Genetic Algorithms Assignment 1

R11921038 Tu-Chin Chiang

April 24, 2024

1 Trap Function

Let's check how often deception occurs in a random function.

(a) Consider the case with two genes. By assigning fitness values for $f(00), f(01), f(10)$, and $f(11)$, does deception ever occur?

No, deception never occurs. The inequality (Ackley, 1985) is defined by

$$\frac{a}{b} \geq \frac{2-(\ell-z)^{-1}}{2-z^{-1}},$$

where $a < b$, ℓ is the problem size, and z is the number of unitation. The inequality is never satisfied when $\ell = 2$ and $z = 1$:

$$\frac{2-(2-1)^{-1}}{2-1^{-1}} = 1 \not\leq \frac{a}{b}.$$

Therefore, deception never occurs in the case with two genes. □

(b) Consider the case with three genes, randomly assign the fitness values for $f(000)$, $f(001)$, $f(010)$, $f(011)$, $f(100)$, $f(101)$, $f(110)$, and $f(111)$ with uniform distribution from 0 to 1. Repeat the experiments 10^6 times. What's the probability that 3-deception occurs?

(c) Repeat (b), but with 4 genes.

	k-deception	Probability
(b)	3	**0.005791**
(c)	4	**0.000286**

Table 1: Probability of 3-deception and 4-deception

□

(d) For a problem with *ell* genes (problem size), the probability that k-deception does NOT occur among any k genes is roughly $(1-p)^{\binom{ell}{k}}$, where p is recorded in (b) and (c). What's the problem size that makes 3-deception occur with probably 0.5? What's that for 4-deception? When does 3-deception occur more often than 4-deception or the other way around?

	Scenario	Problem size *ell*
(d.1)	$Prob$(3-deception) > 0.5	$ell \geq$ **10**
(d.2)	$Prob$(4-deception) > 0.5	$ell \geq$ **18**
(d.3)	3-deception occurs more often than 4-deception	$ell \leq$ **84**

Table 2: Problem size for the occurrence of different scenarios

(d.1)

$$
\begin{aligned}
1-(1-p)^{\binom{ell}{3}} &= 0.5 \\
\Rightarrow (1-0.005791)^{\binom{ell}{3}} &= 0.5 \\
\Rightarrow \binom{ell}{3} \times \log(0.994209) &= \log(0.5) \\
\Rightarrow \binom{ell}{3} = \frac{ell(ell-1)(ell-2)}{6} &= \frac{\log(0.5)}{\log(0.994209)} = 119.36763 \\
\Rightarrow ell &= 9.98429, -3.49215 \pm 7.71619i
\end{aligned}
$$

(d.2)

$$
\begin{aligned}
1-(1-p)^{\binom{ell}{4}} &= 0.5 \\
\Rightarrow (1-0.000286)^{\binom{ell}{4}} &= 0.5 \\
\Rightarrow \binom{ell}{4} \times \log(0.999714) &= \log(0.5) \\
\Rightarrow \binom{ell}{4} = \frac{ell(ell-1)(ell-2)(ell-3)}{24} &= \frac{\log(0.5)}{\log(0.999714)} = 2423.24495 \\
\Rightarrow ell &= 17.0696, -14.0696, 1.5 \pm 15.4891i
\end{aligned}
$$

(d.3)

$$
\begin{aligned}
1-(1-p_3)^{\binom{ell}{3}} &> 1-(1-p_4)^{\binom{ell}{4}} \\
\Rightarrow \binom{ell}{3} \times \log(1-0.005791) &< \binom{ell}{4} \times \log(1-0.000286) \\
\Rightarrow ell - 3 &< 4 \times \frac{\log(1-0.005791)}{\log(1-0.000286)} \\
\Rightarrow ell &< 84.2027
\end{aligned}
$$

In summary, the problem size at which 3-deception is likely to occur with a probability of 0.5 is $ell \geq 10$, while for 4-deception, it is $ell \geq 18$. When the problem size is less than 84, 3-deception occurs more frequently than 4-deception. This is attributed to the fact that the combinations of genes for 4-deception increase more rapidly as the problem size grows.

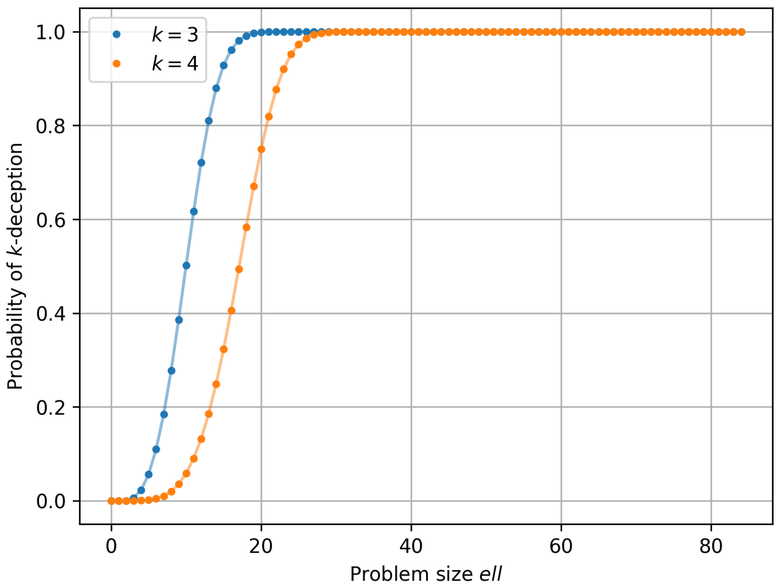


Figure 1: Probabilities of 3 and 4-deception

However, a concern arises when examining the probabilities of 3 and 4-deception, particularly as the problem size becomes larger, as illustrated in Figure 1. The probabilities of both

3 and 4-deception approach 1 when the problem size is large, with probabilities greater than 0.999 observed at problem sizes of 21 and 30, respectively. Therefore, despite the maximum problem size at which 3-deception occurs more often being 84, the results indicate that both 3 and 4-deception occur with high probabilities when the problem size reaches 84.

□

2 Convergence Time

Thierens' convergence-time model assumes perfect mixing. Verify the model with (1) one-point XO, (2) uniform XO, and (3) population-wise shuffling: the set of genes at a particular position of offspring is a random shuffle of the set of genes at the same position of parents.

(a) Experiment these three XO's with a SGA on the OneMax problem with different sizes (ell) of 50, 100, 150, 200, 250, 300, 350, 400, 450, and 500. Plot the results on a figure with problem size versus convergence time. (repeat = 30, population size = $4 \times ell \times ln(ell)$, and tournament size = 2.)

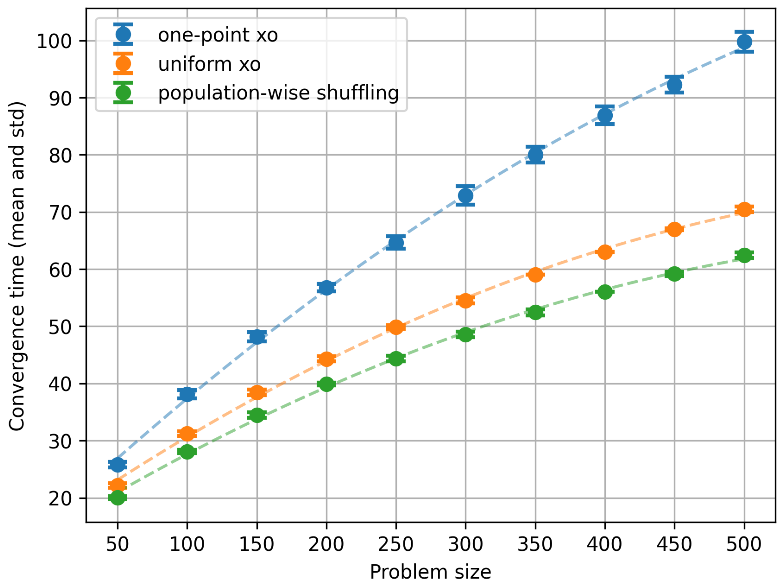


Figure 2: Convergence time of different XO methods

Due to the independent variables of the problem size, I plot the average convergence time versus the problem size of three XO methods in Figure 2. The results show that the convergence time of the one-point XO is the longest, followed by the uniform XO, and the population-wise shuffling has the shortest convergence time under the same problem size. □

(b) What's the theoretical value of selection intensity? How does that compare with the selection intensity measured in your experiments? Is selection intensity really invariant with generation?

(b.1)

$$
\begin{aligned}
\text{Selection intensity } I &= \sqrt{2 \times \Big(ln(s) - ln\big(\sqrt{4.14 \times ln(s)}\big)\Big)} \\
&= \sqrt{2 \times \Big(ln(2) - ln\big(\sqrt{4.14 \times ln(2)}\big)\Big)} \\
&\approx 0.576291
\end{aligned}
$$

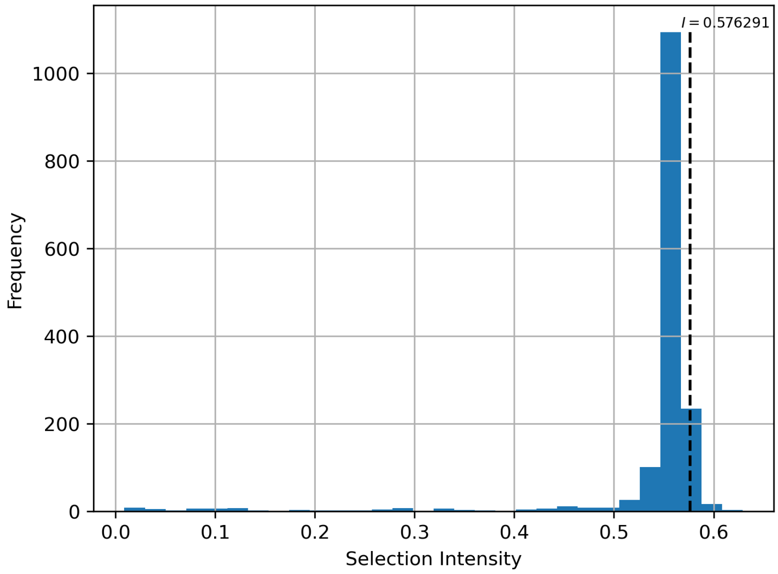


Figure 3: Histograms of selection intensity

(b.2, b.3)

The histogram of the selection intensities of three XO methods with 1 run is shown in Figure 3. The results show that 68.98% of the selection intensities of three XO methods are between 0.546493 and 0.567167, which implies, on one hand, the selection intensities of the tournament slection are **generally invariant with generation.** On the other hand, most of the selection intensities are **slightly less than the theoretical value of 0.576291**.

Furthermore, a few portion of the selection intensities that are less than 0.5 occur in the last 2 to 3 generations. This is likely because that the population diversity decreases dramatically in the last few generations, which leads to a lower selection differential ($i.e.$, $\bar{f}_{t+1} - \bar{f}_t$) and thus a lower selection intensity. □

(c) How does Thierens' model compare with your results?

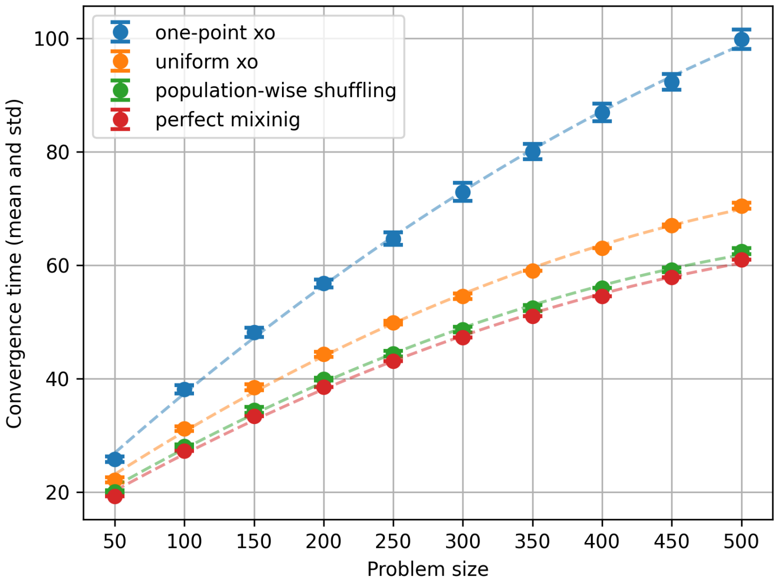


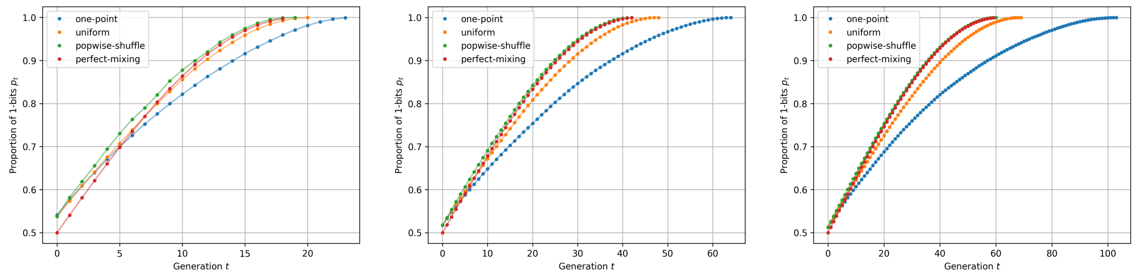
Figure 4: Convergence time of different XO methods and perfect mixing

The main difference between Thierens' model and the experiments lies in the mixing method. Thierens' model assumes perfect mixing, which employs a crossover operator that significantly disrupts the population and establishes an underlying distribution of fitness. To compare the results of the three XO methods with Thierens' model, we first evaluate the convergence time of the three XO methods relative to perfect mixing. Then, we assess disruption by examining the uniformity and standard deviation of the fitness values within the population.

The values and the trends of convergene time are illustrated in Figure 4. The convergence time of the perfect mixing is computed from the equation of Thierens' model:

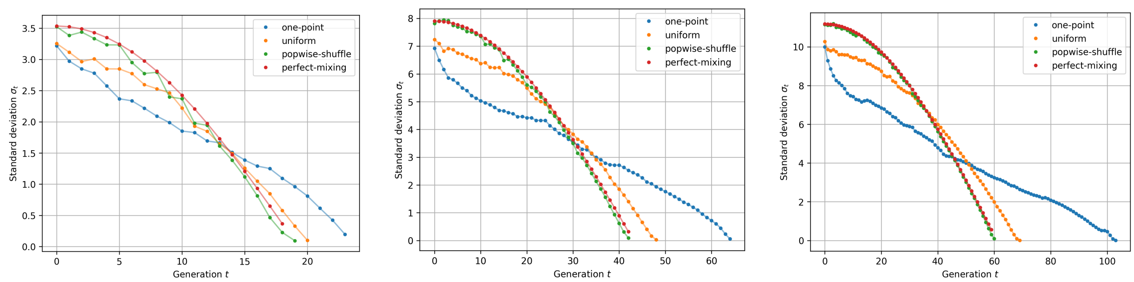
$$t_{conv} = \frac{\pi}{2} \times \frac{\sqrt{ell}}{I},$$

where I is the theorectical value 0.576291 of selection intensity. The results show that the convergence time of the perfect mixing is the shortest for each problem size.



(a) $ell = 50$ (b) $ell = 250$ (c) $ell = 500$

Figure 5: Unitation of different XO methods



(a) $ell = 50$ (b) $ell = 250$ (c) $ell = 500$

Figure 6: Standard deviation of different XO methods

The proportion of the unitation and the standard deviation of the fitness values with problem size 50, 250, and 500 are shown in Figures 5 and 6, respectively. In Figure 5, one-point XO executes more generation to reach the proportion of the unitation 1 than any other XO methods. This reflects the curves of the convergence time in Figure 2 that the one-point XO has the longest convergence time. In Figure 6, the standard deviation of the one-point XO and the uniform XO are lower than the perfect mixing and the population-wise shuffling at the beginning of the generations. However, the decreasing ratio of the one-point XO and the uniform XO are slower than the perfect mixing and the population-wise shuffling. This indicates that the one-point XO and the uniform XO have a lower degree of disruption than the perfect mixing and the population-wise shuffling.

In summary, among the three XO methods, the curve trends of perfect mixing and population-wise shuffling exhibit the highest similarity. Particularly for large problem sizes, the curve trends of perfect mixing and population-wise shuffling nearly coincide, suggesting that population-wise shuffling closely resembles perfect mixing in terms of unitation and standard deviation of fitness values. Therefore, in the OneMax problem executed by SGA, utilizing the XO method with greater disruption leads to shorter convergence times. □