Outline

• Selection Methods
  – Stochastic Selection
  – Deterministic Selection

When to Select

• Selection can occur at two stages of evolution:
  – Parent selection: select individuals from the current population to take part in mating.
  – Survivor selection: select individuals from offspring and/or current population to generate new population.
• Selection operators work on the entire individual, i.e. they are representation-independent.
## How to Select

- **Stochastic Selection (probabilistic)**
  - Random (stochastic uniform)
  - Fitness proportionate
  - Rank (linear vs. non-linear)
  - Tournament (binary vs. more)
- **Deterministic Selection**
  - Deterministic uniform
  - Deterministic replacement
  - Truncation (the best n)

## Selection Pressure

- Selection methods are characterized by their **selection pressure**, also referred to as the **takeover time**, which relates to the time it requires to produce a uniform population (all individuals in the population are identical).
- This can be estimated by repeated application of the selection method alone (w/o genetic operation).
- Higher selection pressure: population diversity (the number of unique individual) is decreased more rapidly.

### Random Selection

- Non-fitness based selection.
- Each individual in the population has the **same probability** to be selected (stochastically random).
- Random selection has the lowest selection pressure among all selection methods.
- This can be used to pair with a non-random selection for parent/offspring selection.

### Random Selection - Continued

- Even with uniform probability distribution, the best individual in the population may never gets chosen and the worst individual gets selected multiple times.
- Under the “sampling error”, the population diversity will decrease. Consequently, the population would still converge.
Fitness Proportionate (Roulette Wheel)

- The probability of an individual $i$ to be selected from a population of $n$ individuals is
  \[ p_i = \frac{f_i}{\sum_{i=1}^{n} f_i} \]
- The expected number of each individuals to be selected is
  \[ E_i = \frac{f_i}{\bar{f}} \]
  - $f_i$: non-negative fitness of individual $i$
  - $\bar{f}$: average population fitness
  - Assume maximization fitness.
  - To convert minimization to maximization, multiply -1 to the fitness.

Example

<table>
<thead>
<tr>
<th>id</th>
<th>f(x)</th>
<th>$p_i$</th>
<th>$E_i$</th>
<th>Actual Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>169</td>
<td>0.09204793</td>
<td>0.368191721</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1276</td>
<td>0.694989107</td>
<td>2.779956427</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>0.016339869</td>
<td>0.065359477</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>361</td>
<td>0.196623094</td>
<td>0.786492375</td>
<td>1</td>
</tr>
<tr>
<td>sum</td>
<td>1836</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>average</td>
<td>459</td>
<td>0.25</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>max</td>
<td>1276</td>
<td>0.694989107</td>
<td>2.779956427</td>
<td>3</td>
</tr>
</tbody>
</table>

over select the super individual 2.

Shortcoming

- During early stage of evolution, super individuals may take over a significant proportion of the population and cause premature convergence.
- Later state of evolution, the fitness of all individuals are close to each other, which makes the fitness-based selection method act like random selection (weak selection pressure).

Fitness Scaling

- Linear scaling: \( f'(x) = af(x) + b \)
  - The average population members contribute one expected count to the next generation
  - The best population members contributes $c$ expected counts to the next generation, \( 1.2 \leq c \leq 2 \)
  \[
  a = \frac{(c - 1) \cdot \text{avg}(f)}{\text{max}(f) - \text{avg}(f)} \\
  b = \frac{\text{avg}(f) \cdot (\text{max}(f) - c \cdot \text{avg}(f))}{\text{max}(f) - \text{avg}(f)}
  \]
Linear Scaling

<table>
<thead>
<tr>
<th>id</th>
<th>f(x)</th>
<th>f(x)+b</th>
<th>pi</th>
<th>Ei</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>169</td>
<td>296.07</td>
<td>0.16</td>
<td>0.645</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1276</td>
<td>918</td>
<td>0.5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>217.98</td>
<td>0.1187</td>
<td>0.4749</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>361</td>
<td>403.94</td>
<td>0.22</td>
<td>0.88</td>
<td>1</td>
</tr>
<tr>
<td>sum</td>
<td>1836</td>
<td>1836</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td></td>
<td></td>
<td>0.25</td>
<td>1</td>
</tr>
<tr>
<td>max</td>
<td>1276</td>
<td>918</td>
<td>0.5</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

c=2;

Rank Selection

- Selection probabilities are based on relative rather than absolute fitness.
- Individuals are ranked according to their fitness.
- The probability of an individual to be selected is proportionate to its rank.
- Ranking introduces a uniform scaling across the population and provides a simple and effective way of controlling selective pressure.

Linear Rank Selection

- For population size n, rank 1 has the best fitness and rank n has the worst fitness:
  \[ p_i = \frac{2}{n} - \varepsilon \]
  \[ p_n = \varepsilon \]
  \[ p_i = \left( \frac{2}{n} - \varepsilon \right) - \left( \frac{2}{n} - 2\varepsilon \right) \frac{i - 1}{n - 1} \]

\( \varepsilon \) controls the slope of the linear probability distribution by ranging from 0.0 (max slope) to 1/n (0 slope, flat line).

Linear Ranking Selection

<table>
<thead>
<tr>
<th>ind</th>
<th>fitness</th>
<th>rank</th>
<th>( p_i )</th>
<th>( E_i )</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>169</td>
<td>3</td>
<td>0.1667</td>
<td>0.6668</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1276</td>
<td>1</td>
<td>0.4999</td>
<td>1.9996</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>4</td>
<td>1E-04</td>
<td>0.0004</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>361</td>
<td>2</td>
<td>0.3333</td>
<td>1.3332</td>
<td>1</td>
</tr>
<tr>
<td>sum</td>
<td>1836</td>
<td>10</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td></td>
<td>0.25</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>max</td>
<td>1276</td>
<td>4</td>
<td>0.4999</td>
<td>1.9996</td>
<td>2</td>
</tr>
</tbody>
</table>

\( \varepsilon = 0.001 \)
Comparison

Tournament Selection

- A small subset of $n$ individuals are chosen at random (with replacement), and the best individual(s) in this set is/are selected.
- The larger the tournament size $n$, the stronger the selection pressure.
- If pick without replacement (each individual can only be selected once the most), selection pressure is increased or decreased?

Binary-Tournament

- Two individuals are selected randomly and the one with better fitness is the winner.
- Easy to implement, as there is no need to maintain the order of the individuals in the population, which is required by other stochastic selection methods.

Deterministic Selection

- Each individual is assigned a fixed number that corresponds to the number of times they will be selected.
Deterministic Uniform

- In Evolutionary Programming, each parent is selected exactly once to produce once offspring.
- In stochastic uniform (random) selection, the best individual may never get selected while the worst may get selected multiple times.
- Deterministic uniform can avoid such bias.

Deterministic Replacement

- In steady-state model, the replacement of current population individual can be decided deterministically:
  - Age-based: replace the oldest
  - Fitness-based: replace the worst
- Advantage/disadvantage compared to random replacement?

Truncation - Elitism

- The best \( b \) individuals in the population are selected.
- Advantages
  - prevents the best found candidate solution accidentally 'gets lost'
  - preserves the currently best found solution as fix points to create offspring in their vicinity
- Disadvantage
  - May cause premature convergence

Truncation Implementation

- Evolution Strategies (\( \mu, \lambda \))-ES
  - \( \mu \) is the size of parent population while \( \lambda \) (\( \lambda >> \mu \)) is the size of the offspring population.
  - Randomly select parents to generate \( \lambda \) offspring
  - The best \( \mu \) individuals among \( \lambda \) offspring form the new population (non-overlapping population).

\[
P(t) \xrightarrow{\text{reproduction}} P(t+1)
\]

\[
\begin{align*}
P(t) & \quad \mu \text{ parents} \\
\lambda \text{ offspring} & \quad \text{reproduction} \\
P(t+1) & \quad \mu \text{ parents} \\
\end{align*}
\]
Truncation Implement II

- Evolution Strategies ($\mu+\lambda$)-ES
  - $\mu$ is the size of parent population while $\lambda$ is the size of the offspring population.
  - Randomly select parents to generate $\lambda$ offspring.
  - The best $\mu$ individuals among $\lambda+\mu$ form the new population (overlapping population).

Selection – Only Model

- Under fitness-based selection, population fitness improve over the generations.
- However, the best individual fitness at the last generation can never be better than the best individual fitness at the initial generation.
- Moreover, under stochastic selection, the best individual may get lost and the population converged to an individual with worse fitness than that of the initial best individual.

Selection Summary

- Stochastic selection:
  - With a small population (<20), the actual selected subsets can differ quite significantly from the expected ones.
  - The best individual might never been selected while the worst individual might be selected more than once.

Selection Summary

- Stochastic selection:
  - Under a finite population, any stochastic selection method is likely to cause a loss of diversity due to sampling error; even uniform selection results in the population converging to single genotype (assignment 1).
  - In practice, stochastic selection can be used to add “noise” to an EA in a way that improves its “robustness” by decreasing the likelihood of converging to a sub-optimal solution.
Traditional EA Selection Categories

<table>
<thead>
<tr>
<th>EA</th>
<th>parent size n</th>
<th>offspring size m</th>
<th>parent-selection</th>
<th>survival-selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>EP</td>
<td>&lt;20</td>
<td>m = n</td>
<td>deterministic</td>
<td>deterministic</td>
</tr>
<tr>
<td>ES</td>
<td>&lt;10</td>
<td>m &gt;&gt; n</td>
<td>stochastic</td>
<td>deterministic</td>
</tr>
<tr>
<td>GA/GP</td>
<td>&gt;20</td>
<td>m = n</td>
<td>stochastic</td>
<td>deterministic</td>
</tr>
</tbody>
</table>

Selection Summary

fitness-based vs. uniform

- Successful search is achieved by a balance between exploration and exploitation.
- Fitness-based selections exploit known good solutions to find better solutions; they are more effective on relatively smooth, time-invariant landscapes.
- Uniform selection explore unknown territories: they are more effective on multi-modal (rugged) fitness landscapes.