ΠΑΡΑΔΕΙΓΜΑ ΒΑΣΙΚΟΥ ΓΕΝΕΤΙΚΟΥ ΑΛΓΟΡΙΘΜΟΥ

<u>Επιλογή (εξαναγκασμένη ρουλέτα)</u>

Το άθροισμα των αξιολογήσεων των μελών:

$F = Άθροισμα\ των\ αξιολογησεων$

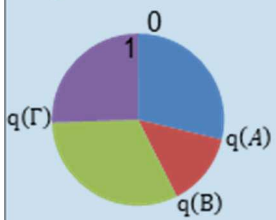
Η πιθανότητα επιλογής των μελών:

- $p(A) = \frac{eval(\text{A})}{F} = \frac{\dots}{F} = \cdots$
- $p(B) = \frac{eval(\text{B})}{F} = \frac{\dots}{F}$
- $p(\Gamma) = \frac{eval(\Gamma)}{F} = \frac{\dots}{F} = \cdots$
- $p(\Delta) = \frac{eval(\Delta)}{F} = \frac{\dots}{F} = \cdots$

Η αθροιστική πιθανότητα των μελών:

- $\text{q}(A) = p(A) = \cdots$
- $\text{q}(B) = q(A) + p(B) = \cdots$
- $\text{q}(\Gamma) = q(\text{B}) + p(\Gamma) = \cdots$
- $\text{q}(\Delta) = q(\Gamma) + p(\Delta) = 1{,}00$

Η ρουλέτα είναι:



(Γυρνάμε τη ρουλέτα τόσες φορές όσες και ο πληθυσμός με βάση τους τυχ. αριθμούς)

0.55: επιλέγεται ο Γ

0.94: επιλέγεται ο Δ

….κ.ο.κ.….

Προσωρινός Πληθυσμός: ***<u>(Γ, Δ, Δ, Β)</u>***

<u>Αναμενόμενος αριθμός αντιγράφων (μόνο εφόσον ζητείται)</u>

Expected_no(A)=POPSIZE * p(A)

Expected_no(B)=POPSIZE * p(B)

….

<u>Διασταύρωση (Μονού Σημείου)</u>

Η συμβολοσειρά που αναπαριστά μια λύση έχει μέγεθος **<u>n</u>**

Τα πιθανά σημεία διασταύρωσης είναι **<u>n-1=….</u>** Θέτουμε κάθε ένα σημείο ισοπίθανο με πιθανότητα **<u>1/(n-1)=….</u>** *(π.χ. 1/8=0,125)*

Συνεπώς το σημείο διαχωρισμού θα επιλέγεται τυχαία με βάση τους τυχαίους αριθμούς και θα επιλέγεται ανάμεσα στις:

- Θέσεις 1-2 μεταξύ <u>*0,000*</u> κ’ <u>*0.125*</u>
- Θέσεις 2-3 μεταξύ <u>*0,125*</u> κ’ <u>*0,250*</u>
- …
- Θέσεις (n-1)-n μεταξύ <u>*0,875*</u> κ’ <u>*1,000*</u>

<u>p_c:Πιθ/τα Διασταύρωσης</u>
Αν είναι 1 τότε διασταυρώνονται όλα τα ζεύγη χωρίς τράβηγμα τυχαίου αριθμού

Για κάθε ζεύγος του προσωρινού πληθυσμού

<u>1ο ζεύγος (***Γ*** και ***Δ***).</u>

Τυχαίος Αριθμός: ***0.21***$\leq p_c$. Διασταυρώνονται!

Τυχαίος Αριθμός: ***0.56***, άρα μεταξύ θέσεων ***<u>4</u>*** και ***<u>5</u>***

- Γ=0001|01 Α’=0001|00
- Δ=0111|00 Β’=0111|01

<u>2ο ζεύγος (***Δ*** και ***Β***).</u>

Τυχαίος Αριθμός: ***0.88***$>p_c$. Δεν Διασταυρώνονται!

Οι γονείς περνάνε στην επόμενη γενιά χωρίς διασταύρωση.

- Δ=011100 Γ’=011100
- Β=000101 Δ’=000101

<u>Ομοίως επαναλαμβάνουμε για όλα τα ζεύγη</u>

<u>p_m:Πιθ/τα Μετάλλαξης</u>
Αν είναι 0 τότε δεν εκτελούμε μετάλλαξη

<u>Μετάλλαξη.</u> Διαδοχικά για κάθε μέλος του πληθυσμού και για κάθε bit χρωμοσώματος του τυχαίου πληθυσμού επιλέγουμε έναν τυχαίο αριθμό

	1ο bit	2ο bit	3ο bit	4ο bit	5ο bit	
Α’=00010	0.77 0	0.23 0	**<u>0.12</u>** **<u>0</u>**	0.93 1	0.28 0	Α’’=00110
Β’=01110	**<u>0.15</u>** **<u>0</u>**	0.82 1	0.34 1	0.32 1	0.44 0	Β’’=11110
Γ’=11101	0.23 1	**<u>0.12</u>** **<u>1</u>**	0.93 1	0.28 0	0.22 1	Γ’’=10101
Δ’=01011	0.82 0	0.34 1	0.32 0	0.44 1	0.77 1	Δ’’=01011

Αν είναι $\leq p_m$ τότε το αντίστοιχο bit αντιστρέφεται!
Αν είναι $>p_m$ τότε το αντίστοιχο bit δεν αντιστρέφεται!

Ένα **πρόβλημα βελτιστοποίησης** μπορεί να **μοντελοποιηθεί** με γενετικό αλγόριθμο:

1. Άτομο (ή Χρωμόσωμα) (αναπαράσταση μίας υποψήφιας λύσης του προβλήματος).

Συνήθως είναι ένας **πίνακας** (π.χ. μονοδιάστατος, διδιάστατος κ.λπ.) με δυαδική, ακέραια ή κωδικοποίηση πραγματικών αριθμών που αναπαριστά τις υποψήφιες λύσεις (ακόμη κι αν δεν σέβονται τους περιορισμούς του προβλήματος).

2. Αντικειμενική Συνάρτηση (ή ικανότητα, ή καταλληλότητα, ή απόδοση, ή αξιολόγηση ή fitness function ή objective function).

Αξιολογεί ένα άτομο και του αποδίδει μία τιμή, έτσι ώστε όσο μεγαλύτερη είναι η τιμή αυτή, τόσο «καλύτερο» είναι το άτομο.

- Αν παίρνει αρνητικές τιμές, τότε προσθέτουμε μία κατάλληλη σταθερά έτσι ώστε να παίρνει μόνο θετικές τιμές >0
- Αν είναι πρόβλημα ελαχιστοποίησης, το μετατρέπουμε σε πρόβλημα μεγιστοποίησης:
  Α' τρόπος: $F(x) = \frac{1}{f(x)+1}$
  Β' τρόπος: $F(x) = -f(x) + C$

3. Τελεστής Επιλογής (Διαδικασία που επιλέγει τα άτομα που θα διασταυρωθούν, ανάλογα με την απόδοσή τους)

Ο τελεστής επιλογής είναι πάντα η εξαναγκασμένη ρουλέτα.
Ελιτισμός: Το καλύτερο άτομο, περνάει απευθείας στην επόμενη γενιά, χωρίς να συμμετέχει στην διαδικασία της επιλογής.

4. Τελεστής Διασταύρωσης (Παράγει 2 παιδιά συνδυάζοντας την γενετική πληροφορία των δύο γονέων).

- <u>Δυαδική Κωδικοποίηση</u>: Διασταύρωση Μονού Σημείου
- <u>Ακέραια Κωδικοποίηση</u>: OX (για προβλήματα μεταθέσεων), αλλιώς διασταύρωση μονού σημείου.
- <u>Κωδικοποίηση Πραγματικών Αριθμών</u>: Διασταύρωση Μονού Σημείου

4. Τελεστής Μετάλλαξης (Προκαλεί «μικρή» τροποποίηση σε ένα άτομο).

- <u>Δυαδική Κωδικοποίηση</u>: Αντιστροφή ενός bit με βάση την p_m.
- <u>Ακέραια Κωδικοποίηση</u>: Ανταλλαγή θέσεων δύο αριθμών (προβλήματα μεταθέσεων), αλλιώς πρόσθεση ή αφαίρεση μίας σταθεράς σε ένα γονίδιο).
- <u>Κωδικοποίηση Πραγματικών Αριθμών</u>: πρόσθεση ή αφαίρεση μίας σταθεράς σε ένα γονίδιο).

Δυαδική Κωδικοποίηση:

Για κάθε bit του ατόμου επιλέγουμε έναν τυχαίο αριθμό στο [0,1]:
- Αν είναι <0.50 θέτουμε το bit ίσο με 0.
- Αν είναι ≥0.50 θέτουμε το bit ίσο με 1.

Ακέραια Κωδικοποίηση (στο Α..Β):

Για κάθε γονίδιο του ατόμου επιλέγουμε έναν τυχαίο αριθμό στο [0,1]:
- Τον πολλαπλασιάζουμε με (Β-Α+1)
- Αποκόπτουμε το δεκαδικό μέρος.
- Θέτουμε το γονίδιο ίσο με τον αριθμό που προέκυψε + Α.

Παράδειγμα: Τυχαίος 0.4394 παράγει τυχαίο στο [5..8]
- 0.4394*(8-5+1)=1.7576
- Αποκοπή Δεκαδικού Μέρους: 1
- Γονίδιο = 1+5 = 6

Ακέραια Κωδικοποίηση (Παραγωγή Μετάθεσης στο 1..Ν):

Παραγωγή μιας μετάθεσης του [1…Ν]:
- Επιλέγουμε τον επόμενο τυχαίο αριθμό στο [0,1].
- Τον πολλαπλασιάζουμε με το Ν
- Τον στρογγυλοποιούμε στον επόμενο ακέραιο
    - Αν ο αριθμός που προέκυψε δεν υπάρχει στο άτομο, τότε θέτουμε την επόμενη κενή θέση του ατόμου ίση με τον αριθμό
    - Αν υπάρχει επαναλαμβάνουμε με τον επόμενο τυχαίο αριθμό.
- Εωσότου απομείνει μία θέση μόνο που την συμπληρώνουμε με τον ακέραιο που λείπει.

Κωδικοποίηση Πραγματικών Αριθμών:

- Ακέραιο Μέρος (όπως παραπάνω)
- Πραγματικό Μέρος (τυχαίος αριθμός στο [0..1))
- Γονίδιο = Ακέραιο + Πραγματικό Μέρος

<u>Ορισμός:</u> Έστω Σ το αλφάβητο των συμβόλων που χρησιμοποιεί ο γενετικός αλγόριθμος για την κωδικοποίηση των χρωμοσωμάτων του πληθυσμού.

Ένα **<u>σχήμα S</u>** (ή πρότυπο S) είναι ένα χρωμόσωμα που χρησιμοποιεί το * (διαβάζεται <u>αδιάφορο σύμβολο</u>) το οποίο μπορεί να αντικατασταθεί από οποιοδήποτε σύμβολο του αλφαβήτου.

Παραδείγματα:

- Στο σχήμα S=11*10 ταιριάζουν οι δύο συμβολοσειρές {11010,11110}
- Στο σχήμα S=*1*1 ταιριάζουν οι τέσσερις συμβολοσειρές {0101,0111,1101,1111}

<u>Σε ένα σχήμα μήκους n στο δυαδικό αλφάβητο:</u>

- Ένα σχήμα με κανένα * θα αναπαριστά μία συμβολοσειρά.
- Ένα σχήμα με k * θα αναπαριστά 2^k συμβολοσειρές
- Ένα σχήμα που αποτελείται μόνο από * θα αναπαριστά 2^n συμβολοσειρές.

Έστω c: πληθάριθμος αλφαβήτου (c=|Σ|) και τα άτομα είναι συμβολοσειρές μήκους n:

- Τα δυνατά σχήματα που μπορούν να κατασκευαστούν είναι $(c+1)^n$
- Μία συμβολοσειρά ταιριάζει σε 2^n διαφορετικά σχήματα.

<u>Ορισμός:</u> **<u>Τάξη</u>** ενός σχήματος **<u>o(S)</u>** είναι αριθμός των θέσεων με 0 και 1

- Προσδιορίζει πόσο ειδικό είναι το σχήμα

<u>Ορισμός:</u> **<u>Οριστικό μήκος</u>** σχήματος **δ(S)**:

- είναι η απόσταση της πρώτης και της τελευταίας σταθερής θέσης

Παραδείγματα:

- Το σχήμα S=11***1 έχει τάξη 3 και οριστικό μήκος 6-1=5
- Το σχήμα S=*0*111* έχει τάξη 4 και οριστικό μήκος 6-2=4

Θεώρημα Σχημάτων: Το αναμενόμενο πλήθος συμβολοσειρών που ταιριάζουν στο σχήμα S στην γενιά t+1:

$$\xi(S,t+1) \geq \xi(S,t) \cdot \frac{eval(S,t)}{\bar{F}(t)} \cdot \left[1 - p_C \frac{\delta(S)}{m-1} - o(S) \cdot p_m\right]$$

«Σχήματα άνω του μέσου όρου απόδοσης, με μικρό οριστικό μήκος και μικρή τάξη λαμβάνουν εκθετικά αυξανόμενες συμβολοσειρές σε διαδοχικές γενιές ενός γενετικού αλγορίθμου»

Όπου:

- $\xi(S,t)$: Πλήθος Ατόμων που ταιριάζουν στο σχήμα S στην γενιά t
- $eval(S,t)$: Μέση Απόδοση ατόμων που ταιριάζουν στο σχήμα S στη γενιά t.
- $\bar{F}(t)$: Μέση Απόδοση του πληθυσμού της γενιάς t
- p_c: Πιθανότητα Διασταύρωσης
- $\delta(S)$: Οριστικό Μήκος Σχήματος
- m: Μήκος Συμβολοσειράς που αναπαριστά ένα άτομο
- $o(S)$: Τάξη Σχήματος
- p_m: Πιθανότητα Μετάλλαξης

Παράδειγμα: Πόσες συμβολοσειρές αναμένεται να ταιριάζουν στο σχήμα S=**0**10** στην γενιά 1 αν p_c=0,75 και p_m=1/9 αν η γενιά 0 είναι η ακόλουθη:

Ατομο	Συμβολοσειρά	Ικανότητα
Α	100101011	25
Β	000010001	10
Γ	010100110	20
Δ	110011001	15
Ε	001001010	5

$\xi(S,0) = 2$ (Α και Δ)

$$eval(S,0) = \frac{eval(A)+eval(\Delta)}{2} = \frac{25+15}{2} = 20$$

$$\bar{F}(0) = \frac{25+10+20+15+5}{5} = 15$$

$$p_c = 0{,}75, \delta(S) = 7-3 = 4, p_m = \frac{1}{9}, o(S) = 3, m = 9$$

Άρα:

$$\xi(S,1) \geq 2 \cdot \frac{20}{15} \cdot \left[1 - 0{,}75\frac{4}{9-1} - 3 \cdot \frac{1}{9}\right] = 0{,}78$$

Πιθανότητα Επιβίωσης Σχήματος:

Επιλογή: $p_S = \frac{eval(S,t)}{\bar{F}(t)}$ **Διασταύρωση:** $p_S = 1 - p_c \frac{\delta(S)}{m-1}$ **Μετάλλαξη:** $p_S = (1-p_m)^{o(S)} \approx 1 - o(S) \cdot p_m$

Πιθανότητα Καταστροφής Σχήματος (αντίστοιχα είναι): $p_D = 1 - p_S$

Ο παρακάτω τύπος δίνει το πλήθος των συμβολοσειρών των συμβολοσειρών που ταιριάζουν στο σχήμα S μετά από K γενιές, αν εφαρμόζεται μόνο επιλογή (όχι διασταύρωση και μετάλλαξη)

$$\boldsymbol{\xi(S, t + \mathrm{K})} = \xi(S, t) \cdot (1 + \varepsilon)^{\mathrm{K}}$$

Όπου ε είναι η επί τοις εκατό απόκλιση της μέσης απόδοσης του πληθυσμού σε σχέση με την μέση απόδοση του σχήματος και δίνεται από τον τύπο:

$$\varepsilon = \frac{eval(S)}{\bar{F}} - 1$$

Όπου eval(S) είναι η μέση απόδοση του σχήματος και $\bar{F}$ είναι η μέση απόδοση του πληθυσμού.

Αν ε>0 **επικρατεί** το σχήμα. Θέτουμε ξ(S,t+K) ίσο με τον συνολικό πληθυσμό για να υπολογίσουμε το k.

Παράδειγμα: ε=0.53 σε έναν πληθυσμό 16 ατόμων όπου στο σχήμα ταιριάζουν 8 άτομα:

$16 = 8 *(1+0.53)^k \Rightarrow$

$16 = 8 *(1.53)^k \Rightarrow$

$2 = 1.53^k \Rightarrow$

$k=\log_{1,53}2 \Rightarrow$

$k=1,63$

Άρα θα επικρατήσει μετά από 2 γενιές.

Αν ε<0 **εξαφανίζεται** το σχήμα. Θέτουμε ξ(S,t+K)<1 για να υπολογίσουμε το k.

Παράδειγμα: ε= - 0.53 σε έναν πληθυσμό 16 ατόμων όπου στο σχήμα ταιριάζουν 8 άτομα:

$8 *(1-0.53)^k < 1 \Rightarrow$

$8 *(0.47)^k < 1 \Rightarrow$

$\log (8 *(0.47)^k) < \log 1 \Rightarrow$

$\log (8) + \log (0.47)^k < 0 \Rightarrow$

$\log (8) + k \log (0.47) < 0 \Rightarrow$

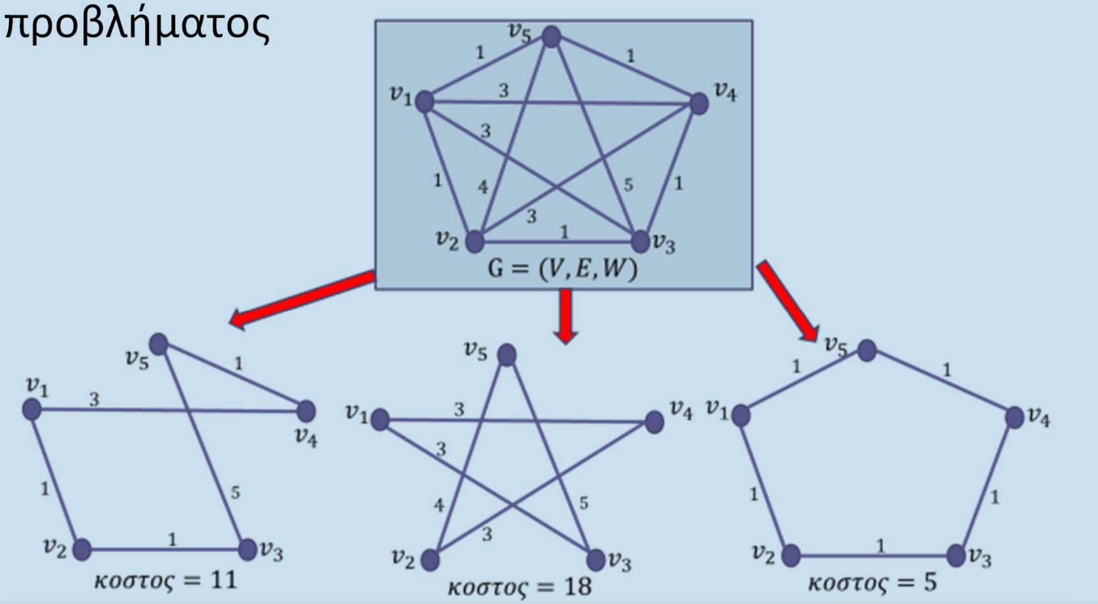
$k>2,73 \Rightarrow$

Άρα θα εξαφανιστεί μετά από 3 γενιές.

Το πρόβλημα του Περιοδεύοντος Πωλητή (Travelling Salesman Problem - TSP):
Δίνονται n πόλεις με τις αντίστοιχες χιλιομετρικές τους αποστάσεις. Ζητείται να κατασκευαστεί ένας περίπατος του πωλητή στις πόλεις, ο οποίος:

- Θα περνάει από όλες τις πόλεις ακριβώς μία φορά.
- Θα ξεκινάει και θα τελειώνει στην ίδια πόλη.
- Θα έχει το ελάχιστο κόστος (άθροισμα χιλιομετρικών αποστάσεων)

Ένα στιγμιότυπο και 3 υποψήφιες λύσεις του προβλήματος



Κωδικοποίηση:

- ένα διάνυσμα ακεραίων που απεικονίζει την σειρά επίσκεψης των κόμβων (π.χ.: $[v_1,v_2,v_3,v_5,v_4]$)

Αξιολόγηση: F(x)= - f(x)+ C όπου:

- f(x)=Άθροισμα Βαρών Ακμών που χρησιμοποιεί η λύση
- C: (Πόλεις) x (Μέγιστη Απόσταση δύο πόλεων)

Γενετικοί Τελεστές:

- **Τελεστής Επιλογής:** Εξαναγκασμένη Ρουλέτα
- **Τελεστής Διασταύρωσης:** Τελεστής OX
- **Τελεστής Μετάλλαξης:** Τυχαία Ανταλλαγή δύο πόλεων στην διάταξη

Παράδειγμα Εφαρμογής Τελεστή OX: A = (1 2 3 |4 5 6 7| 8 9) και B = (4 5 2 |1 8 7 6| 9 3) (**δύο** σημεία διασταύρωσης)

<u>1ος απόγονος Α’:</u>

- Παίρνω τα μεσαία του 1ου γονέα Α’ = (x x x |4 5 6 7| x x)
- Καταγράφω τα στοιχεία που λείπουν με αφετηρία το 2ο σημείο διασταύρωσης του B = (4 5 **2** |**1 8** 7 6 **| 9 3**) (➔9 3 2 1 8)
- Συμπληρώνω τα στοιχεία του Α’ με αφετηρία το 2ο σημείο διασταύρωσης Α’ = (2 1 8 |4 5 6 7| 9 3)

<u>2ος απόγονος Β’</u>: Αντίστοιχα κρατάω το μεσαίο κομμάτι του Β και συμπληρώνω με αφετηρία το 2ο σημείο διασταύρωσης του Α

Το πρόβλημα της ικανοποιησιμότητας (Satisfiability - SAT):

- Είσοδος: Δίνεται φόρμουλα φ σε κανονική συζευκτική μορφή (n: πλήθος μεταβλητών, m: πλήθος προτάσεων).
- Ερώτημα: Είναι η φ ικανοποιήσιμη;

Παράδειγμα:

Η φόρμουλα SAT:

$$\varphi_1 = (x_1 \vee \neg x_2) \wedge (\neg x_1 \vee x_2 \vee x_3) \wedge (\neg x_2 \vee x_3)$$

είναι ικανοποιήσιμη, για παράδειγμα με την αποτίμηση $x_1 = \text{Α}, x_2 = \text{Α}, x_3 = \text{Α}$

Κωδικοποίηση:

- Ένα άτομο αναπαρίσταται με μία δυαδική συμβολοσειρά μήκους n.

Π.χ. το διάνυσμα ακεραίων 1110 αντιστοιχεί στην ανάθεση των τιμών στις μεταβλητές: $x_1 = \text{Α}$, $x_2 = \text{Α}, x_3 = \text{Α}, x_4 = \Psi$

Αξιολόγηση: Πλήθος των προτάσεων (παρενθέσεων) που ικανοποιούνται από την αποτίμηση.

Η ελάχιστη τιμή είναι 0 και η μέγιστη τιμή είναι m (αν η φόρμουλα είναι ικανοποιήσιμη)

Γενετικοί Τελεστές:

- **Τελεστής Επιλογής:** Εξαναγκασμένη Ρουλέτα
- **Τελεστής Διασταύρωσης:** Διασταύρωση Μονού Σημείου
- **Τελεστής Μετάλλαξης:** Αλλαγή ενός bit με βάση την πιθανότητα μετάλλαξης.