Genetic Algorithms
CONVERGENCE VELOCITY
Part 1
A Convergence Velocity Theory of GAs

• Question: How fast is a GA?

• Convergence velocity:
  \[ \varphi = \mathbb{E}(f_{\text{max}}(P(t+1)) - f_{\text{max}}(P(t))) \]
  (expected improvement from $t \to t+1$)

• Stochastic process \(\Rightarrow\) Expectation

• Tools: Probability theory, order statistics, Markov processes

• Simple $f$, simple algorithm variants
Counting Ones (1)

Basic Definitions:

• Success probability:
  \[ p^+_a = P\{ f(m(\tilde{a})) > f(\tilde{a}) \} \]

• \( k \)-step success probability (\( 0 \leq k \leq k_{\text{max}} \)):
  \[ p^+_a(k) = P\{ f(m(\tilde{a})) = f(\tilde{a}) + k \} \]

• Optimal mutation rate \( p^* \)?

• By maximizing the convergence velocity

\[
\varphi_{(1+1)}(l, \tilde{a}, p) = \sum_{k=0}^{k_{\text{max}}} k \cdot p^+_a(k)
\]
Counting Ones (2)

Analysis done for three objective functions:

• Simple bit counting problem $f$

• Standard binary code $g$

• Gray code $h$
Counting Ones (3)

An analyzable function: *Counting Ones*

\[
f(f) = \sum_{i=1}^{l} a_i
\]

Mutation rate \(p, q = 1 - p\), \(f_a := f(f)\)

\[
p^+_a = \sum_{i=0}^{f_a} \binom{f_a}{i} p^i q^{f_a-i} \cdot \sum_{j=i+1}^{l-f_a} \binom{l-f_a}{j} p^j q^{l-f_a-j}
\]

\[
p^-_a = \sum_{i=0}^{l-f_a} \binom{l-f_a}{i} p^i q^{l-f_a-i} \cdot \sum_{j=i+1}^{f_a} \binom{f_a}{j} p^j q^{f_a-j}
\]
Counting Ones (4)

Mutation rate \( p, q = 1 - p, f_a := f(\bar{a}) \)

\[
p_{\bar{a}}^0 = \sum_{i=0}^{f_a} \binom{f_a}{i} \binom{l - f_a}{i} p^{2i} q^{l-2i}
\]

\[
p_{\bar{a}}^+(k) = \sum_{i=0}^{f_a} \binom{f_a}{i} \binom{l - f_a}{i+k} p^{2i+k} q^{l-2i-k}, \quad 0 \leq k \leq l - f_a
\]

\[
p_{\bar{a}}^-(k) = \sum_{i=0}^{l-f_a} \binom{l - f_a}{i} \binom{f_a}{i+k} p^{2i+k} q^{l-2i-k}, \quad 0 \leq k \leq f_a
\]
Counting Ones (5)

Proof idea (for $p_a^+$):

$$\vec{a} = (1001011000) \quad f_a = 4$$

Independent from the position $\Rightarrow$

$$\vec{a}' = (\underbrace{000000}_{l-f_a} \underbrace{1111}_{f_a})$$
Counting Ones (6)

Two cases:

• None of the $f_a$ ones mutates
  $\Rightarrow 1, \ldots, l - f_a$ zeroes need to mutate

• $i \in \{1, \ldots, f_a\}$ ones mutate
  $\Rightarrow j \in \{i+1, \ldots, l - f_a\}$ zeroes need to mutate

• $i$: No. of mutations $1 \rightarrow 0$ (harmful ones)
• $j$: No. of mutations $0 \rightarrow 1$ (useful ones)

• $p^0$: $j = i$, $p^+(k)$: $j = k$ in $p^+$, $p^-(k)$: $j = k$ in $p^-$
Counting Ones (7)

Convergence velocity of a (1+1)-GA:

\[ \varphi_{(1+1)} = \sum_{k=0}^{l-f_a} k \cdot p_a^+(k) \]
Counting Ones (8)

Convergence velocity $\varphi_{(1+1)}$ as a function of $p$ for different values of $f_a$ ($l = 100$):
Counting Ones (9)

**Optimum mutation rate as a function of** $f_a$:
Counting Ones (10)

Approximation of optimum mutation rate (no analytical solution):

$$p^* \approx \frac{1}{2(f_a + 1) - l}$$

([Mühlenbein 1992]: \( p^* \approx 1/l \))
Counting Ones (11)

Time to absorption (Markov-chain-analysis):
$l + 1$ states: $z_k = \{ f(\tilde{a}) = l - k \} (0 \leq k \leq l)$

Transition probabilities

$$p_{ij} = P\{ f(m(\tilde{a})) = l - j \mid f(\tilde{a}) = l - i \}:$$

$$p_{ij} = \begin{cases} 
  p_{l-i}^+(i-j), & i > j \\
  1 - \sum_{k=1}^{j} p_{l-i}^+(k), & i = j \\
  0, & i < j 
\end{cases}$$

- State 0 ($l$ bits are correct) is absorbing
- $\tau = \{ 1, \ldots, l \}$ transient class of state
Counting Ones (12)

Absorption times: Numerically, from the transition matrix in block form, according to

\[ P = \begin{pmatrix} I & 0 \\ R & Q \end{pmatrix} \]

where \( N = (n_{ij}) = (I - Q)^{-1} \) (\( T: \text{transient states} \))

- \( I \) denotes a unity matrix of appropriate size
- Here: \( Q \) will be an \((l - 1) \times (l - 1)\)-matrix
- \( E_i(t) \): Expected time to absorption if started in state \( i \)
Counting Ones (13)

Time to absorption as a function of $l$, for both mutation rate control schemes

- Order of magnitude $\mathcal{O}(l \cdot \ln l)$

- Main problem: Further progress for $f_a \approx l$
Counting Ones (14)

Approximation [[Mühlenbein 1992]]: \( p \ll 1 \Rightarrow \)

\[
p_{\tilde{a}}^+(p) \approx \left(1 - p\right)^{f(\tilde{a})} \cdot \left(1 - (1 - p)^{l-f(\tilde{a})}\right)
\]

no 1 changes \hspace{2cm} at least one 0 changes

Compare: From equation for \( p_{\tilde{a}}^+ \) for \( i = 0 \):

\[
p_{\tilde{a}}^+(p) = (1 - p)^{f(\tilde{a})} \sum_{j=1}^{l-f(\tilde{a})} \binom{l-f(\tilde{a})}{j} p^j (1 - p)^{l-f(\tilde{a})-j}
\]

\[= (1 - (1 - p)^{l-f(\tilde{a})})\]
Counting Ones (15)

Question: Optimal mutation rate $p^*$?

Mühlenbein: Maximize succes probability

$\left( l - f(\bar{a}) \ll l \right)$

\[
p^* = 1 - \left( 1 - \frac{l - f(\bar{a})}{l} \right)^{\frac{1}{l-f(\bar{a})}} \approx \frac{1}{l}
\]
Counting Ones (16)
Impact of other mutation rate settings: Time to absorption as a function of $l$ for different mutation rates
Counting Ones (17)

Effect different mutation rate settings:

- $p_m$ too large: Exponential complexity
  
  (Evolution → random search)

- $p_m$ too small:
  
  Time to absorption almost constant