Evolutionary Algorithms:
Estimation of Distribution Algorithms

May 28, 2013
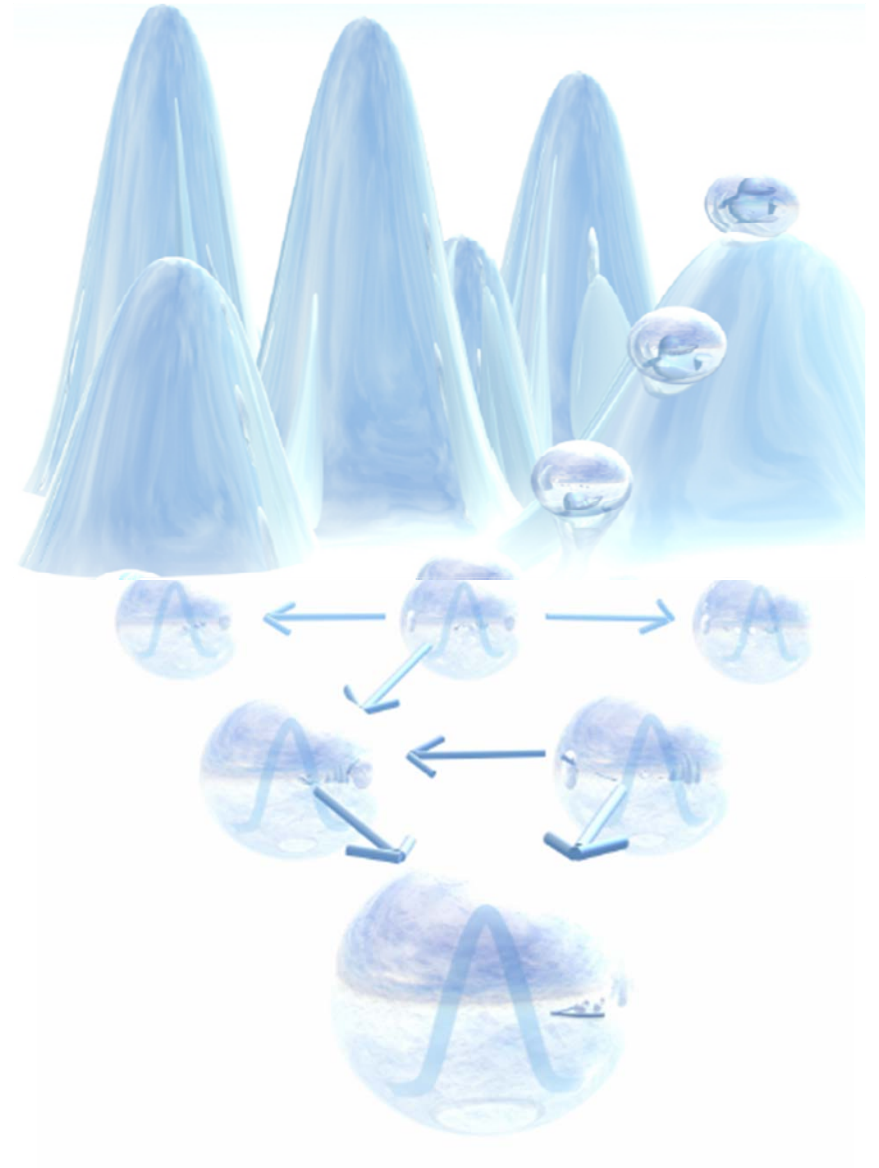
Prof. Chang Wook Ahn

Sungkyunkwan Evolutionary Algorithms Lab.

School of Info. & Comm. Eng.

Sungkyunkwan University
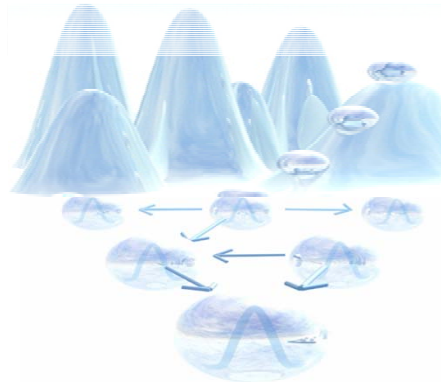
❖ **Estimation of Distribution Algorithms (EDAs)**

➢ **A new paradigm** in terms of **intelligent optimization tools**

➢ **Cross-fertilization:** Evolutionary computation + (Probabilistic) Machine learning

➢ **Practical use:** Automatic learning, Robust & scalable performance
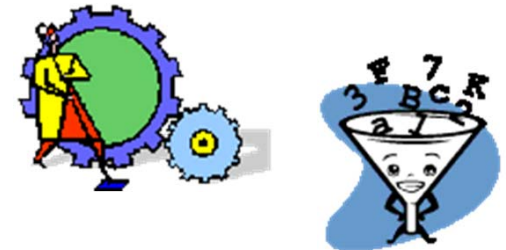
Evolutionary Computation	EDA (since 1998)	Machine Learning

• Ad hoc Operators	• Multiple Search of EC	• Efficient for Extracting Information from Data
• Great for Toy Problems	• Problem Knowledge of ML	• Powerful for Learning & Recognition
• Not Scale Well	• Intelligent Operators by EC and ML	• BNs, GM, Classifier, …
• Powerful for Search	• Quadratic Scalability	
• GAs, GP, ESs, …		

Key Principle

❖ **Evolution of Probability Distribution of Population**

➢ **Selection:** Leads to **probability density evolution**

➢ **Sampling:** Generates new individuals

✓ It replaces **crossover** and **mutation**!

✓ But the probability density itself is identical!

☞ **Note:** Incorporating **problem knowledge** offers better performance.

Prob. Density Estimation

Prob. Density Estimation

Selection: superior individuals!

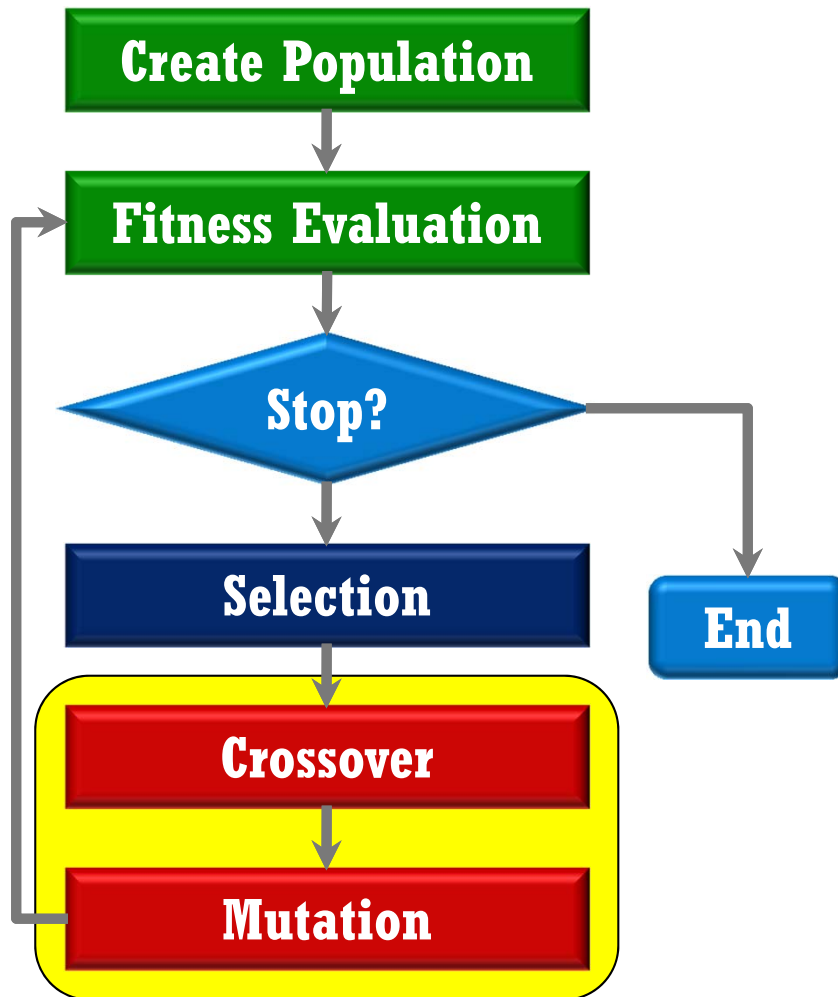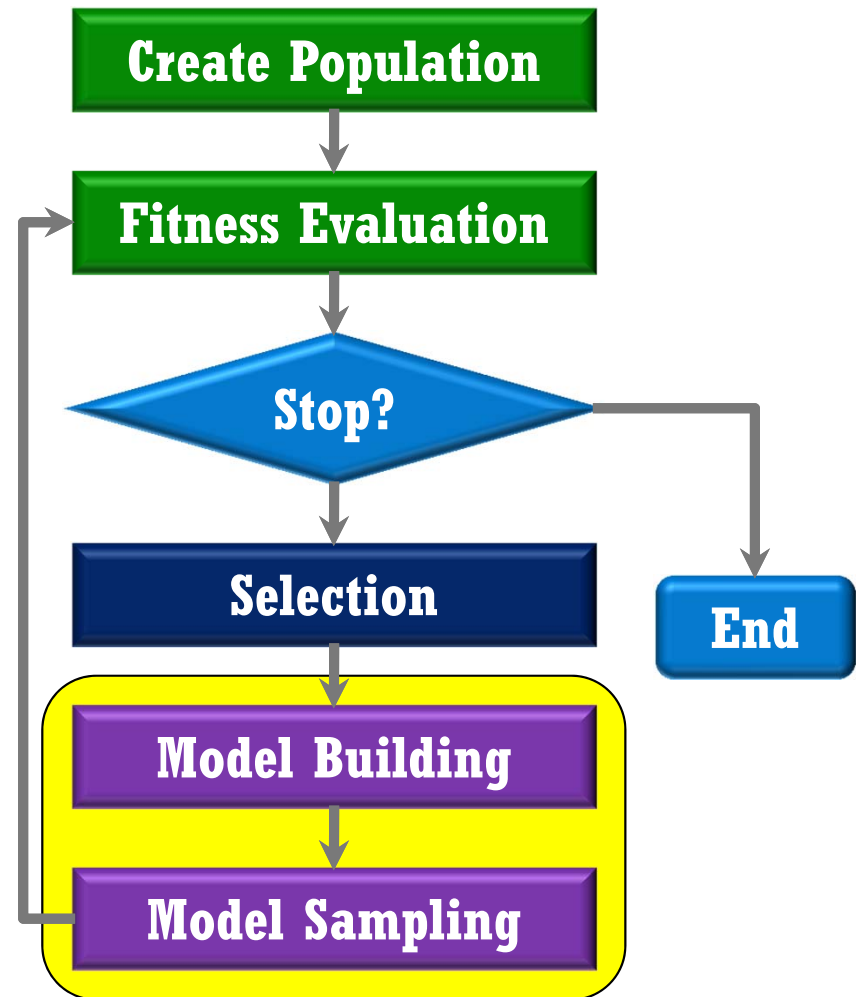Sampling: new individuals!
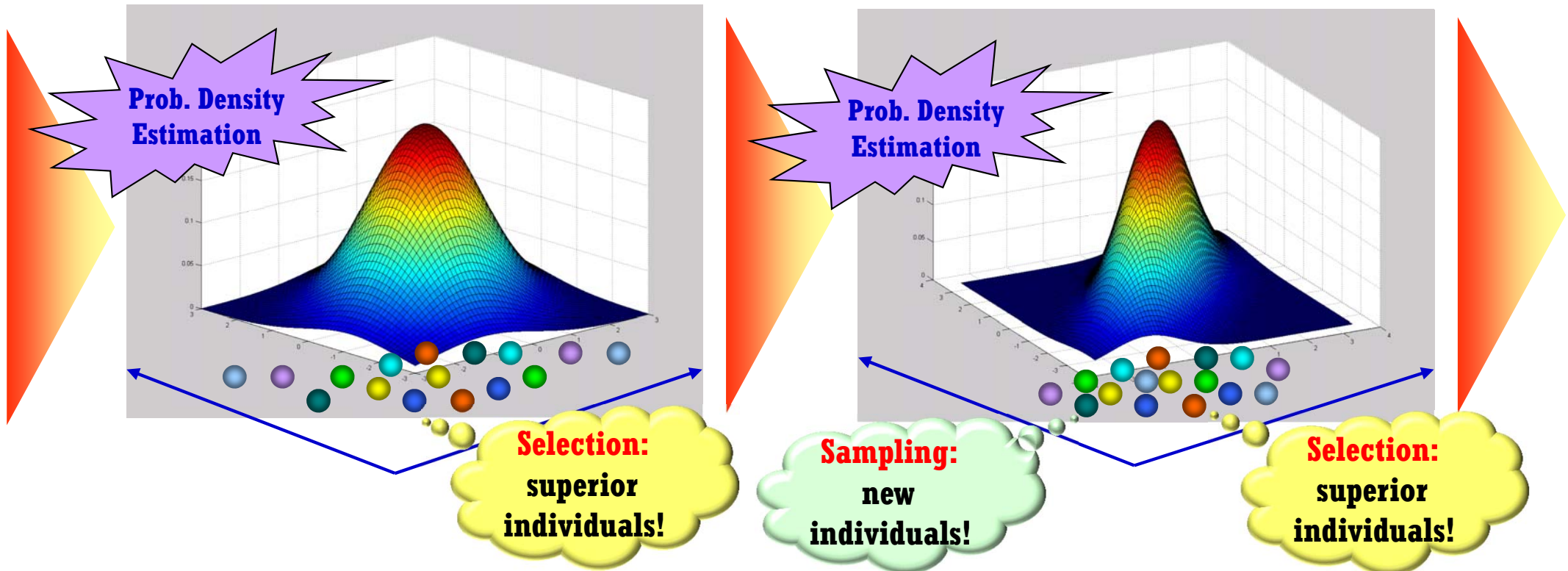
Selection: superior individuals!

❖ **EDAs conduct GAs with Probability Distribution**

➤ **GAs:** Crossover & Mutation are applied to explicit individuals

➤ **EDAs:** Mixing process is conducted by sampling the *PDF* of population
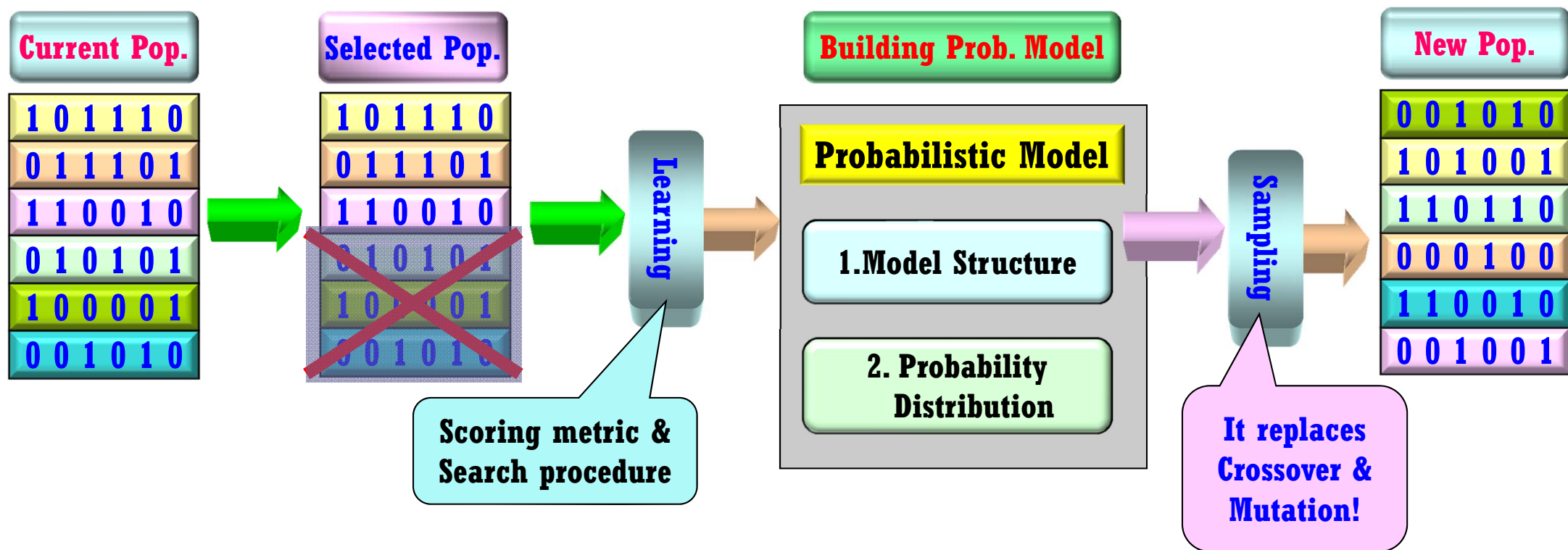
General Procedures

❖ **Details of Estimation of Distribution Algorithms**

➢ Powerful **Intelligent Stochastic Optimization Technique** inspired from both **Genetic Algorithms** and **(Probabilistic) Machine Learning**

➢ **Automatic identification** and **exploitation** of problem regularities

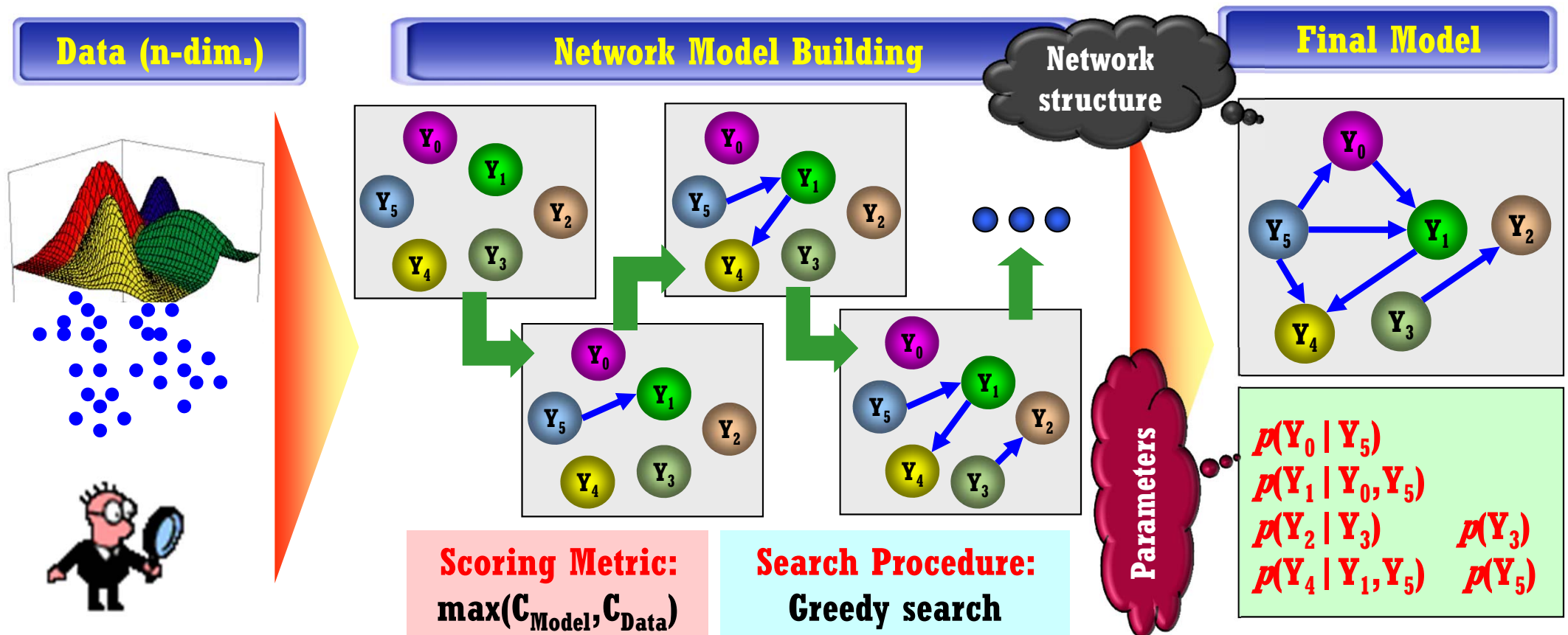➢ Robust and **Scalable** performance on broad classes of challenging problems

Probabilistic Machine Learning

❖ **Graphical Model** (A Representative Probabilistic Machine Learning Technique)

➢ **Representation:** **Dependencies** among variables by a structure

➢ **Structure:** A directed acyclic graph that represents a **factorization** of variables

➢ **Factorization:** A joint distribution by a **product** of **conditional distributions**

➢ **Scoring metric** & **Search procedure** are required

Data (n-dim.)

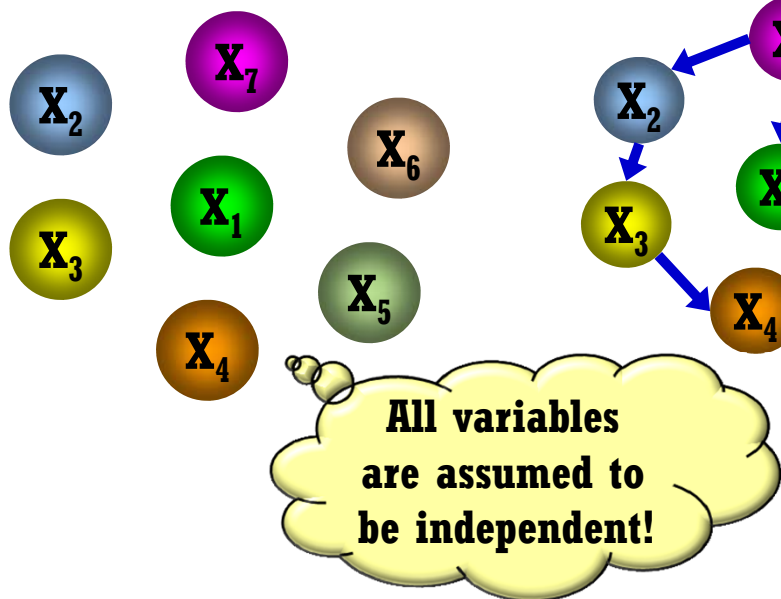Network Model Building

Network structure

Final Model

Scoring Metric: $\max(C_{Model}, C_{Data})$

Search Procedure: Greedy search

Parameters

$p(Y_0 \mid Y_5)$
$p(Y_1 \mid Y_0, Y_5)$
$p(Y_2 \mid Y_3)$ $p(Y_3)$
$p(Y_4 \mid Y_1, Y_5)$ $p(Y_5)$

Classification

❖ **Categories of EDAs**

➢ **Three Classes** exist in accordance with **Model Dependencies**

Univariate Model

X_7 X_2 X_6 X_1 X_3 X_5 X_4

All variables are assumed to be independent!

- PBIL (Baluja '95)
- UMDA (Muhlenbein '96)
- Compact GA (Harik '98)
- QEA (K.H. Han '02)

Bivariate Model

X_7 X_2 X_6 X_3 X_1 X_5 X_4

X_7 X_2 X_6 X_3 X_1 X_5 X_4
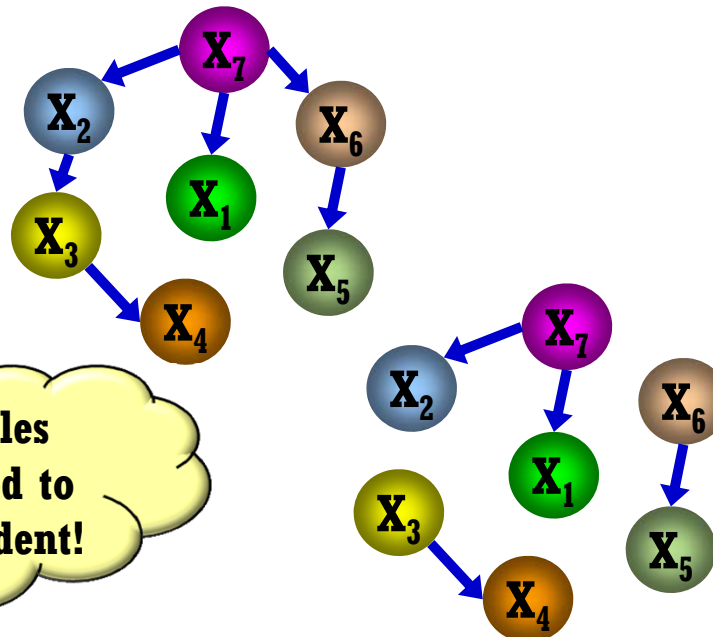
- MIMIC (DeBonet '96)
- COMIT (Baluja '97)
- BMDA (Pelikan '98)

Multivariate Model

X_7 X_2 X_6 X_1 X_3 X_5 X_4

X_7 X_2 X_6 X_1 X_3 X_5 X_4
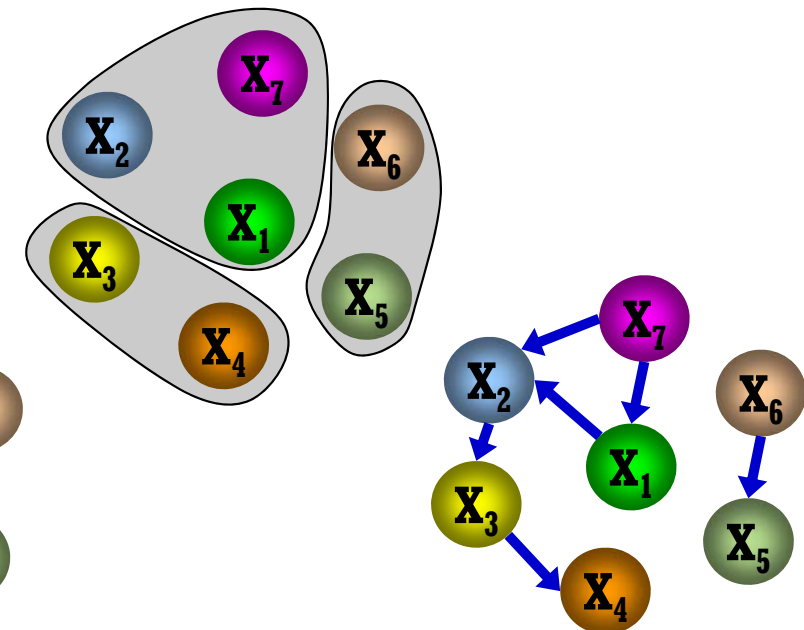
- FDA (Muhlenbein '99)
- ECGA (Harik '99) , mIDEA(Bosman '03)
- BOA (Pelikan '99), rBOA (C.W. Ahn '04)
- Mixed BOA (Ocenassek '02)

Example: UMDA

Simulation by Hand
- Population size, N = 6
- Individual length, l = 6
- Problem: OneMax (Optimum=111111)
- Probability model, $p = (p_1, \ldots, p_l)$

Current (sorted) population

Selected population

Univariate model

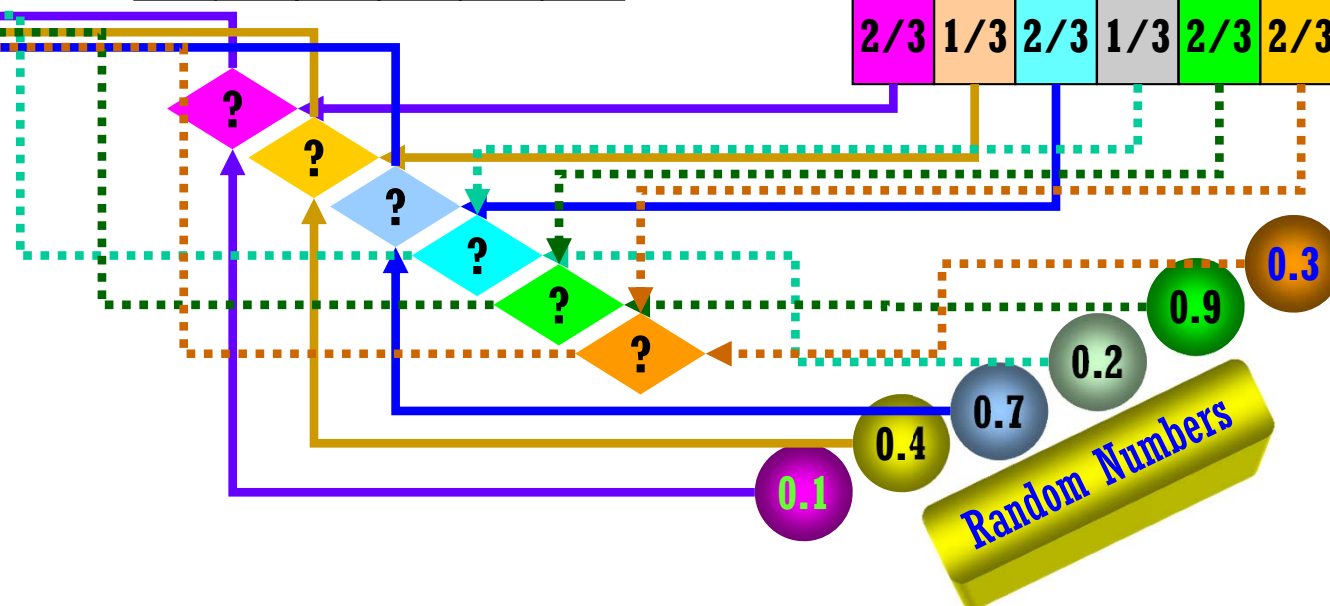| 2/3 | 1/3 | 2/3 | 1/3 | 2/3 | 2/3 |

Next population

Random Numbers

SEAL.

❖ Why Higher-Order Statistics?

➢ **One-Max Problem: Optimum is '111...1'**
  ✓ Single bit '1' is better than '0' **on average**.
➢ **Ex) 5-bit Trap Function: optimum is '11111'**
  ✓ But, f(1****)=1.375, f(0****)=2;
  ✓ Single bits are **misleading**!
  ✓ Thus, consider **5-bit statistics**!
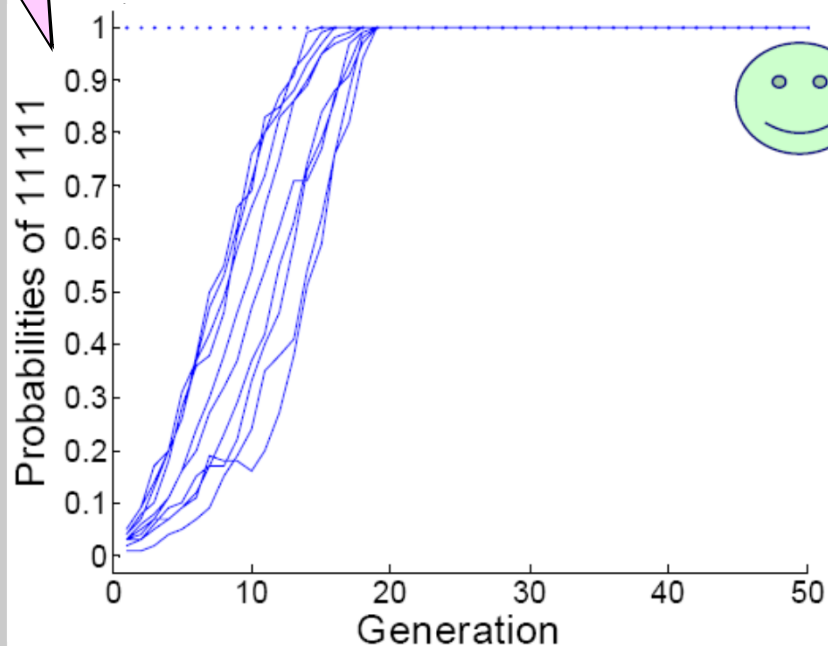
Ex) Artificial Deceptive Problem; i.e., Concatenated Traps
$$F(X) = \sum f(B)$$

Trap1   Trap2   Trap*k*

Via 5-bit statistics!

Probabilities of 11111 vs Generation
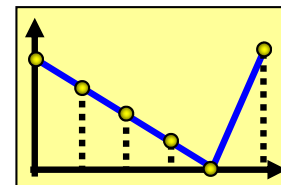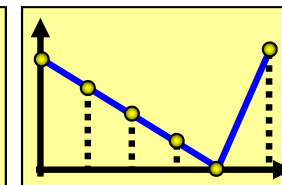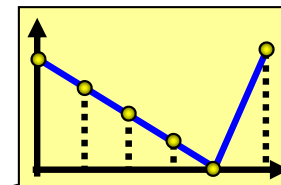
Ex) 5-bit Trap Function; if u(B) = 4, then f(B) = 5; else f(B) = 4 − u(B) ;

$f(B)$ vs $u(B)$
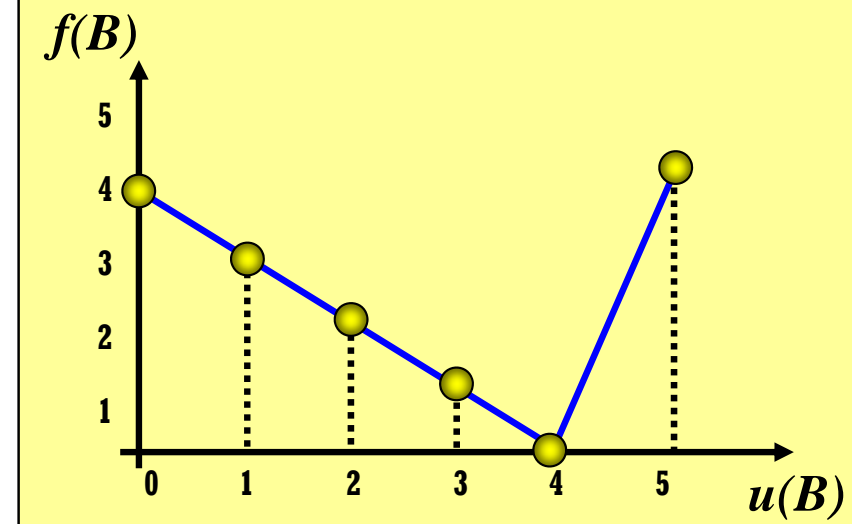
❖ **Combined Optimizers with Mutual Information Trees [Baluja&Davis'97]**

➢ Use a **tree model**

✓ Find a tree that **maximizes mutual information** between connected nodes

$$I(X_i, X_j) = \sum_{a,b} P(X_i = a, X_j = b) \log \frac{P(X_i = a, X_j = b)}{P(X_i = a)P(X_j = b)}$$

➢ **Prim algorithm** is used for finding a maximum spanning tree
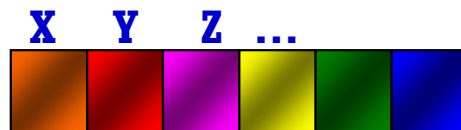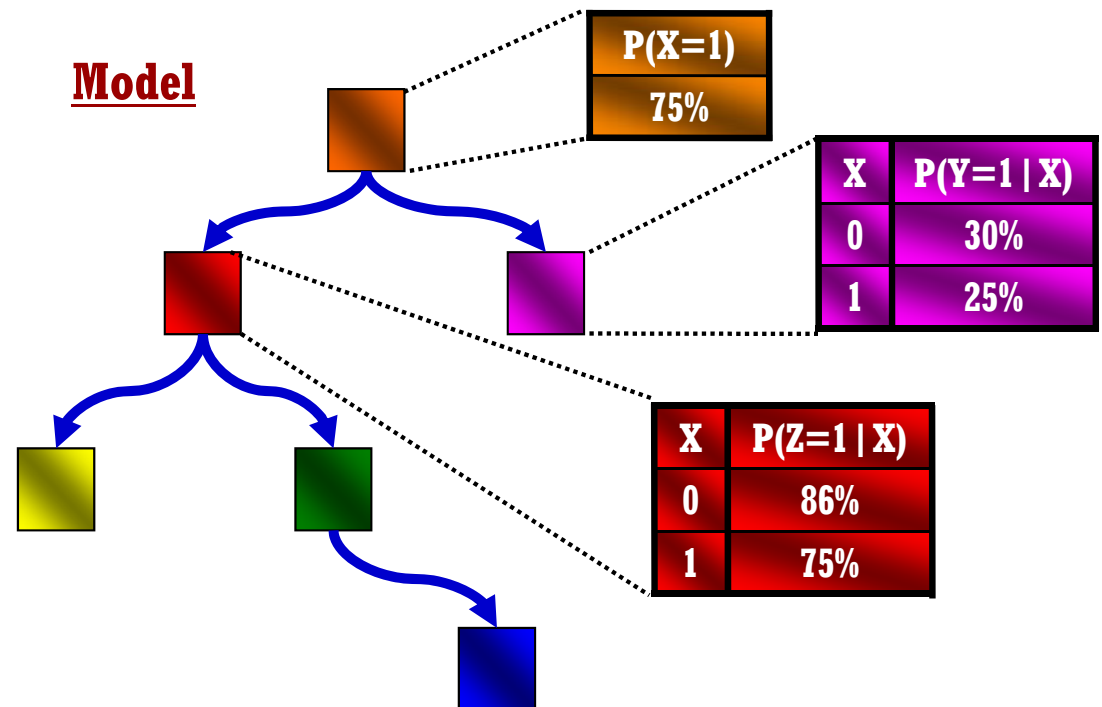
Example: ECGA

❖ **Extended Compact GA** [Harik'99]
  - ➤ **Consider groups of genes**
    - ✓ Use **Minimum Description Length** $\{ \log(N)\sum_I 2^{S[I]} + N\sum Entropy(M_I) \}$
  - ➤ **Greedy algorithm is used for model search**

Model

String

00	16%
01	45%
10	35%
11	4%

0	75%
1	25%

000	12%
001	5%
...	...
111	25%

Selected population

```
[0][1][2][3]
 1  0  0  0
 1  1  0  1
 0  1  1  1
 1  1  0  0
 0  0  1  0
 0  1  1  1
 1  0  0  0
 1  0  0  1
```

Model	MDL
[0] [1] [2] [3]	44

Best Model!

Model	MDL
[0 1] [2] [3]	46.7
[0 2] [1] [3]	39.8
[0 3] [1] [2]	46.7
[0] [1 2] [3]	46.7
[0] [1 3] [2]	45.6
[0] [1] [2 3]	46.7

Model	MDL
[0 2 1] [3]	48.6
[0 2 3] [1]	48.6
[0 2] [1 3]	41.4
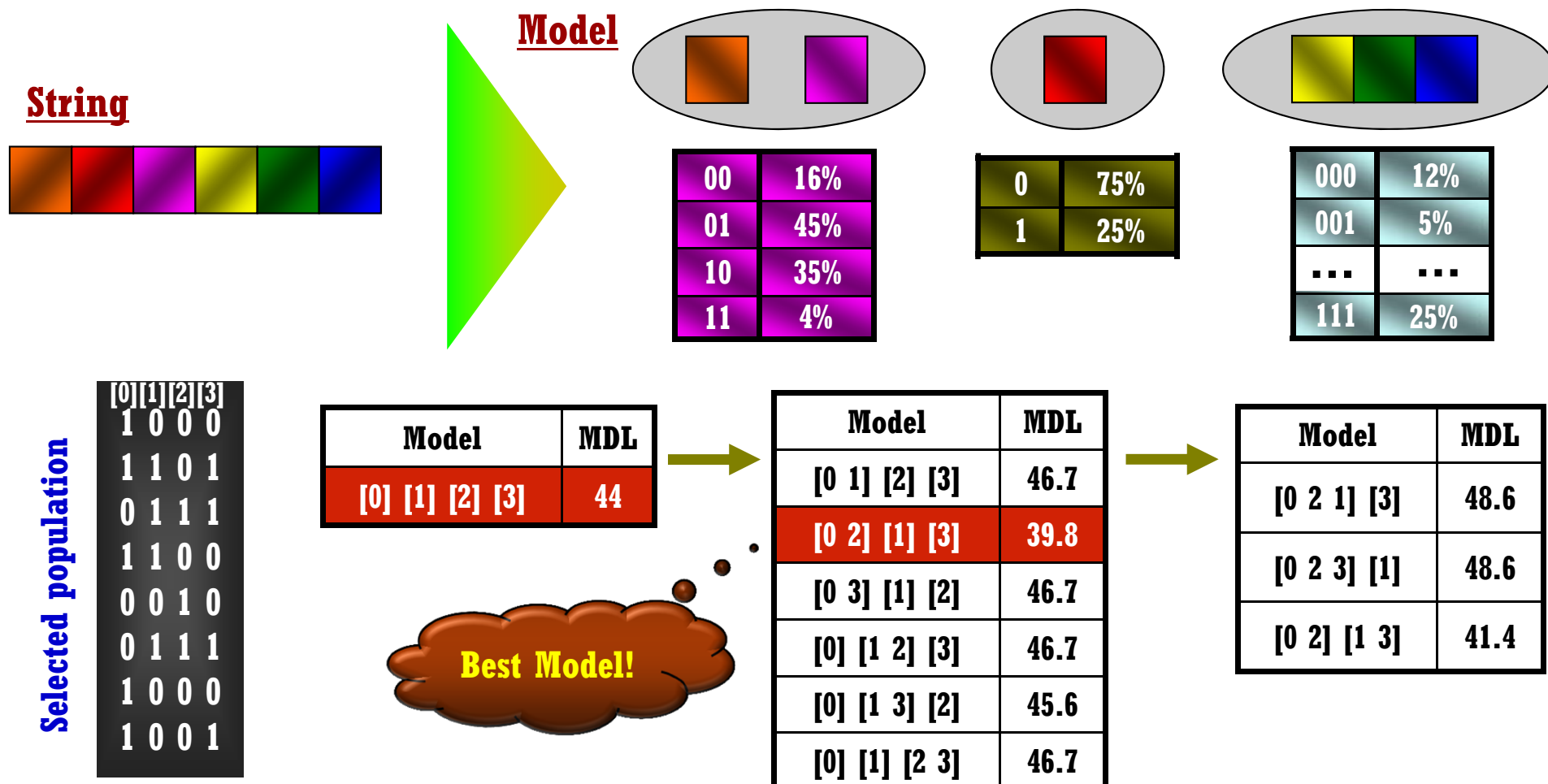
❖ **Bayesian Optimization Algorithm: $O(n^{1.55}) \sim O(n^2)$** [Pelikan'99]

➢ **Exploiting multiple interactions using Bayesian Networks**
  ✓ **Learning the structure**: 1) Scoring metrics, 2) Search procedure
  ✓ **Learning the conditional probabilities**
➢ **Sampling the Bayesian network based on forward simulation**

Current Pop.

1	0	1	1	1	0
0	1	1	1	0	1
1	1	0	0	1	0
0	1	0	1	0	1
1	0	0	0	0	1
0	0	1	0	1	0

Selection

Selected Pop.

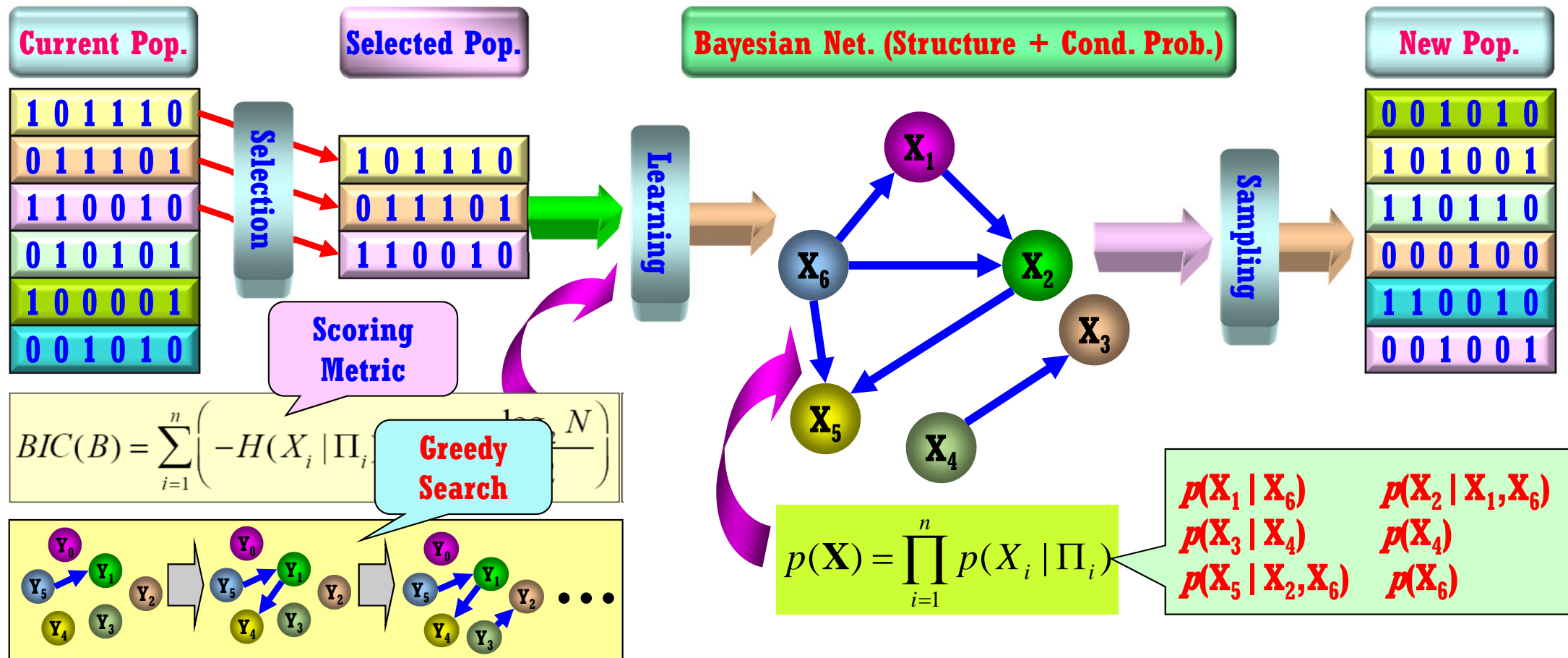|					
1	0	1	1	1	0
0	1	1	1	0	1
1	1	0	0	1	0

Learning

Bayesian Net. (Structure + Cond. Prob.)

New Pop.

1	0	0	1	0	1
1	0	1	0	0	1
1	1	0	1	1	0
0	0	0	1	0	0
1	1	1	0	0	1
0	0	1	0	0	1

Sampling

Scoring Metric

$$BIC(B) = \sum_{i=1}^{n} \left(-H(X_i \mid \Pi_i) \cdots \frac{\log N} \cdots \right)$$

Greedy Search

$$p(\mathbf{X}) = \prod_{i=1}^{n} p(X_i \mid \Pi_i)$$

$p(X_1 \mid X_6)$ $p(X_2 \mid X_1, X_6)$
$p(X_3 \mid X_4)$ $p(X_4)$
$p(X_5 \mid X_2, X_6)$ $p(X_6)$

❖ **Estimation of Distribution Algorithms (EDAs)**

➢ **A new paradigm of EAs: Evolutionary Computation + Machine Learning**

➢ **They discover problem knowledge and then exploit it for evolution**

➢ **They are very powerful for solving quite hard problems.**

☞ **EDAs are the most notable research issue in EC community.**