

Algoritmos de Estimativa de Distribuição

Cassio Pennachin

LCE – PPGEE – UFMG

Setembro de 2010

Roteiro

- Motivação e Precursores
- Algoritmos para Variáveis Binárias
 - cGA
 - ECGA
 - BOA
- Algoritmos para Variáveis Contínuas
 - Univariados
 - Multivariados
- Referências

Contexto

Algoritmos Genéticos em 1 Minuto

- Como um GA tradicional funciona:
 - Seleção
 - Recombinação (crossover)
 - Mutação
 - Repita até resolver o problema ou desistir

Diferentes operadores em cada caso, mas a pergunta é a mesma: dado o que eu sei sobre as soluções existentes, que soluções devo tentar?

O Problema

- Entrada: soluções conhecidas e sua fitness

ID	Solução	Fitness
1	00100	1
2	11001	4
3	01101	0
4	10111	3

- Pergunta: que solução gerar agora?
- Expectativa: ao final da execução, solução(ões) ótima(s)

GAs: Que Solução Gerar Agora?

- Cruzamento: tentar gerar uma solução que combine schemas desejáveis a partir de pais escolhidos em função de bom fitness
- Explora, de forma *implícita*, uma distribuição de probabilidade de fitness em função dos valores de cada gene.

EDAs: Que Solução Gerar Agora?

- Algoritmos de Estimativa de Distribuição fazem uma modelagem *explícita* dessa distribuição.
- A cada geração:
 - Seleciono um subconjunto de soluções de forma proporcional ao fitness (roleta, torneio, etc)
 - Estimo uma distribuição de probabilidades
 - Crio novas soluções por amostragem

Motivação

Quão Bem GAs Funcionam?

- **Depende.**
- Teorema Fundamental dos GAs: schemas
 - Decomposição do problema em peças fundamentais
 - Capacidade de identificar schemas
 - Capacidade de combinar schemas sem destruí-los

Quão Bem GAs Funcionam?

- **Depende.**
- Identificação de schemas: implícita; depende do espaço de busca.
- Combinação não-destrutiva: depende dos operadores e da codificação.

Dependência do Espaço de Busca

- Função OneMax: $f(X) = \sum_{i=1}^N x_i$

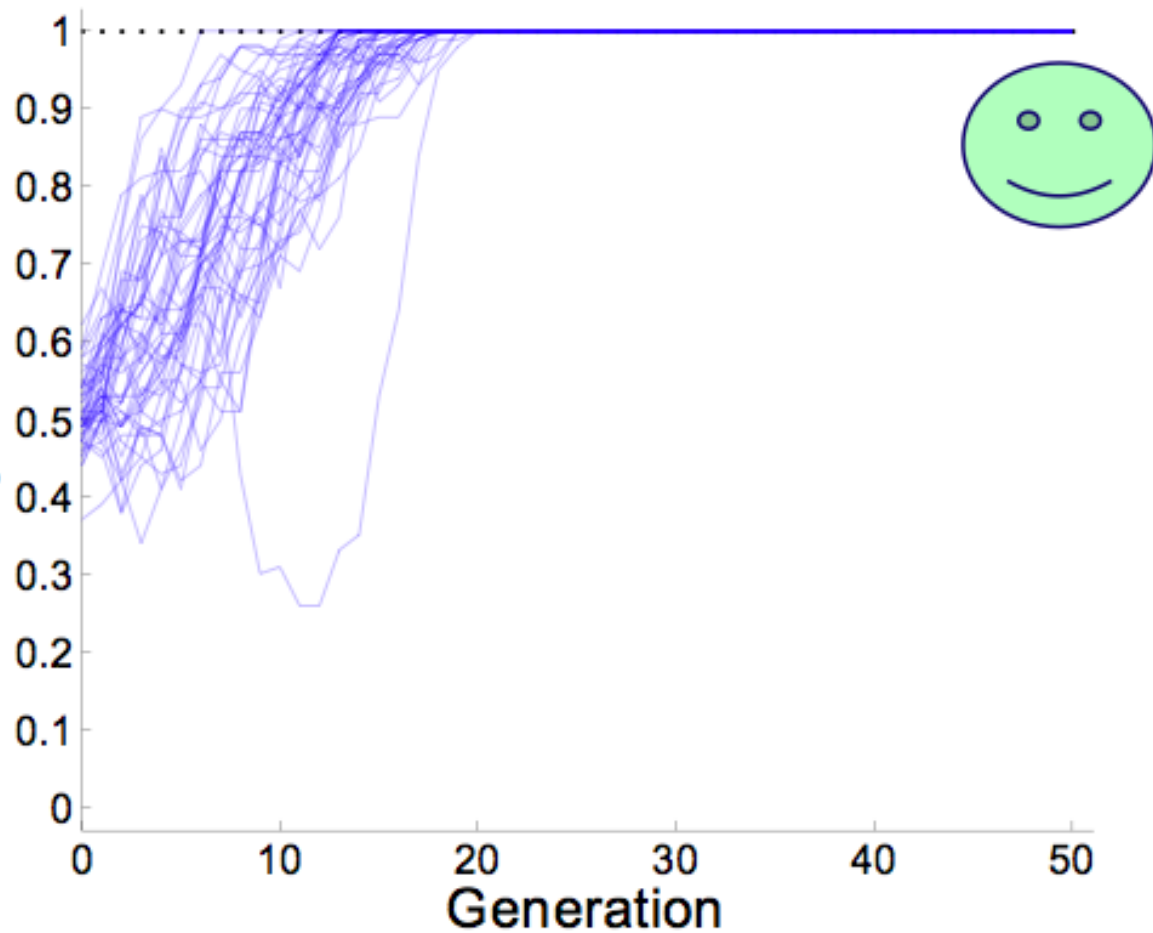
- Função Trap Concatenada:

$$f(X) = \sum_{i=1}^{N/5} \text{trap5}(x_{5i}, x_{5i+1}, \dots, x_{5i+4})$$

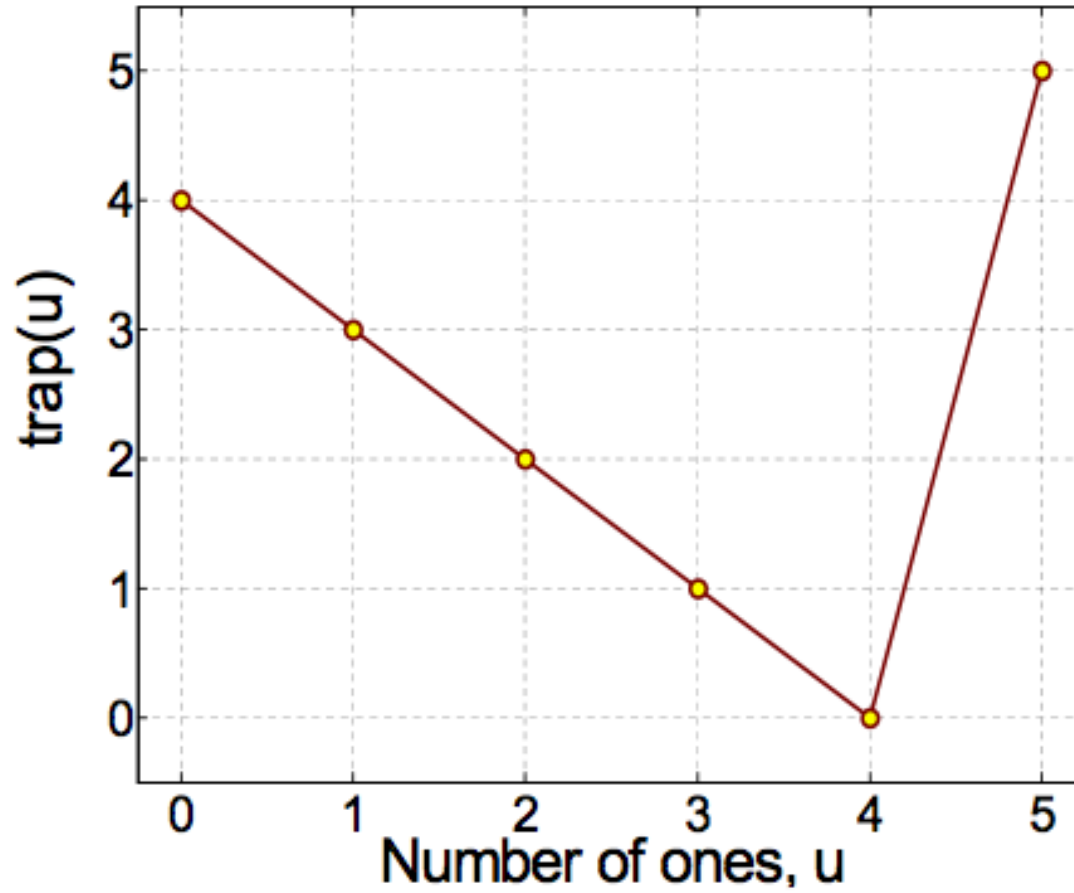
$$\text{trap5}(x_1, \dots, x_5) = \begin{cases} 5 & \text{se } \sum x_i = 5 \\ 4 - \sum x_i & \text{caso contrário} \end{cases}$$

- Ótimos: $X = \{1, 1, \dots, 1\}$; $f(X) = N$ nos dois casos

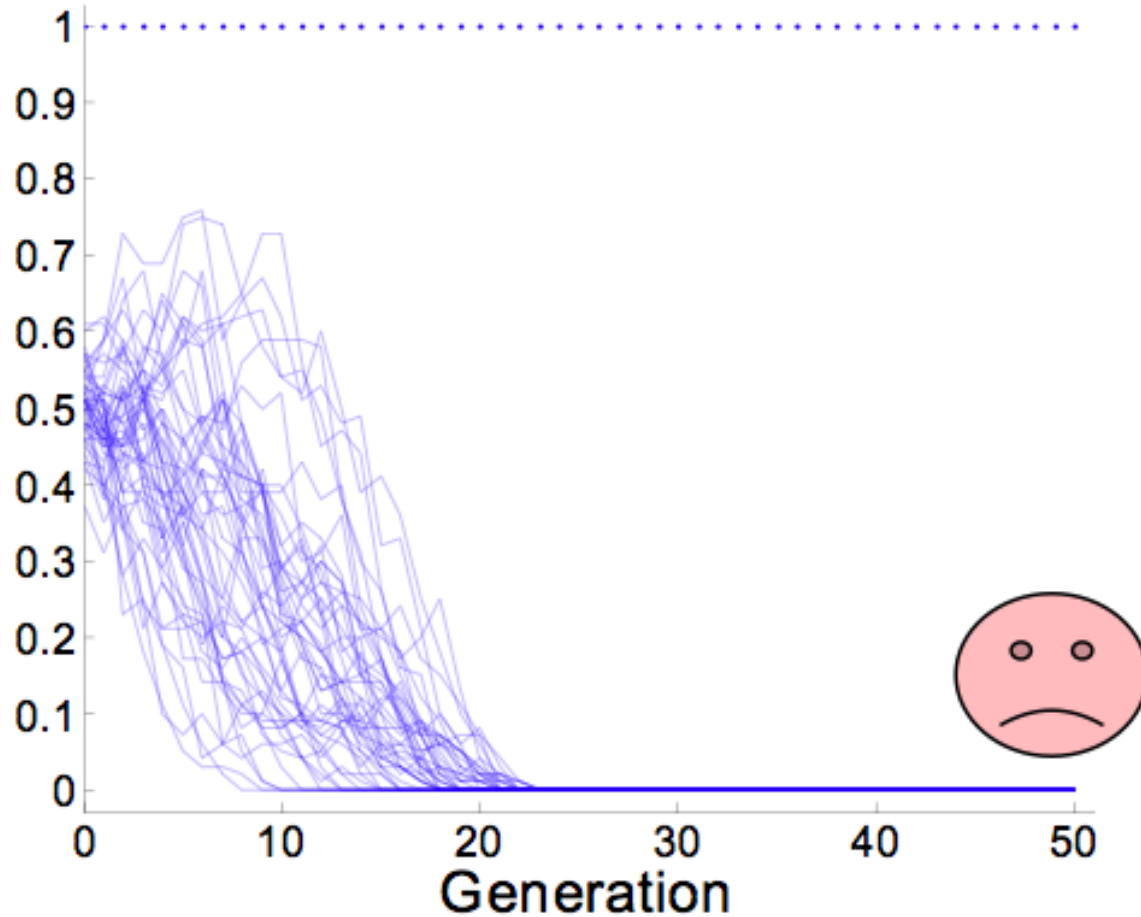
SGA para OneMax: $P(x_i=1)$



Trap5: Uma Armadilha para GAs



SGA para Trap Concat.: $P(x_i=1)$



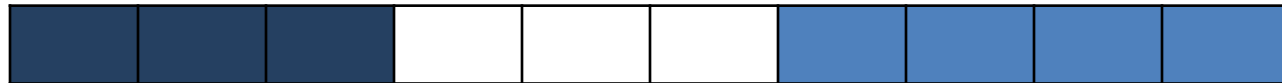
Quão Bem GAs Funcionam?

- Avaliações de fitness necessárias para atingir o ótimo:
 - OneMax: $O(N \log N)$
 - Trap Concatenada: $O(2^N)$

Espaço de busca pode impedir a identificação de schemas.

Dependência de Operadores e Codificação

- Ex 1: Dois schemas, cruzamento de um ponto



Schema 1

Schema 2

Alta probabilidade de preservação
e combinação dos schemas

Dependência de Operadores e Codificação

- Ex 2: Os mesmos schemas com outra ordem de variáveis, cruzamento de um ponto:



Mau casamento operadores-codificação
pode impedir a preservação e
combinação de schemas.

Análise

GAs funcionam bem com espaços de busca e codificações favoráveis, mas podem exigir um número exponencial de avaliações de fitness em casos desfavoráveis.

O que fazer?

Melhorando o Desempenho

- Alternativa tradicional: desenvolvimento de novas codificações e operadores especializados.
- Resultado depende do problema; tentativa e erro; “arte” da aplicação da técnica.

Melhorando o Desempenho

- Nova alternativa: Algoritmos de Estimativa de Distribuição.
- Probabilidades identificam schemas através de dependência entre variáveis.
- Novas soluções são geradas por amostragem, que não depende da posição das variáveis.

Algoritmos de Estimativa de Distribuição: Variáveis Binárias

Como Modelar a Distribuição?

- Mais simples: distribuições univariadas
 - Vetor de probabilidades para cada posição x_i no cromossomo
 - Simples de estimar
 - Não aprende schemas contendo múltiplas variáveis
 - Amostragem preserva e combina schemas

Como Modelar a Distribuição

- Mais informativas: distribuições multivariadas
 - Diferentes formas de representação
 - Estimativa pode ser cara
 - Aprende a *estrutura* do problema, ou quais variáveis pertencem a quais schemas importantes

Vetores de Probabilidades

- Dado um cromossomo com N variáveis, crio o vetor $p = (p_1, \dots, p_n)$, onde $p_i = P(x_i = 1)$.
- Estimativa: conto a proporção de 1's em cada x_i .
- Amostragem: para cada x_i , seto seu valor de acordo com p_i .

Vetores de Probabilidades

ID	Solução	Fitness
1	00100	1
2	11001	4
3	01101	0
4	10111	3



i	$P(x_i = 1)$
1	0.5
2	0.5
3	0.75
4	0.25
5	0.75

Vetores de Probabilidades

- Diversas variações:
 - PBIL (Baluja, 1995)
 - Compact GA (Harik et al, 1998)
 - UMDA (Mühlenbein & Paass, 1996)
- Todos apresentam desempenho similar.

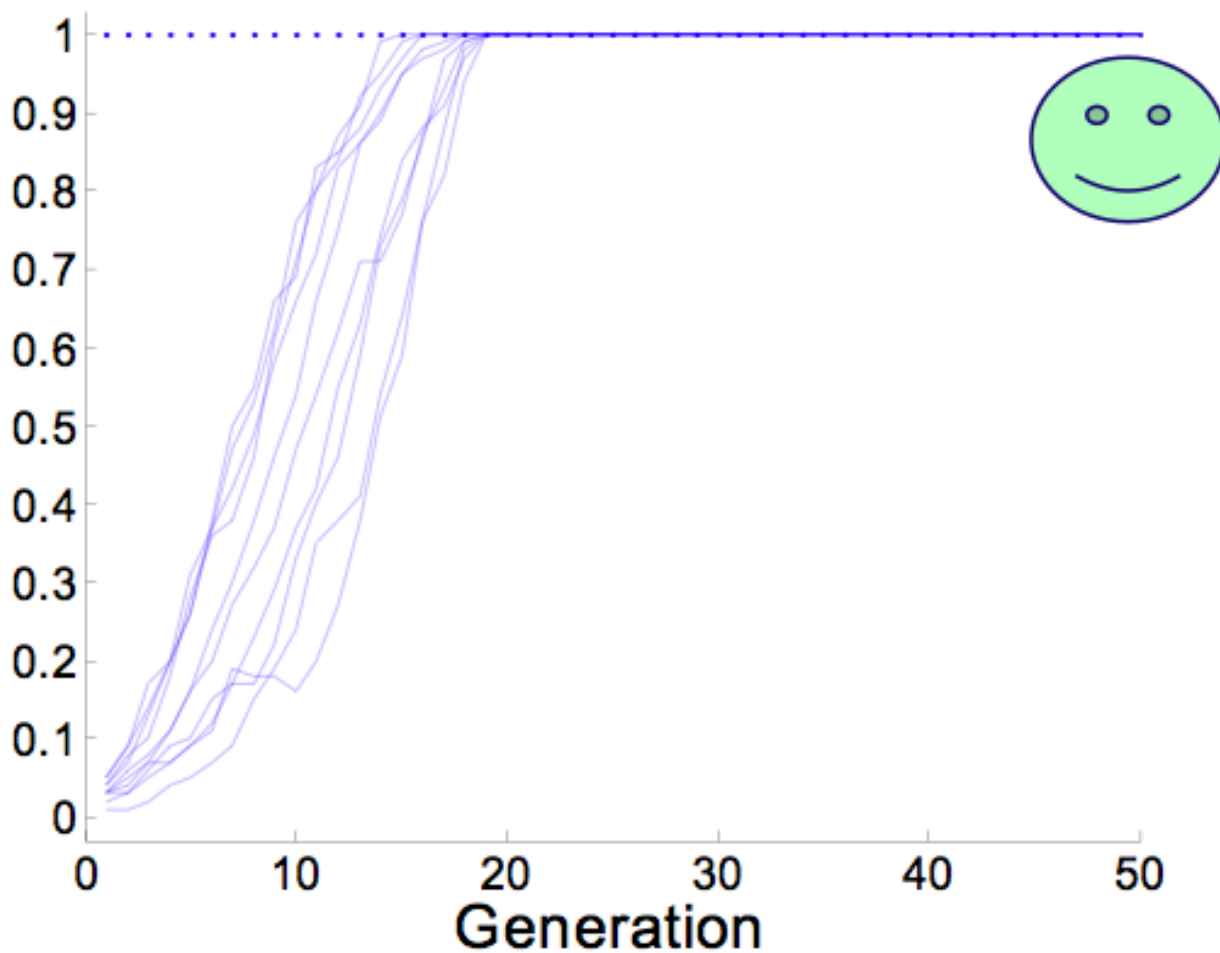
Compact GA (cGA)

1. Inicialize uma população aleatória
2. Repita até critério de parada
 1. Avalie a fitness da população
 2. Selecione um subconjunto da população por roleta
 3. Construa um vetor de probabilidades a partir do subconjunto selecionado
 4. Gere uma nova população para a próxima geração por amostragem do vetor

Vetores de Probabilidades

- Desempenho não depende da ordem das variáveis no cromossomo.
- Desempenho equivalente ao SGA na Trap Concatenada. Por quê? Schemas de tamanho 5 não são encontrados.
- Solução: criar probabilidades para conjuntos de cinco variáveis consecutivas, ao invés de uma.

cGA com Vetores de 5 Variáveis na Trap Concatenada



Conclusão?

- Amostragem de distribuições encontra o máximo se a distribuição modelar a decomposição correta do problema em schemas.
- Próximo passo? **Aprender a decomposição do problema automaticamente.**

Distribuições Multivariadas

- Schema = dependências entre variáveis.
- Decomposição do problema = identificação de subconjuntos de X com maior dependência entre seus membros.

Distribuições Multivariadas

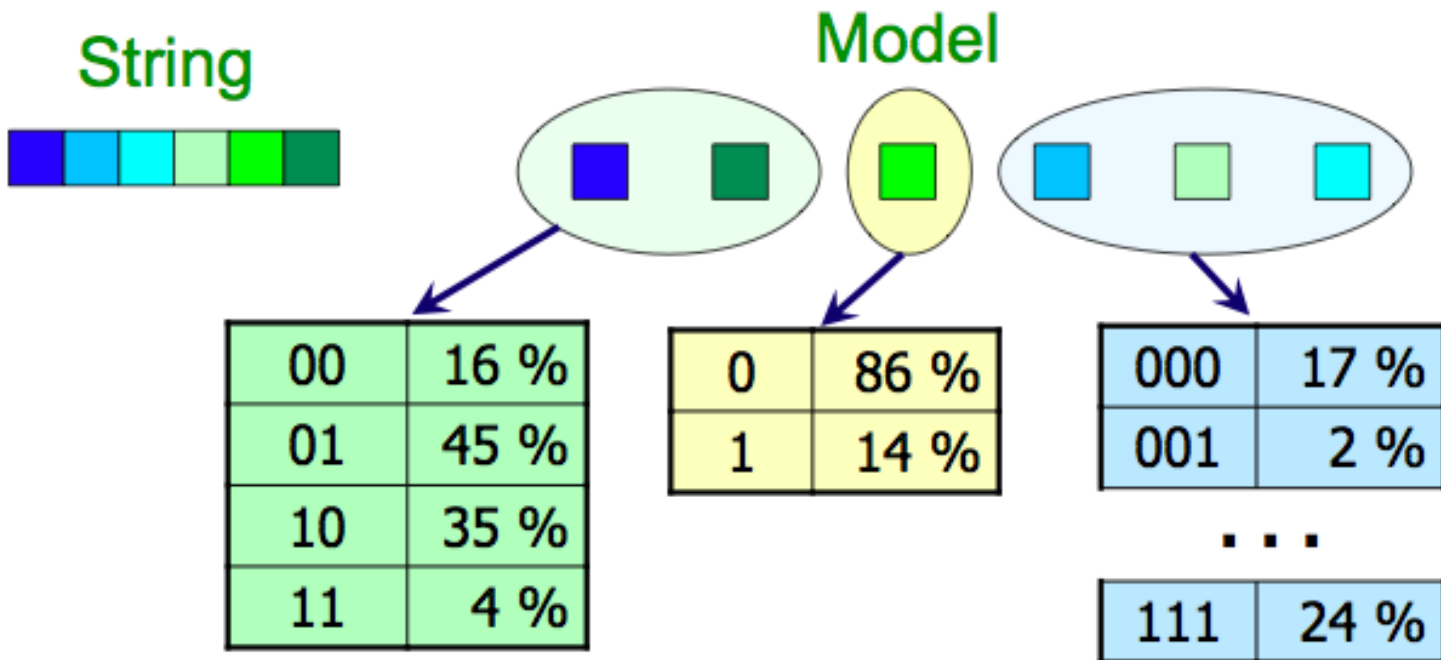
- Diversos algoritmos refletem diferentes maneiras de representar as dependências
- Partição: Extended Compact Genetic Algorithm (ECGA; Harik, 1999)
- Redes Bayesianas: Bayesian Optimization Algorithm (BOA; Pelikan, 2005)
- Redes de Dependência (EDNA; Gamez et al, 2007)

Extended Compact Genetic Algorithm

1. Inicialize uma população aleatória
2. Repita até critério de parada
 1. Avalie a fitness da população
 2. Selecione os melhores X% da população
 3. Particione as variáveis em conjuntos de variáveis relacionadas e crie uma tabela de probabilidades para cada conjunto.
 4. Gere uma nova população para a próxima geração por amostragem de cada tabela.

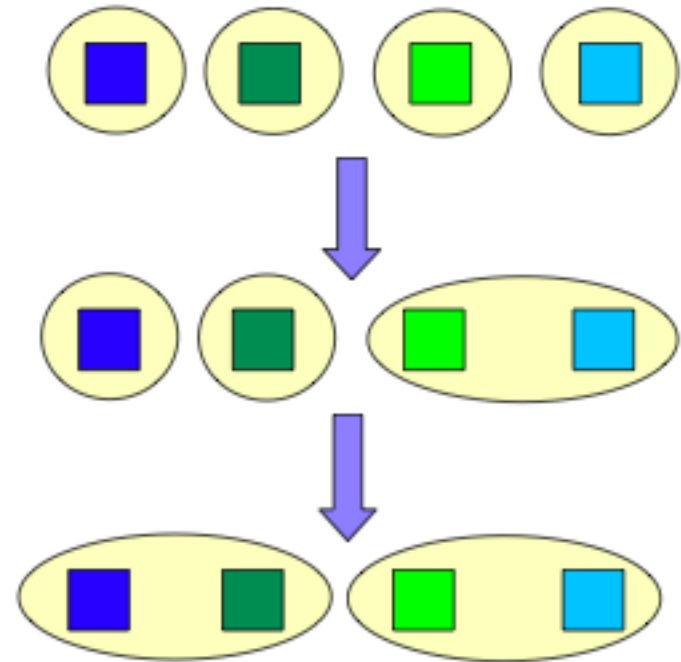
Extended Compact Genetic Algorithm

- Partição do espaço de variáveis em subconjuntos distintos. Para cada conjunto, tenho uma tabela de probabilidade de valores.



Extended Compact Genetic Algorithm

- Geração da partição:
 - Começamos com N conjuntos, um com cada variável.
 - A cada iteração, combinamos dois conjuntos de forma a maximizar a qualidade da distribuição.
 - Paramos quando não há combinação que melhore a qualidade.



Extended Compact Genetic Algorithm

- Qualidade da partição: *Minimum Description Length (MDL)*, ou seja, escolhemos a partição com menor custo.
- Para uma dada partição P as N variáveis são divididas em k subconjuntos $\{c_1, \dots, c_k\}$, cada um de tamanho $t(c_i)$, e cada um com uma distribuição marginal $M(c_i)$ correspondente à sua tabela de frequências.

Extended Compact Genetic Algorithm

$$\text{Custo}(P) = \text{Custo}(\text{Modelo}) + \text{Custo}(\text{Dados})$$

$$\text{Custo}(\text{Modelo}) = \log(N) \sum_{i=1}^k 2^{t(ci)}$$

$$\text{Custo}(\text{Dados}) = N \sum_{i=1}^k \text{entropia}(M(ci))$$

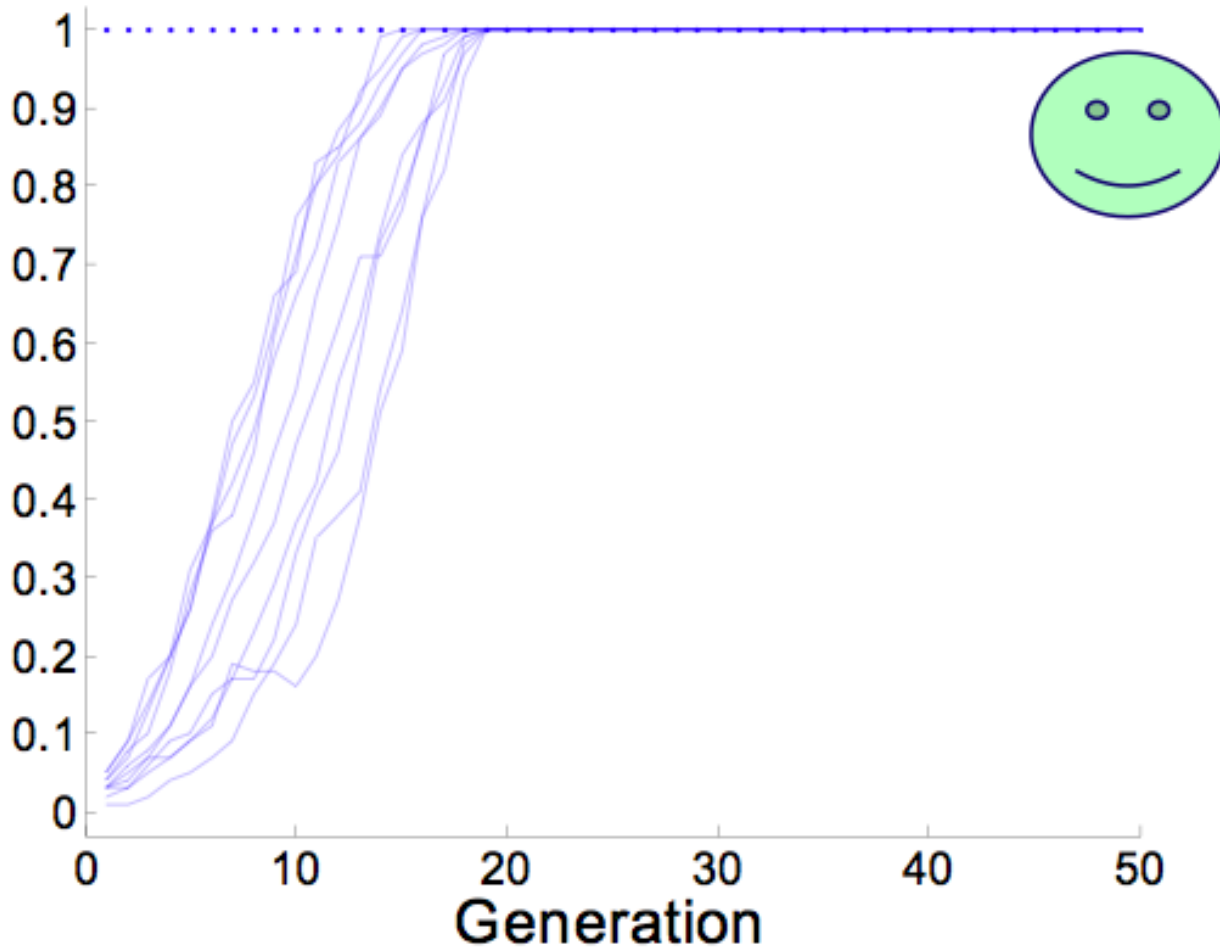
O custo do modelo indica quantos bits são precisos para representar a constituição de cada conjunto.

O custo dos dados estima quantos bits são precisos para representar a população, dadas as tabelas de probabilidade de cada conjunto. Quanto mais uniformes forem as tabelas, maior sua entropia e maior o custo.

Extended Compact Genetic Algorithm

- Amostragem:
 - Percorro os subconjuntos em qualquer ordem
 - Em cada subconjunto, seleciono uma configuração para todas as suas variáveis ao mesmo tempo, de acordo com a tabela de probabilidades.
- Schemas identificados pelo particionamento são sempre preservados.

ECGA na Trap Concatenada



Extended Compact Genetic Algorithm

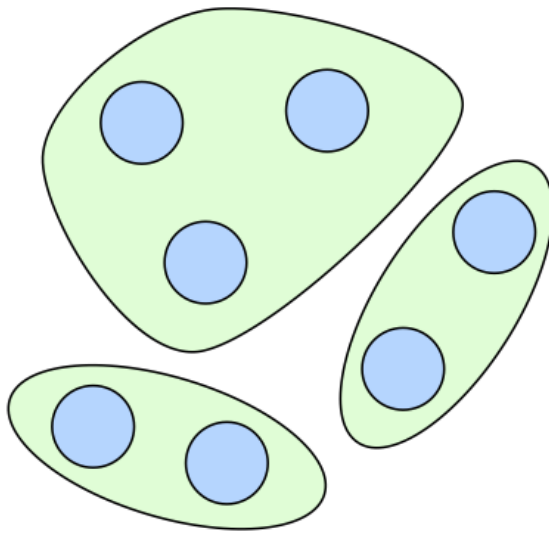
- Avaliações de fitness necessárias para atingir o ótimo:
 - OneMax: $O(N \log N)$
 - Trap Concatenada: **$O(N^2)$**
 - Trap Concatenada com variáveis permutadas (cada função trap5 recebe entradas não contíguas no cromossomo): **$O(N^2)$**

Distribuições Mais Complexas

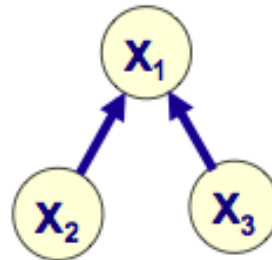
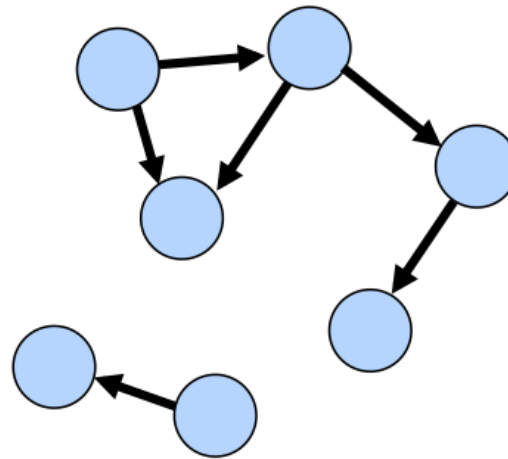
- Limitação do ECGA: cada variável só pertence a um schema.
- Bayesian Optimization Algorithm (BOA, Pelikan 2000): Rede Bayesiana de dependências e distribuições

Bayesian Optimization Algorithm

ECGA



BOA



X_2X_3	$P(X_1=0 X_2X_3)$
00	26 %
01	44 %
10	15 %
11	15 %

Bayesian Optimization Algorithm

- Geração da rede:
 - Começamos com N conjuntos, um com cada variável.
 - A cada iteração, escolhemos uma operação de forma a maximizar a qualidade da distribuição.
 - Paramos quando não há combinação que melhore a qualidade.
 - Operações: inserção, remoção e inversão de aresta.

Algoritmos de Estimativa de Distribuição: Variáveis Contínuas

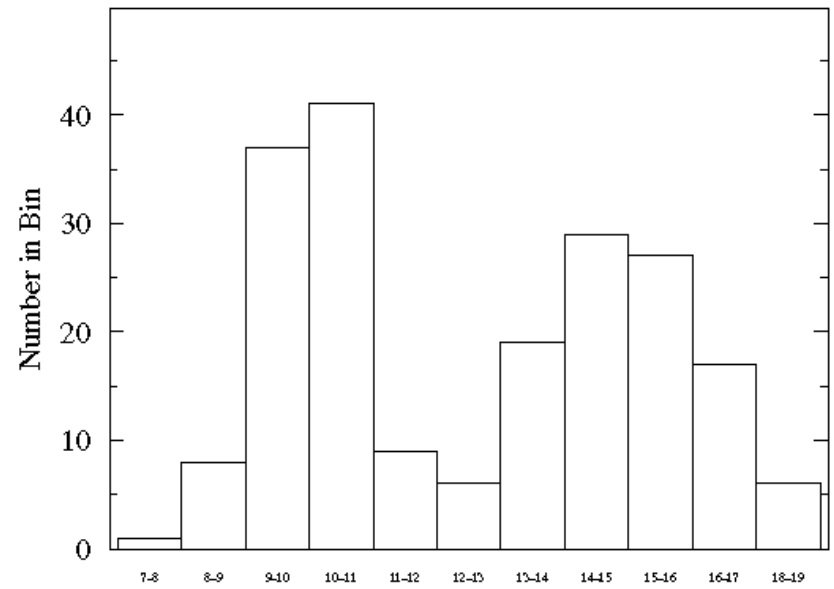
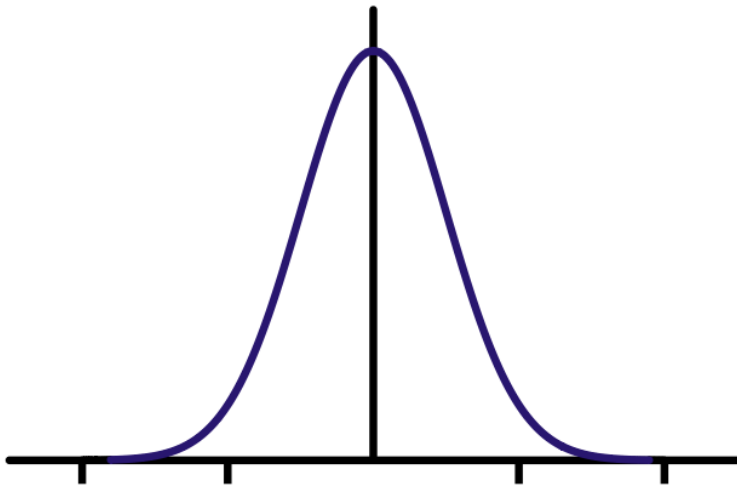
Variáveis Contínuas

- Domínio infinito
- Possibilidades:
 - Discretizar espaço de busca e aplicar EDA binário ou discreto.
 - Estimar *função de densidade de probabilidade*.
 - Variáveis independentes: estimativas univariadas.
 - Dependências entre variáveis: estimativas multivariadas.

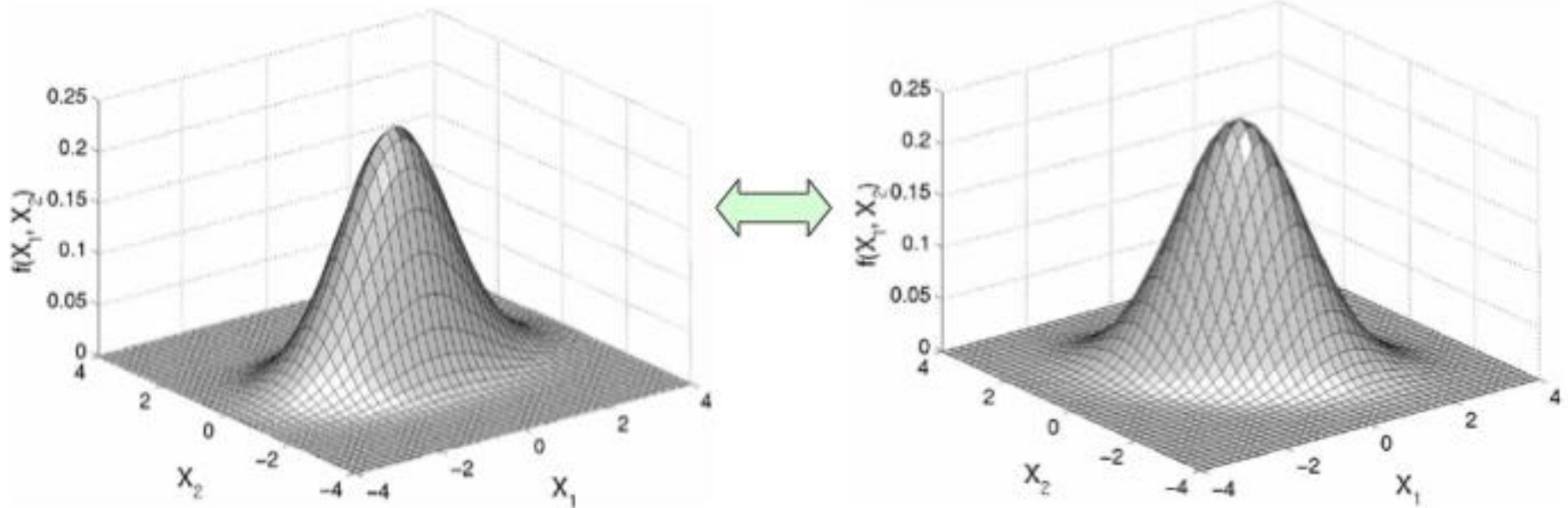
Estimando Variáveis Contínuas

- Formato/família da distribuição
 - Gaussiana
 - Histogramas
- Covariâncias
 - Ignorar
 - Particionar
 - Matriz completa
- Uma distribuição ou várias? Um pico ou vários?
- Heurística de estimativa

Distribuição



Covariâncias



Dependências: $\text{CoVar}(x_1, x_2) > 0$

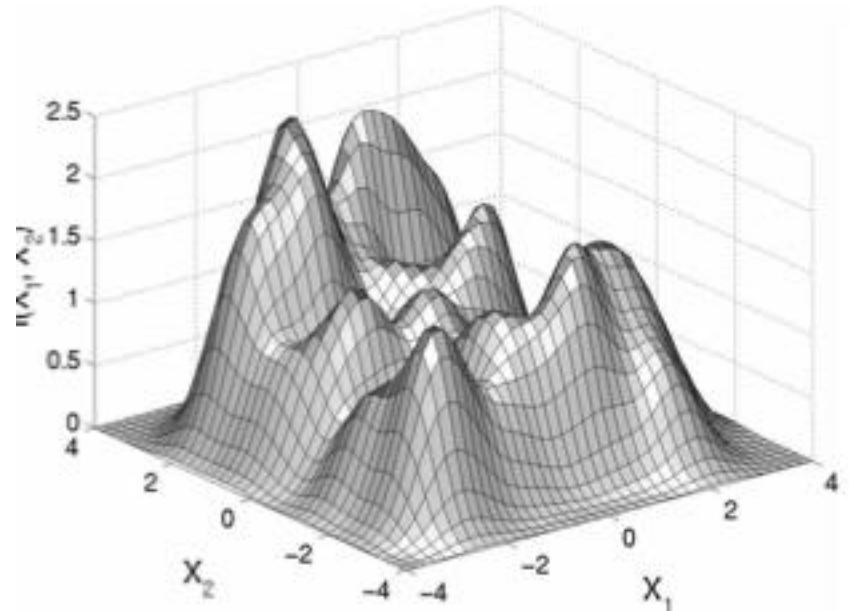
Rotação das distribuições

Kernels e Misturas

Problema: variáveis multimodais

Kernels: precisos mas de difícil estimativa

Misturas: soma ponderada de distribuições unimodais



Gaussian UMDA

- Distribuição: Normal
- Covariâncias: Nenhuma
- Estimativa:
 - Seleção por truncagem ou torneio
 - Uma distribuição por variável
 - Estimativa de máxima verossimilhança para média e variância
- Larrañaga et al., 2000, tb. Sun, Zhang e Tsang 2005 e outros

Estimation of Gaussian Networks Algorithm (EGNA)

- Distribuição: Normal
- Covariâncias: Grupos de variáveis
- Estimativa:
 - Seleção por truncagem ou torneio
 - Redes Gaussianas = Redes Bayesianas para variáveis contínuas
 - Estimativa da rede por heurística similar à do BOA, iniciando com grafo completo
- Larrañaga et al., 2000 e outros

Iterative Density Estimation Evolutionary Algorithm (IDEA)

- Distribuição: Normal
- Covariâncias: Grupos de variáveis
- Estimativa:
 - Seleção por truncagem ou torneio
 - Redes Gaussianas = Redes Bayesianas para variáveis contínuas
 - Estimativa da rede por heurística similar à do BOA, iniciando com variáveis independentes
- Bosman e Thierens, 2000

IDEA Baseado em Misturas

- Distribuição: Mistura de Normais
- Covariâncias: Pares de variáveis ou toda a matriz
- Estimativa:
 - Seleção por truncagem ou torneio
 - Variáveis divididas por agrupamento (k-Means)
 - Heurística iterativa para cada agrupamento
- Bosman e Thierens, 2002

Histogram EDA

- Distribuição: Histogramas
- Covariâncias: Variáveis independentes
- Estimativa:
 - Seleção por truncagem
 - Frequência definida por amostragem
 - Tamanho do histograma fixo por variável
- Zhang et al, 2004 e outros

Técnicas Híbridas

- Busca Local (Zhang et al., 2004)
- Diferencial Evolution (Sun, Zhang e Tsang, 2005)
- Gradientes Conjugados (Bosman e Thierens, 2002)
- etc

Conclusões

Conclusões

- EDAs aprendem a estrutura de schemas de um problema.
- São robustos à ordem das variáveis.
- Têm melhor escalabilidade que GAs.

Quando Usar?

- Custo: a estimativa de distribuição é cara., e deve ser usada se ao menos um dos fatores a seguir for verdadeiro:
 - Avaliações da fitness são caras
 - Problema tem alta dimensionalidade
 - GAs apresentam desempenho insatisfatório

Também Existem EDAs Para...

- Variáveis inteiras
- Otimização combinatória
- Otimização multiobjetivo
- Mineração de dados (através de evolução de programas como em GP)
- Etc

Referências: EDAs Binários

- Harik, G; Lobo, F e Goldberg, D (1998) “The Compact Genetic Algorithm”, IEEE CEC.
- Baluja, S (1995) “Population Based Incremental Learning: A Method for Integrating Genetic Search Based Function Optimization and Competitive Learning”, Tech Report, Computer Science Department, CMU.
- Harik, G (1999) “Linkage Learning via Probabilistic Modeling in the ECGA”, Technical Report 99010, IlliGAL, University of Illinois at Urbana-Champaign.
- Pelikan, M (2005), “Hierarchical BOA: Towards a New Generation of Evolutionary Algorithms”. Springer.

Referências: EDAs Contínuos

- Larrañaga et al (2000) “Optimization in continuous domains by learning and simulation of Gaussian networks”, GECCO.
- Bosman e Thierens (2000) “Expanding from Discrete to Continuous Estimation of Distribution Algorithms: the IDEA”. Parallel Problem Solving from Nature VI, Springer.
- Zhang et al (2004) “Hybrid Estimation of Distribution Algorithm for Global Optimization”, Engineering Computations.
- Sun, Zhang e Tsang (2005) “DE/EDA: A New Evolutionary Algorithm for Global Optimization”, Information Sciences.

Referências: Livros

- Goldberg, D (2002) “The Design of Innovation”, Springer.
- Pelikan, Sastry, Cantu-Paz, eds (2007). “Scalable Optimization via Probabilistic Modeling: From Algorithms to Applications”, Springer.
- Lozano, Larrañaga, Inza, Bengoetxea, eds (2007). “Towards a New Evolutionary Computation: Advances on Estimation of Distribution Algorithms”, Springer.

Código Fonte

- BOA (C++): <http://medal.cs.umsl.edu/>
- ECGA (C++, Matlab):
<http://www.illigal.uiuc.edu/web/>
- MATEDA (Matlab):
<http://www.sc.ehu.es/ccwbayes/members/rsantana/software/matlab/MATEDA.html>

Obrigado!