ELE093 - Modelos Estatísticos e Inferência
Trabalho Final

Um comparativo do desempenho de operadores em um algoritmo de otimização

Davi Pinheiro Viana
Douglas William Araujo da Silva
Letícia Diniz da Cruz
Paulo Cirino Ribeiro Neto
Priscila Pires de Carvalho Rocha

04 de Maio de 2016

Conteúdo

Capítulo 1

Introdução

Uma aplicação possível para a análise de dados obtidos através de um experimento estatístico é a avaliação da influência de determinado fator no resultado de determinado processo. O objetivo deste trabalho é, através de um experimento desse tipo, analisar a influência de um operador no desempenho de determinado algoritmo.

O algoritmo a ser testado é o método conhecido como evolução diferencial. Ele é do tipo otimização, baseado em população e consiste de um ciclo iterativo, no qual um conjunto de soluções-candidatas ao problema são repetidamente sujeitas a operadores de variação e seleção, de forma a promover uma exploração do espaço de variáveis do problema em busca de um ponto de ótimo (máximo ou mínimo) de uma dada função-objetivo.

Os operadores aos quais será analisada a influência no algoritmo são operadores de recombinação. Toda a análise deve ser feita em torno das seguintes perguntas:

> *Há alguma diferença no desempenho médio do algoritmo quando equipado com estes diferentes operadores, para o problema de teste utilizado? Caso haja, qual o melhor operador em termos de desempenho médio (quanto menor o valor retornado, melhor o algoritmo), e qual a magnitude das diferenças encontradas? Há algum operador que deva ser recomendado em relação aos demais?*

Espera-se, com a realização do trabalho, poder responder às questões que norteiam a análise dos dados e também, revisar e fixar os conceitos estatísticos vistos em sala praticando a utilização da ferramenta apresentada que é a linguagem R.

Capítulo 2

Planejamento do experimento

Para execução do experimento, utiliza-se o pacote ExpDE da linguagem R, que permite executar o algoritmo de evolução diferencial com diferentes operadores de recombinação e retorna uma lista contendo a população final (ordenadas pela performance) em que é possível obter a melhor solução. O experimento realizado é um experimento aleatorizado.

A aleatoriedade do experimento se dá pela ordem das chamadas do método de avaliação do desempenho do algoritmo com o operador corrente (ExpDE), que é aleatória. Ele é realizado desta maneira visando reduzir ao máximo o efeito da variável de ruído que, nesse caso, é o poder de processamento da máquina em que o algorítimo está sendo rodado.

O modelo dos dados do experimento é do tipo com efeitos fixos, pois os tratamentos (operadores) foram escolhidos previamente pelo professor. Assim, desejamos testar hipóteses acerca do desempenho médio do algoritmo com cada operador e as conclusões obtidas não podem ser estendidas para operadores similares que não tenham sido considerados.

O experimento possui um único fator (método de recombinação) e quatro níveis, ou tratamentos (operadores de recombinação), do fator. Os operadores analisados são:

- Recombinação aritmética;
- Recombinação binomial;
- Recombinação “Blend Alpha Beta” ;
- Recombinação baseada em um vetor de prioridades (Eigen).

Capítulo 3

Aquisição dos Dados

3.1 Cálculo do Tamanho Amostra

Inicialmente, foi proposto que, para o tamanho do efeito, existem dois níveis simétricos em relação a média geral, e outros dois iguais a zero, conforme é apresentado abaixo:

$$\tau = \{-\frac{\sigma^*}{2}, \frac{\sigma^*}{2}, 0, 0\} \tag{3.1}$$

Utilizando o *power.anova.test*, definiu-se o tamanho amostral para cada nível de 395 (resultado obtido = 394.6142). Ou seja, cada operador deve ser executado 395 vezes e, em seguida, serão efetuados testes e análises.

O tamanho amostral encontrado apresenta um valor alto devido à alta variância entre os grupos e também entre os tratamentos, definidos como parâmetros para o projeto. A variância entre os grupos foi definida como a variância do vetor τ, com $\delta^* = 2,5$ e a variância entre os tratamentos (σ^2), sendo $\sigma = 10$, para manter a proporção d=$\frac{\delta^*}{\sigma} = 0,25$ como especificado.

Durante o semestre foram realizados alguns estudos de caso e como sugestões para próximos experimentos, foi proposto o aumento do tamanho amostral. Esse aumento influencia a potência do teste realizado. Assim, optou-se por manter o tamanho amostral de 395 unidades.

3.2 Coleta dos Dados

A coleta de dados foi feita por meio dos resultados obtidos pela função *ExpDE* com seus devidos operadores.

Uma preocupação que tivemos com a função criada, foi a de aleatorizar a ordem das chamadas da função com cada operador, garantindo que para cada operador diferente a função sera chamada exatamente **n** vezes.

3.2.1 BoxPlot

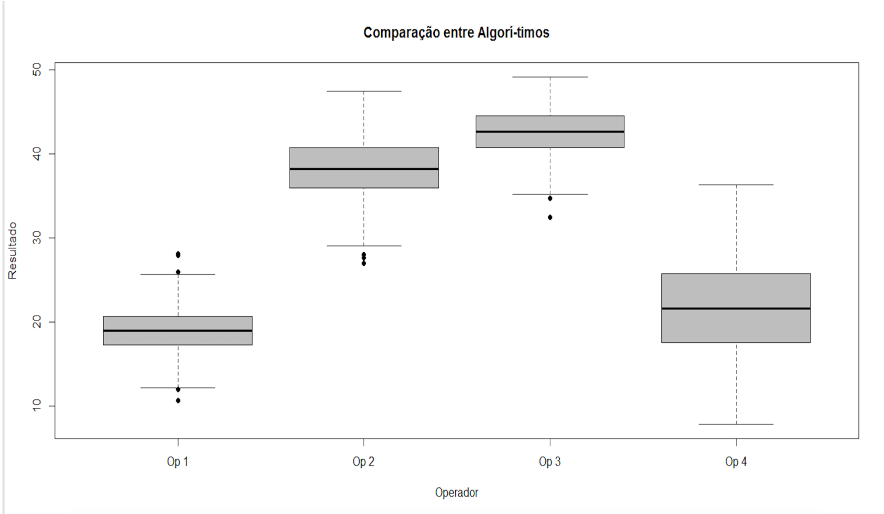


Figura 3.1: *BoxPlot* dos Dados

Após a coleta de dados foi realizado um diagrama de caixa com os dados obtidos. Este diagrama é capaz de apresentar a variabilidade das observações dentro de um nível e a variabilidade entre níveis.

A análise do diagrama permitiu identificar que a distribuição para cada nível aparenta ser aproximadamente simétrica e, portanto, a média e a mediana de cada nível devem ser valores próximos. Dessa forma, mesmo antes de realizar os testes, é possível inferir que, na média, existem operadores melhores que outros na análise realizada.

Com a análise do diagrama percebeu-se também a existência de possíveis valores atípicos (outliers) para os três primeiros níveis.

Capítulo 4

Formulação da Hipótese e Escolha dos Parâmetros

4.1 Formulação da Hipótese

Estamos interessados em testar a igualdade das médias dos 4 grupos, $\mu_1, \mu_2, \mu_3, \mu_4$. Isso é equivalente a testar as seguintes hipóteses:

$$H_0 : \tau_i = 0 \forall i \in \{1, 2, 3, 4\} \tag{4.1}$$

$$H_1 : \exists \tau_i \neq 0 \tag{4.2}$$

Onde τ_i é o efeito do i-ésimo nível.

4.2 Parâmetros

Para este estudo foram utilizados os seguintes parâmetros, dados na proposta do trabalho.

- Mínima diferença de importância prática entre qualquer par de algoritmos (padronizada):(d = d^*) = 0,25;
- Nível de confiança do teste de 95% (α=0,05)
- Potência mínima desejada (para o caso d = d^*): 85% (β = 0,15).

Capítulo 5

Teste das Hipóteses

O método escolhido para testar as hipóteses deste experimento estatístico foi o **ANOVA**.

Este método de comparações múltiplas é uma coleção de modelos estatísticos usada para inferir sobre as diferenças entre as médias dos grupos, utilizando uma análise das variâncias. É conceitualmente similar ao **t-teste** de duas amostras, entretanto é menos conservativo e pode ser utilizado em uma gama maior de problemas.

Se usássemos o teste t para esse mesmo experimento teríamos que fazer seis testes (combinação de quatro grupos, dois a dois), e além disso o nível de confiança final seria 0,735 ($0,954^4$). Utilizando o **ANOVA** realizamos apenas um teste mantemos nosso nível de confiança em 0,95.

5.1 Premissas do Teste

O modelo ANOVA é baseado em três suposições sobre o comportamento dos resíduos e para confirmar essas premissas faram realizados alguns testes no software R:

5.1.1 Independência entre as amostras (Teste Durbin-Watson)

O Teste de Durbin-Watson é usado para testar a presença de autocorrelação (ou dependência) nos resíduos de uma análise de regressão. A autocorrelação significa que observações adjacentes são correlacionadas.

Quando essas observações são correlacionadas, a regressão de mínimos quadrados subestima o erro padrão dos coeficientes e os preditores podem parecer significantes quando na verdade não o são.

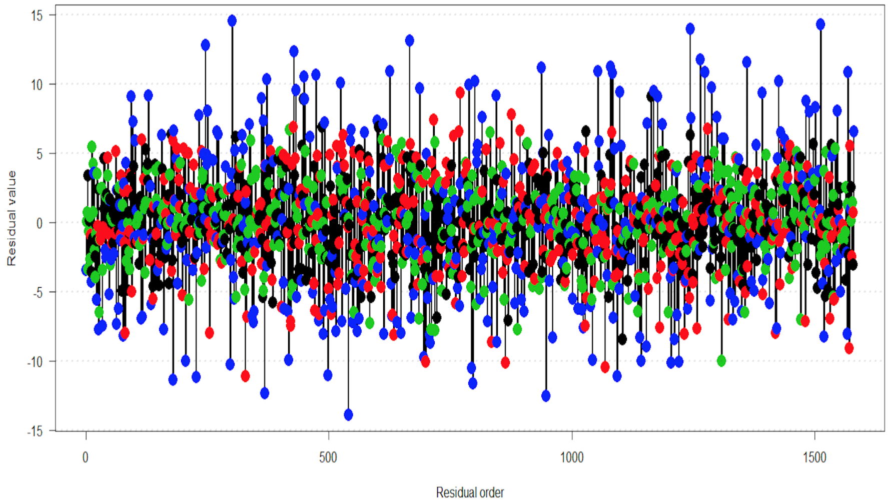


Figura 5.1: *Durbin-Watson Plot*

5.1.2 Homogeneidade na variância dos grupos (Teste Fligner-Killeen)

O Teste Fligner-Killeen é um teste não paramétrico para homogeneidade de um grupo de variâncias baseado na classificação. É um teste útil quando os dados são não normais ou quando há valores atípicos (outliers).

O resultado encontrado para os resíduos das amostras coletadas é apresentado abaixo, onde cada nível corresponde à uma coluna de pontos.

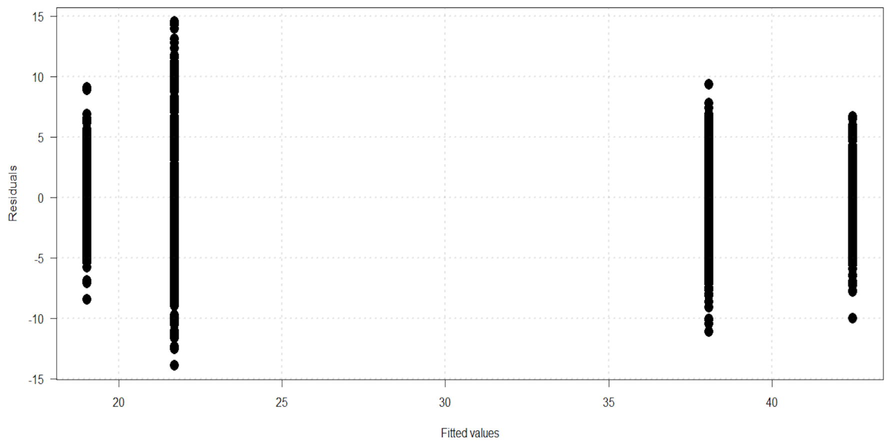


Figura 5.2: *Fligner-Killeen Plot*

5.1.3 Normalidade (Teste de Shapiro Wilk)

A hipótese nula desse teste é que a população tem distribuição normal. Se o valor de p é menor do que o valor escolhido de alfa, então a hipótese nula é rejeitada e há evidência de que os dados testados não são de uma população com distribuição normal. Caso o valor de p seja maior do que o alfa escolhido, então a hipótese nula de que os dados vieram de uma distribuição normal não pode ser rejeitada.

O teste de Shapiro Wilk aplicado para os resíduos nessa análise é apresentado graficamente na figura abaixo.

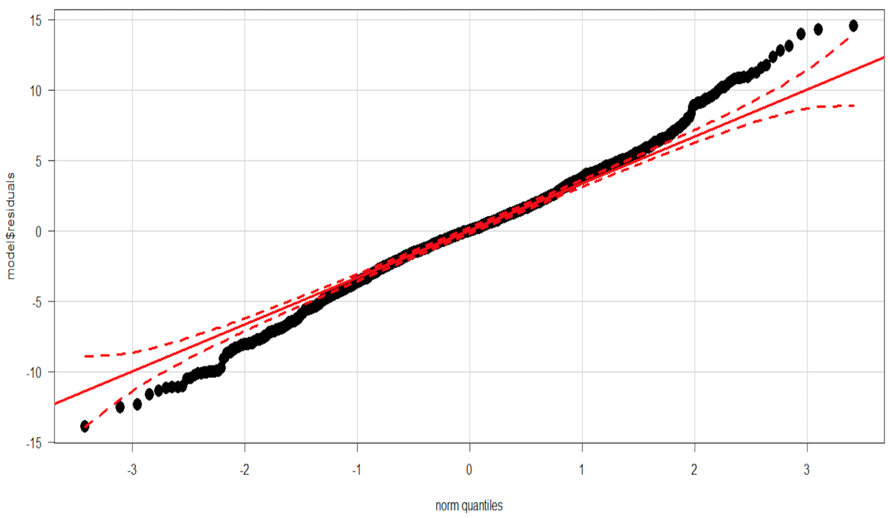


Figura 5.3: *Shapiro Wilk Plot*

É possível perceber que nas extremidades do gráfico existe violação da normalidade. Todavia, o teste ANOVA é relativamente robusto para um tamanho amostral grande o suficiente, como é o caso neste trabalho.

5.2 Resultado Teste de Hipóteses

Usando a função *aov*, do pacote *stat*, que realiza a análise de variância, obtivemos os seguintes resultados:

	Graus de Liberdade	Soma Quadrática	Média Quadrática	$F_critico$	Valor P
Operador	3	161918	53973	3506	$< 2e-16$
Residuais	1576	24260	15		

A partir dos dados de saída do ANOVA podemos calcular o f_0 pela seguinte equação :

$$F_0 = \frac{MS_Level}{MS_E} = \frac{53973}{15} \cong 3598 \tag{5.1}$$

A hipótese nula é rejeitada em relação à alternativa para nível de confiança de 95%, uma vez que o F_0 encontrado é maior do que o $F_critico$. O baixo Valor P também sugere a rejeição da hipótese nula.

Capítulo 6

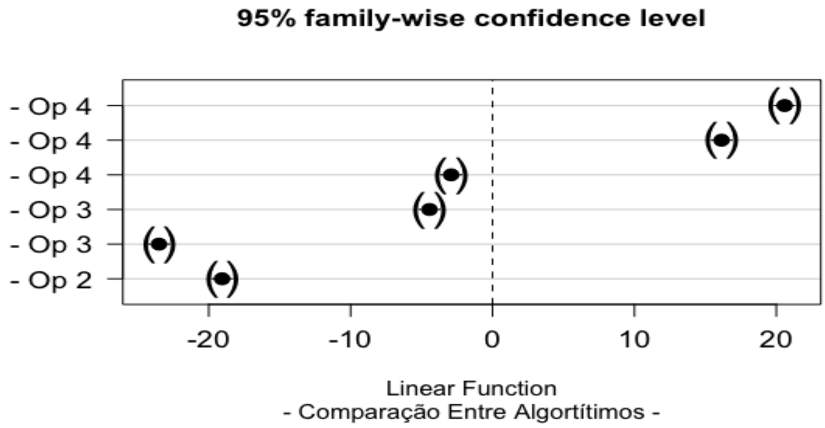
Comparações Múltiplas

Quando a hipótese nula é rejeitada na **ANOVA**, sabemos que pelo menos uma das médias dos grupos é diferente, no entanto, a **ANOVA** não identifica qual/quais média(s) é/são diferente(s).

Para investigar esse diferença utilizamos os métodos de comparações múltiplas, neste experimento usamos o teste de Tukey, usando a função *glht* do pacote *multcomp* do software R.

Os resultados podem ser vistos abaixo pela tabela e gráfico:

	Estimado	Mínimo	Máximo
Op 3 - Op 4	20.58	19.86	21.31
Op 2 - Op 4	16.14	15.41	16.87
Op 1 - Op 4	-2.92	-3.65	-2.19
Op 2 - Op 3	-4.44	-5.17	-3.72
Op 1 - Op 3	-23.50	-24.23	-22.77
Op 1 - Op 2	-19.06	-19.79	-18.33



Os resultados nos mostram que todas as médias são diferentes entre si, pois nenhum dos intervalos de confiança das comparações múltiplas inclui o zero. Entretanto isso ainda não nos mostra qual o melhor operador em termos de desempenho médio. Um procedimento simples, porem efetivo, é comparar as médias amostrais que vão indicar qual desempenho é melhor, sabendo que todas diferem entre si.

Abaixo podemos ver no gráfico as médias amostrais do desempenho dos quatro operadores:

Figura 6.1: Media dos Resultados

O gráfico mostra que o Operador 1, na média, possui um desempenho melhor que os outros.

Capítulo 7

Conclusões

Diante dos resultados encontrados, podemos afirmar que, dentro de um nível de confiança de 95% a hipótese nula do teste inicial foi refutada. Isso indica que o tamanho de efeito obtido para cada nível, pelo menos para um nível, é maior do que o intervalo de confiança centralizado na média geral das amostras

O valor da potência de teste de 85% era esperado, uma vez que o mesmo teste foi realizado de duas maneiras diferentes: A primeira, considerando a potência desejada para encontrar o tamanho amostral (N) necessário, e a segunda, utilizando o N obtido, para validar que a potência desejada é obtida.

Além disso, baseando-se nas questões que nortearam todos os testes, podemos construir as seguintes conclusões:

- Baseado na refutação da hipótese nula, os operadores de recombinação influenciam no desempenho médio do algoritmo;

- Baseado nos testes das médias, o melhor operador é o Operador 1: Recombinação aritmética.

7.1 Discussão sobre possíveis limitações do estudo e sugestões de melhorias

Uma possível sugestão de melhoria, seria executar o experimento em uma máquina específica para esse fim, sem a interferência de aplicações externas que gerassem um ruído no poder de processamento durante a execução.