

Projeto de Computação Evolucionária

Experiência de Benchmark entre Diferentes Métodos de Seleção de Progenitores

André Clemêncio e Joel Pires

25 de Maio de 2018

Resumo

A convergência prematura é um problema bastante conhecido no que diz respeito a algoritmos evolucionários e, de modo a ser reduzida, é necessário garantir uma grande diversidade da população. Neste documento iremos testar e comparar diferentes algoritmos para a seleção dos progenitores. Um é baseado na escolha por torneio, outro usando uma amostragem estocástica uniforme e, por último, usamos o método de seleção por roleta. O objetivo será perceber qual deles resulta tanto numa maior diversidade da população assim como numa maior qualidade do resultado final em termos de performance.

1 Introdução

Na área de Computação Evolucionária é importante perceber de que forma a alteração nos indivíduos de uma população vai afetar a qualidade dos resultados que serão posteriormente obtidos. No nosso caso específico iremos testar e comparar diferentes métodos para a seleção dos progenitores. O objetivo desta seleção será fazer com que a geração seguinte seja mais adaptado, ou seja, obtenha melhores resultados relativos à resolução de um determinado problema. Para isso é preciso garantir que haja diversidade suficiente para procurar em novas zonas do espaço de soluções e assim evitar a convergência prematura.

Com este trabalho iremos tentar comparar três diferentes métodos para a seleção dos progenitores que são: seleção por torneio, seleção por roleta e ainda seleção através de uma amostragem estocástica uniforme. O objetivo será, então, obter resultados quanto à performance de cada um dos métodos (relativamente à qualidade do resultado e ao grau de diversidade da população) e tentar perceber qual deles será o mais indicado para a resolução do problema.

2 Problema de Benchmark

2.1 Escolha do Problema

Optámos pelo clássico *Travelling Salesman Problem (TSP)* - ou problema do Caixeiro-Viajante em português. É um problema subejamente conhecido, mas não deixa de ser por isso um problema complexo e cuja otimização ainda seja passível de melhorias. Por isso, fomos levados a querer saber que resposta os algoritmos evolucionários dariam a este emblemático problema.

2.2 Descrição Geral

O objetivo do TSP é encontrar o caminho mais curto relativamente a um determinado número de cidades, começando num determinado ponto e passando por todas as cidades regressando, no fim, ao ponto inicial.

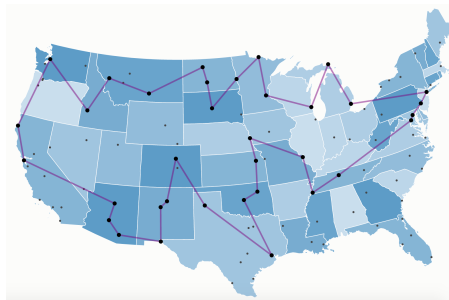


Figura 1: Exemplo de uma instância do problema de TSP

2.3 Representação do Problema

Para resolução deste problema pode-se adicionar ou excluir novas cidades, ou pode-se alterar a distância entre pares de cidades. Foi implementada esta última opção, pois corresponde melhor a uma situação simples do mundo real. Cada indivíduo é constituído por um *array* de dois elementos. O primeiro elemento diz respeito à sequência de cidades a visitar e o segundo elemento diz respeito à medida de *fitness* avaliada. O fenótipo é um inteiro que caracteriza uma determinada cidade e o seu genótipo é um conjunto de duas coordenadas cartesianas.

2.4 Função de *Fitness*

A função de *fitness* consiste na distância euclidiana entre a sequência de cidades apresentada por cada solução/indivíduo. Atenção que o TSP é um problema de minimização e, por isso, quanto menor for o valor de *fitness* melhor adaptado será esse indivíduo ao problema.

2.5 Função de Diversidade

Para medir a diversidade usámos a distância de Hamming. Na prática, faz-se uma comparação entre dois indivíduos (que têm naturalmente o mesmo comprimento) e conta-se qual é o número de posições nas quais eles diferem entre si. Fazemos isso para a população toda de uma geração, depois para todas as gerações, de seguida fazemos a média de todas as medidas de diversidade de todas as gerações e chega-se a um só valor representativo da diversidade daquela *run*. Os valores podem variar entre 0 (diversidade nula) e 2 (diversidade total).

3 Métodos de seleção dos progenitores

3.1 Torneio

Na escolha por torneio são selecionados, aleatoriamente, um determinado número de elementos da população. Quanto maior for o tamanho do torneio menos probabilidade existe de indivíduos mais fracos serem escolhidos visto que muito provavelmente existem indivíduos mais fortes no mesmo torneio. É escolhido como progenitor o melhor entre os selecionados. O mesmo processo é repetido para a escolha dos outros progenitores.

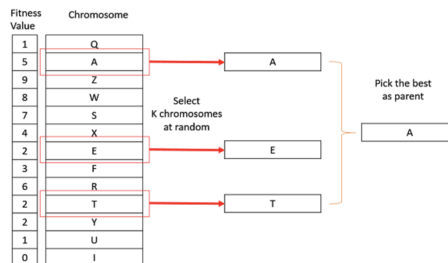


Figura 2: Torneio

3.2 Roleta

Na seleção pelo método da roleta é atribuído uma percentagem a um indivíduo sendo calculada através da valor de *fitness* do indivíduo sobre a soma de todos os valores de *fitness* dos elementos da população. Em seguida é gerado um valor aleatório entre 0 e 1. Vai sendo acumulado ao *fitness* a percentagem de cada indivíduo. Será escolhido o primeiro indivíduo em que o valor da soma acumulada ultrapasse o valor gerado aleatoriamente.

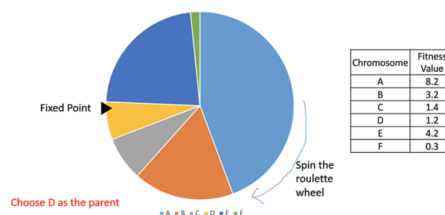


Figura 3: Roleta

3.3 Amostragem estocástica uniforme

O método de escolha através da amostragem estocástica uniforme é bastante parecido ao método pela roleta. A diferença está no facto de que, neste caso, são gerados de uma só vez todos os pontos fixos (a uma distância de 1/elementos a escolher) e, consequentemente, todos os progenitores são escolhidos de uma só vez.

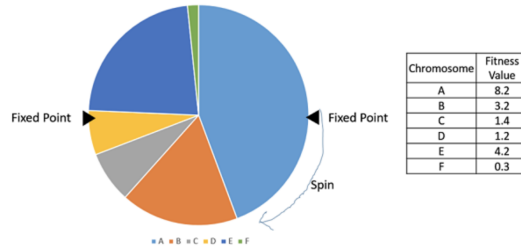


Figura 4: Amostragem Estocástica Uniforme

4 Ambiente Experimental

Em relação ao ambiente em que testámos o nosso programa tivemos em conta inúmeros parâmetros. Como o nosso objetivo será perceber a influência que o algoritmo da escolha dos progenitores tem nos valores de *fitness* e da diversidade de uma população, variámos apenas este parâmetro testando para os três algoritmos já referidos anteriormente. No caso do algoritmo do Torneio variámos ainda o valor do tamanho do torneio - um caso com o tamanho igual a 10 e um outro caso com o tamanho igual a 50. Esta última variação pareceu-nos interessante fazer para também tirar-mos algumas conclusões sobre a influência que o tamanho do torneio tem na variável dependente.

Foram usados dois computadores para os testes. Um deles possuía o sistema operativo Windows e outro MacOS. O primeiro tinha um processador i5 e o segundo tinha um i7. Corremos então o algoritmo evolucionário 30 vezes para cada método de seleção de progenitores e para uma instância do problema do TSP de 259 cidades. Em média, o programa finalizava 30 *runs* ao fim de 10 horas, precisando cerca de 40 horas para finalizar as 120 *runs*. Para cada *run* registámos as seguintes métricas: melhor indivíduo de todas as gerações dessa *run*; média dos melhores indivíduos ao longo das gerações dessa *run*; média dos valores de diversidade registados ao longo das gerações da *run*.

4.1 Ficheiro de teste

Foi utilizado um ficheiro de teste *zi929.tsp* que tem informação relativa a coordenadas cartesianas de 259 cidades do Zimbabwe. Poderíamos ter feito ainda com mais cidade, mas o tempo que levaria a correr o programa aumenta grandemente dada também a complexidade do problema em questão.

4.2 Formalização do Problema

O problema ou questão principal a que nos propomos responder durante este estudo estatístico é:

- Qual é o método de seleção de progenitores que promove mais a diversidade das populações?

Importante dizer que, como nós queremos combater a convergência prematura, decidimos apenas levar em consideração a medida da diversidade para os testes estatísticos. No entanto, faremos também uma análise mais superficial usando os valores de *fitness* de cada grupo de dados para tentar saber se a promoção de diversidade coincide com o surgimento dos indivíduos mais adaptados ao problema.

4.3 Variáveis Dependentes

Embora, como já mencionado, estejamos a retirar três métricas diferentes, nomeadamente: melhor indivíduo, média dos melhores indivíduos e média das diversidades ao longo das gerações de uma *run*; apenas uma delas é considerada a nossa variável dependente para avaliar a diversidade dos métodos de seleção, a saber, a média de diversidade registada ao longo das gerações.

4.4 Variáveis Independentes

As variáveis que nós vamos fazer variar ao longo da experiência são unicamente o método de seleção de progenitores e terá os seguintes níveis: seleção por Torneio de tamanho igual a 10; seleção por Torneio de tamanho igual a 50; seleção por Roleta; seleção por Amostragem Estocástica Uniforme.

4.5 Variáveis de Controle

Muitas são as nossas variáveis de controle, precisamente porque queríamos isolar e avaliar especificamente a influência dos métodos de seleção de progenitores nos resultados. Segue-se a lista dessas variáveis:

- Instância do problema de TSP;
- Número de gerações;
- Tamanho da população;
- Probabilidade de mutação;
- Probabilidade de crossover;
- Tamanho da seleção de elitismo;
- Algoritmo de recombinação
- Algoritmo de mutação
- Método de seleção dos sobreviventes

4.6 Valores das Variáveis de Controle

Antes de avançar com a experiência foram realizados alguns testes de forma apurar valores para os parâmetros que se mostrassem razoáveis, i.e., valores que nos permitissem ter gerações e populações suficientemente grandes para retirar conclusões significativas e simultaneamente que resultassem num bom compromisso entre tempo de execução do programa e valor de *fitness* dos indivíduos. Em baixo apresenta-se uma tabela com os valores dos parâmetros de controle que foram usados para a experiência:

Número de Gerações	Tamanho da População	Probabilidade de Mutação	Probabilidade de Crossover	Porcentagem da elite
250	70	5%	75%	10%

Tabela 1: Valores Iniciais dos parâmetros

4.6.1 População

A população é gerada de forma randomizada. Para além disso, para cada teste diferente que se efetue, uma nova população é gerada, não havendo duas populações iguais entre testes.

4.6.2 Método de Seleção dos Sobreviventes

O método usado como já previamente deduzido foi a seleção elitista. Esta seleção evita que os indivíduos de melhor qualidade se percam, assim sendo permanece um determinado número dos melhores indivíduos para a geração seguinte. O resto dos indivíduos obtêm-se utilizando outro método de seleção, atuando sobre todo a população inicial.

4.6.3 Operadores de Variação

Algoritmo de Recombinação: O algoritmo de recombinação genética usado foi o *ordered crossover* - *OX*. Neste algoritmo, uma parte de um dos progenitores é mapeada para uma parte do cromossoma do outro progenitor. Substituindo-se essa porção, preenche-se o resto pelos genes que ficaram; genes já presentes são omitidos e a ordem é preservada.

Algoritmo de Mutação: O algoritmo usado foi o *swap mutation*. Neste caso, selecionam-se aleatoriamente duas posições no cromossoma e trocam-se simplesmente os valores.

5 Análise da Estatística

5.1 Descrição dos Dados

Os dados são compostos por 4 grupos relativamente a quatro métodos de seleção de progenitores diferentes, a saber: Torneio 10, Torneio 50, Roleta, Amostragem Estocástica Uniforme. Cada grupo de dados é constituído por 30 entradas diferentes correspondentes a 30 runs diferentes e cada entrada regista a média do valores de diversidade ao longo das gerações dessa *run* - medido através da distância de Hamming. Todos estes dados podem ser consultados no **Anexo A**. Fazendo uma análise descritiva dos dados, conclui-se o seguinte em relação aos 4 grupos de dados:

- O valor **mínimo** de diversidade registado foi de **0.67918**
- O valor **máximo** de diversidade registado foi de **1.99841**
- A **média** de valores é de **1.46176**
- A **mediana** é **1.56893**
- A **mediana** é **1.56893**
- A **variância** é de **0.47452**
- O **desvio padrão** é de **0.22517**
- A **moda** de cada grupo de dados é, pela ordem que se apresentam no **Anexo A**, a seguinte: **1.1064**, **0.6791**, **1.9942** e **1.6869**
- A medida de **assimetria** de cada grupo de dados é, pela ordem que se apresentam no **Anexo A**, a seguinte: **0.18404209**, **0.17815938**, **-0.48756217** e **-0.71024764**
- A **curtose** de cada grupo de dados é, pela ordem que se apresentam no **Anexo A**, a seguinte: **-0.6042826**, **-0.65389361**, **-0.63450864** e **0.4371124**
- Valor do **primeiro quartil**: **1.0534**
- Valor do **segundo quartil**: **1.56893**
- Valor do **terceiro quartil**: **1.8844**

Para termos uma noção de como os valores de **diversidade** de cada grupo de dados se encontram **distribuídos** foi elaborado um diagrama de caixa que pode ser visto na **Figura 6**.

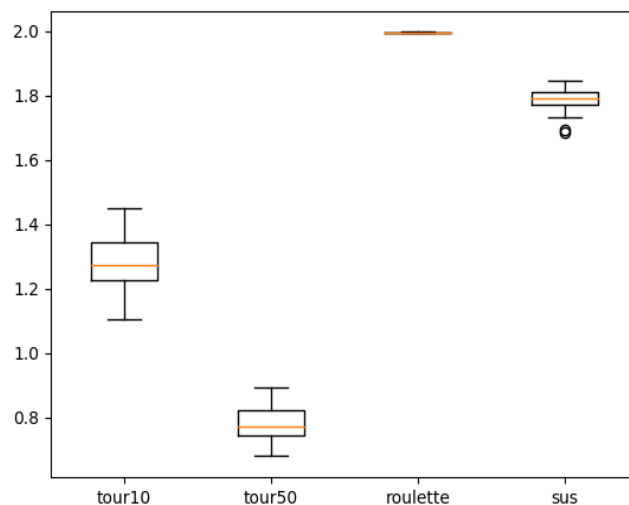


Figura 6: Diagrama de caixa com os dados distribuídos pelos 4 métodos

Além disso, elaborámos histogramas para cada grupo de dados que nos ajudaram a ver a **frequência** dos valores de cada grupo de dados. As **Figuras 7, 8, 9 e 10** representam as imagens obtidas.

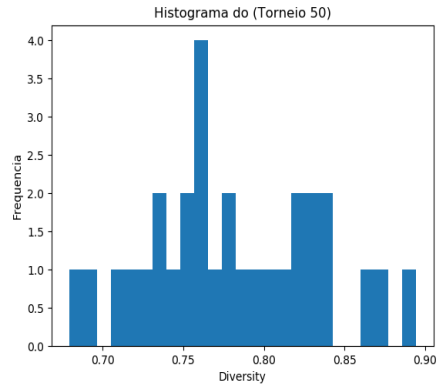
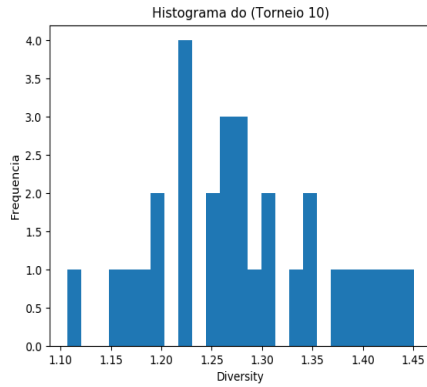


Figura 7: Histograma para o método do Torneio (10) Figura 8: Histograma para o método do Torneio (50)

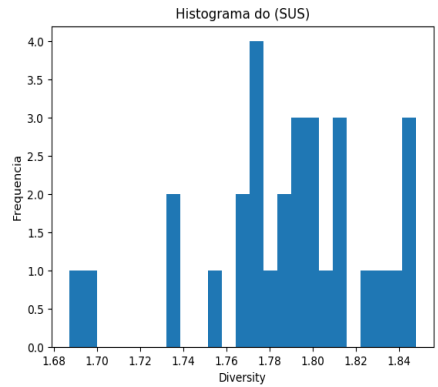
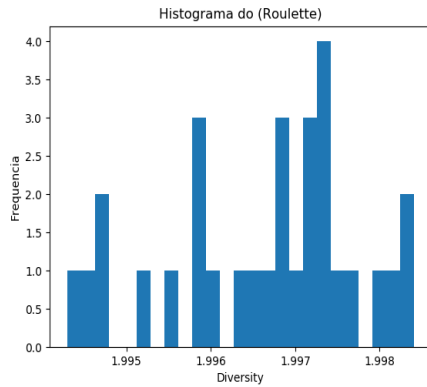


Figura 9: Histograma para o método da Roleta

Figura 10: Histograma para o método por Amostragem Estocástica Uniforme

5.2 Análise e Conclusões Preliminares

Ao analisar os dados não nos podemos esquecer que estamos a falar de um problema de **minimização**. Consequentemente, quanto menor for a pontuação de fitness do indivíduo, melhor estará este adaptado ao problema. Através de uma análise preliminar dos resultados que se encontram discriminados no **Anexo A**, poderemos concluir que:

- É no **método de seleção por torneio** onde se pode obter o **melhor indivíduo possível** pois o valor **mínimo** (118824.125), **máximo** (151823.154) e **médio** (132945.331) de fitness do melhor indivíduo supera qualquer um dos outros três métodos.
- É também através deste método que nós conseguimos obter **em média os melhores** indivíduos dado que o valor **mínimo** (157079.81675), **máximo** (183643.215341) e **médio** (167088.011934) - na coluna *média dos melhores* - superando de igual forma qualquer um dos outros três métodos.
- O **tamanho** do método de torneio é tão importante que nos permite passar da **melhor pontuação possível** (118824.125) para a **pior pontuação obtida** (175915.662)
- Curiosamente, é também quando aumentamos o tamanho deste método que obtemos os **valores de diversidade mais baixos** registados relativamente aos outros três métodos, quer o valor **mínimo** (0.679182), **máximo** (0.894503) ou **médio** (0.7809888) de diversidade.
- Quando usámos o método por **roleta**, obtivámos a **pior média dos melhores** indivíduos, com os valores de **mínimo**, **máximo** e **médio** de 188151.8556, 210284.8843 e 200170.9901, respetivamente.

- Esse método também registou o **pior valor de *fitness*** de todos relativamente ao **mínimo** (146225.6996) e **médio** (158510.7202) do **melhor** indivíduo.
- Curiosamente, é também quando usámos esse método que observámos valores **mínimos**(1.99429), **máximos**(1.99841) e **médios** (1.99659) de **diversidade mais altos**.

5.3 Inferência

Antes de procedermos para os testes estatísticos, temos que nos debruçar sobre qual será o teste estatístico apropriado. O nosso conjunto de resultados é constituído por **quatro** grupos diferentes correspondentes aos quatro diferentes métodos de seleção dos progenitores. Para além disso, as populações alvo da experiência são **diferentes** entre cada grupo de testes. Posto isto, falta saber se deveremos optar por testes paramétricos ou não. Para isso, estudemos a **normalidade** e **homogeneidade** dos dados primeiro. Os outputs dos nossos testes foram os seguintes: **melhor indivíduo** (indivíduo com o menor valor de fitness), **média dos melhores** indivíduos de cada uma das gerações e **média das medidas diversidade** ao longo das gerações.

Dado que queremos saber qual o método que melhor evita a **convergência prematura** dos resultados, achamos por bem realizar os testes estatísticos relativamente à medida de **diversidade**. Da visualização do gráfico da **Figura 6** percebemos que os dados distribuem-se de forma bem distinta e, presumivelmente, será desnecessário estar a realizar todos os pares de testes entre os quatro métodos de seleção. Contudo, para corroborar estatisticamente e cientificamente as nossas conclusões, é exatamente isso que vamos fazer. Prossigamos então para a realização de testes de normalidade e homogeneidade, que se serão descritos a seguir, a fim de determinar se optaremos por testes paramétricos ou não.

5.3.1 Normalidade

Antes de procedermos ao teste de hipótese relativamente à normalidade, foram elaborados histogramas com a normal sobreposta de forma a termos um apoio visual nas conclusões a retirar quanto à normalidade de cada grupo de dados. As **Figuras 11, 12, 13 e 14** representam os histogramas obtidos.

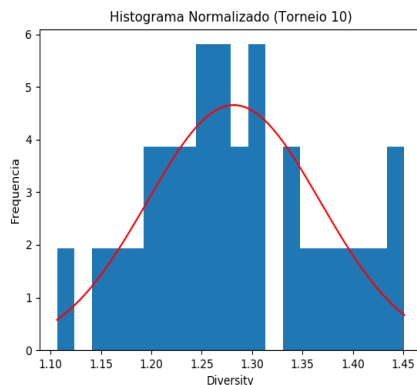


Figura 11: Histograma normalizado (Torneio 10)

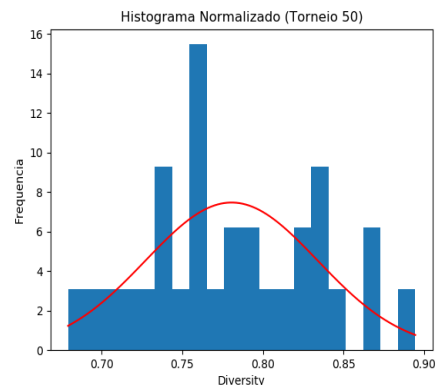


Figura 12: Histograma normalizado (Torneio 50)

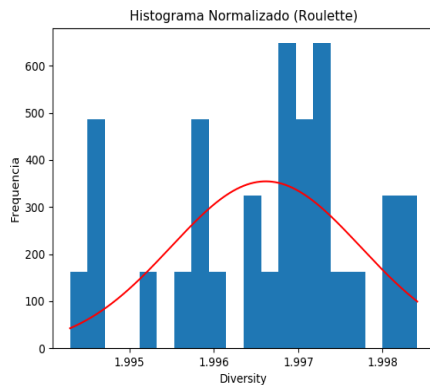


Figura 13: Histograma normalizado (Roleta)

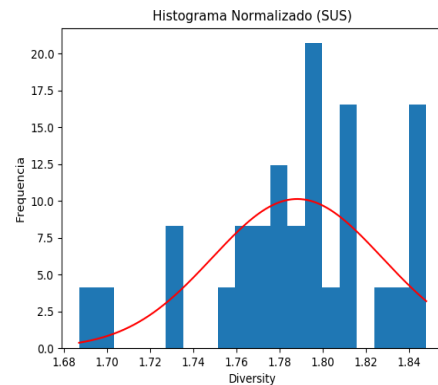


Figura 14: Histograma normalizado (SUS)

Relativamente ao teste da normalidade, admitamos as seguintes hipóteses para um nível de significância de **0.05**:

- **Hipótese Nula (H0):** Os dados seguem uma distribuição normal.
- **Hipótese Alternativa (H1):** Os dados não seguem uma distribuição normal.

O teste de **Kolmogorov-Smirnov** pode ser usado para testar a qualidade de um ajuste. No nosso caso, ajuda-nos porque as amostras serão padronizadas e comparadas com uma distribuição normal, isto é, a média e a variância das amostras ficarão iguais às da distribuição normal. Usando a função *kstest()*, obtiveram-se os seguintes valores para cada um dos quatro métodos de seleção:

Torneio 10		Torneio 50		Roleta		Estocástico	
Teste estatístico	p-valor	Teste estatístico	p-valor	Teste estatístico	p-valor	Teste estatístico	p-valor
0.38031	0.000207	0.43905	8.79996 e-06	0.44838	5.07018 e-06	0.33976	0.00135

Tabela 1: Resultados do Teste de Kolmogorov-Smirnov

De acordo com os dados observados, conclui-se que existe **pelo menos um** grupo de dados que não verifica normalidade. De facto, **nenhum deles verifica normalidade** dado que o p-valor de cada um deles é menor do que o valor de **alfa** e, portanto, **rejeitamos H0**.

A seguir vamos prosseguir para outro teste de normalidade que assume as mesmas hipóteses formuladas acima - é o teste de **Shapiro-Wilk**. Este teste tem a mesma finalidade, contudo revela-se mais preciso que o de Kolmogorov-Smirnov para grupos de dados pequenos. Fazendo uso da função *shapiro()* do módulo *scipy*, obtiveram-se os seguintes valores:

Torneio 10		Torneio 50		Roleta		Estocástica	
Teste Estatístico	p-valor	Teste Estatístico	p-valor	Teste Estatístico	p-valor	Teste Estatístico	p-valor
0.9812	0.8565	0.9822	0.8810	0.9458	0.1310	0.9414	0.0997

Tabela 2: Resultados do Teste de Shapiro-Wilk

Através deste teste nós concluímos **algo diferente**. Uma vez que os p-valores de todos os grupos de dados são superiores ao nível de significância, então **aceitamos a hipótese nula**.

Dado que foram efetuados dois testes de normalidade com conclusões contraditórias e depois de uma análise das **Figuras 11, 12, 13 e 14**, concluímos que os dados **não seguem uma distribuição normal**. Posto isto, avançar para o teste da homogeneidade torna-se irrelevante, contudo iremos proceder ao teste da variância para efeitos de aprendizagem e corroboração dos resultados.

5.3.2 Homogeneidade

Tal como o teste de normalidade, elaboremos igualmente as seguintes hipóteses assumindo um nível de significância de **0.05**:

- **Hipótese Nula (H0):** A variância entre distribuições é homogênea.
- **Hipótese Alternativa (H1):** A variância entre distribuições não é homogênea.

O teste de **Levene** é um teste apropriado nesta situação. O teste de Levene é um teste estatístico inferencial que avalia a igualdade de variâncias de uma variável em dois ou mais grupos de dados. Existem **três variantes** do teste de Levene:

- usando a **mediana**: apropriado para distribuições não normais.
- usando a **média**: apropriada para distribuições simétricas.
- usando a função **trimmed**: ideal para distribuições de cauda longa.

O nosso caso enquadra-se na primeira variante, ou seja, temos de usar a mediana no teste de levene. Usando a função *levene()* obtiveram-se os seguintes resultados:

Teste estatístico	p-valor
21.18666	5.157345e-11

Tabela 3: Resultados do Teste de Levene

De acordo com os dados acima, vemos que o p-valor é **menor** que o valor de significância (0.05) e, por isso, **rejeitamos H₀**, concluindo-se que **não se verifica homogeneidade**.

5.3.3 Análise a todos os grupos

Dado que temos mais que dois grupos de dados e que não verificam nem a normalidade nem a homogeneidade, é necessário testar de forma não-paramétrica todas as amostras simultaneamente de forma a verificar se há amostras que se originam da mesma distribuição. Para esse efeito, procederemos ao **teste de Kruskal-Wallis**. Este teste pode indicar então que existe uma amostra que domina estocasticamente uma outra amostra. Contudo, não é referenciado onde essa dominância ocorre ou para quantos pares de grupos se obtém essa dominância. Elaboremos então novamente duas hipóteses com um nível de significância de **0.05**:

- **Hipótese Nula (H₀):** A média de todos os grupos de dados é igual.
- **Hipótese Alternativa (H₁):** A média de todos os grupos de dados é diferente.

Usando a função *kruskal()* obtiveram-se os seguintes resultados:

Teste estatístico	p-valor
111.570247	5.03946e-24

Tabela 4: Resultados do Teste de Kruskal-Wallis

Analisando a tabela acima, verifica-se que o p-valor é **menor** que 0.05 e, por isso, **rejeita-se H₀**, ou seja, a média de todos os grupos de dados é diferente e há realmente **diferenças estatísticas significativas entre os dados**.

5.3.4 Análise a cada par de grupos de dados

Conforme já mencionado, a rejeição da hipótese nula no teste de kruskal não indica qual dos grupos é diferente. Teremos então de efetuar comparações post-hoc entre pares de grupos de dados para realmente determinar quais grupos diferem e de que forma. O nosso objetivo não é provar que determinado método é melhor que os outros todos a nível de diversidade. Queremos sim fazer uma **análise comparativa entre todos os quatro métodos de seleção de progenitores quanto à sua diversidade** conforme foi explicitado na secção da formalização do problema. Conforme também já foi mencionado na análise preliminar dos resultados que fizemos anteriormente, através da visualização da **Figura 6** conclui-se que a análise de todos os pares de dados é desnecessária. Contudo, queremos confirmar a nossa análise preliminar com uma análise estatística e rigorosa. Assim sendo, iremos, no limite, realizar seis testes não-paramétricos, nomeadamente:

- Torneio 10 vs Torneio 50
- Torneio 10 vs Roleta
- Torneio 10 vs Estocástica Uniforme
- Torneio 50 vs Roleta
- Torneio 50 vs Estocástica Uniforme
- Roleta vs Estocástica Uniforme

À medida que fizermos cada teste pode-se revelar desnecessário fazer outros, pelo que, na prática, poderá ser redundante ter de fazer exatamente seis análises comparativas. O teste que se revela mais apropriado para os múltiplos testes aos pares de grupos é o teste não-paramétrico de **Mann-Whitney**. Este é um teste apropriado para duas **amostras independentes** e é considerado a versão da rotina de teste não-paramétrico de t-Student. Antes de avançar, debruçemo-nos um pouco sobre um cuidado a ter nas múltiplas comparações - a correção de Bonferroni.

5.3.5 Correção de Bonferroni

De forma a evitar o acumular de erros resultante das múltiplas comparações, se quisermos manter a mesma probabilidade global de erros de tipo I, temos de corrigir o valor de alfa para cada comparação. Uma forma de o fazer é recorrendo à correção de Bonferroni. Funciona da seguinte forma: o novo valor de alfa é obtido dividindo o valor original (no nosso caso 0.05) pelo número de comparações que queremos fazer (no nosso caso 6). O novo valor de significância, i.e., valor de alfa, será de **0,0083**.

5.3.6 Torneio 10 vs Torneio 50

Admitamos as seguintes duas hipóteses:

- **Hipótese Nula (H0):** O valor médio de diversidade do Torneio 50 é maior ou igual que o do Torneio 10.
- **Hipótese Alternativa (H1):** O valor médio de diversidade do Torneio 50 é menor que o do Torneio 10.

Realizando o teste de Mann-Whitney, obtiveram-se os seguintes valores:

Teste estatístico	p-valor
0.0	9.05957e-11

Tabela 5: Torneio 10 vs Torneio 50

O p-valor é **menor** que o nível de significância e, portanto, **rejeitamos H0**, concluindo-se que o **valor médio de diversidade do Torneio 10 é maior que o do Torneio 10**.

5.3.7 Torneio 10 vs Roleta

Admitamos as seguintes duas hipóteses:

- **Hipótese Nula (H0):** O valor médio de diversidade da Roleta é menor ou igual que o do Torneio 10.
- **Hipótese Alternativa (H1):** O valor médio de diversidade da Roleta é maior que o do Torneio 10

Depois de efetuar o teste de Mann-Whitney, chegaram-se aos seguintes valores:

Teste estatístico	p-valor
900.0	9.05957e-11

Tabela 6: Torneio 10 vs Roleta

O p-valor mais uma vez é **menor** que o nível de significância e, por isso, **rejeita-se H0** e conclui-se que o **valor médio de diversidade da Roleta é maior que o do Torneio 10**

5.3.8 Torneio 10 vs Estocástica Uniforme

Admitamos as seguintes duas hipóteses:

- **Hipótese Nula (H0):** O valor médio de diversidade da Amostragem Estocástica Uniforme é menor ou igual que o do Torneio 10.
- **Hipótese Alternativa (H1):** O valor médio de diversidade da Amostragem Estocástica Uniforme é maior que o do Torneio 10.

O teste de Mann-Whitney devolveu-nos os seguintes resultados:

Teste estatístico	p-valor
900.0	9.05957e-11

Tabela 7: Torneio 10 vs Estocástica Uniforme

O p-valor mais uma vez é **menor** que o nível de significância e, por isso, **rejeita-se H0** e conclui-se que o valor médio de diversidade da Amostragem Estocástica Uniforme é maior que o do Torneio 10. Neste momento torna-se evidente que os dois testes que planeávamos fazer a seguir tornam-se **redundantes**, a saber: Torneio 50 vs Roleta e Torneio 50 vs Amostragem Estocástica Uniforme. Isto porque dos testes anteriores já se pode inferir que o valor médio de diversidade do Torneio 50 é de facto inferior ao da Roleta e da Amostragem Estocástica Uniforme. Resta-nos então fazer uma análise comparativa entre o método da Roleta e o da Amostragem Estocástica Uniforme.

5.3.9 Roleta vs Estocástica Uniforme

Admitamos as seguintes duas hipóteses:

- **Hipótese Nula (H0):** O valor médio de diversidade da Amostragem Estocástica Uniforme é maior ou igual ao da Roleta.
- **Hipótese Alternativa (H1):** O valor médio de diversidade da Amostragem Estocástica Uniforme é menor ao da Roleta.

Do teste de Mann-Whitney obtivemos o seguinte:

Teste estatístico	p-valor
0.0	9.05957e-11

Tabela 8: Roleta vs Estocástica Uniforme

E novamente nos deparamos com a situação em que o p-valor é **menor** que o nível de significância e, consequentemente, **rejeitamos H0** e concluímos que o **valor médio de diversidade da Amostragem Estocástica Uniforme é menor que a da Roleta**.

5.3.10 Conclusões Finais

Para além das conclusões retiradas a secção de Análise e Conclusões Preliminares, depois dos testes estatísticos feitos pode-se concluir que:

- O método que evita melhor a convergência prematura é a Roleta.
- O método que pior lida com a convergência prematura é o Torneio de tamanho 50.
- De facto, querendo fazer uma comparação completa, desde o que promove mais diversidade para o que promove menos, temos que: Roleta > SUS > Torneio 10 > Torneio 50.
- Contudo verifica-se algo interessante: a promoção de diversidade não é sinónimo de obtenção dos melhores indivíduos. Prova disso é o método de roleta que, embora o seu valor médio de diversidade seja superior ao dos outros métodos, possui os piores indivíduos a nível de *fitness*.
- Contudo, não se pode concluir que há uma proporcionalidade inversa entre diversidade da população e *fitness* dos indivíduos. Para isso vejamos a seguir em mais detalhe dois métodos: o Torneio de tamanho 10 e o torneio de tamanho 50.
- O Torneio de tamanho 50 é, segundo os testes estatísticos, aquele cujo valor médio de diversidade é mais baixo, contudo não é o que apresenta os piores valores de *fitness*.
- O torneio de tamanho 10 é aquele que contém os melhores valores de *fitness* e aparente ter valores de diversidade razoáveis, de facto, não é o que apresenta pior valor médio de diversidade de acordo com os testes estatísticos.
- A Amostragem Estocástica Uniforme mostrou-se com valores razoáveis quer de performance, quer de diversidade, não atingindo nenhum extremo, i.e., não consegue ser nem o melhor nem o pior em nenhuma das métricas.
- O tamanho do método do torneio importa. De facto, de acordo com os testes, tanto maior o tamanho do torneio, tanto menor será o seu valor da diversidade; contudo, olhando para a tabela de resultados tanto menor será também o *fitness* dos seus indivíduos.

Nota: Todos os valores dos testes podem ser consultados no **Anexo B**.

6 Conclusão

Os resultados vão de encontro às nossas expectativas. Debruçando-nos um pouco sobre a natureza dos métodos da seleção de progenitores, vemos que é natural que os métodos por torneio promovam menor diversidade que os métodos de roleta e da amostragem estocástica uniforme. No caso da amostragem estocástica uniforme é parecida com a roleta mas reduz a aleatoriedade ao uniformizar as probabilidades de escolha. A aleatoriedade presente nos métodos do torneio é drasticamente reduzida em relação aos outros métodos dado que de um grupo de indivíduos seleciona-se o melhor (não é uma escolha aleatória). Naturalmente, por causa disto, será no método do torneio onde encontraremos os melhores indivíduos, embora não exploremos outras zonas do espaço de soluções que poderão ser adequadas ao problema.

No balanço geral, esta experiência de Benchmark foi útil para aprofundar conhecimentos na modelação e representação de problemas reais em algoritmos evolucionários. Sobretudo foi essencial para realizar um estudo estatístico rigoroso aplicado ao campo da Computação Evolucionária e contribuir cientificamente para uma questão tão importante nesta área como é a de saber quais são os métodos de seleção de progenitores que melhor evitam a convergência prematura.

7 Bibliografia

- Introduction to Evolutionary Computation, A. Eiben and J. Smith (2 ed.), Springer, 2015
- Ernesto Costa, Evolutionary Computation Slides Code, DEI-FCTUC, University of Coimbra, 2018
- www.docs.scipy.org

Anexo A

	Torneio (10)			Torneio (50)			Roleta			Amostragem Estocástica Uniforme		
	Melhor	Média dos Melhores	Diversidade	Melhor	Média dos Melhores	Diversidade	Melhor	Média dos Melhores	Diversidade	Melhor	Média dos Melhores	Diversidade
	127225,161	164701,7333	1,40261917	148475,662	192406,2753	0,89450306	164683,948	203691,8446	1,99452219	156915,666	195860,156	1,77936425
	127474,007	160288,4941	1,18542244	155029,148	199214,9083	0,7584591	156358,071	190331,3055	1,99559979	151436,625	195891,4217	1,75145634
	118824,125	160081,3267	1,38505159	157666,613	192462,5652	0,76716306	148926,1	191856,8529	1,9968969	160380,712	202211,0986	1,76661927
	138013,235	173347,1102	1,37842966	159556,564	191631,9501	0,82074384	168056,261	205048,7357	1,99676054	151226,249	196409,2852	1,7932733
	125739,344	165244,5708	1,45088844	152906,781	186113,6617	0,82347444	151716,594	193266,6455	1,99716963	160950,91	196289,5307	1,69525805
	146460,966	183643,2153	1,10649472	150990,982	195125,2168	0,76149235	166174,663	210284,8843	1,9984102	154295,466	194772,1345	1,83436188
	139988,329	168262,2185	1,26778252	163429,756	189207,6785	0,67918215	150828,214	193720,7981	1,99695677	147409,181	190837,3248	1,81246726
	126210,499	161835,9402	1,19329492	160474,458	188090,511	0,69059676	157202,246	198990,6302	1,99726608	154985,886	199487,3969	1,77589197
	136266,906	168484,5521	1,27049648	154308,278	186267,615	0,83032918	156356,26	199150,3715	1,99816408	153090,099	194624,7789	1,73485661
	122465,794	159505,0905	1,28521373	153647,417	190452,8178	0,80301662	154672,88	199439,656	1,99602884	158708,495	197421,6187	1,77275229
	132431,916	165094,3002	1,22938462	158988,309	192079,264	0,83291011	156299,678	202604,3334	1,99681043	154290,785	189357,9069	1,8103819
	129176,401	162829,6266	1,22489461	148554,541	185521,3243	0,84220942	163768,696	209628,3888	1,99578272	151754,496	188806,6521	1,7922988
	128782,535	167637,9417	1,22821389	140558,834	176510,944	0,83614957	153034,342	192302,9873	1,99715965	166588,308	203435,0595	1,76718801
	130883,086	167230,261	1,34503397	148119,236	184422,7158	0,86373153	158464,484	199051,4506	1,99737251	163087,066	197808,1756	1,68698978
	126457,92	161736,8384	1,34801733	158620,642	185173,2524	0,71672196	146941,222	196258,7853	1,99747562	157992,142	199119,1685	1,82489295
	120977,507	157079,8168	1,29632235	147708,982	194130,4312	0,78932375	166935,259	208613,2175	1,99591908	166255,857	201093,3774	1,77665694
	127216,468	167785,0747	1,41980759	154696,242	193181,8285	0,81643344	164918,805	205966,3952	1,99578937	155208,259	202901,4025	1,77542302
	125573,848	159601,4547	1,33771026	150835,758	184550,1811	0,74336601	153919,784	196213,2777	1,99647451	142872,235	187888,3514	1,78787198
	147957,121	175841,2352	1,26131358	172902,743	199991,1184	0,70594594	146225,7	194778,0176	1,99729269	152930,341	193296,6031	1,79695843
	132112,021	165949,0305	1,27666941	175915,662	199493,6393	0,72949687	164841,522	204510,8996	1,99804435	158557,122	205211,6618	1,8001347
	135579,738	166401,1882	1,24582471	154233,746	188568,3257	0,87283794	160766,294	201803,5404	1,99711309	159963,345	198982,6856	1,84536739
	142232,452	173607,7525	1,24950651	161567,458	195110,0968	0,77675838	159625,046	199684,2194	1,99827717	159081,411	192732,5149	1,80971671
	133455,873	160639,9987	1,19706985	167168,706	197419,4959	0,75484713	165612,003	204047,9963	1,99769845	149705,879	191357,325	1,7992367
	129570,433	166683,6404	1,30431124	143102,341	173194,065	0,76191141	167157,064	208806,0398	1,99692351	151146,497	192096,3568	1,80804377
	139530,687	166544,3116	1,15045357	159549,102	189738,9191	0,79134925	166920,311	205678,9457	1,99525389	150369,112	196198,07	1,84136297
	131429,613	170335,9854	1,3043445	151086,121	187414,4723	0,7339769	154506,788	193999,9896	1,99729602	154490,422	196719,7371	1,79391853
	151823,154	180223,9507	1,22063076	161330,549	187076,8902	0,75733493	158086,135	200131,4025	1,99465855	148302,463	193749,3494	1,84449267
	126770,093	166464,3666	1,43660688	162539,838	192049,8638	0,74909659	148643,208	188151,8556	1,99467851	153087,44	198126,2881	1,7329442
	140559,247	168031,9963	1,16533712	146812,71	180951,3445	0,73846359	156267,81	199512,1769	1,99429602	159441,831	194438,4421	1,78429328
	142414,844	170980,3282	1,28549311	153276,631	193631,5166	0,7761331	169103,98	209509,3007	1,9963581	147590,488	194541,4823	1,84787847
Min	118824,125	157079,8168	1,10649472	140558,834	173194,065	0,67918215	146225,7	188151,8556	1,99429602	142872,235	187888,3514	1,68698978
Max	151823,154	183643,2153	1,45088844	175915,662	199991,1184	0,89450306	169103,98	210284,8843	1,9984102	166588,308	205211,6618	1,84787847
Média	132945,331	167088,0119	1,28156321	155954,009	189511,5023	0,78098886	158510,72	200170,9901	1,99659861	155049,229	196086,4178	1,78678815

ANEXO B

----- KS RESULTS -----

testResult (tour10):
0.38031026138301904

p-value (tour10):
0.00020714827908219213

testResult (tour50):
0.43905391017150597

p-value (tour50): 8.79996977132258e-06

testResult (roulette):
0.44838373192828873

p-value (roulette):
5.070190466138769e-06

testResult (sus): 0.33976505559653647

p-value (sus): 0.0013577877192305632

----- SHAPIRO RESULTS -----

testResult (tour10): 0.981200635433

p-value (tour10): 0.856505811214

testResult (tour50): 0.982221364975

p-value (tour50): 0.881036639214

testResult (roulette): 0.945886135101

p-value (roulette): 0.131064146757

testResult (sus): 0.941488981247

p-value (sus): 0.0997612401843

----- HOMOGENEITY RESULTS -----

testResult: 21.186661483285526

p-value: 5.157345446254471e-11

----- KRUSKAL RESULTS -----

testResult: 111.57024793388427

p-value: 5.0394632691027984e-24

----- MANN WHITNEY RESULTS -----

tour10 vs tour50

testResult: 0.0

p-value: 9.059578077486472e-11

tour10 vs roulette

testResult: 900.0

p-value: 9.059578077486472e-11

tour10 vs Sus

testResult: 900.0

p-value: 9.059578077486472e-11

tour50 vs roulette

testResult: 900.0

p-value: 9.059578077486472e-11

tour50 vs Sus

testResult: 900.0

p-value: 9.059578077486472e-11

Roulette vs Sus

testResult: 0.0

p-value: 9.059578077486472e-11