

Projeto de Inovação com Algoritmos Genéticos: discussões sobre as aulas

Márcio da Silva Arantes

marcio@icmc.usp.br

Conteúdo

1. Dedução do tamanho da população	3
1.1. Prova padrão mais detalhada (adaptado de Arturo)	3
1.2. Impacto dos Erros nas aproximações.....	5
1.3. Resolver o problema diretamente ou evoluir uma população de soluções?	7
2. Dedução do numero de gerações	9
2.1. Problema UmMax	9
3. Questionando Aplicabilidade do Teorema Fundamental de Fischer para UmMax	14
3.1. Analisando Tamanho da População	14
3.2. Analisando Tamanho do Problema	15
3.3. Analisando Taxa de Truncamento.....	16
4. Nova dedução do numero de gerações	18
4.1. Problema UmMax	18
4.2. Analisando se a equação encontrada condiz com dados experimentais.....	20
4.2.1. Tamanho da População	20
4.2.2. Tamanho do Problema	20
4.2.3. Taxa de Truncamento.....	21
5. Erros das Hipóteses	21
5.1. Erro com o aumento do tamanho da população	22
5.2. Erro através das gerações	22
5.3. Erro com o aumento do alfa no truncamento	23
6. Problema BinInt.....	23
6.1. Analisando se equação encontrada condiz com dados experimentais.	24
6.1.1. Tamanho da População	24
6.1.2. Tamanho do Problema	25
6.1.3. Taxa de Truncamento.....	25
7. Problema UmMax com Seleção por Torneio	26
7.1. Sensibilidade do Torneio	26

7.2.	Sensibilidade do Tamanho do Problema.....	27
7.3.	Sensibilidade do Tamanho da População.....	27
8.	Problema UmMax com Seleção Proporcional.....	27
8.1.	Sensibilidade do Tamanho do Problema.....	28
8.2.	Sensibilidade do Tamanho da População.....	29
9.	Dedução BinInt para Seleção Proporcional.....	29
9.1.	Sensibilidade do Tamanho do Problema.....	31
9.2.	Sensibilidade do Tamanho da População.....	33
10.	Formula Geral para Encontrar o Fitness Médio e Variância – GA simples.....	33
10.1.	Provando para UmMax	34
10.2.	Provando para Trap.....	37
11.	Prova da Composição das Variâncias	40
12.	Formula Geral para Encontrar o Fitness Médio – ECGA simples	42
12.1.	Problema Trap.....	43
12.2.	Dedução número de gerações para Trap utilizando seleção por truncamento.	44
13.	Analisando a sensibilidade nos parâmetros.....	46
13.1.	Problema UmMax com seleção truncamento	46
13.2.	Problema Trap com seleção truncamento.....	53
13.3.	Problema BinInt com seleção proporcional	57
13.4.	Resolver o problema diretamente ou evoluir uma população de soluções?	58

1. Dedução do tamanho da população

1.1. Prova padrão mais detalhada (adaptado de Arturo)

Apresentamos o cálculo do tamanho da população necessária para que cada instância de um BB seja representada pelo menos uma vez na população de tamanho n . Como exemplo, considera-se uma solução de 4 *bits* (x_1, x_2, x_3, x_4) em que os dois primeiros *bits* constituem um BB. Se os valores das duas primeiras variáveis constituem um BB então uma partição desse problema é representada como $x_1 x_2 **$. As instancias dessa partição ou esquemas são: 00 **, 01 **, 10 **, 11 **. É preciso então determinar o tamanho n da população inicial de forma que esta possua pelo menos uma de cada possível instancia dessa partição. Isto é, a população deve conter os esquemas acima mencionados.

Seja χ a cardinalidade do alfabeto, portanto, $\chi = \{0,1\}$.

O * representa um valor qualquer para o *bit*.

O k é o tamanho do BB, $k = 2$.

O n é o tamanho da população.

A probabilidade de ocorrência ou não ocorrência de uma instancia da partição é representada na seguinte tabela:

Instancias da partição $x_1 x_2 **$ (esquemas)	A probabilidade de que uma instancia ocorra uma vez	A probabilidade de que uma instancia não ocorra uma vez
00**	$\frac{1}{\chi^k}$	$1 - \frac{1}{\chi^k}$
01**	$\frac{1}{\chi^k}$	$1 - \frac{1}{\chi^k}$
10**	$\frac{1}{\chi^k}$	$1 - \frac{1}{\chi^k}$
11**	$\frac{1}{\chi^k}$	$1 - \frac{1}{\chi^k}$

O tamanho de espaço de busca n_k do BB da partição é χ^k (ou também o número de possíveis instancias do BB), portanto, tem-se, $n_k = \chi^k = 2^2 = 4$.

A probabilidade de que não ocorra nenhuma instancia da partição na população de tamanho n é

$$\left[1 - \frac{1}{\chi^k}\right]^n \quad (1.1)$$

A probabilidade p_k de que uma instancia dessa partição ocorra uma ou mais vezes, é dada pela seguinte equação.

$$p_k = 1 - \left[1 - \frac{1}{\chi^k}\right]^n \quad (1.2)$$

$$p_k = 1 - \left[1 - \frac{n}{n \cdot \chi^k}\right]^n \quad (1.3)$$

Substituindo $\frac{n}{\chi^k}$ por r , tem-se:

$$p_k = 1 - \left[1 - \frac{r}{n}\right]^n \quad (1.4)$$

Utilizando a aproximação $\left(1 - \frac{p}{q}\right)^q \approx e^{-p}$, valida para $p \ll q$, tem-se

$$p_k = 1 - e^{-r} \quad (1.5)$$

Verificando $p \ll q$ é verdade quando temos blocos maiores.

$$r \ll n$$

$$\frac{n}{\chi^k} \ll n$$

$$1 \ll \chi^k$$

Como $r = \frac{n}{\chi^k}$ a Equação 1.5 é rescrita como:

$$p_k = 1 - e^{-\left(\frac{n}{\chi^k}\right)} \quad (1.6)$$

Sabe-se que $n_k = \chi^k$ e substituindo na Equação 1.6:

$$p_k = 1 - e^{-\left(\frac{n}{n_k}\right)} \quad (1.7)$$

Ao considerar uma inicialização aleatória das soluções em que as ocorrências de esquemas, representando BBs de mesmo tamanho, possuam a mesma taxa de sucesso, a probabilidade de pelo menos uma ocorrência de cada um dos possíveis esquemas na população inicial é dada por $p_s = p_k^{n_k}$, substituindo p_k pela Equação 1.7.

$$p_s = \left(1 - e^{-\left(\frac{n}{n_k}\right)}\right)^{n_k} \quad (1.8)$$

$$p_s = \left(1 - \frac{n_k \cdot e^{-\left(\frac{n}{n_k}\right)}}{n_k}\right)^{n_k} \quad (1.9)$$

Seja $t = n_k e^{-\frac{n}{n_k}}$, tem-se:

$$p_s = \left(1 - \frac{t}{n_k}\right)^{n_k} \quad (1.10)$$

Utilizando novamente a aproximação $\left(1 - \frac{p}{q}\right)^q \approx e^{-p}$, obtêm-se:

$$p_s = e^{-t} = e^{-n_k e^{-\frac{n}{n_k}}} \quad (1.10)$$

Verificando $p \ll q$ é verdade quando temos uma população grande.

$$t \ll n_k$$

$$n_k e^{-\frac{n}{n_k}} \ll n_k$$

$$e^{-\frac{n}{n_k}} \ll 1$$

$$\ln e^{-\frac{n}{n_k}} \ll 0$$

$$-\frac{n}{\chi^k} \cdot \ln e \ll 0$$

$$-n \ll 0$$

Para estimar então o tamanho de uma população que amostrasse adequadamente todas as instâncias de um BB com uma dada probabilidade p_s , é necessário isolar n na Equação (1.11).

Seja α a chance de que um BB não tenha todas as suas instâncias representadas. Então, $p_s = 1 - \alpha$ e, portanto,

$$1 - \alpha = e^{-n_k e^{-\frac{n}{n_k}}} \quad (1.12)$$

Aplicando logaritmo na expressão anterior

$$\ln(1 - \alpha) = -n_k e^{-\frac{n}{n_k}} \quad (1.13)$$

Sendo $\ln(1 - x) \approx -x$, para pequenos valores de x ,

$$-\alpha = -n_k e^{-\frac{n}{n_k}}$$

$$\frac{\alpha}{n_k} = e^{-\frac{n}{n_k}}$$

Aplicando logaritmo na expressão acima

$$\begin{aligned}\ln\left(\frac{\alpha}{n_k}\right) &= -\frac{n}{n_k} \\ \ln(\alpha) - \ln(n_k) &= -\frac{n}{n_k} \\ n &= n_k \cdot [\ln(n_k) - \ln(\alpha)]\end{aligned}$$

Sabe-se $n_k = \chi^k$ e substituindo na equação acima,

$$n = \chi^k \cdot [k \cdot \ln(\chi) - \ln(\alpha)] \quad (1.14)$$

Admitindo como tolerável o erro de único BB, isto é, no máximo um BB não estar representado pelas soluções da população inicial, adota-se $\alpha = \frac{1}{m}$. Reescrevendo a Equação acima. Obtém-se:

$$n = \chi^k \cdot [k \cdot \ln(\chi) + \ln(m)] \quad (1.15)$$

1.2. Impacto dos Erros nas aproximações

A aproximação abaixo utilizada:

$$\left(1 - \frac{r}{n}\right)^n \approx e^{-r}$$

Vem da seguinte formula.

$$\lim_{n \rightarrow \infty} \left(1 - \frac{r}{n}\right)^n = e^{-r}$$

Seja então a aproximação:

$$p_k = 1 - \left(1 - \frac{r}{n}\right)^n \approx 1 - e^{-r}$$

Logo podemos definir o erro, como o valor utilizado menos o valor real:

$$\varepsilon = [1 - e^{-r}] - \left[1 - \left(1 - \frac{r}{n}\right)^n\right]$$

$$\varepsilon = \left(1 - \frac{r}{n}\right)^n - e^{-r}$$

$$\varepsilon = \left(1 - \frac{r}{n}\right)^n - \lim_{x \rightarrow \infty} \left(1 - \frac{r}{x}\right)^x$$

Seja a função $f(x) = \left(1 - \frac{r}{x}\right)^x$ definida para valores de x maior que 1 esta função é monótona crescente e seu valor máximo é (e^{-r}) quando x tende a infinito. Abaixo foi reescrito o erro utilizando a função definida (veja que o erro é uma função de n).

$$\varepsilon = f(n) - f(\infty)$$

Como a função é monótona crescente e o argumento $n < \infty$ logo $f(n) < f(\infty)$ gerando um erro negativo. Logo:

$$\varepsilon < 0$$

Como o erro utilizado é negativo, o valor de probabilidade utilizado é menor que a probabilidade p_k que existe de fato, então estamos subestimando a probabilidade o que nos

levará a exigir uma população maior para garantir que exista pelo menos uma instância de cada BB na população gerada.

Daqui para frente será desenvolvida toda a prova anterior, mas considerando todos os erros de arredondamento.

$$r = \frac{n}{n_k}$$

$$n_k = \chi^k$$

$$p_k = 1 - e^{-r} - \varepsilon$$

$$\varepsilon = \left(1 - \frac{r}{n}\right)^n - e^{-r} < 0$$

$$p_s = (p_k)^{n_k} = \left(1 - e^{-\frac{n}{n_k}} - \varepsilon\right)^{n_k}$$

$$p_s = \left(1 - \frac{t}{n_k}\right)^{n_k}$$

$$t = n_k \left(e^{-\frac{n}{n_k}} + \varepsilon\right)$$

$$\delta = \left(1 - \frac{t}{n}\right)^n - e^{-t} < 0$$

$$p_s = e^{-t} + \delta = e^{-n_k \left(e^{-\frac{n}{n_k}} + \varepsilon\right)} + \delta = e^{-n_k \left(e^{-\frac{n}{n_k}} + \varepsilon\right)} + \delta$$

$$1 - \alpha = e^{-n_k \left(e^{-\frac{n}{n_k}} + \varepsilon\right)} + \delta$$

$$\ln(1 - \alpha - \delta) = -n_k \left(e^{-\frac{n}{n_k}} + \varepsilon\right)$$

Fazendo $\gamma = \ln(1 - (\alpha + \delta)) + (\alpha + \delta) < 0$

$$\gamma + \varepsilon \cdot n_k - (\alpha + \delta) = -n_k \left(e^{-\frac{n}{n_k}}\right)$$

$$\frac{(\alpha + \delta) - (\gamma + \varepsilon \cdot n_k)}{n_k} = e^{-\frac{n}{n_k}}$$

$$\ln[(\alpha + \delta) - (\gamma + \varepsilon \cdot n_k)] - \ln[n_k] = -\frac{n}{n_k}$$

$$\ln[n_k] - \ln[(\alpha + \delta) - (\gamma + \varepsilon \cdot n_k)] = \frac{n}{n_k}$$

$$n = \chi^k \cdot (k \cdot \ln[\chi] - \ln[(\alpha + \delta) - (\gamma + \varepsilon \cdot \chi^k)])$$

$$\alpha = \frac{1}{m}$$

$$n = \chi^k \cdot \left(k \cdot \ln[\chi] - \ln\left[\frac{(1 + m \cdot \delta) - (m \cdot \gamma + m \cdot \varepsilon \cdot \chi^k)}{m}\right]\right)$$

$$n = \chi^k \cdot (k \cdot \ln[\chi] + \ln[m]) - \chi^k \cdot (\ln[(1 + m \cdot \delta) - (m \cdot \gamma + m \cdot \varepsilon \cdot \chi^k)])$$

$$erro\ total = -\chi^k \cdot (\ln[(1 + m \cdot \delta) - (m \cdot \gamma + m \cdot \varepsilon \cdot \chi^k)])$$

1.3. Resolver o problema diretamente ou evoluir uma população de soluções?

Seja a equação abaixo que estima o tamanho da população que é necessário para que cada instância de um BB seja representada pelo menos uma vez na população.

$$n = \chi^k \cdot (k \cdot \ln \chi + \ln m)$$

O tamanho do problema é dado por.

$$l = k \cdot m$$

Onde:

- n Tamanho da população
- l Tamanho do problema
- k Tamanho do BB
- l Número de BB's
- χ Cardinalidade do alfabeto

Aqui será feita uma análise simples dos custos teóricos para a solução de um problema qualquer, supondo que este problema seja separável, composto de m BB de tamanho k . Aqui já estamos supondo que sabemos os BB's do problema dado por alguma técnica.

O custo de resolver o problema diretamente, por um algoritmo exaustivo (A) dado que temos os BB's será:

$$A = m \cdot \chi^k$$

Que representa a obter o ótimo exaustivamente de todos os BB's e combinar os m BB's em uma solução ótima global.

O custo de utilizar um algoritmo evolutivo (B) que trabalha com a população n dada acima será:

$$B = g \cdot \chi^k \cdot (k \cdot \ln \chi + \ln m)$$

Onde g é o numero de gerações necessárias para o método convergir.

A pergunta aqui é qual método é melhor? Para isto vamos avaliar o crescimento de cada um dos termos em relação aos demais nestes dois algoritmos. Aqui também não será analisada quantas gerações g o método B levará até convergir, deixaremos isto em aberto.

Na demonstração abaixo temos uma desigualdade inicial onde estamos procurando valores para satisfazer a situação onde utilizar uma abordagem direta é mais vantagem.

Partindo de:

$$m \cdot \chi^k < g \cdot \chi^k \cdot (k \cdot \ln \chi + \ln m)$$

Cortando o termo exponencial nos dois lados da equação:

$$m < g \cdot (k \cdot \ln \chi + \ln m)$$

Isolando o termo k :

$$m < g \cdot (k \cdot \ln \chi + \ln m)$$

$$m < g \cdot k \cdot \ln \chi + g \cdot \ln m$$

$$m - g \cdot \ln m < g \cdot k \cdot \ln \chi$$

$$\frac{m - g \cdot \ln m}{g \cdot \ln \chi} < k$$

1º Vemos que:

$$\lim_{g \rightarrow \infty} \left(\frac{m - g \cdot \ln m}{g \cdot \ln \chi} \right) = \lim_{g \rightarrow \infty} \left(\frac{-g \cdot \ln m}{g \cdot \ln \chi} \right) = \lim_{g \rightarrow \infty} \left(\frac{-\ln m}{\ln \chi} \right) = \frac{-\ln m}{\ln \chi} \leq 0 < k$$

Logo se o numero de gerações for alto o método direto (A) é melhor sempre independentemente dos demais parâmetros χ e m ; bem isto é obvio.

2º Vemos que:

$$\lim_{\chi \rightarrow \infty} \left(\frac{m - g \cdot \ln m}{g \cdot \ln \chi} \right) = 0 < k$$

Logo se o alfabeto também for muito grande o método direto (A) é melhor sempre independentemente dos demais parâmetros k e m .

Mas 3º vemos que:

$$\lim_{m \rightarrow \infty} \left(\frac{m - g \cdot \ln m}{g \cdot \ln \chi} \right) = \lim_{m \rightarrow \infty} \left(\frac{m}{g \cdot \ln \chi} \right) = \infty < k$$

Neta situação quando o número de BB's cresce o método direto (A) só será melhor quando este crescer muito também. Ou seja, o método evolutivo começa a valer mais a pena quando temos um problema que pode ser dividido em vários subproblemas pequenos.

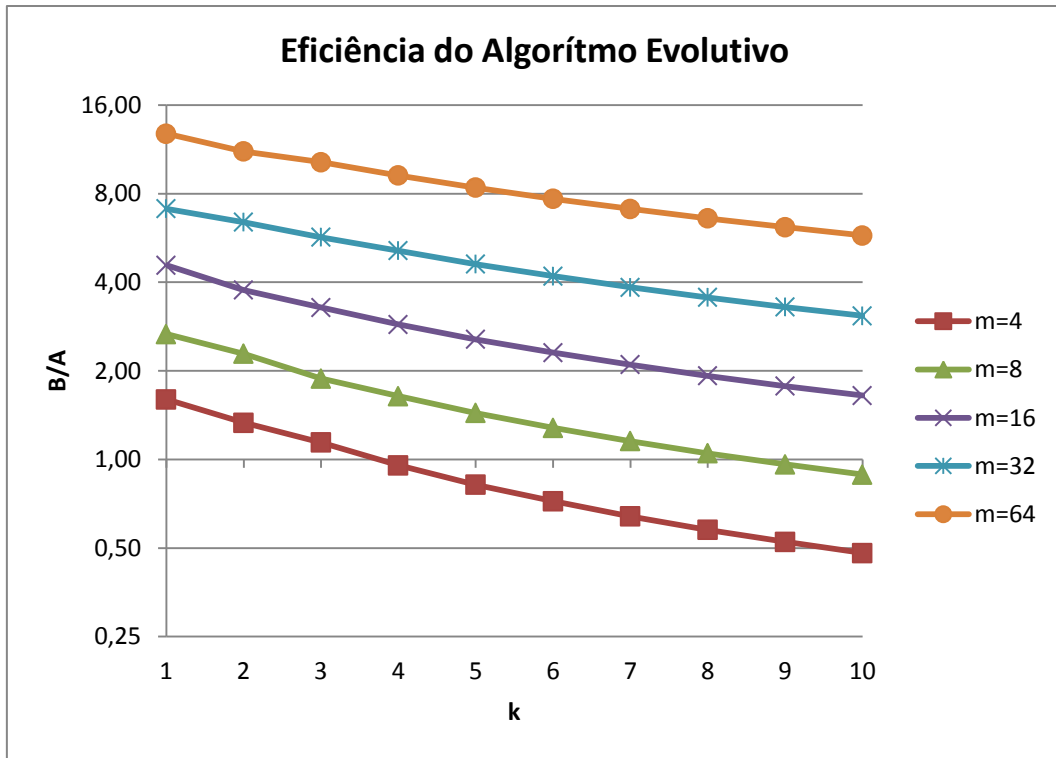
Onde podemos considerar também um limite para o numero de gerações como abaixo, ao isolarmos g :

$$\frac{m}{(k \cdot \ln \chi + \ln m)} < g$$

O numero de gerações g tem que ser maior do que a equação acima para o método direto (A) ser melhor que um algoritmo evolutivo. Ou seja, isto impõe um limite de gerações para o algoritmo evolutivo do qual vale mais apenas resolver o problema diretamente. Como se pode perceber não é um valor grande. Como exemplo suponha resolver: 8 x trap5

$$\left(\frac{8}{5 \cdot \ln 2 + \ln 8} \right) = 1.44 \text{ gerações}$$

Gráfico mostrando como se comporta a eficiência do Algoritmo Evolutivo (B) comparado com o método Direto (A), quando se varia os valores de k . No gráfico temos uma linha para cada valor de m . Veja que à medida que k aumenta a eficiência cai (quando chega abaixo do valor 1 o método B perde para o método A). Vemos também que para maiores valores de m a eficiência é maior, mas sempre decrescente com o aumento de k .



2. Dedução do numero de gerações

Teorema fundamental de Fisher da seleção natural

$$\overline{f_{t+1}} - \overline{f_t} = \frac{\sigma_t^2}{\overline{f_t}}$$

Onde:

σ_t^2 Variância do fitness da população na geração t .

$\overline{f_t}$ Média do fitness da população na geração t .

Utilizando seleção por truncamento teremos um fator I próximo a uma constante.

$$I = \frac{\sigma_t}{\overline{f_t}}$$

$$\overline{f_{t+1}} - \overline{f_t} = I \cdot \sigma_t$$

2.1. Problema UmMax

$$f(x) = \sum_{i=1}^l x_i$$

p_t Probabilidade de cada bit da string l valer 1.

l Tamanho do problema.

$$\overline{f_t} = l \cdot p_t$$

$$\sigma_t^2 = l \cdot p_t \cdot (1 - p_t)$$

Verificando hipótese de I ser constante:

Substituindo σ_t e \bar{f}_t na equação da constante

$$I_t = \frac{\sigma_t}{\bar{f}_t} = \frac{\sqrt{l \cdot p_t \cdot (1 - p_t)}}{l \cdot p_t} = \sqrt{\frac{(1 - p_t)}{l \cdot p_t}}$$

Vemos que a “constante” pode variar ao longo das gerações de acordo com p_t , para o problema UmMax é natural acreditar que:

$$p_1 = \frac{1}{2}$$

$$p_g = 1 - \frac{1}{n}$$

Na primeira geração inicia-se com igual probabilidade de ser 0 ou 1, e no final quando o algoritmo converge é natural acreditar que todos os indivíduos atingiram o ótimo e tem probabilidade quase certa de ser 1. Sendo assim podemos ver que a “constante” na verdade varia de:

$$I_1 = \sqrt{\frac{(1 - 1/2)}{l \cdot 1/2}} = \sqrt{\frac{1/2}{l \cdot 1/2}} = \frac{1}{\sqrt{l}}$$

$$I_g = \sqrt{\frac{(\frac{1}{n})}{l \cdot (1 - \frac{1}{n})}} = \frac{1}{\sqrt{l \cdot (n - 1)}} \approx 0$$

Prova considerando o I constante:

$$\bar{f}_{t+1} - \bar{f}_t = I \cdot \sigma_t$$

$$l \cdot p_{t+1} - l \cdot p_t = I \cdot \sqrt{l \cdot p_t \cdot (1 - p_t)}$$

$$p_{t+1} - p_t = \frac{I}{\sqrt{l}} \cdot \sqrt{p_t \cdot (1 - p_t)}$$

Aproximando por uma equação diferencial, teremos:

$$\frac{\partial p_t}{\partial t} = \frac{I}{\sqrt{l}} \cdot \sqrt{p_t \cdot (1 - p_t)}$$

$$\frac{\partial p_t}{\sqrt{p_t \cdot (1 - p_t)}} = \frac{I \cdot \partial t}{\sqrt{l}}$$

$$\int \frac{\partial p_t}{\sqrt{p_t \cdot (1 - p_t)}} = \frac{I}{\sqrt{l}} \cdot \int \partial t$$

$$\int_{p_1}^{p_g} \frac{\partial p_t}{\sqrt{p_t \cdot (1 - p_t)}} = \frac{I}{\sqrt{l}} \cdot \int_1^g \partial t$$

Temos que

$$\int \frac{dx}{\sqrt{ax^2 + bx + c}} = -\frac{1}{\sqrt{-a}} \cdot \arcsin\left(\frac{2ax + b}{\sqrt{b^2 - 4ac}}\right) \quad \text{para } a < 0 \text{ e } 4ac - b^2 < 0$$

Logo

$$\left[-\arcsin\left(\frac{-2x + 1}{\sqrt{1}}\right) \right]_{p_1}^{p_g} = \frac{I}{\sqrt{l}} \cdot [t]_1^g$$

$$\arcsin(1 - 2 \cdot p_1) - \arcsin(1 - 2 \cdot p_g) = \frac{I}{\sqrt{l}} \cdot (g - 1)$$

Utilizando

$$p_1 = \frac{1}{2}$$

$$p_g = 1 - \frac{1}{n}$$

$$\frac{I}{\sqrt{l}} \cdot (g - 1) = \arcsin(1 - 2 \cdot p_1) - \arcsin(1 - 2 \cdot p_g)$$

$$\frac{I}{\sqrt{l}} \cdot (g - 1) = \arcsin(0) - \arcsin\left(\frac{2}{n} - 1\right)$$

$$\arcsin(x) = \sum_{n=0}^{\infty} \frac{(2n)!}{4^n (n!)^2 (2n+1)} \cdot x^{2n+1} \approx x + \frac{x^3}{6} + \frac{3x^5}{40}$$

$$\arcsin(-x) = -\arcsin(x)$$

É natural acreditar também que

$$\left(1 - \frac{2}{n}\right) \leq \arcsin\left(1 - \frac{2}{n}\right) \leq \frac{\pi}{2}$$

$$\left(\frac{2}{n} - 1\right) \geq \arcsin\left(\frac{2}{n} - 1\right) \geq \frac{-\pi}{2}$$

Considerando limite inferior (apenas o primeiro termo):

$$\frac{I}{\sqrt{l}} \cdot (g - 1) = 0 - \left(\frac{2}{n} - 1\right) = 1 - \frac{2}{n}$$

$$(g - 1) = \frac{\sqrt{l}}{I} \cdot \left(1 - \frac{2}{n}\right)$$

Substituindo as constantes

$$I_1 = \frac{1}{\sqrt{l}} \quad , \quad I_g = \frac{1}{\sqrt{l \cdot (n-1)}}$$

$$l \cdot \left(1 - \frac{2}{n}\right) \cdot \sqrt{(n-1)} \geq (g - 1) \geq l \cdot \left(1 - \frac{2}{n}\right)$$

Considerando limite superior (população muito grande):

$$\frac{I}{\sqrt{l}} \cdot (g - 1) = 0 - \left(\frac{-\pi}{2} \right) = \frac{\pi}{2}$$

$$(g - 1) = \frac{\sqrt{l}}{I} \cdot \frac{\pi}{2}$$

$$\frac{l \cdot \pi}{2} \cdot \sqrt{(n - 1)} \geq (g - 1) \geq \frac{l \cdot \pi}{2}$$

Consideramos então de todas estas estimativas a maior delas onde temos o pior caso do numero de gerações.

$$g = 1 + \frac{l \cdot \pi}{2} \cdot \sqrt{(n - 1)}$$

$$g = 1 + l \cdot \left(1 - \frac{2}{n} \right)$$

Prova considerando o I variável:

$$\overline{f_{t+1}} - \overline{f_t} = \frac{\sigma_t^2}{\overline{f_t}}$$

$$l \cdot p_{t+1} - l \cdot p_t = \frac{l \cdot p_t \cdot (1 - p_t)}{l \cdot p_t}$$

$$p_{t+1} - p_t = \frac{(1 - p_t)}{l}$$

Aproximando por uma equação diferencial, teremos:

$$\frac{\partial p_t}{\partial t} = \frac{(1 - p_t)}{l}$$

Veja que para o problema UmMax é simples de resolver esta equação diferencial:

$$\frac{\partial p_t}{(1 - p_t)} = \frac{\partial t}{l}$$

$$\int \frac{\partial p_t}{(1 - p_t)} = \int \frac{\partial t}{l}$$

$$\int_{p_1}^{p_g} \frac{\partial p_t}{(1 - p_t)} = \int_1^g \frac{\partial t}{l}$$

$$[-\ln|1 - p_t|]_{p_1}^{p_g} = \frac{g - 1}{l}$$

$$l \cdot \ln|1 - p_1| - l \cdot \ln|1 - p_g| + 1 = g$$

$$p_1 = \frac{1}{2}$$

$$p_g = 1 - \frac{1}{n}$$

$$g = 1 + l \cdot \ln|1 - p_1| - l \cdot \ln|1 - p_g|$$

$$g = 1 + l \cdot \ln\left|\frac{1}{2}\right| - l \cdot \ln\left|\frac{1}{n}\right|$$

$$g = 1 + l \cdot \ln\left|\frac{n}{2}\right|$$

Percebemos que é linear em função do tamanho do problema l tal como também é linear quando consideramos o I constante. Mas em relação ao tamanho da população de indivíduos notamos que esta é uma função que cresce de forma logarítmica, porém a estimativa com I constante tem como limite superior um valor que cresce na raiz do tamanho da população isto mostra uma inconsistência entre os dois modelos e mais ainda em relação a prova anteriormente mostrada em aula. Na prova mostrada em aula o tempo de convergência não depende do tamanho da população isto é estranho, pois é natural acreditar que uma população grande de indivíduos demorará mais para convergir. De fato existe um erro considerável quando consideramos, por exemplo, uma população de 1000.000 indivíduos veja:

Prova da aula (linear em l , mas independente de n)

$$g = \frac{l \cdot \pi}{2} \approx l * 1.57$$

Prova com I constante

$$g = 1 + \frac{l \cdot \pi}{2} \cdot \sqrt{(n-1)} \approx l * 1570$$

$$g = l \cdot \left(1 - \frac{2}{n}\right) \approx l$$

Prova com I variável

$$g = 1 + l \cdot \ln\left|\frac{n}{2}\right| = 1 + l \cdot \ln|500.000| \approx l * 13.2$$

Aqui defendo que a estimativa mais rigorosa é a que considera I variável, pois não trabalha com suposições e também não tem arredondamentos e utiliza apenas uma aproximação apenas durante a iteração entre as gerações por uma equação diferencial. A prova da aula utiliza uma aproximação como solução para a equação diferencial (isto não é necessário, pois a equação diferencial tem solução exata), também considera o valor de I , como o valor da população inicial que é uma estimativa menos rigorosa por isto estima o numero de gerações como um valor abaixo do valor mais correto.

3. Questionando Aplicabilidade do Teorema Fundamental de Fischer para UmMax

Aqui será considerado que o Teorema fundamental de Fisher da seleção natural não funciona para o problema UmMax, logo vamos também descobrir o que funciona para este problema. Alguns aspectos que o teorema de Fischer não trata é o fator de truncamento por exemplo no problema UmMax se truncarmos nos 10% melhores é natural acreditar que teremos uma convergência rápida(mas pode ocorrer deriva), se utilizarmos 50% melhores é natural acreditar que vai demorar mais para convergir devido a maior aceitação de indivíduos piores. Na equação de Fischer isto é ignorado, mas para um problema de otimização isto é bem relevante.

Em uma análise empírica utilizando crossover uniforme e seleção por truncamento obteve-se as seguintes evoluções da distribuição do fitness através das gerações.

Os parâmetros utilizados são:

- n Tamanho da população.
- l Tamanho do problema.
- α Desvios padrões acima da média para truncamento.

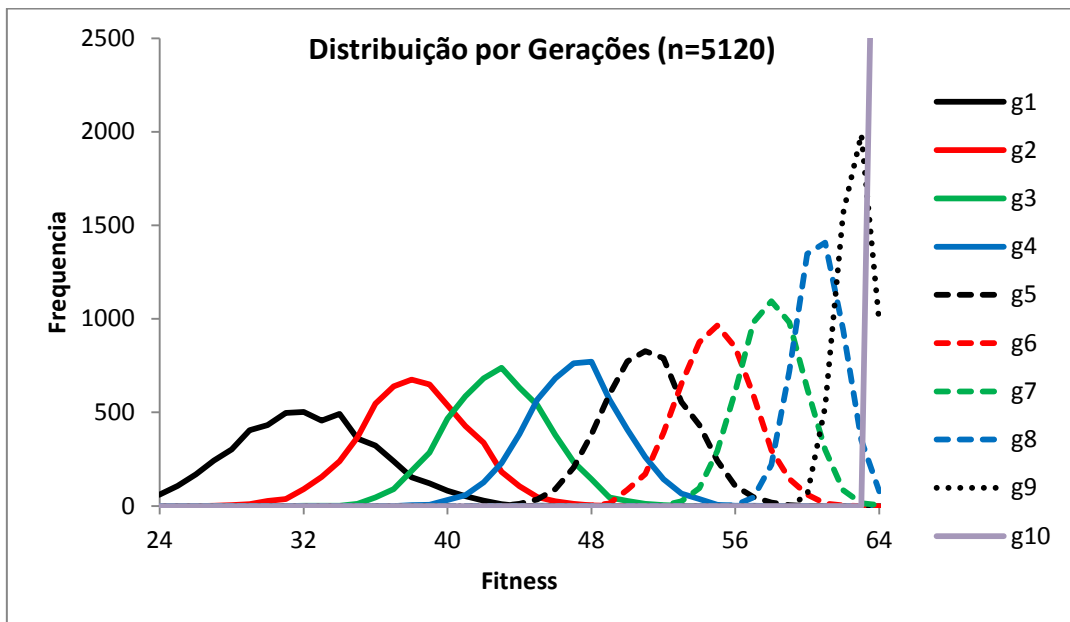
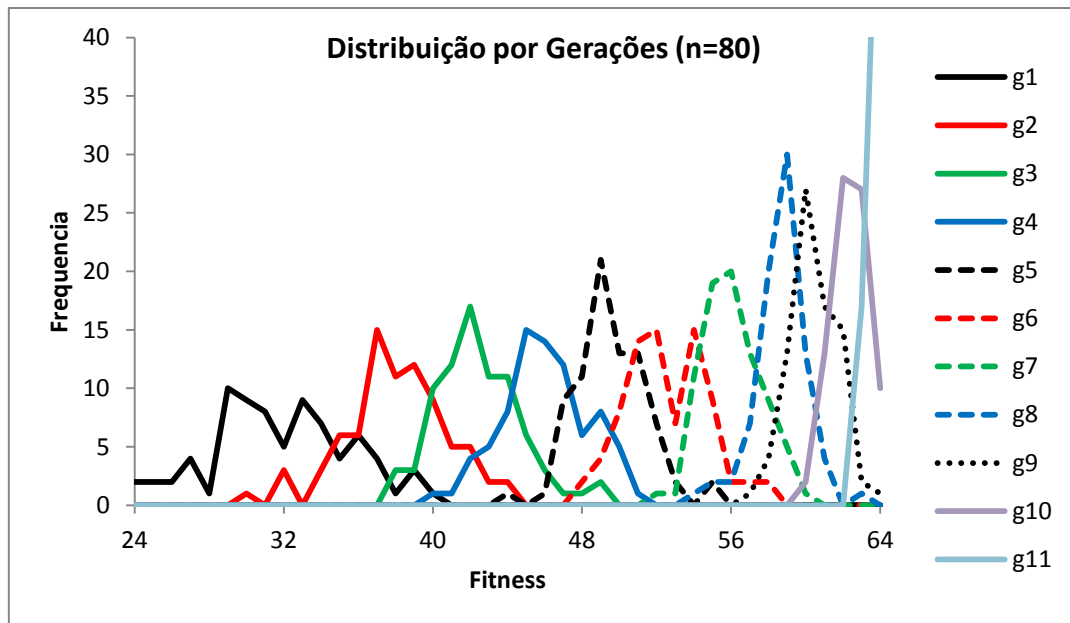
3.1. Analisando Tamanho da População

Na tabela abaixo temos o numero de gerações para convergir para vários tamanhos de população, com $l = 64$ e $\alpha = 1$

n	Gerações
40	Not Solution
80	11
160	11
320	11
640	11
1280	10
2560	11
5120	10

Pode-se reparar que o aumento do tamanho da população não afetou o tempo de convergência para o problema de uma forma considerável na medida em que se aumentou o tamanho da população em escala exponencial, isto é uma informação relevante.

Nos dois gráficos abaixo, temos a evolução das distribuições normais para a população de tamanho 80 e 5120 respectivamente. Abaixo nos vemos que as distribuições normais para uma população maior ficam mais estáveis (com pouco ruído), mas o quanto avança nos dois casos parece ser igual. Ou seja, parece que o numero de indivíduos na população não afeta a convergência.



3.2. Analisando Tamanho do Problema

Na tabela abaixo temos o numero de gerações para convergir para vários tamanhos de problema, com $n = l \cdot 40$ e $\alpha = 1$

l	n	Gerações
8	320	4
16	640	5
32	1280	7
64	2560	11
128	5120	15

Apesar de aumentar o tamanho do problema mais o tamanho da população o crescimento aparenta ser racional, pois cada vez de dobramos o problema as gerações crescem

aproximadamente $\sqrt{2}$. Aqui não mostramos os gráficos, mas na evolução, as gerações continuam a ter distribuições normais de fitness. Abaixo mostra que o mesmo padrão ocorre quando fixamos o n .

l	n	Gerações
8	320	4
16	320	5
32	320	7
64	320	11
128	320	15

3.3. Analisando Taxa de Truncamento

Abaixo é mostrado varias taxas de truncamento dos $Z(\%)$ melhores e seu respectivo α para distribuições normais. O α representa quantos desvios padrões acima da média está o corte do truncamento.

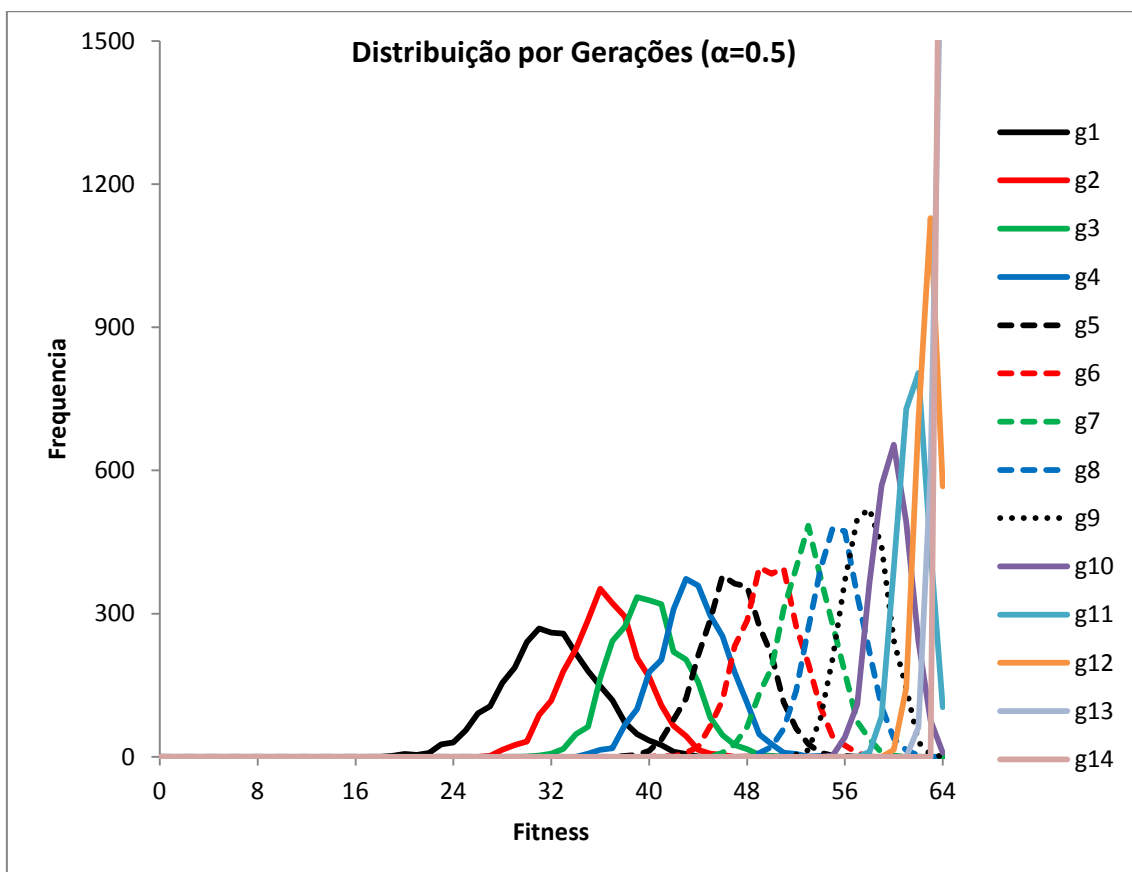
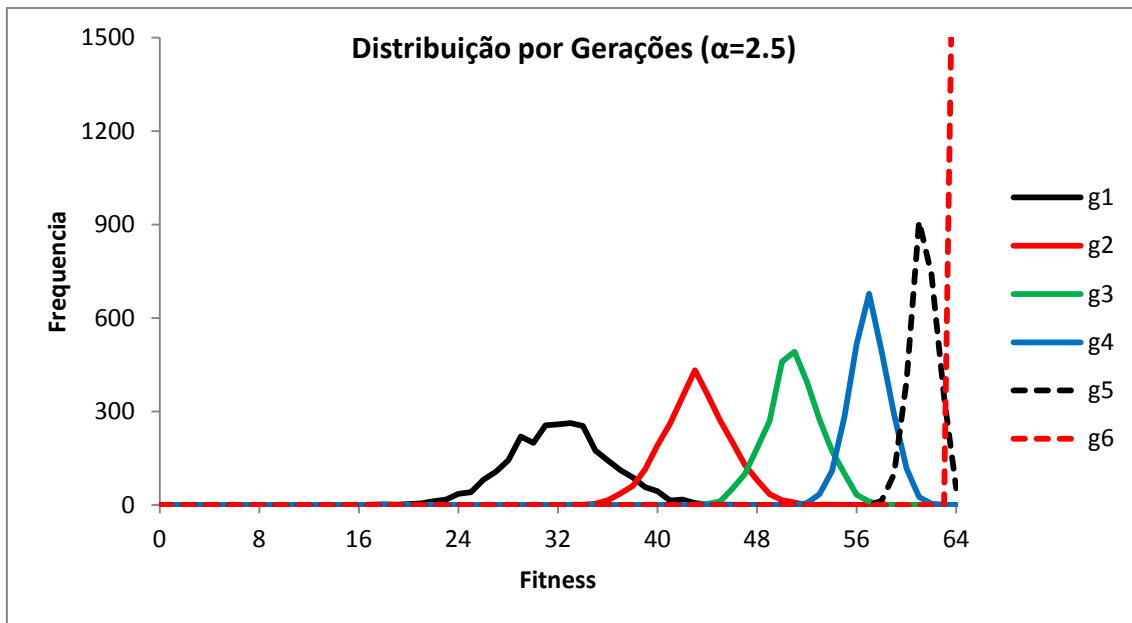
$Z(\%)$	α
05	1.65
10	1.29
20	0.85
30	0.53
40	0.26
50	0.00

$Z(\%)$	α
50.00	0.0
30.85	0.5
15.87	1.0
06.68	1.5
02.28	2.0
00.62	2.5
00.13	3.0

Na tabela abaixo temos o numero de gerações para convergir para vários valores de α , com $l = 64$ e $n = 2560$

$Z(\%)$	α	Gerações
50.00	0.0	19
30.85	0.5	14
15.87	1.0	11
06.68	1.5	9
02.28	2.0	7
00.62	2.5	6
00.13	3.0	Not Solution

É bem visível que quanto mais bem selecionada é a população mais rapidamente temos uma convergência. Nos gráficos abaixo temos com avançou as distribuições normais utilizando α de 2.5 e 0.5 respectivamente.



Nota-se claramente que utilizando uma seleção mais forte avança-se bem mais no fitness médio possibilitando uma convergência rápida para o método.

Notamos que a distribuição dos fitness através das gerações segue aproximadamente uma distribuição normal cuja média aumenta gradativamente e cuja variância diminui também gradativamente. É natural acreditar então que este crescimento pode ser modelado pela equação de Fisher, bem é isto que será verificado teoricamente mais tarde.

4. Nova dedução do numero de gerações

Como na equação de Ficher não levamos em conta o peso do truncamento aqui será utilizada uma tentativa de gerar uma equação baseada no levantamento de dados anterior e que leve em consideração este fator também. Utilizando um truncamento nos $Z(\%)$ melhores estaremos selecionando exatamente os fitness que estão a mais que uma quantia α de desvios padrões da média. Onde encontramos que a média do fitness esperado para a próxima geração será:

$$\overline{f_{t+1}} \geq \overline{f_t} + \alpha \cdot \sigma_t$$

Onde α é uma constante a ser calculada olhando para tabela da distribuição normal de probabilidades, tendo uma relação direta entre α e $Z(\%)$ melhores. A equação acima considera então a forma mais lenta que o fitness deverá crescer. Resultando em um limitante superior para o numero de gerações. Abaixo foi incorporada uma constante $\varepsilon \geq 0$ e transformado a desigualdade em igualdade (tal constante deve ser descoberta para o problema em questão).

$$\overline{f_{t+1}} = \overline{f_t} + (\alpha + \varepsilon) \cdot \sigma_t$$

Esta equação acima será utilizada como uma concorrente para a equação de Fischer quando se utilizar seleção por truncamento, mas não temos qualquer indicação de que esta seja melhor. Sabemos apenas do fato de que a taxa de truncamento afeta a o tempo da convergência.

4.1. Problema UmMax

Dedução das gerações:

$$\begin{aligned}\overline{f_t} &= l \cdot p_t \\ \sigma_t^2 &= l \cdot p_t \cdot (1 - p_t)\end{aligned}$$

$$l \cdot p_{t+1} = l \cdot p_t + (\alpha + \varepsilon) \cdot \sqrt{l \cdot p_t \cdot (1 - p_t)}$$

$$p_{t+1} - p_t = \frac{(\alpha + \varepsilon)}{\sqrt{l}} \cdot \sqrt{p_t \cdot (1 - p_t)}$$

Aproximando por uma equação diferencial, teremos:

$$\frac{\partial p_t}{\partial t} = \frac{(\alpha + \varepsilon)}{\sqrt{l}} \cdot \sqrt{p_t \cdot (1 - p_t)}$$

$$\int \frac{\partial p_t}{\sqrt{p_t \cdot (1 - p_t)}} = \frac{(\alpha + \varepsilon)}{\sqrt{l}} \cdot \int \partial t$$

$$\int_{p_1}^{p_g} \frac{\partial p_t}{\sqrt{p_t \cdot (1 - p_t)}} = \frac{(\alpha + \varepsilon)}{\sqrt{l}} \cdot \int_1^g \partial t$$

Temos que

$$\int \frac{dx}{\sqrt{ax^2 + bx + c}} = -\frac{1}{\sqrt{-a}} \cdot \arcsen\left(\frac{2ax + b}{\sqrt{b^2 - 4ac}}\right) \quad \text{para } a < 0 \text{ e } 4ac - b^2 < 0$$

Logo

$$\left[-\arcsen\left(\frac{-2 \cdot p_t + 1}{\sqrt{1}}\right) \right]_{p_1}^{p_g} = \frac{(\alpha + \varepsilon)}{\sqrt{l}} \cdot [t]_1^g$$

$$\arcsen(1 - 2 \cdot p_1) - \arcsen(1 - 2 \cdot p_g) = \frac{(\alpha + \varepsilon)}{\sqrt{l}} \cdot (g - 1)$$

Utilizando

$$p_1 = \frac{1}{2}$$

$$p_g = 1 - \frac{1}{n}$$

$$\frac{(\alpha + \varepsilon)}{\sqrt{l}} \cdot (g - 1) = \arcsen(1 - 2 \cdot p_1) - \arcsen(1 - 2 \cdot p_g)$$

$$\frac{(\alpha + \varepsilon)}{\sqrt{l}} \cdot (g - 1) = \arcsen(0) - \arcsen\left(\frac{2}{n} - 1\right)$$

$$\arcsen(x) = \sum_{n=0}^{\infty} \frac{(2n)!}{4^n (n!)^2 (2n+1)} \cdot x^{2n+1} \approx x + \frac{x^3}{6} + \frac{3x^5}{40}$$

É natural acreditar também que

$$\left(\frac{2}{n} - 1\right) \geq \arcsen\left(\frac{2}{n} - 1\right) \geq \frac{-\pi}{2}$$

Considerando limite inferior (apenas o primeiro termo, população pequena):

$$\frac{(\alpha + \varepsilon)}{\sqrt{l}} \cdot (g - 1) = 0 - \left(\frac{2}{n} - 1\right) = 1 - \frac{2}{n}$$

$$(g - 1) = \frac{\sqrt{l}}{(\alpha + \varepsilon)} \cdot \left(1 - \frac{2}{n}\right)$$

Considerando limite superior (população muito grande):

$$\frac{(\alpha + \varepsilon)}{\sqrt{l}} \cdot (g - 1) = 0 - \left(\frac{-\pi}{2}\right) = \frac{\pi}{2}$$

$$(g - 1) = \frac{\sqrt{l}}{(\alpha + \varepsilon)} \cdot \frac{\pi}{2}$$

Consideramos então de todas estas estimativas a maior delas onde temos o pior caso do numero de gerações.

$$g = 1 + \frac{\sqrt{l}}{(\alpha + \varepsilon)} \cdot \frac{\pi}{2}$$

4.2. Analisando se a equação encontrada condiz com dados experimentais.

Seja então as equações anteriores que são baseadas na equação de Fisher:

Eq1 – Prova da aula (linear em l , mas independente de n)

$$g = \frac{l \cdot \pi}{2}$$

Eq2 – Prova com I constante

$$g = 1 + \frac{l \cdot \pi}{2} \cdot \sqrt{(n-1)}$$

Eq3 – Prova com I variável

$$g = 1 + l \cdot \ln \left| \frac{n}{2} \right|$$

Eq4 – Nova estimativa utilizando ($\varepsilon = 0.5$)

$$g = 1 + \frac{\sqrt{l}}{(\alpha + \varepsilon)} \cdot \frac{\pi}{2}$$

4.2.1. Tamanho da População

Na tabela abaixo temos o número de gerações que levou para convergir e a estimativa dada por cada uma das equações acima para vários tamanhos de população, com $l = 64$, $\alpha = 1$ e $\varepsilon = 0.5$

n	Gerações	Eq1	Eq2	Eq3	Eq4
40	Not Solution	101	629	193	10
80	11	101	895	238	10
160	11	101	1269	282	10
320	11	101	1797	326	10
640	11	101	2543	371	10
1280	10	101	3597	415	10
2560	11	101	5087	459	10
5120	10	101	7194	504	10

As duas equações que apresentam um comportamento aceitável são Eq1 e Eq4 visto que elas não mudam com a mudança do tamanho da população. Apesar de Eq1 ser bem superior isto pode ser causado por alguma constante.

4.2.2. Tamanho do Problema

Na tabela abaixo temos o número de gerações para convergir para vários tamanhos de problema, com $n = 40 \cdot l$, $\alpha = 1$ e $\varepsilon = 0.5$

l	n	Gerações	Eq1	Eq2	Eq3	Eq4
8	320	4	13	226	42	4
16	640	5	26	637	94	6
32	1280	7	51	1799	208	7
64	2560	11	101	5087	459	10
128	5120	15	202	14387	1006	13

Aqui vemos que a equação Eq1 não serve mais, pois ao dobrar o problema ela prediz que dobrará as gerações, já a equação Eq4 se mantém firme pois ao dobrar o problema aumenta em $\sqrt{2}$ que acompanha próximo ao numero de gerações encontrada.

Abaixo mostra que o mesmo padrão ocorre quando fixamos o $n = 320$. Onde Eq4 mantém mais correta.

l	n	Gerações	Eq1	Eq2	Eq3	Eq4
8	320	4	13	226	42	4
16	320	5	26	450	83	6
32	320	7	51	899	164	7
64	320	11	101	1797	326	10
128	320	15	202	3593	651	13

4.2.3. Taxa de Truncamento

Na tabela abaixo temos o numero de gerações para convergir para vários valores de α , com $l = 64$ e $n = 2560$

$Z(\%)$	α	Gerações	Eq1	Eq2	Eq3	Eq4
50.00	0.0	19	101	5087	459	27
30.85	0.5	14	101	5087	459	14
15.87	1.0	11	101	5087	459	10
06.68	1.5	9	101	5087	459	8
02.28	2.0	7	101	5087	459	7
00.62	2.5	6	101	5087	459	6
00.13	3.0	Not Solution	101	5087	459	5

A equação 4 é a única que considera a força da seleção por truncamento por isto é a única que consegue prever o numero de gerações quando modificamos este parâmetro. Vemos que para valores baixos de α a equação estima muito alto o numero de gerações isto se deve ao fato de que pela desigualdade abaixo quando $\alpha \rightarrow 0$ não existe previsão para a melhoria do fitness.

$$\overline{f_{t+1}} \geq \overline{f_t} + \alpha \cdot \sigma_t$$

5. Erros das Hipóteses

Aqui será verificado a veracidade das hipóteses abaixo:

$$\overline{f_{t+1}} - \overline{f_t} = \frac{\sigma_t^2}{\overline{f_t}} \quad H1$$

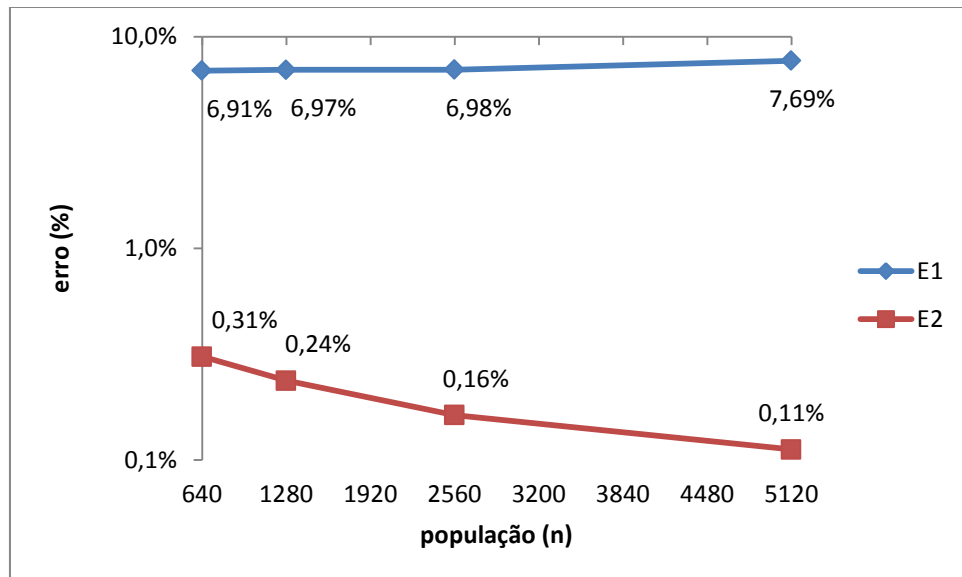
$$\overline{f_{t+1}} - \overline{f_t} = (\alpha + \varepsilon) \cdot \sigma_t \quad H2$$

Para tanto será confrontado os dados experimentais medidos e a teoria, calculamos um erro em percentual por estas duas hipóteses dado por: Abaixo segue os erros de H1 e H2. Será visto que H2 é melhor, a justificativa para isto é o fato de que está foi criada exatamente para funcionar com o truncamento. É provável que H1 seja melhor para outros tipos de seleção. Aqui também estamos olhando apenas para um problema (UmMax).

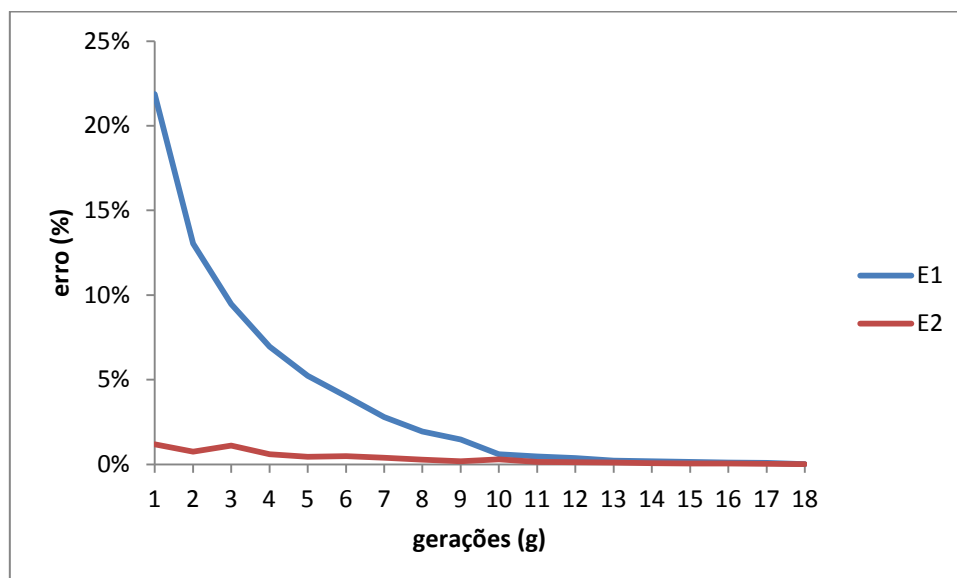
$$\epsilon_1 = \frac{(\overline{f_{t+1}} - \overline{f_t}) - \frac{\sigma_t^2}{\overline{f_t}}}{\overline{f_t}}$$

$$\epsilon_2 = \frac{(\overline{f_{t+1}} - \overline{f_t}) - (\alpha + \varepsilon) \cdot \sigma_t}{\overline{f_t}}$$

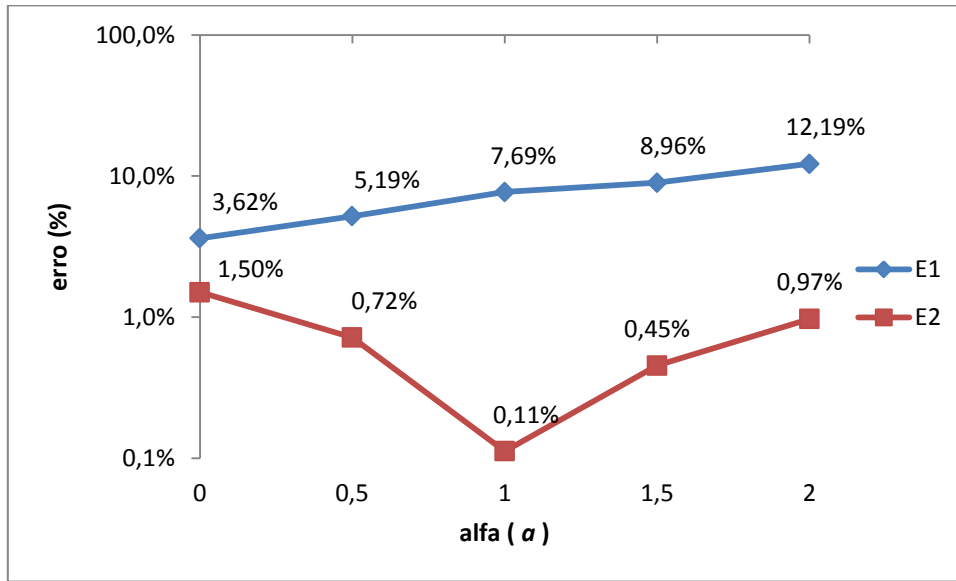
5.1. Erro com o aumento do tamanho da população



5.2. Erro através das gerações



5.3. Erro com o aumento do alfa no truncamento



6. Problema BinInt

Esta equação abaixo será utilizada, pois apresentou um resultado mais coerente quando se utilizou seleção por truncamento. Vamos verificar se também é boa para o problema BinInt.

$$\overline{f_{t+1}} = \overline{f_t} + (\alpha + \varepsilon) \cdot \sigma_t$$

Definição do problema, fitness médio e a variância.

$$f(x) = \sum_{i=1}^l 2^{l-i} \cdot x_i$$

l Tamanho do problema.

λ Número de bits que convergiram em 1.

Fitness médio.

$$\overline{f(\lambda)} = 2^l - 2^{l-\lambda} + 2^{l-\lambda-1} = 2^l \cdot \left(1 - \frac{1}{2 \cdot 2^\lambda}\right)$$

Variância.

$$\sigma(\lambda) = \frac{2^{l-\lambda-1}}{\sqrt{3}} = \frac{2^l}{2 \cdot \sqrt{3}} \cdot \left(\frac{1}{2^\lambda}\right)$$

Substituindo em:

$$\overline{f_{t+1}} = \overline{f_t} + (\alpha + \varepsilon) \cdot \sigma_t$$

$$2^l \cdot \left(1 - \frac{1}{2 \cdot 2^{\lambda_{t+1}}}\right) - 2^l \cdot \left(1 - \frac{1}{2 \cdot 2^{\lambda_t}}\right) = (\alpha + \varepsilon) \cdot \frac{2^l}{2 \cdot \sqrt{3}} \cdot \left(\frac{1}{2^{\lambda_t}}\right)$$

$$\left(\frac{1}{2^{\lambda_{t+1}}}\right) - \left(\frac{1}{2^{\lambda_t}}\right) = (\alpha + \varepsilon) \cdot \frac{-1}{\sqrt{3}} \cdot \left(\frac{1}{2^{\lambda_t}}\right)$$

$$\mu_t = \frac{1}{2^{\lambda_t}}$$

$$\mu_{t+1} - \mu_t = \frac{-(\alpha + \varepsilon)}{\sqrt{3}} \cdot \mu_t$$

Aproximando por uma equação diferencial.

$$\frac{\partial \mu_t}{\partial t} = \frac{-(\alpha + \varepsilon)}{\sqrt{3}} \cdot \mu_t$$

$$\int_{\mu_1}^{\mu_g} \frac{d\mu_t}{\mu_t} = \frac{-(\alpha + \varepsilon)}{\sqrt{3}} \cdot \int_1^g dt$$

$$[\ln \mu_t]_{\mu_1}^{\mu_g} = \frac{-(\alpha + \varepsilon)}{\sqrt{3}} \cdot (g - 1)$$

$$\mu_1 = \frac{1}{2^{\lambda_1}} = \frac{1}{2^0} = 1$$

$$\mu_g = \frac{1}{2^{\lambda_g}} = \frac{1}{2^l}$$

$$\ln \left| \frac{1}{2^l} \right| - \ln 1 = \frac{-(\alpha + \varepsilon)}{\sqrt{3}} \cdot (g - 1)$$

$$\ln |2^l| = \frac{(\alpha + \varepsilon)}{\sqrt{3}} \cdot (g - 1)$$

$$(g - 1) = \frac{\ln |2| \cdot \sqrt{3} \cdot l}{(\alpha + \varepsilon)}$$

6.1. Analisando se equação encontrada condiz com dados experimentais.

$$(g - 1) = \frac{\ln |2| \cdot \sqrt{3} \cdot l}{(\alpha + \varepsilon)}$$

6.1.1. Tamanho da População

Na tabela abaixo temos o numero de gerações que levou para convergir e a estimativa dada por cada uma das equações acima para vários tamanhos de população, com $l = 32$, $\alpha = 1$ e $\varepsilon = 0.5$

Na última coluna temos uma o resultado da previsão dividido por uma constante e que se aproxima dos dados experimentais.

n	Gerações	Eq	2.0
40	not	27	13
80	not	27	13
160	13	27	13

320	12	27	13
640	12	27	13
1280	13	27	13
2560	13	27	13
5120	13	27	13

A equação apresenta um comportamento aceitável visto que ela não muda com a mudança do tamanho da população.

6.1.2. Tamanho do Problema

Na tabela abaixo temos o numero de gerações para convergir para vários tamanhos de problema, com $\alpha = 1$ e $\varepsilon = 0.5$

l	n	Gerações	Eq	1.8
4	20	5	5	3
8	40	5	8	5
16	80	8	14	8
32	160	15	27	15

Abaixo mostra que o mesmo padrão ocorre quando fixamos o $n = 160$. Onde Eq4 mantém mais correta.

l	n	Gerações	Eq	1.8
4	160	3	5	3
8	160	5	8	5
16	160	8	14	8
32	160	15	27	15

6.1.3. Taxa de Truncamento

Na tabela abaixo temos o numero de gerações para convergir para vários valores de α , com $l = 32$ e $n = 5120$

$Z(\%)$	α	Gerações	Eq	2.7
50.00	0.0	29	78	29
30.85	0.5	18	40	15
15.87	1.0	13	27	11
06.68	1.5	9	21	8
02.28	2.0	7	17	7
00.62	2.5	7	14	6
00.13	3.0	not	12	5

Um fato a ser notado aqui é que para o problema BinInt a equação prevê corretamente apenas o crescimento assintótico. Não conseguindo dizer com exatidão o valor absoluto do número de gerações. Outro fato é que a equação define sempre um limite superior para o número de gerações. Acredito que isto se deve ao fato de termos feito uma suposição conservadora durante a prova: foi suposto que os bits mais importantes (com maior peso na função objetivo) irão convergir primeiro e só depois os demais convergirão. Bem isto não é necessariamente verdade, tendo liberdade, portanto para o método convergir bem antes do previsto. Mas é notável que para um problema bem mais difícil o tempo de convergência é apenas linear. Relativo ao tamanho do problema.

7. Problema UmMax com Seleção por Torneio

Para o problema UmMax temos a dedução no número de gerações utilizando seleção por truncamento.

$$g = 1 + \frac{\sqrt{l}}{(\alpha + \varepsilon)} \cdot \frac{\pi}{2}$$

Esta equação será adaptada para utilizar Torneio. O parâmetro α da equação indica quantos desvios padrões está o truncamento acima da média. Como na seleção Roleta utilizamos outro parâmetro γ que informa de quantos indivíduos é o torneio. Teremos como hipótese que existe uma equivalência entre γ e α dado como abaixo.

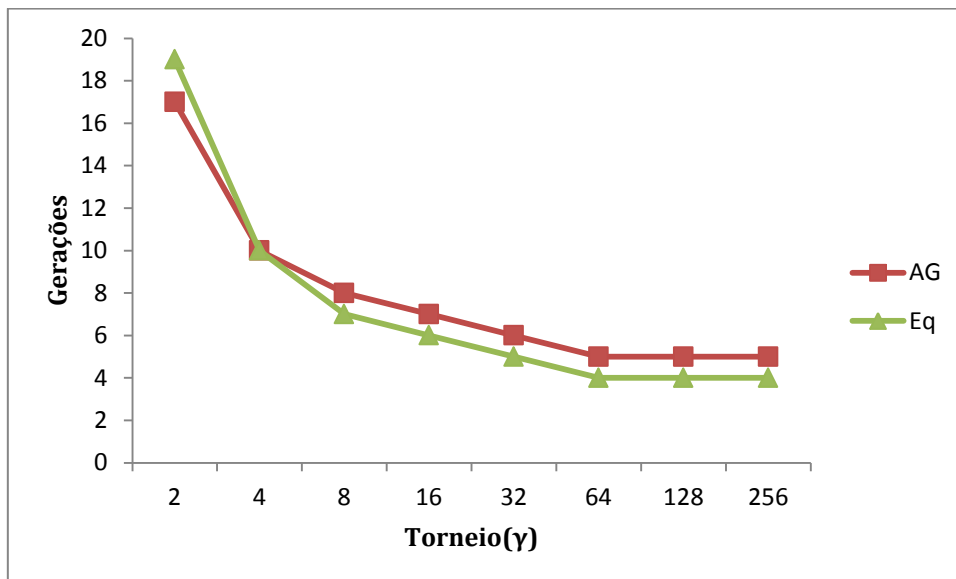
$$\alpha = \frac{(\log_2|\gamma| - 1)}{2}$$

Logo a equação para torneio no problema UmMax será:

$$g = 1 + \frac{\pi \cdot \sqrt{l}}{(\log_2|\gamma| - 1 + 2 \cdot \varepsilon)}$$

7.1. Sensibilidade do Torneio

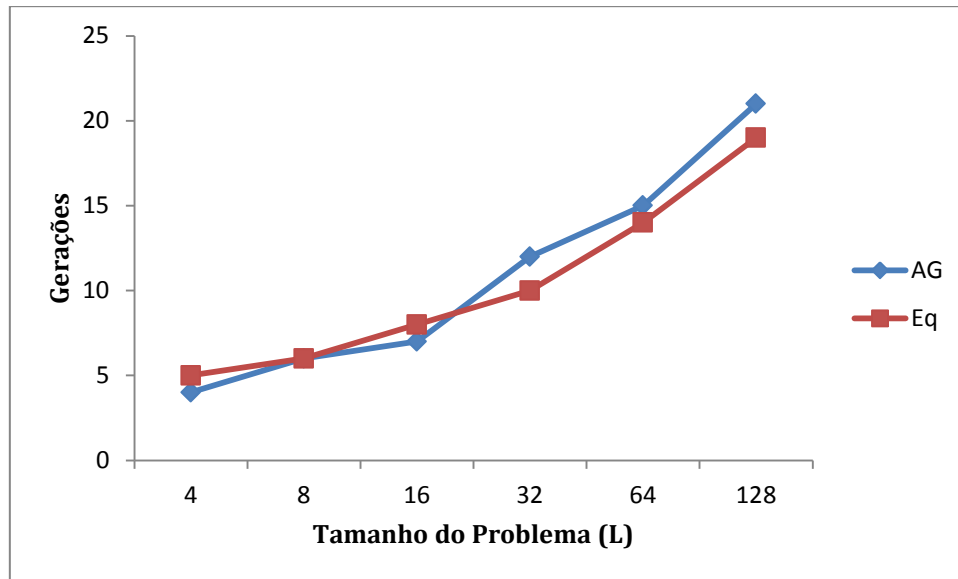
No gráfico e tabela abaixo temos o número de gerações que o AG levou para convergir e a estimativa dada pela equação adaptada para torneio. Foi utilizado $l = 32$, $n = 5120$ e $\varepsilon = 0.5$



γ	α	AG	Eq
2	0	17	19
4	0.5	10	10
8	1	8	7
16	1.5	7	6
32	2	6	5
64	2.5	5	4
128	3	5	4
256	3.5	5	4

7.2. Sensibilidade do Tamanho do Problema

No gráfico e tabela abaixo verificamos o que ocorre quando aumentamos rapidamente o tamanho do problema, onde foi fixado $l = 32$, $n = 5 \cdot l$, $\gamma = 4$ e $\varepsilon = 0.5$



<i>n</i>	<i>l</i>	AG	Eq
20	4	4	5
40	8	6	6
80	16	7	8
160	32	12	10
320	64	15	14
640	128	21	19

7.3. Sensibilidade do Tamanho da População

Na tabela abaixo temos o numero de gerações para convergir para vários tamanhos de de população, com $l = 32$, $\gamma = 4$ e $\varepsilon = 0.5$

<i>n</i>	AG	Eq
20	not	10
40	11	10
80	11	10
160	11	10
320	10	10
640	10	10

Percebemos que a adaptação funcionou muito bem para prever com bastante exatidão o número de gerações, uma situação interessante a ser observada é que a equação não só está prevendo bem o comportamento assintótico, mas informando em termos absolutos o número de gerações até a convergência.

8. Problema UmMax com Seleção Proporcional

Nos testes com seleção proporcional os resultados foram bem diferentes. Em termos de comportamento assintótico em relação aos tipos de seleção anteriormente utilizados. Mas aqui resgataremos a equação que foi deduzida na seção 2.

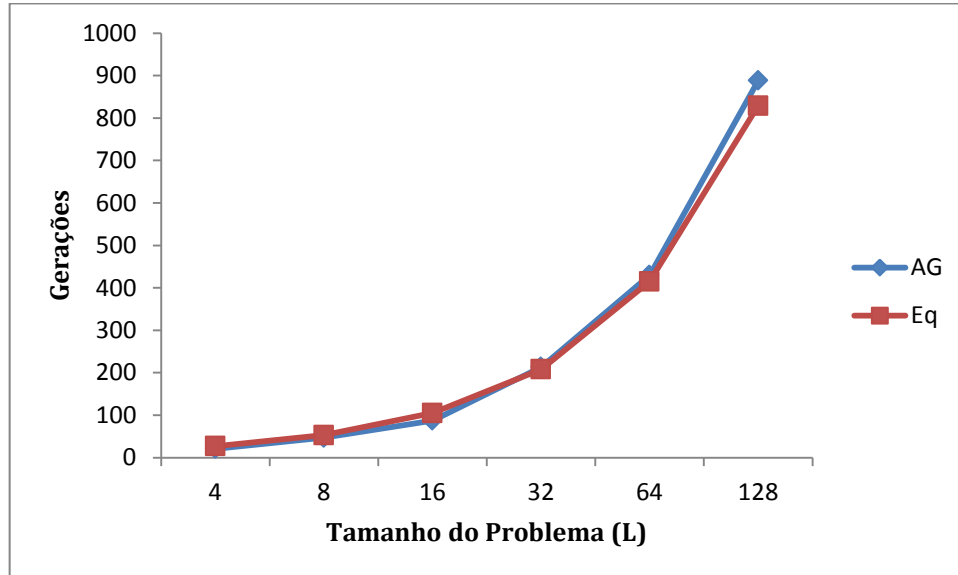
$$g = 1 + l \cdot \ln \left| \frac{n}{2} \right|$$

Esta equação é a mais correta para seleção proporcional pois sua dedução baseou-se apenas no teorema de Fisher e que foi construído para seleção proporcional.

Na equação acima vemos que está considera que o tamanho do problema influencia com crescimento linear para o problema UmMax(as anteriores o crescimento era em raiz de l), esta equação considera também que o tamanho da população influencia em escala logarítmica. Os resultados a seguir mostram que isto está correto. Nos resultados notou-se que a seleção proporcional além de demorar muito mais tempo para convergir exigia uma população de indivíduos muito alta para resolver o problema. Devido também a uma maior instabilidade verificada pela execução da seleção proporcional, o AG foi executado 10 vezes e obtido a média (AG-avg) e desvio padrão (AG-sd) do número de gerações. O desvio padrão está mostrado dividido pela média por isto está em porcentagem.

8.1. Sensibilidade do Tamanho do Problema

No gráfico abaixo verificamos o que ocorre quando aumentamos o tamanho do problema, onde foi fixado $n = 1280$. Nota-se pelo gráfico que a previsão é bem exata em termos de valores absolutos e pela tabela vemos que o crescimento é linear (não se confunda com o gráfico: o eixo do tamanho do problema tem crescimento exponencial).



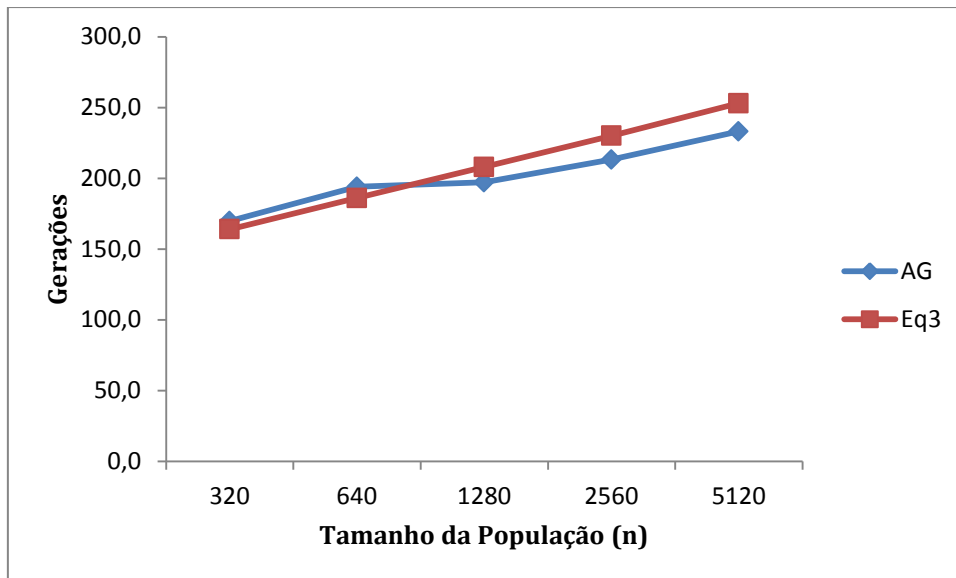
Abaixo é mostrada a tabela com os valores de média (AG-avg) e desvio padrão (AG-sd) obtidos pelo AG e os valores previstos pela equação (Eq).

l	AG-avg	AG-sd	Eq
4	21.1	18%	27
8	46.9	14%	53
16	87.4	11%	105

32	213.1	19%	208
64	429.8	17%	415
128	888.3	30%	829

8.2. Sensibilidade do Tamanho da População

No gráfico abaixo temos o numero de gerações para convergir para vários tamanhos de população, com $l = 32$



Abaixo é mostrada a tabela com os valores de média (AG-avg) e desvio padrão (AG-sd) obtidos pelo AG e os valores previstos pela equação (Eq).

n	AG-avg	AG-sd	Eq
320	169.7	30%	164
640	194.0	35%	186
1280	197.1	17%	208
2560	213.2	16%	230
5120	233.1	8%	253

Percebemos que a adaptação funcionou muito bem para prever com bastante exatidão o número de gerações, uma situação interessante a ser observada é que a equação não só está prevendo bem o comportamento assintótico, mas informando em termos absolutos o número de gerações até a convergência.

9. Dedução BinInt para Seleção Proporcional

Visto que para o problema UmMax com seleção proporcional a dedução correta foi supor uma constante na equação de Ficher. Aqui também faremos a prova teórica para o problema BinInt considerando l variável.

Fitness médio.

$$\overline{f(\lambda)} = 2^l - 2^{l-\lambda} + 2^{l-\lambda-1} = 2^l \cdot \left(1 - \frac{1}{2 \cdot 2^\lambda}\right)$$

Variância.

$$\sigma(\lambda) = \frac{2^{l-\lambda-1}}{\sqrt{3}} = \frac{2^l}{2 \cdot \sqrt{3}} \cdot \left(\frac{1}{2^\lambda}\right)$$

Substituindo em:

$$\overline{f_{t+1}} - \overline{f_t} = \frac{\sigma_t^2}{\overline{f_t}}$$

$$2^l \cdot \left(1 - \frac{1}{2 \cdot 2^{\lambda_{t+1}}}\right) - 2^l \cdot \left(1 - \frac{1}{2 \cdot 2^{\lambda_t}}\right) = \frac{\left(\frac{2^l}{2 \cdot \sqrt{3}} \cdot \left(\frac{1}{2^{\lambda_t}}\right)\right)^2}{2^l \cdot \left(1 - \frac{1}{2 \cdot 2^{\lambda_t}}\right)}$$

Desenvolvendo termo quadrático

$$2^l \cdot \left(1 - \frac{1}{2 \cdot 2^{\lambda_{t+1}}}\right) - 2^l \cdot \left(1 - \frac{1}{2 \cdot 2^{\lambda_t}}\right) = \frac{\frac{2^{2 \cdot l}}{4 \cdot 3} \cdot \left(\frac{1}{2^{2 \cdot \lambda_t}}\right)}{2^l \cdot \left(1 - \frac{1}{2 \cdot 2^{\lambda_t}}\right)}$$

Cortando 2^l e simplificando

$$\frac{1}{2 \cdot 2^{\lambda_{t+1}}} - \frac{1}{2 \cdot 2^{\lambda_t}} = \frac{\frac{-1}{4 \cdot 3} \cdot \left(\frac{1}{2^{2 \cdot \lambda_t}}\right)}{\left(1 - \frac{1}{2 \cdot 2^{\lambda_t}}\right)}$$

Multiplicando $\left(1 - \frac{1}{2 \cdot 2^{\lambda_t}}\right)$ nos dois lados

$$\left(\frac{1}{2 \cdot 2^{\lambda_{t+1}}} - \frac{1}{2 \cdot 2^{\lambda_t}}\right) \cdot \left(1 - \frac{1}{2 \cdot 2^{\lambda_t}}\right) = \frac{-1}{4 \cdot 3} \cdot \left(\frac{1}{2^{2 \cdot \lambda_t}}\right)$$

Desenvolvendo

$$\left(\frac{1}{2 \cdot 2^{\lambda_{t+1}}} - \frac{1}{2 \cdot 2^{\lambda_t}}\right) - \left(\frac{1}{4 \cdot 2^{\lambda_{t+1}} \cdot 2^{\lambda_t}} - \frac{1}{4 \cdot 2^{\lambda_t} \cdot 2^{\lambda_t}}\right) = \frac{-1}{4 \cdot 3} \cdot \left(\frac{1}{2^{\lambda_t} \cdot 2^{\lambda_t}}\right)$$

$$\frac{1}{2 \cdot 2^{\lambda_{t+1}}} - \frac{1}{2 \cdot 2^{\lambda_t}} - \frac{1}{4 \cdot 2^{\lambda_{t+1}} \cdot 2^{\lambda_t}} + \frac{1}{4 \cdot 2^{\lambda_t} \cdot 2^{\lambda_t}} + \frac{1}{4 \cdot 3 \cdot 2^{\lambda_t} \cdot 2^{\lambda_t}} = 0$$

$$\frac{1}{2 \cdot 2^{\lambda_{t+1}}} - \frac{1}{2 \cdot 2^{\lambda_t}} - \frac{1}{4 \cdot 2^{\lambda_{t+1}} \cdot 2^{\lambda_t}} + \frac{1}{3 \cdot 2^{\lambda_t} \cdot 2^{\lambda_t}} = 0$$

Definindo μ_t e substituindo

$$\mu_t = \frac{1}{2^{\lambda_t}}$$

$$\frac{\mu_{t+1}}{2} - \frac{\mu_t}{2} - \frac{\mu_{t+1} \cdot \mu_t}{4} + \frac{\mu_t \cdot \mu_t}{3} = 0$$

$$6 \cdot \mu_{t+1} - 6 \cdot \mu_t - 3 \cdot \mu_{t+1} \cdot \mu_t + 4 \cdot \mu_t \cdot \mu_t = 0$$

$$\mu_{t+1} = \frac{6 \cdot \mu_t - 4 \cdot \mu_t \cdot \mu_t}{(6 - 3 \cdot \mu_t)}$$

Subtraindo μ_t nos dois lados e aproximando por uma equação diferencial

$$\frac{\partial \mu_t}{\partial t} = \frac{6 \cdot \mu_t - 4 \cdot \mu_t \cdot \mu_t}{(6 - 3 \cdot \mu_t)} - \mu_t$$

Simplificando

$$\begin{aligned} \frac{\partial \mu_t}{\partial t} &= \frac{6 \cdot \mu_t - 4 \cdot \mu_t \cdot \mu_t - 6 \cdot \mu_t + 3 \cdot \mu_t \cdot \mu_t}{(6 - 3 \cdot \mu_t)} \\ \frac{\partial \mu_t}{\partial t} &= \frac{\mu_t^2}{(3 \cdot \mu_t - 6)} \end{aligned}$$

Montando uma equação integral

$$\int_{\mu_1}^{\mu_g} \frac{(3 \cdot \mu_t - 6) \partial \mu_t}{\mu_t^2} = \int_1^g \partial t$$

Resolvendo a integral

$$\begin{aligned} 3 \cdot \int_{\mu_1}^{\mu_g} \frac{\partial \mu_t}{\mu_t} - 6 \cdot \int_{\mu_1}^{\mu_g} \frac{\partial \mu_t}{\mu_t^2} &= \int_1^g \partial t \\ 3 \cdot [\ln |\mu_t|]_{\mu_1}^{\mu_g} - 6 \cdot \left[\frac{-1}{\mu_t} \right]_{\mu_1}^{\mu_g} &= \int_1^g \partial t \\ 3 \cdot \left(\ln \left| \frac{\mu_g}{\mu_1} \right| \right) + 6 \cdot \left(\frac{1}{\mu_g} - \frac{1}{\mu_1} \right) &= (g - 1) \end{aligned}$$

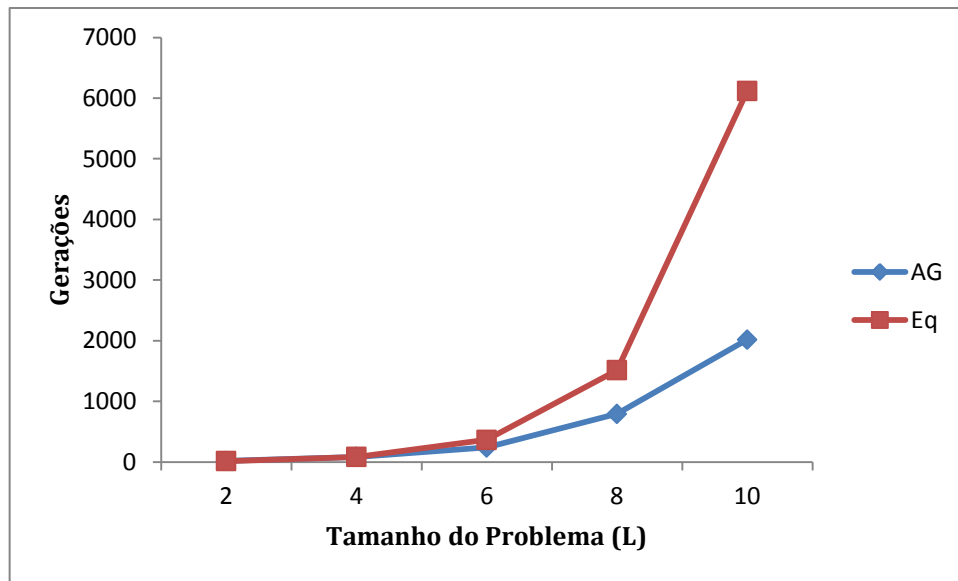
Tomando os valores abaixo e substituindo.

$$\begin{aligned} \mu_1 &= \frac{1}{2^{\lambda_1}} = \frac{1}{2^0} = 1 \\ \mu_g &= \frac{1}{2^{\lambda_g}} = \frac{1}{2^l} \\ 3 \cdot \left(\ln \left| \frac{1}{2^l} \right| \right) + 6 \cdot (2^l - 1) &= (g - 1) \\ (g - 1) &= 6 \cdot (2^l - 1) - 3 \cdot \ln |2| \cdot l \\ g &= 3 \cdot (2^{l+1} - \ln |2| \cdot l) - 5 \end{aligned}$$

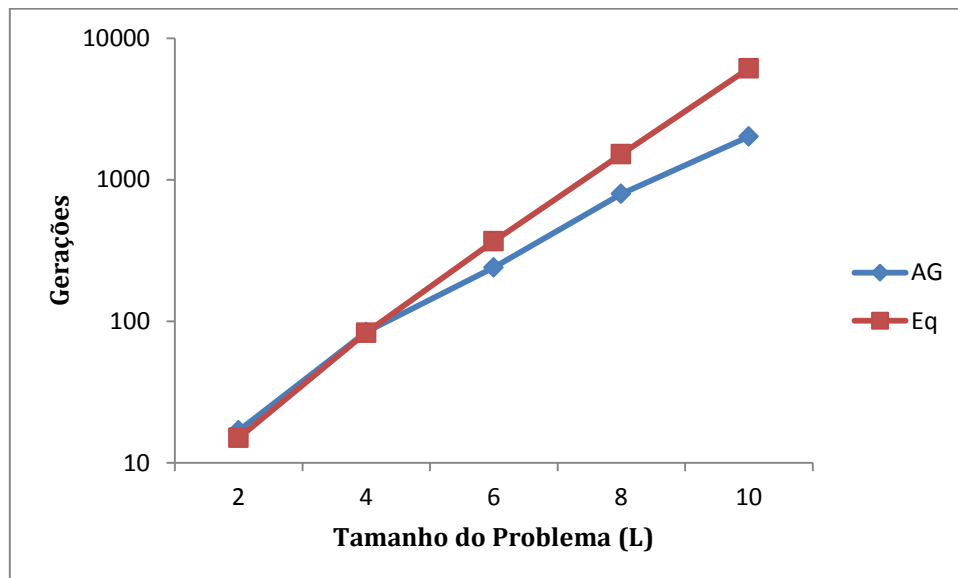
Verificando notamos que para o problema BinInt a equação prevê que este levará um número de gerações exponenciais com o tamanho do problema. Aqui notamos que o tamanho da população não afetará o tempo de convergência. Em seguida será verificado isto experimentalmente.

9.1. Sensibilidade do Tamanho do Problema

No gráfico abaixo verificamos o que ocorre quando aumentamos o tamanho do problema, onde foi fixado $n = 2560$.



O AG parece ter um crescimento exponencial também, em seguida o mesmo gráfico em escala exponencial.



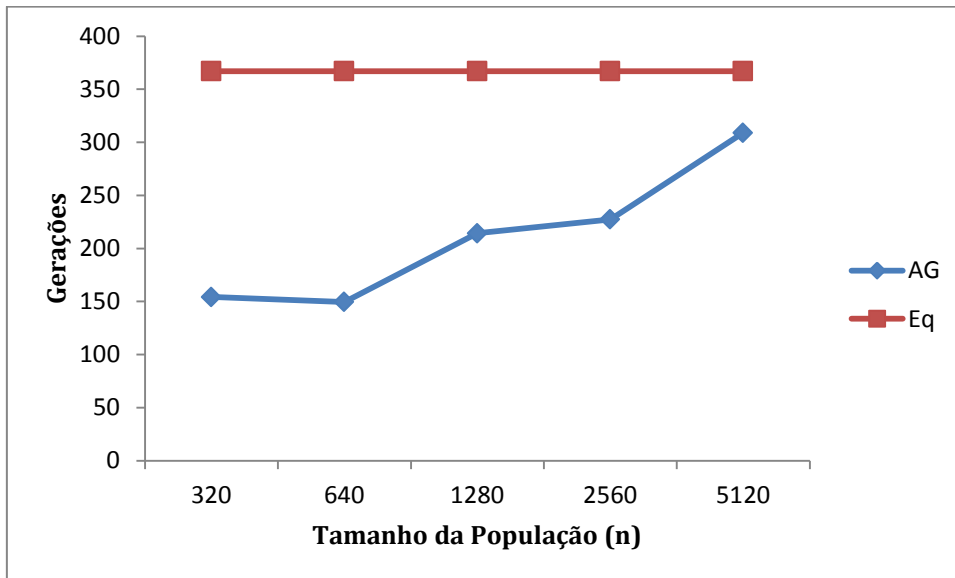
Como temos aproximadamente uma reta descrita pelo AG vemos um crescimento muito rápido das gerações com um aumento suave do tamanho do problema. Como inclinação da reta descrita pelo AG é menor do que a da Equação isto sugere que o fator exponencial do AG deve ser uma base um pouco menor que 2.

Abaixo é mostrada a tabela com os valores de média (AG-avg) e desvio padrão (AG-sd) obtidos pelo AG e os valores previstos pela equação (Eq).

l	AG-avg	AG-sd	Eq
2	16.9	16%	15
4	83.9	24%	83
6	239.6	28%	367
8	793.3	51%	1515
10	2018.0	22%	6119

9.2. Sensibilidade do Tamanho da População

No gráfico abaixo temos o numero de gerações para convergir para vários tamanhos de população, com $l = 6$



Vemos que a equação não prevê crescimento, mas de fato parece existir um crescimento baixo. Abaixo é mostrada a tabela com os valores de média (AG-avg) e desvio padrão (AG-sd) obtidos pelo AG e os valores previstos pela equação (Eq).

<i>n</i>	AG-avg	AG-sd	Eq
320	154.3	41%	367
640	149.6	26%	367
1280	214.3	28%	367
2560	227.4	39%	367
5120	309.0	27%	367

Percebemos que a adaptação funcionou bem para prever o crescimento assintótico para o tamanho do problema, mas acabou ignorando o fato de que existe um crescimento também baseado no tamanho da população.

10. Formula Geral para Encontrar o Fitness Médio e Variância – GA simples

Dado um problema qualquer cuja função de avaliação $f(u)$ dependa apenas do número de bits configurados com 1, definiremos então:

p_t Probabilidade de cada bit da string l valer 1.

l Tamanho do problema.

$f(u)$ Função de avaliação.

$$u = \sum_{i=1}^l x_i$$

O fitness médio pode ser encontrado somando todos os fitness possíveis ponderados pela sua probabilidade de ocorrer, logo:

$$\bar{f}_t = \sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1 - p_t)^{l-u}) \cdot f(u)$$

Veja que tomamos uma função de avaliação que depende apenas do numero de uns e zeros não tendo importância a posição em que o bit foi encontrado. Somando para todo o u uma combinação do numero de ocorrências deste padrão possível multiplicado pela probabilidade de que os u primeiros bits sejam 1 e os $(l-u)$ bits seja zero, multiplicado ainda pela função de avaliação.

A variância de n amostras de valor y_i com frequência F_i de ocorrer e média μ pode ser obtida por.

$$\sigma^2 = \frac{\sum_{i=1}^n F_i \cdot (y_i - \mu)^2}{\sum_{i=1}^n F_i}$$

Logo a variância para nosso caso é:

$$n = l$$

$$\mu = \bar{f}_t$$

$$y_i = f(u)$$

$$F_i = \binom{l}{u} \cdot (p_t^u \cdot (1 - p_t)^{l-u})$$

$$\sum_{i=1}^l F_i = 1$$

Substituindo temos:

$$\sigma^2 = \sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1 - p_t)^{l-u}) \cdot (f(u) - \bar{f}_t)^2$$

10.1. Provando para UmMax

Para o problema temos que:

$$f(u) = u$$

$$u = \sum_{i=1}^l x_i$$

Encontrando o fitness médio.

$$\bar{f}_t = \sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1 - p_t)^{l-u}) \cdot f(u)$$

$$\bar{f}_t = \sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1-p_t)^{l-u}) \cdot u$$

Não encontrou-se solução para a equação acima mas por análise empírica acredita-se que a solução abaixo é válida. (Foi confirmado também pelo Wolfram, porém não foi mostrado os passo para a solução: <http://www.wolframalpha.com>).

$$\sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1-p_t)^{l-u}) \cdot u = l \cdot p_t$$

Encontrando a variância.

$$\sigma^2 = \sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1-p_t)^{l-u}) \cdot (f(u) - \bar{f}_t)^2$$

$$\sigma^2 = \sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1-p_t)^{l-u}) \cdot (u - l \cdot p_t)^2$$

$$\sigma^2 = \sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1-p_t)^{l-u}) \cdot (u^2 - 2 \cdot l \cdot p_t \cdot u + (l \cdot p_t)^2)$$

$$\sigma^2 = S_1 + S_2 + S_3$$

$$S_1 = \sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1-p_t)^{l-u}) \cdot (u^2)$$

$$S_2 = \sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1-p_t)^{l-u}) \cdot (-2 \cdot l \cdot p_t \cdot u)$$

$$S_3 = \sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1-p_t)^{l-u}) \cdot (l \cdot p_t)^2$$

Resolvendo S_1

$$S_1 = \sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1-p_t)^{l-u}) \cdot (u^2)$$

Novamente não encontrou-se solução para a equação acima mas por análise empírica acredita-se que.(Também foi confirmado pelo Wolfram)

$$\sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1-p_t)^{l-u}) \cdot (u^2) = (l \cdot p_t)^2 + l \cdot p_t \cdot (1-p_t)$$

Substituindo

$$S_1 = (l \cdot p_t)^2 + l \cdot p_t \cdot (1-p_t)$$

Resolvendo S_2

$$S_2 = \sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1-p_t)^{l-u}) \cdot (-2 \cdot l \cdot p_t \cdot u)$$

$$S_2 = -2 \cdot l \cdot p_t \cdot \sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1-p_t)^{l-u}) \cdot (u)$$

Como antes pela solução empírica.

$$\sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1-p_t)^{l-u}) \cdot u = l \cdot p_t$$

$$S_2 = -2 \cdot (l \cdot p_t)^2$$

Resolvendo S_3

$$S_3 = \sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1-p_t)^{l-u}) \cdot (l \cdot p_t)^2$$

$$\sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1-p_t)^{l-u}) \cdot (l \cdot p_t)^2$$

$$S_3 = (l \cdot p_t)^2 \cdot \sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1-p_t)^{l-u})$$

Sabe-se que

$$\sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1-p_t)^{l-u}) = 1$$

Substituindo

$$S_3 = (l \cdot p_t)^2$$

Juntando tudo

$$S_1 = (l \cdot p_t)^2 + l \cdot p_t \cdot (1-p_t)$$

$$S_2 = -2 \cdot (l \cdot p_t)^2$$

$$S_3 = (l \cdot p_t)^2$$

$$\sigma^2 = S_1 + S_2 + S_3$$

$$\sigma^2 = (l \cdot p_t)^2 + l \cdot p_t \cdot (1-p_t) - 2 \cdot l \cdot p_t \cdot l \cdot p_t + (l \cdot p_t)^2$$

$$\sigma^2 = l \cdot p_t \cdot (1-p_t)$$

10.2. Provando para Trap

$$f(x) = \sum_{i=1}^m w(u)$$

$$u = \sum_{j=1}^k x_{ij}$$

$$w(u) = \begin{cases} k & ; u = k \\ k - 1 - u & ; u < k \end{cases}$$

p_t Probabilidade de cada bit da string l valer 1.

l Tamanho do problema.

m Número de BBs

k Tamanho dos BBs

$$l = k \cdot m$$

Encontrando o fitness médio.

Adaptando a equação geral para trabalhar com m BB's de tamanho k .

$$\bar{f}_t = \sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1 - p_t)^{l-u}) \cdot f(u)$$

Como os m problemas são independentes teremos então:

$$\bar{f}_t = \sum_{i=1}^m \sum_{u=0}^k \binom{k}{u} \cdot (p_t^u \cdot (1 - p_t)^{k-u}) \cdot w(u)$$

Substituindo $w(u)$ e separando o primeiro termo quando $u = k$

$$\bar{f}_t = \sum_{i=1}^m \left(\binom{k}{k} \cdot (p_t^k \cdot (1 - p_t)^{k-k}) \cdot k + \sum_{u=0}^{k-1} \binom{k}{u} \cdot (p_t^u \cdot (1 - p_t)^{k-u}) \cdot (k - 1 - u) \right)$$

$$\bar{f}_t = m \cdot k \cdot p_t^k + \sum_{i=1}^m \left(\sum_{u=0}^{k-1} \binom{k}{u} \cdot (p_t^u \cdot (1 - p_t)^{k-u}) \cdot ([k - 1] - [u]) \right)$$

Quebrando em dois problemas S_1 e S_2 .

$$\bar{f}_t = m \cdot k \cdot p_t^k + \sum_{i=1}^m (S_1 - S_2)$$

$$S_1 = \sum_{u=0}^{k-1} \binom{k}{u} \cdot (p_t^u \cdot (1 - p_t)^{k-u}) \cdot (k - 1)$$

$$S_2 = \sum_{u=0}^{k-1} \binom{k}{u} \cdot (p_t^u \cdot (1 - p_t)^{k-u}) \cdot (u)$$

Resolvendo S_1 , temos que:

$$\sum_{u=0}^k \binom{k}{u} \cdot (p_t^u \cdot (1-p_t)^{k-u}) = 1$$

Multiplicando por $(k-1)$

$$\sum_{u=0}^k \binom{k}{u} \cdot (p_t^u \cdot (1-p_t)^{k-u}) \cdot (k-1) = (k-1)$$

Separando o ultimo termo, quando $u = k$

$$\binom{k}{k} \cdot (p_t^k \cdot (1-p_t)^{k-k}) \cdot (k-1) + \sum_{u=0}^{k-1} \binom{k}{u} \cdot (p_t^u \cdot (1-p_t)^{k-u}) \cdot (k-1) = (k-1)$$

$$p_t^k \cdot (k-1) + S_1 = (k-1)$$

Isolando S_1

$$S_1 = (k-1) - p_t^k \cdot (k-1) = (k-1) \cdot (1-p_t^k)$$

Resolvendo S_2 , temos que:

$$\sum_{u=0}^k \binom{k}{u} \cdot (p_t^u \cdot (1-p_t)^{k-u}) \cdot (u) = k \cdot p_t$$

Separando o ultimo termo, quando $u = k$

$$\binom{k}{k} \cdot (p_t^k \cdot (1-p_t)^{k-k}) \cdot (k) + \sum_{u=0}^{k-1} \binom{k}{u} \cdot (p_t^u \cdot (1-p_t)^{k-u}) \cdot (u) = k \cdot p_t$$

$$p_t^k \cdot k + S_2 = k \cdot p_t$$

Isolando S_2

$$S_2 = k \cdot p_t - k \cdot p_t^k = k \cdot (p_t - p_t^k)$$

Juntando tudo:

$$\overline{f_t} = m \cdot k \cdot p_t^k + \sum_{i=1}^m (S_1 - S_2) = m \cdot (k \cdot p_t^k + S_1 - S_2)$$

$$S_1 = (k-1) - p_t^k \cdot (k-1) = (k-1) - k \cdot p_t^k + p_t^k$$

$$S_2 = k \cdot p_t - k \cdot p_t^k$$

$$\overline{f_t} = m \cdot (k \cdot p_t^k + (k-1) - k \cdot p_t^k + p_t^k - k \cdot p_t + k \cdot p_t^k)$$

$$\overline{f_t} = m \cdot ((k+1) \cdot p_t^k - k \cdot p_t + (k-1))$$

Para confirmar equacione no Wolfram

$$\begin{aligned} \text{sum}[i = 1..m] & \left(C(k, k) * p^k * k + \text{sum}[u \right. \\ & \left. = 0..k-1] \left(C(k, u) * (p^u * (1-p)^{k-u}) * (k-1-u) \right) \right) \end{aligned}$$

Encontrando a variância.

Adaptando a equação geral para trabalhar com m BB's de tamanho k .

$$\sigma^2 = \sum_{u=0}^l \binom{l}{u} \cdot (p_t^u \cdot (1-p_t)^{l-u}) \cdot (f(u) - \bar{f}_t)^2$$

Definiremos primeiramente a média(\bar{f}_{ti}) e a variância(σ_i^2) do i -esimo BB. $i \in \{1 \dots m\}$

$$\bar{f}_{ti} = \sum_{u=0}^k \binom{k}{u} \cdot (p_t^u \cdot (1-p_t)^{k-u}) \cdot w(u)$$

Por analogia com o anterior, a solução deste fitness médio será:

$$\bar{f}_{ti} = (k+1) \cdot p_t^k - k \cdot p_t + (k-1)$$

A variância parcial é dada por.

$$\sigma_i^2 = \sum_{u=0}^k \binom{k}{u} \cdot (p_t^u \cdot (1-p_t)^{k-u}) \cdot (w(u) - \bar{f}_{ti})^2$$

Deve-se resolver esta variância parcial, e em seguida agrupar em uma variância global. Fiz uma pesquisa, mas não encontrei já pronto uma propriedade da variância que permitia isto. Logo foi feita uma dedução da seguinte propriedade abaixo e cuja prova se encontra na próxima seção. Onde a propriedade vale apenas quando a cardinalidade dos conjuntos das amostras de cada BB é igual. Ou seja a cardinalidade do conjunto que gerou σ_i^2 e μ_i^2 para todo i tem que ser igual. Em nosso caso isto é verdade pois trabalhamos com uma população de n indivíduos que amostram m BB, logo o numero de amostras de cada BB será n .

$$\sigma^2 + \mu^2 = \sum_{i=1}^m \frac{\sigma_i^2 + \mu_i^2}{m}$$

Por esta propriedade vemos que se tivermos a σ_i^2 e μ_i e μ podemos encontrar σ^2 . Logo resolvendo as partes podemos compor tudo por esta propriedade. Para verificar se a propriedade é valida e/ou cometeu-se erros na sua dedução, a propriedade foi verificada com dados empíricos aleatórios. E sua igualdade sempre foi satisfeita.

Aqui não será tera prosseguimento para encontrar a equação da variância da trap, pois os cálculos estão ficando grandes. Com possibilidade de erros e se encontrada uma equação muito complicada no final tudo se torna pouco util. Como já foi visto o próprio fitness médio já resultou em uma equação grande que sugere uma equação diferencial depois de difícil solução.

Obs: Foi verificado posteriormente que esta teoria funciona apenas para um AG simples que não trabalha com o conhecimento dos BB's, se utilizarmos um EDA que descobre os BB's esta teoria não pode ser aplicada.

11. Prova da Composição das Variâncias

Aqui será encontrada e provada uma propriedade para a variância que permite definir a variância de um conjunto a partir dos valores de variância e média das partições do conjunto. Mais especificamente a propriedade vale apenas para compor partições que tenham o mesmo tamanho.

Então vamos definir que existem l partições chamadas de a_j em um conjunto A onde todas as partições têm cardinalidade n

$$|a_j| = n \quad \forall j$$

$$A = a_1 \cup \dots \cup a_j \cup \dots \cup a_l$$

A cardinalidade do conjunto A será então.

$$|A| = \sum_{j=1}^l |a_j| = l \cdot n$$

A média do conjunto A pode ser definida facilmente através da composição das partes veja.

$$\mu = \frac{1}{|A|} \cdot \sum_{j=1}^l |a_j| \cdot \mu_j = \frac{1}{l \cdot n} \cdot \sum_{j=1}^l n \cdot \mu_j = \frac{1}{l} \cdot \sum_{j=1}^l \mu_j$$

Onde μ_j representa a média da partição j .

Iniciando a demonstração...

Dada a definição da variância do conjunto A (conjunto união das partições). Onde μ é a média do conjunto A e y_i são os elementos do conjunto A (Obs: o elemento y_i pertence a uma e somente uma das partições).

$$\sigma^2 = \frac{1}{|A|} \cdot \sum_{i \in A} (y_i - \mu)^2$$

Desenvolvendo e separando os somatórios.

$$\sigma^2 = \frac{1}{l \cdot n} \cdot \sum_{i \in A} (y_i^2 - 2 \cdot y_i \cdot \mu + \mu^2)$$

$$\sigma^2 = \frac{1}{l \cdot n} \cdot \left(\sum_{i \in A} (y_i^2) - 2 \cdot \mu \cdot \sum_{i \in A} (y_i) + \sum_{i \in A} (\mu^2) \right)$$

Seja μ definido como abaixo e substituindo.

$$\mu = \frac{1}{|A|} \cdot \sum_{i \in A} y_i = \frac{1}{l \cdot n} \cdot \sum_{i \in A} y_i = \frac{1}{l \cdot n} \cdot \sum_{j=1}^l \sum_{i \in a_j} y_i = \frac{1}{l \cdot n} \cdot \sum_{j=1}^l n \cdot \mu_j$$

$$\sigma^2 = \frac{1}{l \cdot n} \cdot \left(\sum_{i \in A} (y_i^2) - 2 \cdot \left(\frac{1}{l \cdot n} \cdot \sum_{j=1}^l n \cdot \mu_j \right) \cdot \sum_{i \in A} (y_i) + l \cdot n \cdot \mu^2 \right)$$

Redefinindo como abaixo e substituindo.

$$\sum_{i \in A} (y_i) = \sum_{j=1}^l \left(\sum_{i \in a_j} y_i \right) = \sum_{j=1}^l (n \cdot \mu_j)$$

$$\sigma^2 = \frac{1}{l \cdot n} \cdot \left(\sum_{j=1}^l \sum_{i \in a_j} (y_i^2) - 2 \cdot \left(\frac{1}{l \cdot n} \cdot \sum_{j=1}^l n \cdot \mu_j \right) \cdot \sum_{j=1}^l (n \cdot \mu_j) + l \cdot n \cdot \mu^2 \right)$$

Juntando os dois termos iguais.

$$\sigma^2 = \frac{1}{l \cdot n} \cdot \left(\sum_{j=1}^l \sum_{i \in a_j} (y_i^2) - \frac{2}{l \cdot n} \cdot \left(\sum_{j=1}^l n \cdot \mu_j \right)^2 + l \cdot n \cdot \mu^2 \right)$$

Substituindo a definição da média como soma das partes.

$$\sigma^2 = \frac{1}{l \cdot n} \cdot \left(\sum_{j=1}^l \sum_{i \in a_j} (y_i^2) - \frac{2}{l \cdot n} \cdot \left(\sum_{j=1}^l n \cdot \mu_j \right)^2 + \frac{1}{l \cdot n} \cdot \left(\sum_{j=1}^l n \cdot \mu_j \right)^2 \right)$$

Juntando o segundo termo com o terceiro, pois possuem partes semelhantes.

$$\sigma^2 = \frac{1}{l \cdot n} \cdot \left(\sum_{j=1}^l \sum_{i \in a_j} (y_i^2) - \frac{1}{l \cdot n} \cdot \left(\sum_{j=1}^l n \cdot \mu_j \right)^2 \right)$$

Adicionando informação ao primeiro termo para moldar um significado de variância das partições.

$$\sigma^2 = \frac{1}{l \cdot n} \cdot \left(\sum_{j=1}^l \sum_{i \in a_j} (y_i - \mu_j)^2 - \frac{1}{l \cdot n} \cdot \left(\sum_{j=1}^l n \cdot \mu_j \right)^2 + \sum_{j=1}^l \sum_{i \in a_j} (2 \cdot \mu_j \cdot y_i) - \sum_{j=1}^l \sum_{i \in a_j} (\mu_j^2) \right)$$

Desenvolvendo 1º, 3º e 4º termo.

$$\sigma^2 = \frac{1}{l \cdot n} \cdot \left(\sum_{j=1}^l n \cdot \sigma_j^2 - \frac{1}{l \cdot n} \cdot \left(\sum_{j=1}^l n \cdot \mu_j \right)^2 + \sum_{j=1}^l 2 \cdot n \cdot \mu_j^2 - \sum_{j=1}^l n \cdot \mu_j^2 \right)$$

Unindo 3º e 4º termo

$$\sigma^2 = \frac{1}{l \cdot n} \cdot \left(\sum_{j=1}^l n \cdot \sigma_j^2 - \frac{1}{l \cdot n} \cdot \left(\sum_{j=1}^l n \cdot \mu_j \right)^2 + \sum_{j=1}^l n \cdot \mu_j^2 \right)$$

Cortando o n em todos os termos.

$$\sigma^2 = \frac{1}{l} \cdot \left(\sum_{j=1}^l \sigma_j^2 - \frac{1}{l} \cdot \left(\sum_{j=1}^l \mu_j \right)^2 + \sum_{j=1}^l \mu_j^2 \right)$$

Reordenando os termos.

$$\sigma^2 = \frac{1}{l} \cdot \left(\sum_{j=1}^l \sigma_j^2 + \sum_{j=1}^l \mu_j^2 - \frac{1}{l} \cdot \left(\sum_{j=1}^l \mu_j \right)^2 \right)$$

Substituindo definição abaixo no 3º termo.

$$\mu \cdot l = \sum_{j=1}^l \mu_j$$

$$\sigma^2 = \frac{1}{l} \cdot \left(\sum_{j=1}^l \sigma_j^2 + \sum_{j=1}^l \mu_j^2 - l \cdot \mu^2 \right)$$

Simplificando temos a propriedade desejada.

$$\sigma^2 + \mu^2 = \sum_{j=1}^l \frac{\sigma_j^2 + \mu_j^2}{l}$$

12. Formula Geral para Encontrar o Fitness Médio - ECGA simples

Dado um problema qualquer cuja função de avaliação $f(\cdot)$ dependa apenas do número de bits configurados com 1 e cuja propagação da informação se dá através de BB's que não se alteram ao longo das gerações, definiremos então:

p_{tu}	Probabilidade de cada BB de assumir a u -esima configuração.
l	Tamanho do problema.
k	Tamanho do BB.
m	Número de BB's.
$f(\cdot)$	Função de avaliação.
u	É a soma dos bits do BB.

$$f(\cdot) = \sum_{i=1}^m w(u)$$

O fitness médio pode ser encontrado somando todos os fitness possíveis ponderados pela sua probabilidade de ocorrer, logo:

$$\bar{f}_t = \sum_{i=1}^m \sum_{u=0}^k p_{tu} \cdot w(u)$$

$$\sum_{u=0}^k p_{tu} = 1$$

Obs: Aqui não foi tirado um tempo para tentar encontrar uma definição geral para a variância, deixando isto em aberto.

12.1. Problema Trap

Para a função trap podemos simplificar como o problema da Ruína do Jogador e considerando apenas blocos que convergiram para o ótimo global e o ótimo local.

$$\bar{f}_t = \sum_{i=1}^m (p_{t0} \cdot (k-1) + p_{tk} \cdot k)$$

Onde teremos que:

$$p_{t0} + p_{tk} = 1$$

$$p_{t0} = 1 - p_{tk}$$

Substituído temos.

$$\bar{f}_t = \sum_{i=1}^m ((1 - p_{tk}) \cdot (k-1) + p_{tk} \cdot k)$$

$$\bar{f}_t = \sum_{i=1}^m (k-1 + p_{tk})$$

Chamando $p_{tk} = p_t$, probabilidade do sucesso para o ótimo global do BB.

$$\bar{f}_t = \sum_{i=1}^m (k-1 + p_t)$$

$$\bar{f}_t = m \cdot (k-1 + p_t)$$

Obs: Não foi encontrada uma prova para a variância, mas por análise empírica acreditasse que está tenha valor igual a:

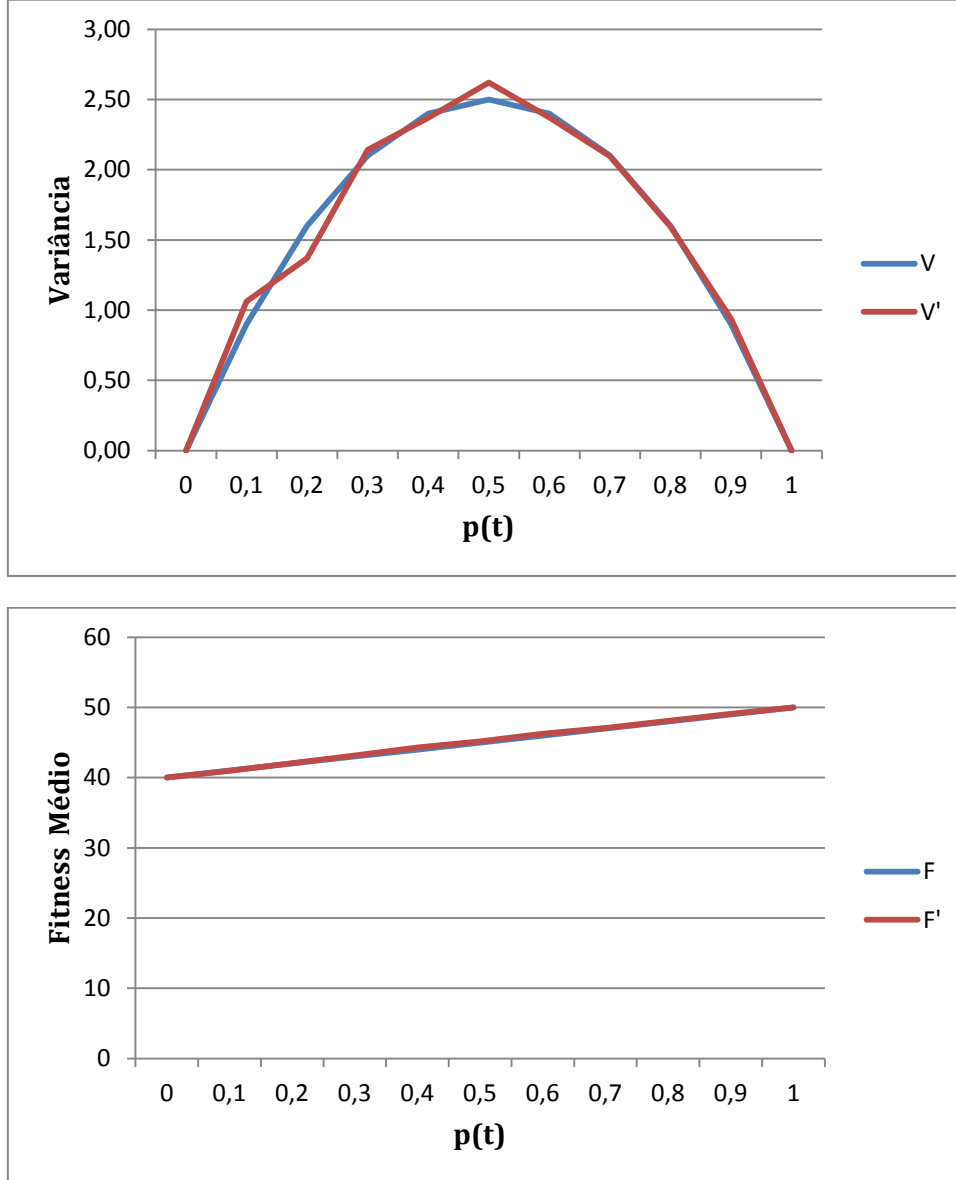
$$\sigma^2 = m \cdot p_t \cdot (1 - p_t)$$

Logo temos:

$$\bar{f}_t = m \cdot (k-1 + p_t)$$

$$\sigma^2 = m \cdot p_t \cdot (1 - p_t)$$

Estas equações foram verificadas experimentalmente, utilizaram-se soluções ($n = 100$) aleatoriamente geradas para ($k = 5$) e ($m = 10$) onde tais soluções já tinham seus BB's gerados como ótimo local ($0_1 \dots 0_k$) com probabilidade $(1 - p_t)$ ou ótimo global ($1_1 \dots 1_k$) com probabilidade p_t . Os gráficos abaixo mostram os valores de fitness médio e variância com valores da equação acima (F, V) e experimentais (F', V').



12.2. Dedução número de gerações para Trap utilizando seleção por truncamento.

$$\overline{f_{t+1}} - \overline{f_t} = (\alpha + \varepsilon) \cdot \sigma_t$$

$$\overline{f_t} = m \cdot (k - 1 + p_t)$$

$$\sigma^2 = m \cdot p_t \cdot (1 - p_t)$$

$$m \cdot (k - 1 + p_{t+1}) - m \cdot (k - 1 + p_t) = (\alpha + \varepsilon) \cdot \sqrt{p_t \cdot (1 - p_t)}$$

$$p_{t+1} - p_t = \frac{(\alpha + \varepsilon)}{\sqrt{m}} \cdot \sqrt{p_t \cdot (1 - p_t)}$$

$$\frac{\partial p_t}{\partial t} = \frac{(\alpha + \varepsilon)}{\sqrt{m}} \cdot \sqrt{p_t \cdot (1 - p_t)}$$

$$\int_{p_1}^{p_g} \frac{\partial p_t}{\sqrt{p_t \cdot (1 - p_t)}} = \frac{(\alpha + \varepsilon)}{\sqrt{m}} \cdot \int_1^g \partial t$$

Temos que

$$\int \frac{dx}{\sqrt{ax^2 + bx + c}} = -\frac{1}{\sqrt{-a}} \cdot \arcsen\left(\frac{2ax + b}{\sqrt{b^2 - 4ac}}\right) \quad \text{para } a < 0 \text{ e } 4ac - b^2 < 0$$

Logo

$$\left[-\arcsen\left(\frac{-2 \cdot p_t + 1}{\sqrt{1}}\right) \right]_{p_1}^{p_g} = \frac{(\alpha + \varepsilon)}{\sqrt{m}} \cdot [t]_1^g$$

$$\arcsen(1 - 2 \cdot p_1) - \arcsen(1 - 2 \cdot p_g) = \frac{(\alpha + \varepsilon)}{\sqrt{m}} \cdot (g - 1)$$

Utilizando

$$p_1 = \frac{1}{2}$$

$$p_g = 1 - \frac{1}{n}$$

$$\frac{(\alpha + \varepsilon)}{\sqrt{m}} \cdot (g - 1) = \arcsen(1 - 2 \cdot p_1) - \arcsen(1 - 2 \cdot p_g)$$

$$\frac{(\alpha + \varepsilon)}{\sqrt{m}} \cdot (g - 1) = \arcsen(0) - \arcsen\left(\frac{2}{n} - 1\right)$$

$$\arcsen\left(\frac{2}{n} - 1\right) \approx \frac{-\pi}{2}$$

$$g = 1 + \frac{\sqrt{m}}{(\alpha + \varepsilon)} \cdot \frac{\pi}{2}$$

O problema UmMax seria:

$$g = 1 + \frac{\sqrt{l}}{(\alpha + \varepsilon)} \cdot \frac{\pi}{2}$$

Veremos que por termos simplificado o problema considerando que os blocos já convergiram para os esquemas ótimo e sub-ótimos a equação que aproxima-se mais dos resultados experimentais é a equação do problema UmMax.

13. Analisando a sensibilidade nos parâmetros

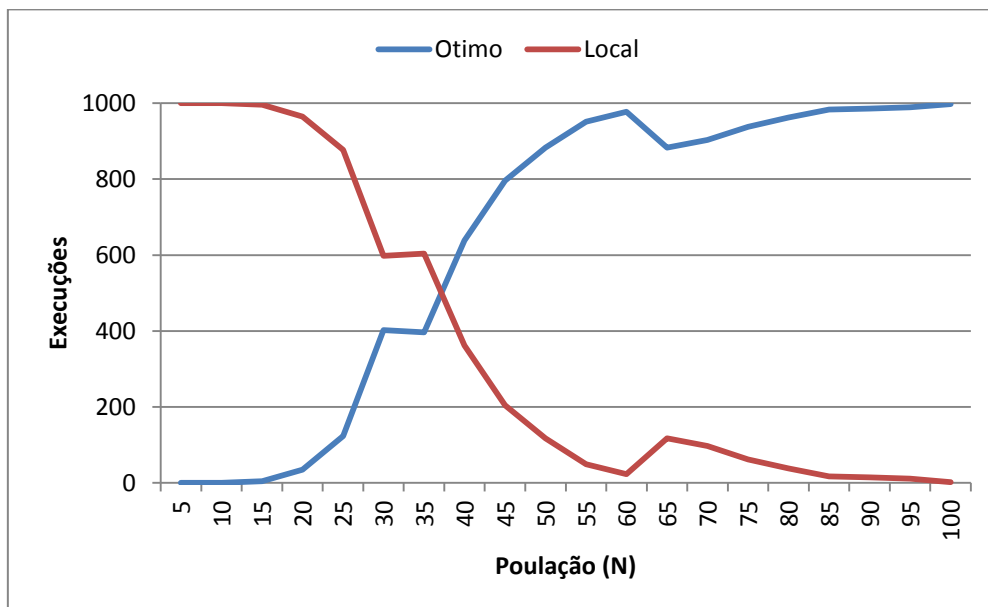
Para ser mais criterioso ao avaliar os experimentos aqui serão executados testes computacionais com 1000 execuções e tirados média, máximo, mínimo e desvio padrão para comparar com as equações.

13.1. Problema UmMax com seleção truncamento

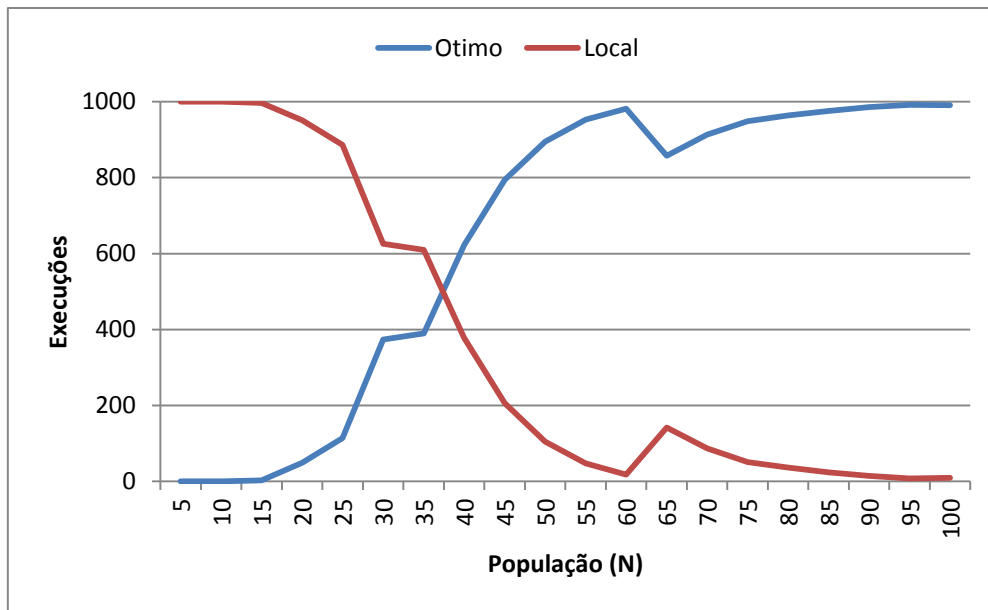
$$g = 1 + \frac{\sqrt{l}}{(\alpha + \varepsilon)} \cdot \frac{\pi}{2} \quad \text{com} \quad \varepsilon = 0.5 \quad (Eq)$$

Sensibilidade do tamanho da população (N)

Aqui se utilizou seleção por truncamento com $l = 64$ e $\alpha = 0 \rightarrow Z(\%) = 50\%$. Abaixo o gráfico do número de sucessos em encontrar o ótimo global do problema. Para uma população maior que 100 o problema sempre é resolvido no ótimo.



Pelo fato de este gráfico ter mostrado-se estranho, visto que contém um padrão não previsto da curva a medida que temos uma população maior. Foi levantada algumas hipóteses da causa: Primeiro, o código pode conter erros e isto leva ao mesmo padrão independente de quantas vezes se execute mesmo partindo de soluções aleatorias. Segundo, o problema reside na geração dos números aleatórios do C (função rand pode não ser aleatória o suficiente). Terceiro este é um padrão de fato do algoritmo visto que este algoritmo é um método iterativo e que depende fortemente semente inicial (fractais possuem esta característica se sensibilidade a parâmetros iniciais). No código não foi encontrado nada que pudesse causar este padrão (mas é passível de que exista erros no código, o ideal seria refazer um AG simples e verificar – isto seria rápido de fazer, mas não foi feito). Para a segunda hipótese foi utilizado um gerador de números aleatórios realmente confiável descrito no livro (O'Reilly book : Secure Programming Cookbook for C and C++). Mas o padrão se repetiu com podemos ver abaixo, foram feito mais testes mas o padrão sempre se repete com bastante precisão.



Não satisfeito foi melhor discretizado o parâmetro do tamanho da população que aqui vai de 5 até 100 num passo de 5 em 5. O gráfico abaixo mostra o range de 10 até 90 indo de 1 em 1 (foi utilizado a rand do C pois não mostrou melhora com o gerador novo e este demorava muito para executar). Vemos que é exatamente o mesmo padrão, apenas agora tem mais detalhes que eram omitidos no padrão anterior mas se for pego apenas os pontos do padrão anterior neste teste ficará uma imagem muito semelhante a anterior. Então será aceita como provável a 3ª hipótese é possível estudar teoricamente as iterações do método e chegar em uma equação iterativa que informe se dada um tamanho de população e indivíduos gerados aleatoriamente qual a probabilidade de convergir para o ótimo ou não. Esta equação iterativa deveria expressar este padrão, mas isto pode não ser possível ou ser difícil de provar.

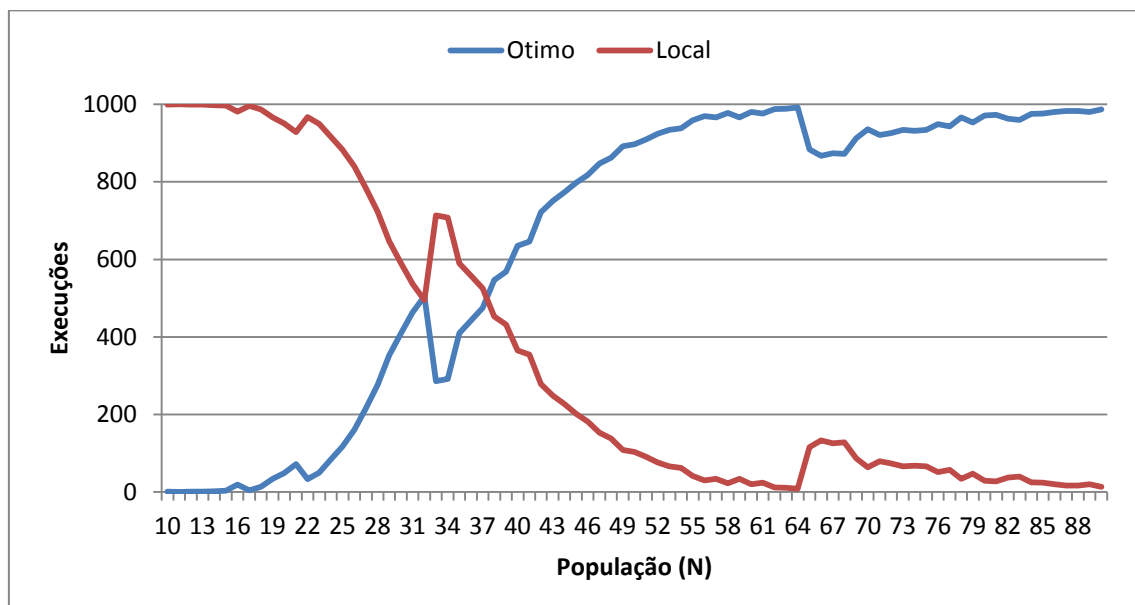


Gráfico do número de gerações médias até a convergência para os casos de convergência para o ótimo e o local. Vemos que após certo tamanho da população o número de gerações se torna constante e o tempo de convergência nos dois casos se torna próximo. A linha tracejada representa o que é previsto pela equação teórica.

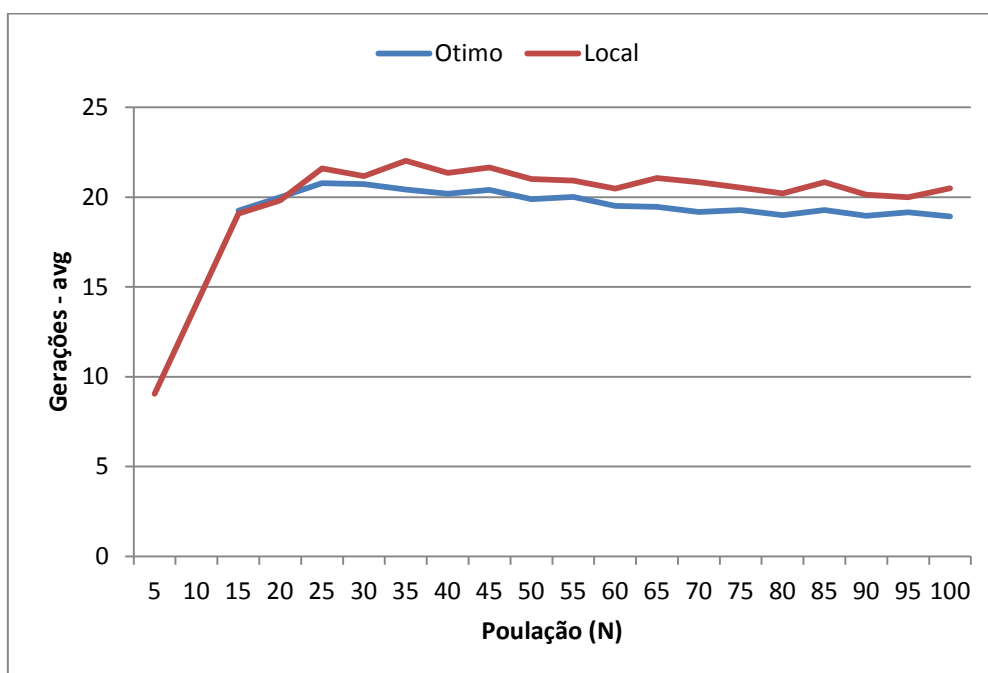
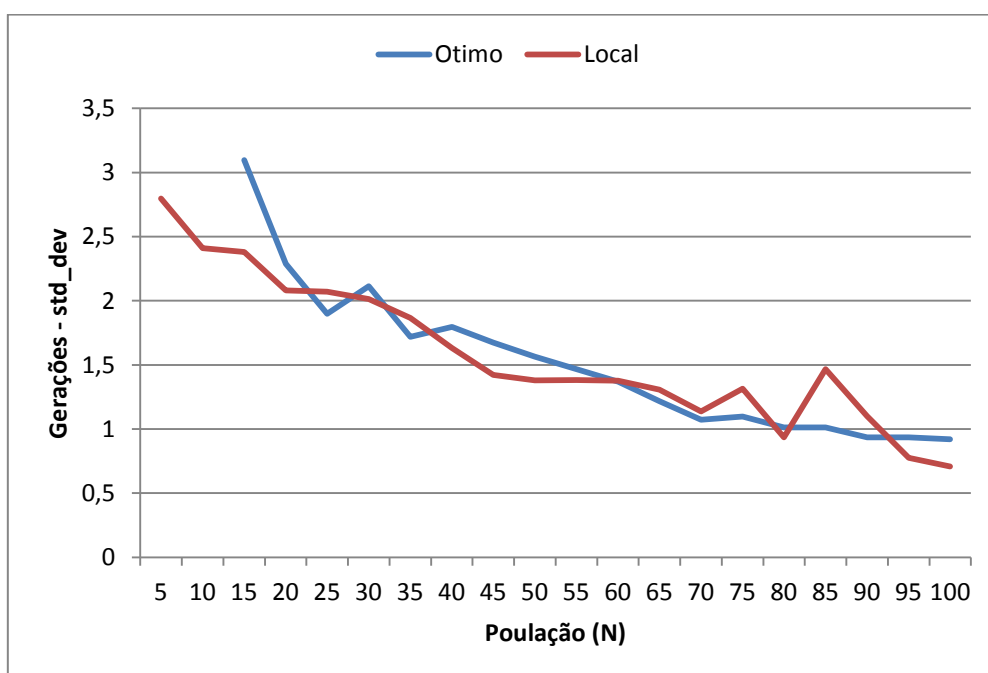
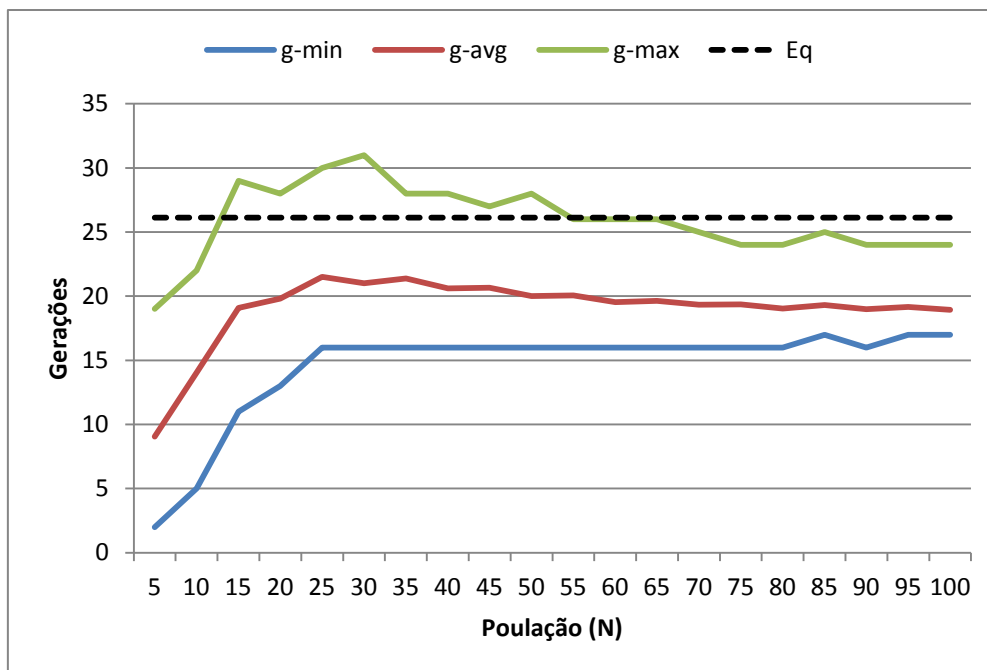


Gráfico do desvio padrão nas gerações médias. Vemos que quanto maior a população mais estável é o número de gerações percebemos que existe um desvio abaixo de 3 gerações o que mostra uma grande estabilidade do método.



No gráfico baixo temos o numero de gerações mínimo (g-min), médio (g-avg) e máximo (g-max). Vemos que após um dado tamanho de população todos os valores ficam compreendidos entre 15 e 25.



Sensibilidade da taxa de truncamento (Z%)

Aqui se utilizou seleção por truncamento com $l = 64$ e $n = 100$.

Gráfico do número de sucessos em encontrar o ótimo global do problema. Vemos que altas taxas causam a deriva do método e rapidamente o método não consegue mais nenhuma solução ótima.

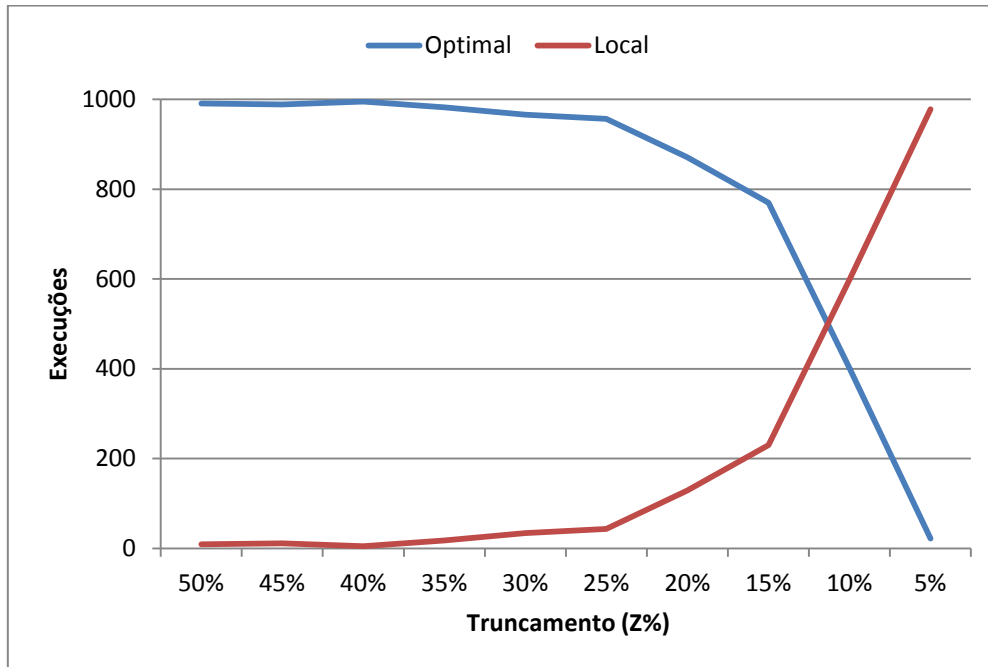
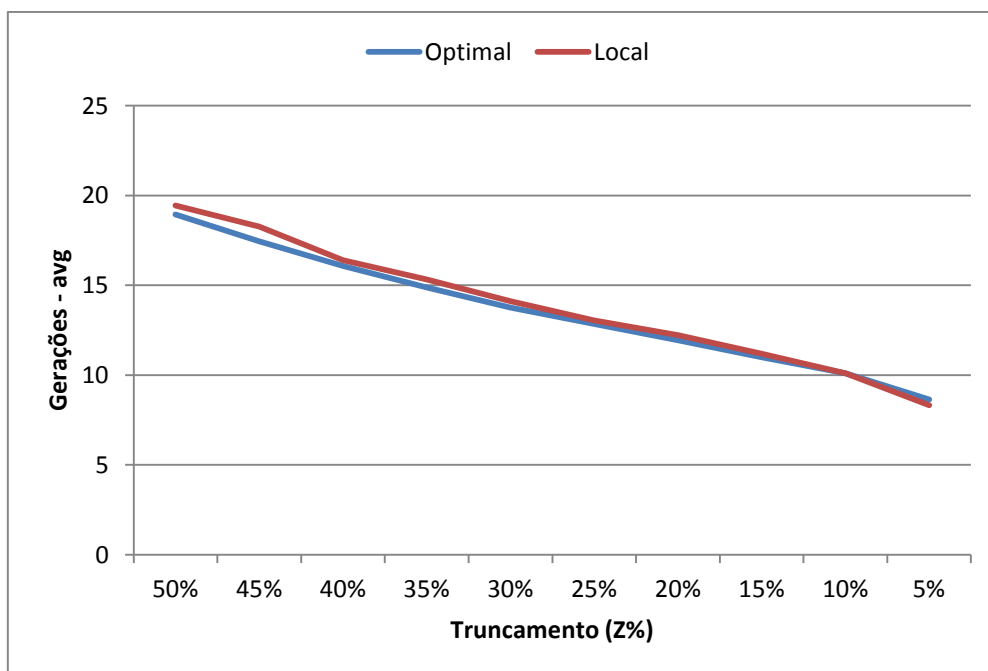
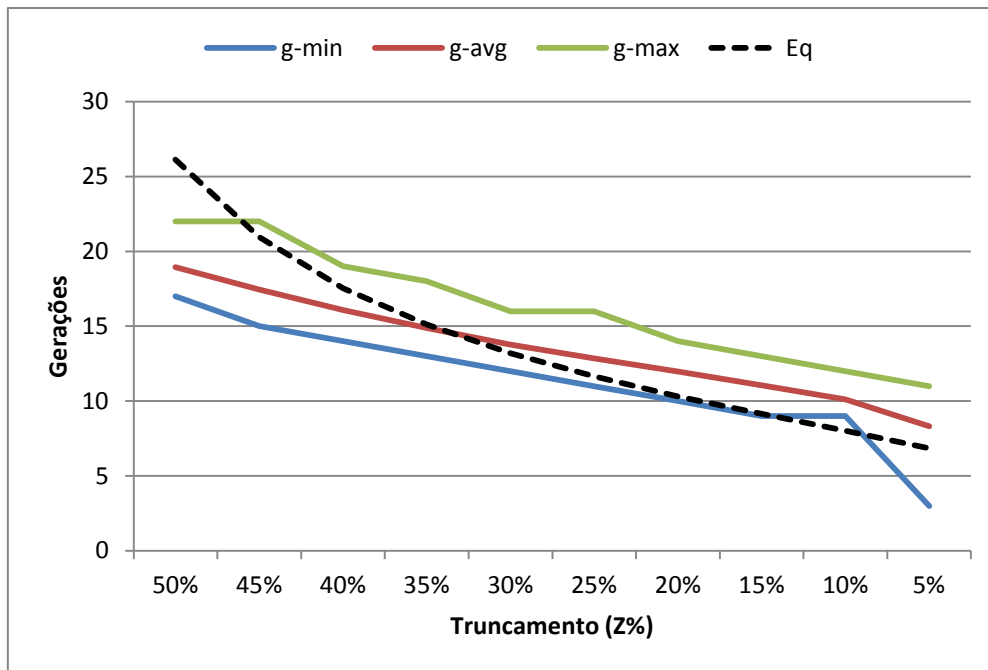
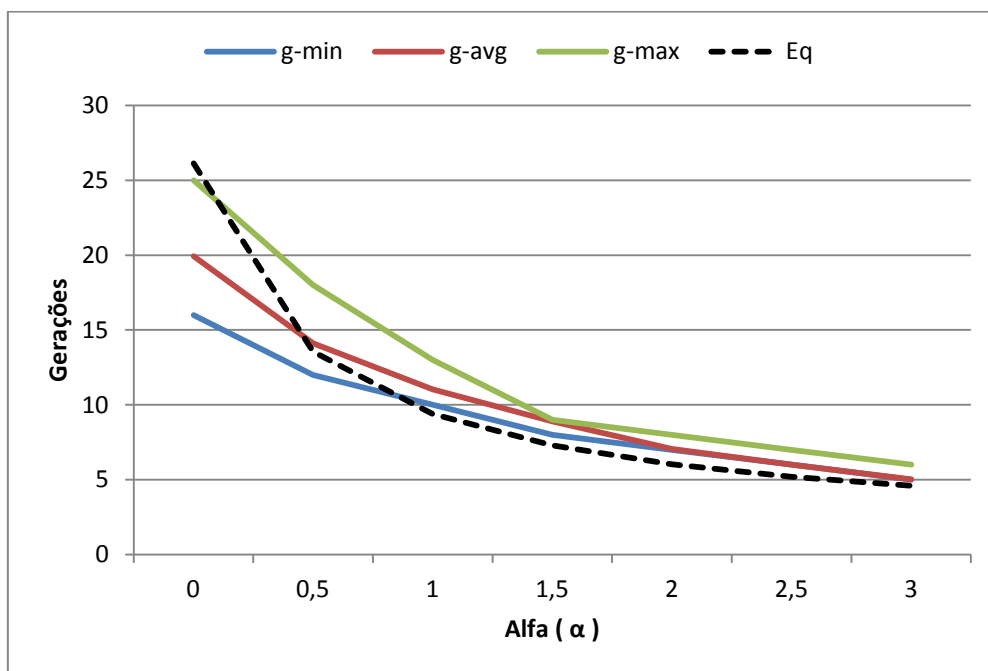


Gráfico do número de gerações médias até a convergência para os casos de convergência para o ótimo e o local. Vemos que decresce linearmente junto com a taxa de truncamento e todos os desvios padrões ficaram abaixo de 1 geração mostrando estabilidade.



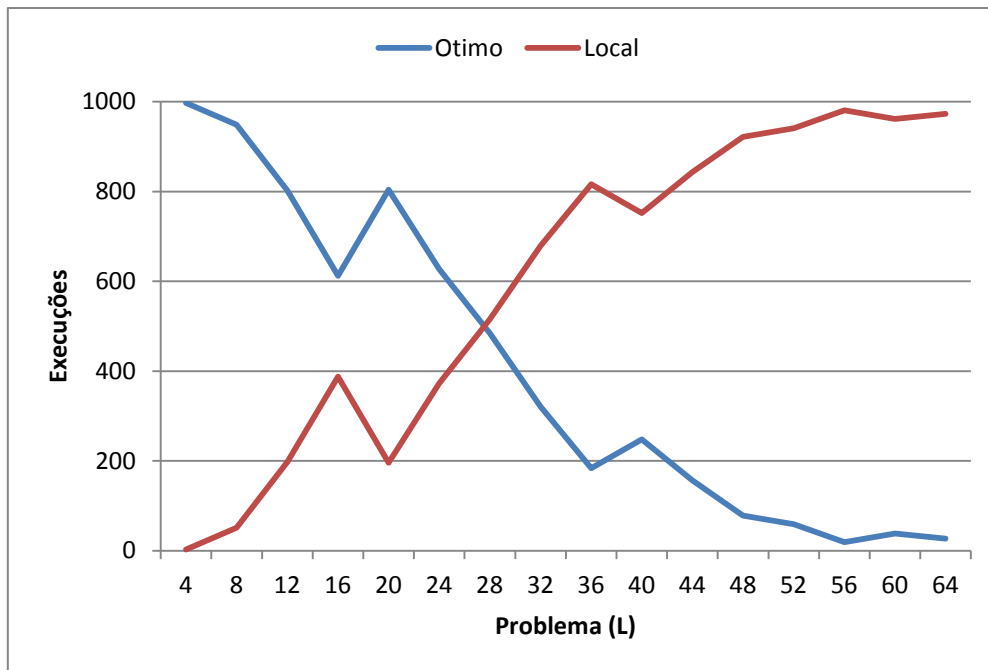


$$n = 25/Z \text{ e } l = 64$$

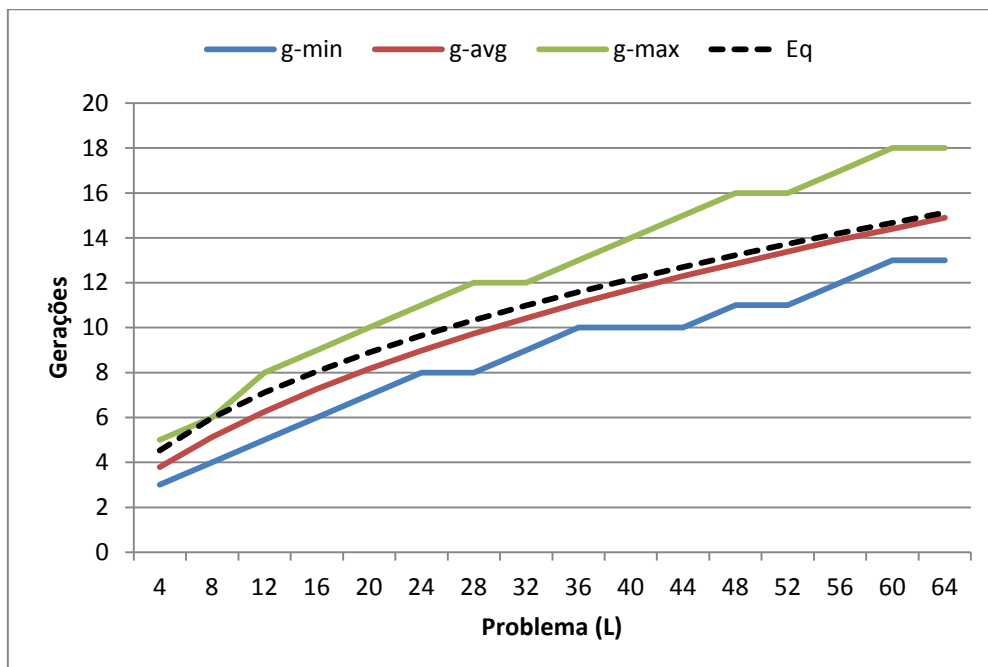


Sensibilidade do tamanho do problema (L)

Primeiramente contaram-se os sucessos com $n = 20$ e $\alpha = 0.39 \rightarrow Z(\%) = 35\%$. No gráfico baixo vemos que à medida que o problema aumenta o tamanho da população de 20 não é suficiente para resolver problema que está crescendo.



Para analisar o número de gerações utilizou-se uma população maior para ter uma maior estabilidade e encontrar mais soluções ótimas para instâncias grandes. Logo foi utilizado $n = 100$ e $\alpha = 0.39 \rightarrow Z(\%) = 35\%$. No gráfico baixo vemos que à medida que o problema aumenta a previsão da equação encontra com bastante precisão o número de gerações. Não está mostrado aqui mais o desvio padrão ficou abaixo de 0.75 gerações o que mostra estabilidade do AG.



13.2. Problema Trap com seleção truncamento

Para a Trap descobriu-se $\varepsilon = 0.35$

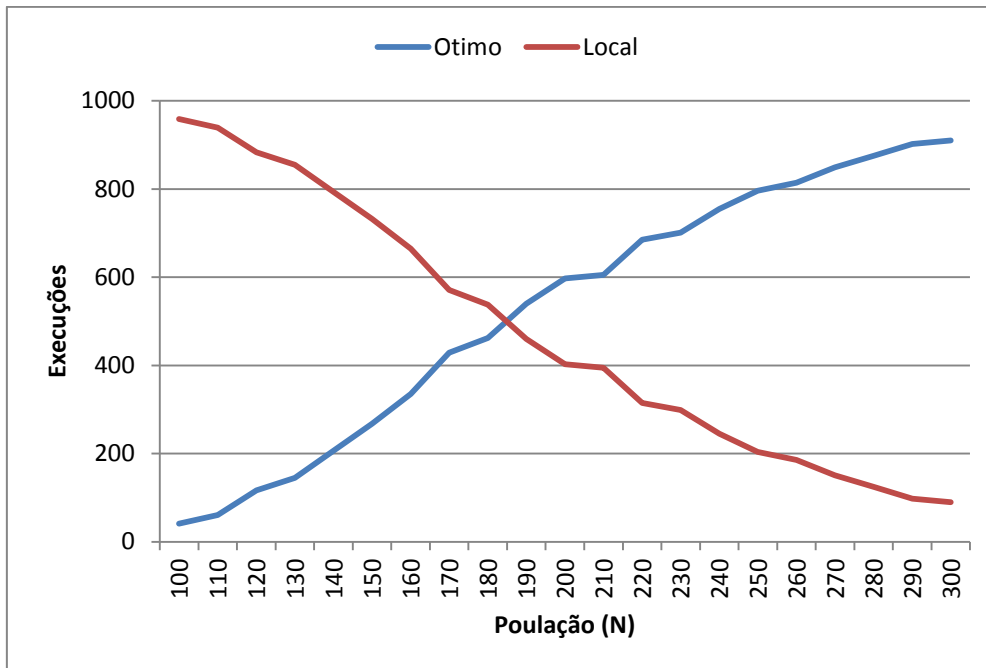
$$g = 1 + \frac{\sqrt{l}}{(\alpha + \varepsilon)} \cdot \frac{\pi}{2} \quad \text{com} \quad \varepsilon = 0.35 \quad (Eq)$$

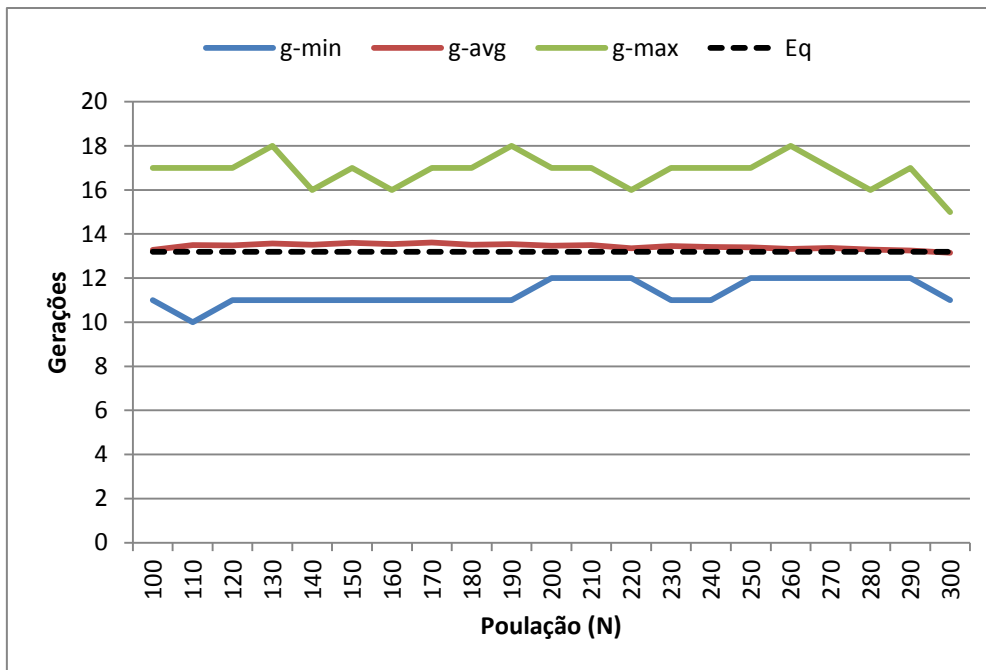
Obs. Veja que a equação que foi utilizada não é a equação da Trap e sim a equação do problema UmMax visto que esta se mostra mais coerente com os dados experimentais. Isto se deve provavelmente ao fato de que na prova da equação da Trap já partia-se da suposição de que os BB's já estavam formados e isto subestimou o tempo de convergencia.

Sensibilidade do tamanho da população (N)

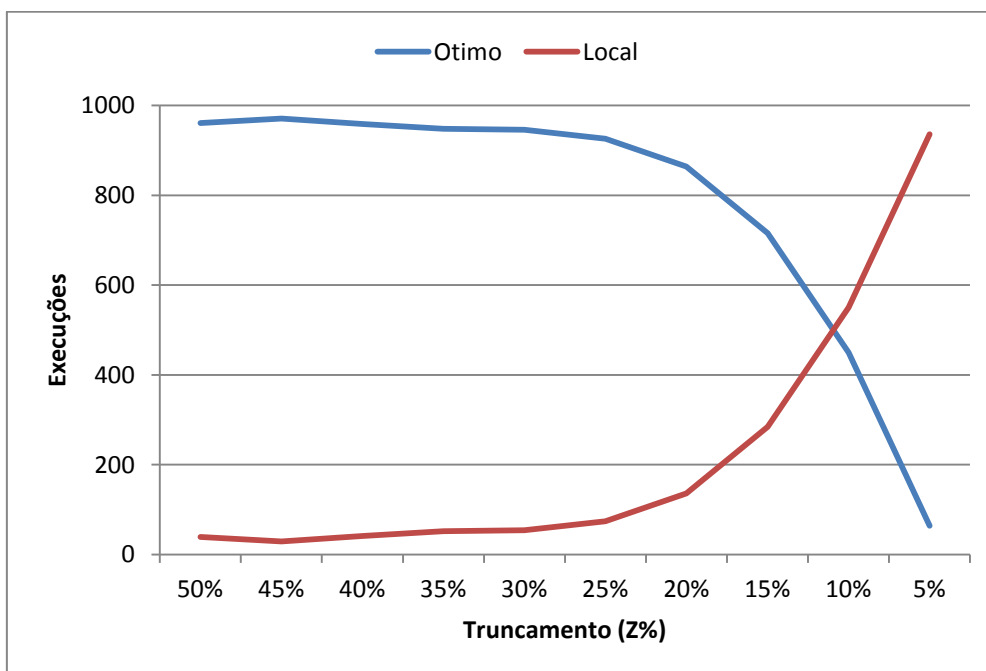
Aqui se utilizou seleção por truncamento com $l = 64$ e $k = 4$ e $m = 16$ e $\alpha = 0.68 \rightarrow Z(\%) = 25\%$. Foram realizados um total de 1000 execuções.

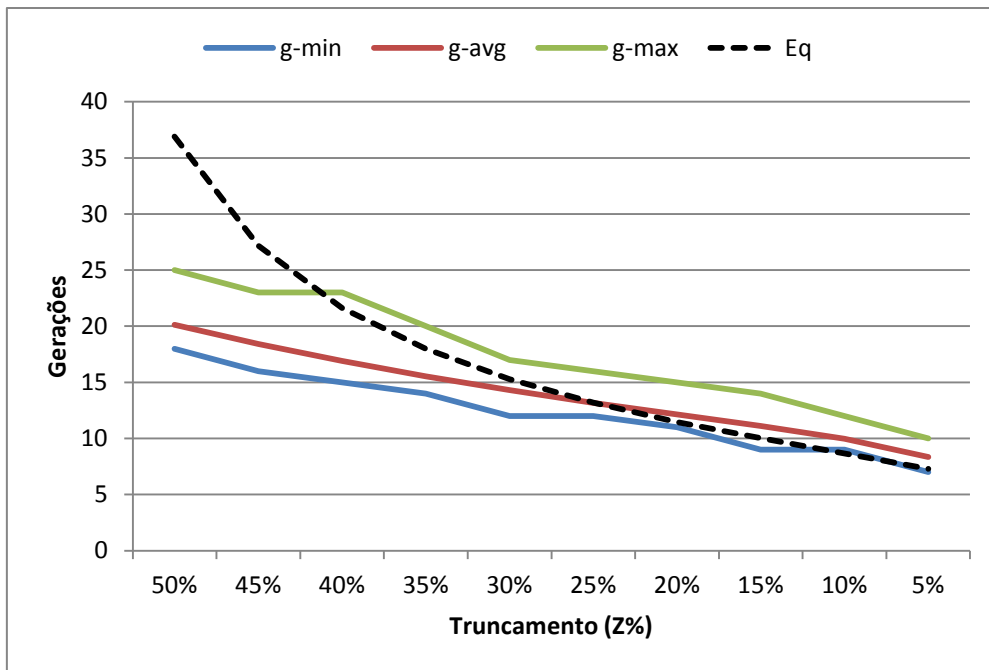
Gráfico do número de sucessos em encontrar o ótimo global do problema. Para uma população maior que 100 o problema sempre é resolvido no ótimo.



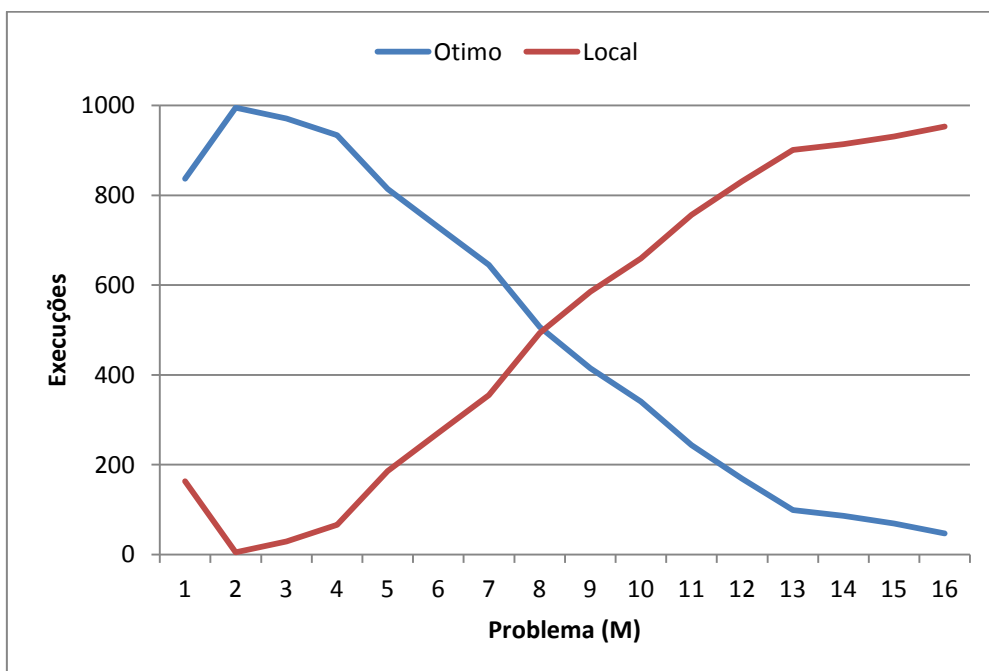


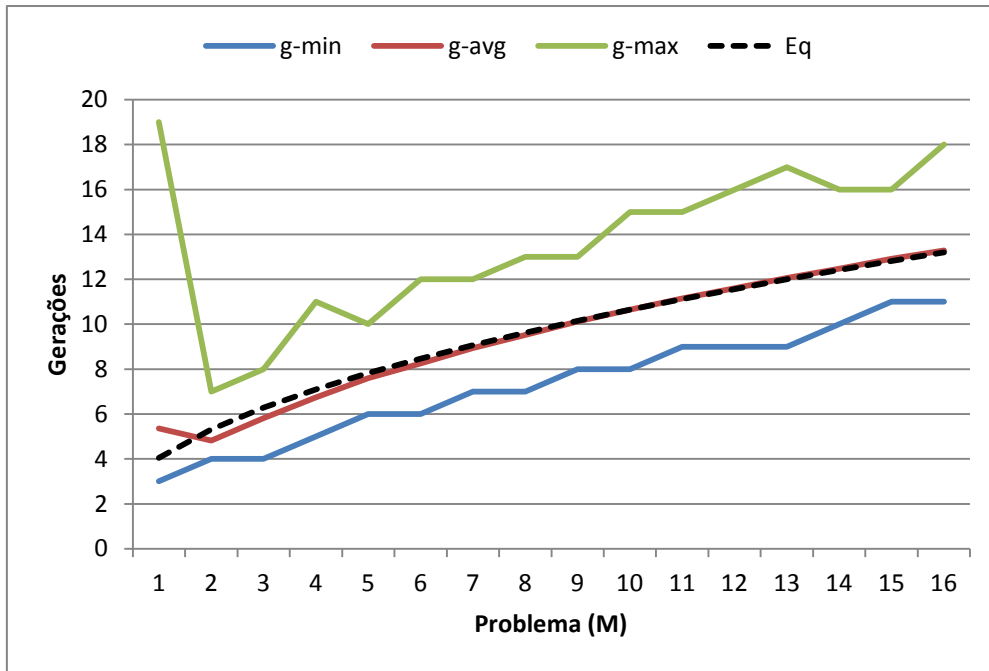
$$n = 64 \text{ e } l = 64 \text{ e } k = 4 \text{ e } m = 16$$



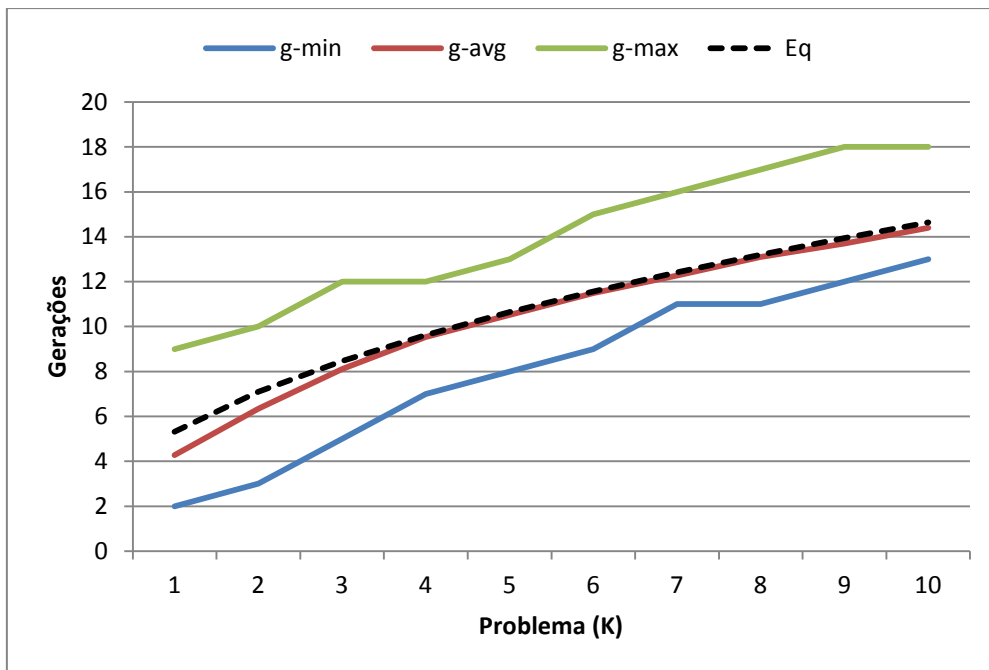


$n = 100$ e $l = k * m$ e $k = 4$ e $\alpha = 0.68 \rightarrow Z(\%) = 25\%$.

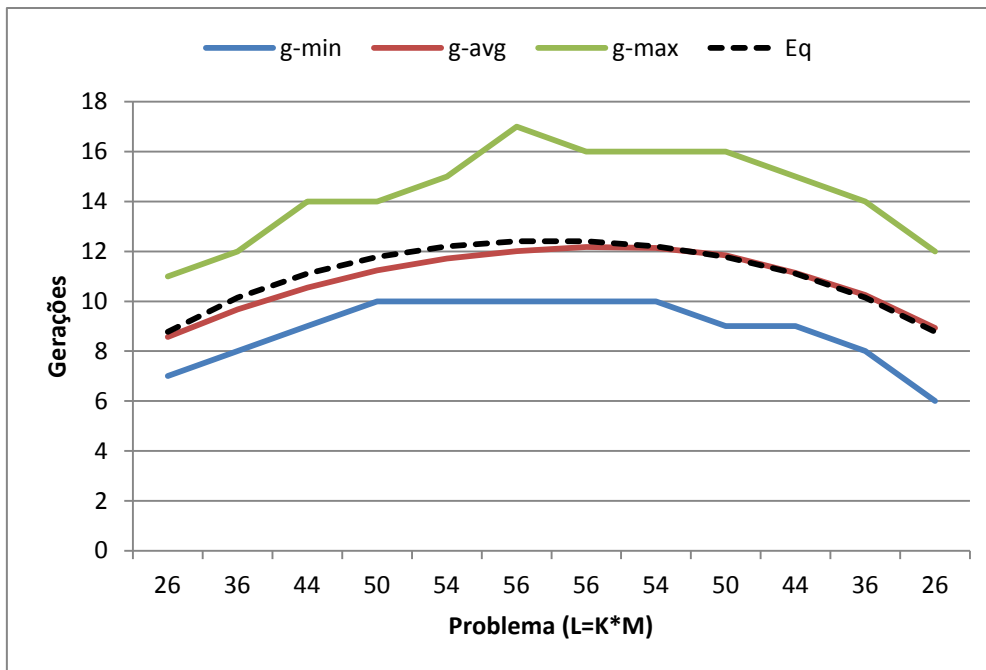




$n = 5 \cdot 2^k$; $l = k \cdot m$; $m = 8$; $\alpha = 0.68 \rightarrow Z(\%) = 25\%$.



$n = 28 \cdot \left(\frac{3}{2}\right)^k$; $l = k \cdot m$; $k + m = 15$; $\alpha = 0.68 \rightarrow Z(\%) = 25\%$.



13.3. Problema BinInt com seleção proporcional

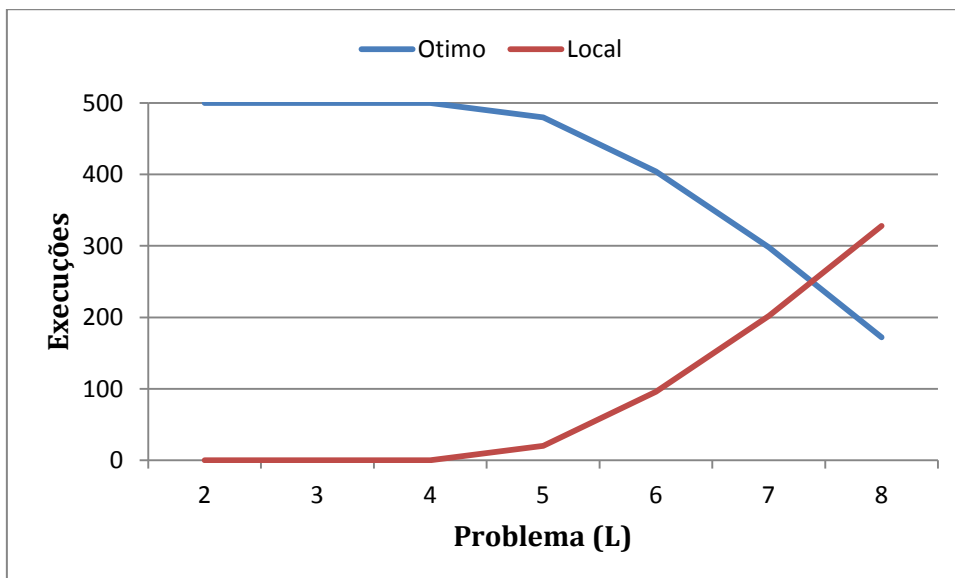
Seleção proporcional

$$g = 3 \cdot (2^{l+1} - \ln|2| \cdot l) - 5 \quad (Eq)$$

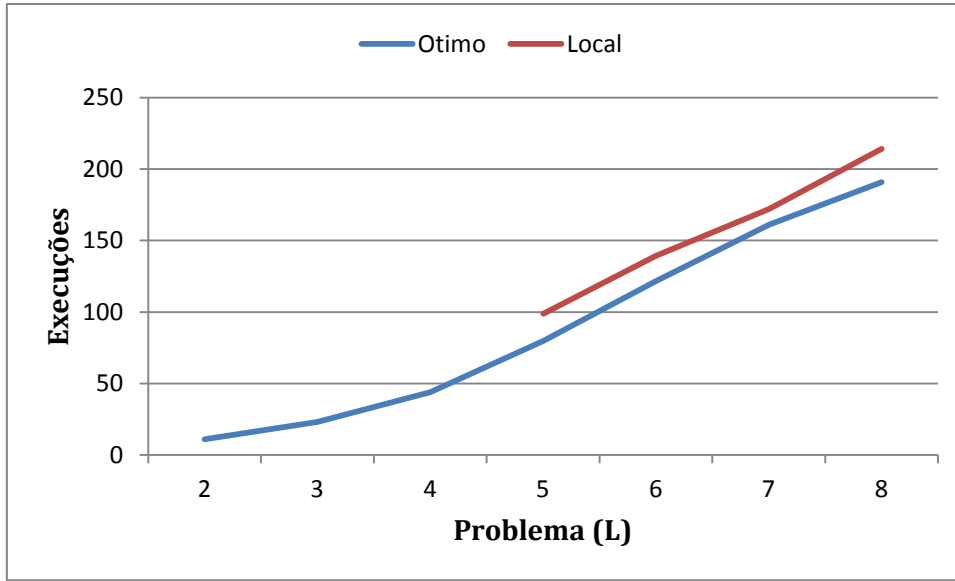
Sensibilidade do tamanho do problema (L)

500 execuções com $n = 200$.

Abaixo temos o numero de ótimos e subótimos encontrados.



Abaixo temos o numero de gerações médias discriminando as execuções que conseguiram ótimos e subótimos.



Obs. Para o problema BinInt fez-se a comparação apenas da sensibilidade do tamanho do problema visto que este é o fator que mais importa.

13.4. Resolver o problema diretamente ou evoluir uma população de soluções?

Voltando ao tópico 1.3 novamente temos a pergunta que agora pode ser melhor respondida para um problema como a trap como exemplo.

Temos então que o método direto é melhor que o um ECGA simples se:

$$m < g \cdot (k \cdot \ln \chi + \ln m)$$

Agora já sabemos o valor do numero de gerações para a trap convergir com truncamento, (equação do UmMax).

$$g = 1 + \frac{\sqrt{l}}{(\alpha + \varepsilon)} \cdot \frac{\pi}{2} \quad \text{com} \quad \varepsilon = 0.35 \quad (Eq)$$

Substituindo temos.

$$m < \left(1 + \frac{\sqrt{l}}{(\alpha + \varepsilon)} \cdot \frac{\pi}{2}\right) \cdot (k \cdot \ln \chi + \ln m)$$

Simplificando aproximadamente ($\ln \chi \approx 0.7$).

$$\sqrt{m} < \frac{\sqrt{k}}{(\alpha + \varepsilon)} \cdot \frac{\pi}{2} \cdot \left(\frac{k \cdot 7}{10} + \ln m\right)$$

Chamando $k = \rho^2$ e $m = \omega^2$

$$\omega < \frac{\rho}{(\alpha + \varepsilon)} \cdot \frac{\pi}{2} \cdot \left(\frac{\rho^2 \cdot 7}{10} + 2 \cdot \ln \omega\right)$$

$$\omega \cdot (\alpha + \varepsilon) < \left(\frac{7\pi \cdot \rho^3}{20} + \pi \cdot \ln \omega \cdot \rho\right)$$

Simplificando aproximadamente ($\frac{7\pi}{20} \approx 1$).

$$\omega \cdot (\alpha + \varepsilon) < (\rho^3 + \pi \cdot \ln \omega \cdot \rho)$$

Visto que ($\pi \cdot \ln \omega = \pi \cdot \ln \sqrt{m} \approx \text{constante} \rightarrow C$)

$$\omega \cdot (\alpha + \varepsilon) < (\rho^3 + C \cdot \rho)$$

Isolando o ω e voltando os termos originais

$$\omega < \frac{(\rho^3 + C \cdot \rho)}{(\alpha + \varepsilon)}$$

$$\sqrt{m} < \sqrt{k} \cdot \frac{(k + C)}{(\alpha + \varepsilon)}$$

$$\alpha = 0.68 \text{ e } \varepsilon = 0.35$$

$$\sqrt{m} < \sqrt{k} \cdot (k + C)$$

Como a constante é positiva o pior caso para o metodo direto é quando ela vale 0.

$$m < k^3$$

Logo o metodo direto é melhor nesta condição acima, ou seja um pouco que cresce o tamanho dos BB's já não compensa utilizar ECGA simples pois o tempo para selecionar os BB's ótimos mais o tempo de convergência da população para o ótimo global já é superior ao tempo de resolver cada BB's exaustivamente e compor tudo em uma solução.

Aqui é claro estamos sendo um pouco simplista pois estamos resolvendo um problema que seja fortemente separavel ou totalmente (para que possamos compor a solução ótima global a partir de suas partes ótimas, BB's). Isto não funcionará para problemas que tenham sobreposição, outro fato é que supomos que temos o modelo que identifica os BB's, poderíamos utilizar uma população aleatória para encontrar isto (assim estaríamos utilizando a mesma tecnica que o ECGA simples para encontrar isto, e como removemos a primeira geração da contagem durante a prova encontraríamos o mesmo resultado).

Assim acredito que é possível criar um ECGA modificado onde temos apenas a primeira geração para gerar o modelo em seguida aplicamos uma busca local (ou busca exaustiva nos BB's montados) conseguindo a solução ótima logo na primeira geração, na aula foi falado que fizeram algo nesta linha, penso que é justificavel pelo que mostro aqui.

Não houve tempo para finalizar o relatório, pois o relatório não convergia e deixou varias coisas em aberto para serem feitas.