθμος, πρέπει να αποφασίσουμε (και να συμπεριλάβουμε) 5 συστατικά:
  - ❑ Τον τρόπο αναπαράστασης (κωδικοποίησης) των λύσεων.
  - ❑ Τον τρόπο δημιουργίας του αρχικού πληθυσμού (**initialization**). Συνήθως αυτό υλοποιείται τυχαία.
  - ❑ Την συνάρτηση καταλληλότητας (**fitness function**).
  - ❑ Την μέθοδο επιλογής των γονέων.
  - ❑ Την διαδικασία αναπαραγωγής.

# Κωδικοποίηση λύσεων

- ▶ Στους κλασικούς ΓΑ, οι λύσεις παριστάνονται ως συμβολοσειρές (**strings**) ενός πεπερασμένου αλφαβήτου.
  - ❑ Συνήθως χρησιμοποιείται το δυαδικό αλφάβητο (0/1) οπότε αναφερόμαστε σε δυαδικές συμβολοσειρές (**bit strings**).
  - ❑ Σε πάρα πολλά προβλήματα συναντώνται μεταβλητές διαφόρων τύπων. Επομένως, πρέπει κάθε φορά να κωδικοποιούμε τις μεταβλητές αυτές σε strings.
  - ❑ Οι κωδικοποιημένη αναπαράσταση των υποψηφίων λύσεων ονομάζεται χρωμόσωμα (**chromosome**).
  - ❑ Τα επιμέρους συστατικά τμήματα της λύσης - χρωμοσώματος, τα οποία αντιστοιχούν στα χαρακτηριστικά της λύσης, καλούνται γονίδια (**genes**).

# Συνάρτηση καταλληλότητας

- ▶ Η συνάρτηση καταλληλότητας (**fitness function**) είναι μια συνάρτηση που δέχεται ως είσοδο ένα χρωμόσωμα (λύση) και αποδίδει (έξοδος) έναν πραγματικό αριθμό ανάλογο της καταλληλότητας - ευρωστίας της λύσης.
  - ❑ Η συνάρτηση καταλληλότητας λαμβάνει συνήθως υπόψη όλους τους περιορισμούς, τόσο τους ανελαστικούς όσο και τους ελαστικούς.
  - ❑ Κατά την διαδικασία υπολογισμού της εξόδου, καταγράφονται οι παραβιάσεις των περιορισμών, οι οποίες ποινικοποιούνται. Η ποινικοποίηση αφορά στην αντιστοίχιση ενός πραγματικού αριθμού ως κόστος για την κάθε παραβίαση.
  - ❑ Το κόστος παραβίασης (δηλαδή ο πραγματικός αριθμός) είναι ανάλογος της βαρύτητας (σημαντικότητας) του περιορισμού που παραβιάζεται.
  - ❑ Η έξοδος της fitness function είναι το άθροισμα του επιμέρους κόστους των διάφορων παραβιάσεων.
  - ❑ Σε ένα πρόβλημα ελαχιστοποίησης, όσο μικρότερη είναι η τιμή της fitness function τόσο πιο εύρωστη είναι η λύση. Προφανώς η ελάχιστη τιμή ισούται με 0.

# Επιλογή γονέων (1)

- ▶ Σύμφωνα με τον Κανόνα Επιλογής, θα πρέπει οι πιο εύρωστες λύσεις να έχουν μεγαλύτερη πιθανότητα να γίνουν γονείς, από την πιθανότητα που θα έχουν οι λιγότερο εύρωστες λύσεις, έτσι ώστε να υπάρχει προοπτική βελτίωσης του επόμενου πληθυσμού έναντι του προηγούμενου. Κατά την επιλογή:
  - Υπολογίζουμε την τιμή της fitness function όλων των ατόμων του τρέχοντος πληθυσμού και να αντιστοιχίζουμε σε κάθε άτομο μια τιμή πιθανότητας επιλογής, ανάλογη της τιμής της fitness function.
  - Τα άτομα με μεγάλη πιθανότητα επιλογής αναμένεται να επιλεγούν περισσότερες από μια φορές ως γονείς, ενώ αυτά με μικρή πιθανότητα αναμένεται να μην επιλεγούν, χωρίς αυτό να είναι βέβαιο.
  - Δεν είναι βέβαιο ότι δύο εύρωστοι γονείς θα γεννήσουν εύρωστους απογόνους, ούτε ότι δύο μη εύρωστοι γονείς δεν μπορεί να γεννήσουν εύρωστους απογόνους. Αυτός είναι ο λόγος που επιβάλλει την πιθανοτική (στοχαστική) επιλογή των γονέων.

## Επιλογή γονέων (2)

- ▶ Κατά την διαδικασία επιλογής κατασκευάζουμε την δεξαμενή ζευγαρώματος (**mating pool**). Σε αυτήν αντιγράφονται οι υποψήφιοι γονείς με πιθανότητα αντιγραφής ανάλογης της καταλληλότητάς τους.
- ▶ Για την επιλογή των γονέων χρησιμοποιούνται αρκετές τεχνικές, όπως η ρουλέτα (**roulette selection**), η αναλογική καταλληλότητα (**proportionate selection**), η διαδικασία με τουρνουά (**tournament selection**).

# Αλγόριθμος επιλογής Roulette Selection

- ▶ 1. Τοποθέτησε την τιμή fitness κάθε χρωμοσώματος στον πίνακα **OrderedFitnesses[]**.
- ▶ 2. Διάταξε τον πίνακα OrderedFitnesses[] κατά **αύξουσα σειρά**.
- ▶ 3. Υπολόγισε το **άθροισμα S** όλων των στοιχείων του πίνακα OrderedFitnesses[].
- ▶ 4. Στο πρώτο κελί ενός πίνακα **Cummulative\_Fitnesses[]** τοποθέτησε την τιμή του πρώτου κελιού του πίνακα OrderedFitnesses[]
- ▶ 5. Σε κάθε κελί k του πίνακα Cummulative\_Fitnesses[], με  $k > 1$ , τοποθέτησε το άθροισμα των κάθε κελιού m του πίνακα OrderedFitnesses[], όπου  $m \leq k$ .
- ▶ 6. Πάραξε έναν **ομοιόμορφα τυχαίο αριθμό random** στο διάστημα **[0, S]**.
- ▶ 7. Σύγκρινε τον random με την τιμή των κελιών του πίνακα Cummulative\_Fitnesses[], αρχίζοντας από το πρώτο και επίστρεψε τον αριθμό του **πρώτου κελιού z για το οποίο ισχύει Cummulative\_Fitnesses[z] >= random**.
- ▶ 8. Το ζητούμενο χρωμόσωμα είναι αυτό που η fitness του είναι τοποθετημένη στο z κελί του πίνακα OrderedFitnesses[].
- ▶ 9. Αν επέλεξες τον ζητούμενο αριθμό χρωμοσωμάτων, τότε τερμάτισε, αλλιώς επανάλαβε από το βήμα 6.

Σημείωση: Ο παραπάνω αλγόριθμος εφαρμόζεται όταν έχουμε πρόβλημα μεγιστοποίησης, όπου πιο εύρωστο χρωμόσωμα θεωρείται αυτό με την μεγαλύτερη τιμή fitness. Σε πρόβλημα ελαχιστοποίησης αρκεί στον πίνακα OrderedFitnesses[] να τοποθετήσουμε την αντίστροφη τιμή της τιμής fitness.

# Παράδειγμα εφαρμογής roulette selection σε μεγιστοποίηση

- ▶ Έστω ένας πληθυσμός  $N = 10$ , όπου τα χρωμοσώματα έχουν τις ακόλουθες τιμές fitness:
- ▶  $F1 = 5, F2 = 7, F3 = 10, F4 = 3, F5 = 8, F6 = 15, F7 = 4, F8 = 5, F9 = 9$  και  $F10 = 11$ . Ο πίνακας **OrderedFitnesses[]** έχει την μορφή:

P4	P7	P1	P8	P2	P95	P9	P3	P10	p6
3	4	5	5	7	8	9	10	11	15

- ▶ Ο πίνακας **Cummulative\_Fitnesses[]** έχει την μορφή:

3	7	12	17	24	32	41	51	62	77
---	---	----	----	----	----	----	----	----	----

- ▶ Έστω ότι θέλουμε να επιλέξουμε 5 γονείς και έστω ότι οι 5 τυχαίοι ομοιόμορφα επιλεγμένοι αριθμοί που παρήγαγε το σύστημα, στο διάστημα  $[0, 77]$ , είναι οι εξής: 10, 22, 38, 25, 8.
- ▶ Τότε, με βάση τον αριθμό 10 θα επιλεγεί ως γονέας το χρωμόσωμα p1 (αφού το 12 είναι ο πρώτος αριθμός που είναι μεγαλύτερος από το 10. Επίσης, ο 12 είναι τοποθετημένος στο 3<sup>ο</sup> κελί, το οποίο αντιστοιχεί στο χρωμόσωμα p1). Με την ίδια λογική, με βάση τον τυχαίο αριθμό 22 θα επιλεγεί το p2, με βάση τον 38 θα επιλεγεί το p9, με βάση τον 25 θα επιλεγεί το p5 και τέλος με βάση τον 8 θα επιλεγεί το p1. Παρατηρείστε ότι το p1 επελέγη 2 φορές. Προφανώς, μια άλλη ακολουθία τυχαίων αριθμών θα είχε ως αποτέλεσμα ενδεχομένως διαφορετική επιλογή γονέων.



# Αλγόριθμος επιλογής proportionate selection

- ▶ Κάθε επιλεγόμενο χρωμόσωμα  $i$  επιλέγεται με πιθανότητα  $P(i) = \frac{fitness(i)}{\text{αθροισμα της fitness όλων των χρωμοσωμάτων}}$ , όπου  $fitness(i)$  είναι η  $fitness$  του χρωμοσώματος  $i$ .
- ▶ 1. Υπολόγισε τις **πιθανότητες** όλων των χρωμοσωμάτων.
- ▶ 2. Βάλε τις πιθανότητες σε έναν πίνακα **Ordered\_Probs[]**.
- ▶ 3. Διάταξε τον πίνακα Ordered\_Probs[] **κατά αύξουσα σειρά**.
- ▶ 4. Στο πρώτο κελί ενός πίνακα **Cumulative\_Probs[]** βάλε την τιμή του πρώτου κελιού του πίνακα **Ordered\_Probs[]**.
- ▶ 5. Σε κάθε κελί  $k$  του πίνακα Cumulative\_Probs[], με  $k > 1$ , βάλε το άθροισμα όλων των κελιών  $m$  του πίνακα Ordered\_Probs[], **όπου  $m \leq k$** .
- ▶ 6. Πάραξε έναν ομοιόμορφα τυχαίο αριθμό random **στο διάστημα [0, 1]**.
- ▶ 7. Σύγκρινε τον random με την τιμή των κελιών του πίνακα Cumulative\_Probs[], αρχίζοντας από το πρώτο και επίστρεψε τον αριθμό του **πρώτου κελιού  $z$  για το οποίο ισχύει  $Cumulative\_Probs[z] \geq random$** .
- ▶ 8. Το ζητούμενο χρωμόσωμα είναι αυτό που η πιθανότητά του είναι τοποθετημένη στο  $z$  κελί του πίνακα Ordered\_Probs[] .
- ▶ 9. Αν επέλεξες τον ζητούμενο αριθμό χρωμοσωμάτων, τότε τερμάτισε, αλλιώς επανάλαβε από το βήμα 6.

# Αλγόριθμος επιλογής tournament selection (1)

- ▶ 1. Θέσε best ίσο με ένα κενό χρωμόσωμα.
- ▶ 2. Για  $k$  φορές επανάλαβε τα ακόλουθα βήματα:
  - 2.1. Επίλεξε ένα τυχαίο χρωμόσωμα από τον πληθυσμό, έστω  $ind$ .
  - 2.2. Αν το best είναι το κενό ή αν  $f(ind) < f(best)$  θέσε  $best = ind$
- ▶ 3. Επίστρεψε το best

# Αλγόριθμος επιλογής tournament selection (2)

## ► Παρατηρήσεις

- 1: τον παραπάνω αλγόριθμο τον εκτελούμε κάθε φορά που θέλουμε να δημιουργήσουμε **έναν** γονέα.
- 2: ο ακέραιος  $k$  χαρακτηρίζει την **ένταση** της συγκεκριμένης **επιλογής**. Προφανώς, αν  $k = 1$ , τότε έχουμε **τυχαία επιλογή**. Αντίθετα, αν η τιμή του  $k$  είναι **αρκετά μεγάλη**, π.χ.  $k = 10N$ , όπου  $N$  το μέγεθος του πληθυσμού, τότε η πιθανότητα επιλογής κάθε χρωμοσώματος τουλάχιστον μια φορά, είναι αρκετά μεγάλη. Καμία από τις δύο παραπάνω επιλογές δεν είναι καλή. Αντίθετα, θέλουμε μια τιμή για το  $k$  ώστε να επιλέγονται συχνότερα τα καλά χρωμοσώματα και σπανιότερα τα λιγότερο καλά, αλλά να μην επιλέγονται συνεχώς τα ίδια καλά χρωμοσώματα. Για  $k = 2$  έχουμε το **binary tournament** το οποίο χρησιμοποιείται συχνότερα.

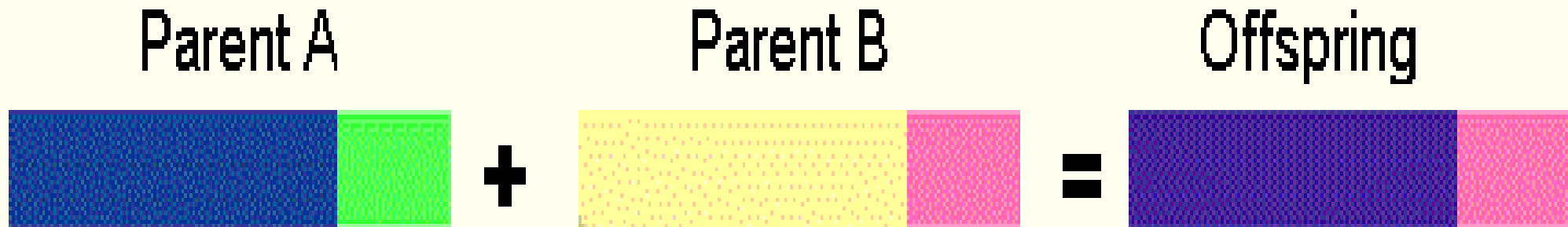
# Διασταύρωση (Crossover ή mating) (1)

- ▶ Διασταύρωση είναι η σύνθεση δύο γονέων με σκοπό την παραγωγή ενός απογόνου. Το είδος της διασταύρωσης εξαρτάται τόσο από την κωδικοποίηση του χρωμοσώματος όσο και από την φύση του προβλήματος.
- ▶ Υπάρχουν τα εξής είδη διασταύρωσης, τα οποία χρησιμοποιούνται συχνότερα:
  - ❑ Διασταύρωση ενός σημείου (single point crossover)
  - ❑ Διασταύρωση δύο σημείων (two point crossover)
  - ❑ Ομοιόμορφη διασταύρωση (uniform crossover)
- ▶ Σημείο διασταύρωσης (cross over point) είναι ένα σημείο το οποίο ορίζει μια νοητή διαχωριστική γραμμή στη δομή του χρωμοσώματος. Προφανώς, κατά την διασταύρωση ενός σημείου στο αρχικό χρωμόσωμα οριοθετούνται δύο τμήματα: το αριστερό και το δεξιό. Αντίθετα, στην διασταύρωση δύο σημείων οριοθετούνται τρία τμήματα: το αριστερό, το μεσαίο και το δεξιό.

# Διασταύρωση (Crossover ή mating) (2)

## ► Διασταύρωση ενός σημείου (one point crossover)

- Σε αυτήν την διασταύρωση επιλέγεται ένα σημείο  $p$ . Το δυαδικό τμήμα του **ενός γονέα από την αρχή μέχρι το  $p$**  (αριστερό) αντιγράφεται στον απόγονο, ενώ το δυαδικό τμήμα του **δεύτερου γονέα από το  $p$  μέχρι το τέλος** (δεξιό) αντιγράφεται στον απόγονο.

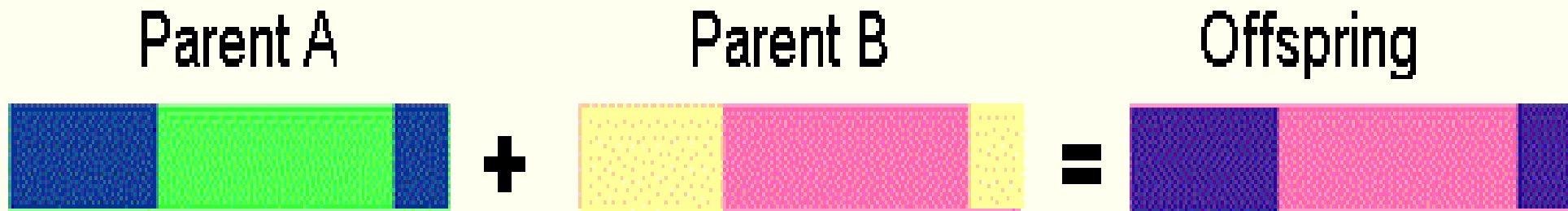


$$11001011 + 11011111 = 11001111$$

# Διασταύρωση (Crossover ή mating) (3)

## ► Διασταύρωση δύο σημείων (two point crossover)

- Σε αυτήν επιλέγονται δύο σημεία τα οποία ορίζουν τρία τμήματα σε κάθε γονέα: το αριστερό, το μεσαίο και το δεξιό. Στον απόγονο αντιγράφεται το αριστερό και το δεξιό τμήμα από τον πρώτο γονέα και το μεσαίο από τον δεύτερο γονέα.

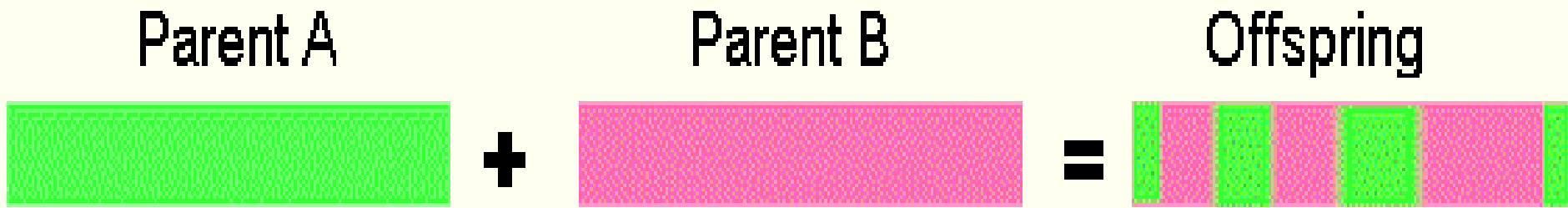


$$11001011 + 11011111 = 11011111$$

# Διασταύρωση (Crossover ή mating) (4)

## ► Ομοιόμορφη διασταύρωση (uniform crossover)

- Σε αυτήν την περίπτωση επιλέγονται ομοιόμορφα τυχαία bits από τους δύο γονείς και αντιγράφονται στον απόγονο.



$$11001011 + 11011101 = 11011111$$



# Μετάλλαξη (mutation)

- ▶ Κατά την μετάλλαξη επιλέγεται κάποιο (ή κάποια) γονίδιο (γονίδια) και η τιμή τους από 1 γίνεται 0 ή από 0 γίνεται 1. Η μετάλλαξη γίνεται με κάποια πιθανότητα, η οποία πρέπει να είναι σχετικά μικρή. Ο σκοπός της είναι η ποικιλομορφία και η διαφοροποίηση.

# Ανανέωση πληθυσμού

- ▶ Σε έναν γενετικό αλγόριθμο, σε κάθε γενεά, ο προηγούμενος πληθυσμός αντικαθίσταται από τον νέο. Ο νέος πληθυσμός περιλαμβάνει τους απογόνους του προηγούμενου πληθυσμού.
  - ❑ Αν ο νέος πληθυσμός αποτελείται μόνο από τους απογόνους, τότε αναφερόμαστε σε **ολική αντικατάσταση** του πληθυσμού.
  - ❑ Αν ο νέος πληθυσμός περιλαμβάνει, εκτός από τους απογόνους και κάποιους από τους γονείς της προηγούμενης γενιάς, τότε αναφερόμαστε σε **μερική αντικατάσταση** του πληθυσμού. Αυτή η περίπτωση είναι και η πιο ρεαλιστική, αφού δύο γενιές συνυπάρχουν. Μάλιστα, δίνεται η ευκαιρία στους απογόνους να ανταγωνιστούν τους γονείς και τελικά να επικρατήσουν οι καλύτεροι.
  - ❑ Τα ερωτήματα που πρέπει να απαντηθούν στην περίπτωση της μερικής αντικατάστασης πληθυσμού είναι:
    - ❖ Ποιοι θα επιλεγούν για να γίνουν γονείς
    - ❖ Ποιοι θα αποχωρήσουν για να αφήσουν χώρο για τους απογόνους (το μέγεθος του πληθυσμού θεωρείται σταθερό).
- ▶ **Ελιτισμός** είναι η πρακτική κατά την οποία αντιγράφονται αυτούσια κάποια εύρωστα χρωμοσώματα στην επόμενη γενεά (πληθυσμό). Αυτά τα χρωμοσώματα μπορούν επίσης να είναι παράλληλα και γονείς. Το σχήμα ελιτισμού δεν είχε αρχικά προταθεί από τον Holland.

# Σύγκλιση

- ▶ Όταν λέμε ότι ένας **αλγόριθμος συγκλίνει** εννοούμε ότι έχει εντοπιστεί η βέλτιστη ή μια περίπου βέλτιστη λύση.
- ▶ Ένας καλά σχεδιασμένος αποδοτικός Γενετικός αλγόριθμος οφείλει να συγκλίνει μετά από έναν αριθμό γενεών (επαναλήψεων). Σε αυτήν την περίπτωση, ένα χρωμόσωμα ή κάποιες λίγο διαφοροποιημένες μορφές του έχουν επικρατήσει στον πληθυσμό.
- ▶ Ένα **γονίδιο συγκλίνει** όταν έχει την ίδια τιμή στο 95% περίπου των χρωμοσωμάτων.
- ▶ Ένας **πληθυσμός συγκλίνει** όταν όλα τα γονίδια των χρωμοσωμάτων του έχουν συγκλίνει.

# Πλεονεκτήματα των Γενετικών αλγορίθμων

- ▶ Περιλαμβάνουν απλές και κατανοητές έννοιες.
- ▶ Επιδέχονται παράλληλη υλοποίηση.
- ▶ Ενσωματώνουν εύκολα διάφορα στοιχεία παραλλαγής-διαφοροποίησης.
- ▶ Υβριδοποιούνται με άλλες μεθόδους βελτιστοποίησης.
- ▶ Υπάρχει μια πληθώρα εφαρμογών - προβλημάτων στα οποία μπορούν να εφαρμοστούν.
- ▶ Τις περισσότερες φορές βρίσκουν ικανοποιητικές λύσεις σχετικά σύντομα.
- ▶ Δεν ενδιαφέρει η σημασία της υπό εξέταση πληροφορίας.
- ▶ Συνδυάζουν σε ικανοποιητική αναλογία την δυνατότητα για εξερεύνηση και την δυνατότητα για εκλέπτυνση. (diversification vs intensification).

# Περιορισμοί των Γενετικών αλγορίθμων

- ▶ Ο πληθυσμός πρέπει να έχει μέγεθος ανάλογο του προβλήματος και συνήθως κυμαίνεται μεταξύ 20-30 ή 50-100.
- ▶ Ο λόγος της διασταύρωσης (crossover rate) πρέπει να είναι περίπου 80% - 95%.
- ▶ Η πιθανότητα μετάλλαξης (mutation rate) πρέπει να είναι μικρή, περίπου 0.5% - 1%.
- ▶ Για κάθε πρόβλημα πρέπει να χρησιμοποιείται η καταλληλότερη μέθοδος επιλογής γονέων.

Σας ευχαριστώ για την προσοχή σας

