Estimation of Distribution Algorithms, an introduction

Ramón Sagarna

Department of Computer Science
University of Birmingham

SEBASE Summer School 2007

Outline

What are EDAs?

- Estimation of Distribution Algorithms (EDAs) = Probabilistic Model-Building Genetic Algorithms = Iterated Density Estimation Algorithms

What are EDAs?

- Estimation of Distribution Algorithms (EDAs) = Probabilistic Model-Building Genetic Algorithms = Iterated Density Estimation Algorithms
- EDAs are usually presented as a family of population based evolutionary algorithms

What are EDAs?

- Estimation of Distribution Algorithms (EDAs) = Probabilistic Model-Building Genetic Algorithms = Iterated Density Estimation Algorithms
- EDAs are usually presented as a family of population based evolutionary algorithms
- Main characteristic: instead of using typical breeding operators, at each step of the search
  - a probability distribution is induced (learnt) from a subset of the population of solutions,
  - this distribution is sampled to generate new solutions

An abstract EDA

GA

1. $D_0 \leftarrow$ Generate and evaluate an initial population of solutions
2. Repeat for $l = 0, 1, 2, ...,$ until a stopping criterion is met
3. $D_l^{Sel} \leftarrow$ Select a subset of solutions from D_l
4. $D_l^{Cro} \leftarrow$ Apply crossover to solutions from D_l^{Sel}
5. $D_l^{Mut} \leftarrow$ Apply mutation to solutions in D_l^{Cro}
6. Evaluate solutions in D_l^{Mut}
7. $D_{l+1} \leftarrow$ Create the new population with solutions from D_l^{Mut} and D_l

An abstract EDA

GA

1. $D_0 \leftarrow$ Generate and evaluate an initial population of solutions
2. Repeat for $l = 0, 1, 2, ...,$ until a stopping criterion is met
3. $D_l^{Sel} \leftarrow$ Select a subset of solutions from D_l
4. $D_l^{Cro} \leftarrow$ Apply crossover to solutions from D_l^{Sel}
5. $D_l^{Mut} \leftarrow$ Apply mutation to solutions in D_l^{Cro}
6. Evaluate solutions in D_l^{Mut}
7. $D_{l+1} \leftarrow$ Create the new population with solutions from D_l^{Mut} and D_l

EDA

1. $D_0 \leftarrow$ Generate and evaluate an initial population of solutions
2. Repeat for $l = 0, 1, 2, ...,$ until a stopping criterion is met
3. $D_l^{Sel} \leftarrow$ Select a subset of solutions from D_l
4. $p_l(\boldsymbol{x}) \leftarrow$ Induce a probability distribution from D_l^{Sel}
5. $D_l^{New} \leftarrow$ Sample solutions from $p_l(\boldsymbol{x})$
6. Evaluate solutions in D_l^{New}
7. $D_{l+1} \leftarrow$ Create the new population with solutions from D_l^{New} and D_l

An abstract EDA

GA

1. $D_0 \leftarrow$ Generate and evaluate an initial population of solutions
2. Repeat for $l = 0, 1, 2, ...,$ until a stopping criterion is met
3. $D_l^{Sel} \leftarrow$ Select a subset of solutions from D_l
4. $D_l^{Cro} \leftarrow$ Apply crossover to solutions from D_l^{Sel}
5. $D_l^{Mut} \leftarrow$ Apply mutation to solutions in D_l^{Cro}
6. Evaluate solutions in D_l^{Mut}
7. $D_{l+1} \leftarrow$ Create the new population with solutions from D_l^{Mut} and D_l

EDA

1. $D_0 \leftarrow$ Generate and evaluate an initial population of solutions
2. Repeat for $l = 0, 1, 2, ...,$ until a stopping criterion is met
3. $D_l^{Sel} \leftarrow$ Select a subset of solutions from D_l
4. $p_l(\boldsymbol{x}) \leftarrow$ Induce a probability distribution from D_l^{Sel}
5. $D_l^{New} \leftarrow$ Sample solutions from $p_l(\boldsymbol{x})$
6. Evaluate solutions in D_l^{New}
7. $D_{l+1} \leftarrow$ Create the new population with solutions from D_l^{New} and D_l

Notation

- X is a random variable, x is a possible value of X
- $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is an n-dimensional random variable, $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ is a possible configuration of $\boldsymbol{X}$

We restrict to discrete variables

Notation

- X is a random variable, x is a possible value of X
- $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is an n-dimensional random variable, $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ is a possible configuration of $\boldsymbol{X}$

We restrict to discrete variables

- $p(x) = p(X = x)$ denotes the probability distribution (probability mass function) of X
- $p(\boldsymbol{x}) = p(\boldsymbol{X} = \boldsymbol{x})$ denotes the joint probability distribution (probability mass function) of $\boldsymbol{X}$
- Given $s \subseteq \{1, \ldots, n\}$, $p(\boldsymbol{x}_s) = p(\boldsymbol{X}_s = \boldsymbol{x}_s)$ is the marginal probability distribution of $\boldsymbol{X}_s$
- Given X_i,X_j, $p(x_i|x_j) = p(X_i = x_i|X_j = x_j)$ denotes the probability of X_i given $X_j = x_j$
- Given $\boldsymbol{X}_i$,$\boldsymbol{X}_j$, $p(\boldsymbol{x}_i|\boldsymbol{x}_j) = p(\boldsymbol{X}_i = \boldsymbol{x}_i|\boldsymbol{X}_j = \boldsymbol{x}_j)$ denotes the probability of $\boldsymbol{X}_i$ given $\boldsymbol{X}_j = \boldsymbol{x}_j$

A *parameter* is a numerical characterisation of the population such that the probability distribution is partially or completely described

Notation

- X is a random variable, x is a possible value of X
- $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is an n-dimensional random variable, $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ is a possible configuration of $\boldsymbol{X}$

We restrict to discrete variables

- $p(x) = p(X = x)$ denotes the probability distribution (probability mass function) of X
- $p(\boldsymbol{x}) = p(\boldsymbol{X} = \boldsymbol{x})$ denotes the joint probability distribution (probability mass function) of $\boldsymbol{X}$
- Given $s \subseteq \{1, \ldots, n\}$, $p(\boldsymbol{x}_s) = p(\boldsymbol{X}_s = \boldsymbol{x}_s)$ is the marginal probability distribution of $\boldsymbol{X}_s$
- Given X_i,X_j, $p(x_i|x_j) = p(X_i = x_i|X_j = x_j)$ denotes the probability of X_i given $X_j = x_j$
- Given $\boldsymbol{X}_i$,$\boldsymbol{X}_j$, $p(\boldsymbol{x}_i|\boldsymbol{x}_j) = p(\boldsymbol{X}_i = \boldsymbol{x}_i|\boldsymbol{X}_j = \boldsymbol{x}_j)$ denotes the probability of $\boldsymbol{X}_i$ given $\boldsymbol{X}_j = \boldsymbol{x}_j$

A *parameter* is a numerical characterisation of the population such that the probability distribution is partially or completely described

Informally, we call *empirical* distribution to the joint probability distribution of X or $\boldsymbol{X}$ in an actual sequence of experiments

A less abstract EDA

EDA

1. $D_0 \leftarrow$ Generate and evaluate an initial population of solutions
2. Repeat for $l = 0, 1, 2, ...,$ until a stopping criterion is met
3. $D_l^{Sel} \leftarrow$ Select a subset of solutions from D_l
4. $p_l(\boldsymbol{x}) \leftarrow$ Induce a probability distribution from D_l^{Sel}
5. $D_l^{New} \leftarrow$ Sample solutions from $p_l(\boldsymbol{x})$
6. Evaluate solutions in D_l^{New}
7. $D_{l+1} \leftarrow$ Create the new population with solutions from D_l^{New} and D_l

A less abstract EDA

EDA

1. $D_0 \leftarrow$ Generate and evaluate an initial population of solutions
2. Repeat for $l = 0, 1, 2, ...,$ until a stopping criterion is met
3. $D_l^{Sel} \leftarrow$ Select a subset of solutions from D_l
4. $p_l(\boldsymbol{x}) \leftarrow$ Induce a probability distribution from D_l^{Sel}
5. $D_l^{New} \leftarrow$ Sample solutions from $p_l(\boldsymbol{x})$
6. Evaluate solutions in D_l^{New}
7. $D_{l+1} \leftarrow$ Create the new population with solutions from D_l^{New} and D_l

A less abstract EDA

EDA

1. $D_0 \leftarrow$ Generate and evaluate an initial population of solutions
2. Repeat for $l = 0, 1, 2, ...,$ until a stopping criterion is met
3. $D_l^{Sel} \leftarrow$ Select a subset of solutions from D_l
4. $p_l(\boldsymbol{x}) \leftarrow$ Estimate a probability distribution from D_l^{Sel}
5. $D_l^{New} \leftarrow$ Sample solutions from $p_l(\boldsymbol{x})$
6. Evaluate solutions in D_l^{New}
7. $D_{l+1} \leftarrow$ Create the new population with solutions from D_l^{New} and D_l

An example

$$f(\boldsymbol{x}) = |\prod_{i=1}^{6} a_i^{x_i} - \prod_{i=1}^{6} a_i^{(1-x_i)}|$$

, with $(a_1, a_2, a_3, a_4, a_5, a_6) = (1.3, 0.2, 6.1, 4.9, 2.5, 8.3)$ and $X_i \in \{0, 1\}$

An example

$$f(\boldsymbol{x}) = |\prod_{i=1}^{6} a_i^{x_i} - \prod_{i=1}^{6} a_i^{(1-x_i)}|$$

, with $(a_1, a_2, a_3, a_4, a_5, a_6) = (1.3, 0.2, 6.1, 4.9, 2.5, 8.3)$ and $X_i \in \{0, 1\}$

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	1	0	0	0	1	0	46.367
2	0	1	0	0	0	1	95.483
3	0	0	0	0	1	1	12.979
4	0	1	1	1	1	1	122.744
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
7	0	0	0	1	1	1	100.089
8	1	0	0	1	1	1	130.958
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
11	0	1	0	0	0	0	806.083
12	1	1	1	0	0	1	100.089
13	0	0	0	0	0	1	11.129
14	1	1	0	1	0	0	125.301
15	0	0	1	1	0	1	247.437
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
18	0	1	0	0	0	1	95.483
19	0	0	0	0	0	0	160.257
20	0	0	1	1	0	0	24.495

An example

$$f(\boldsymbol{x}) = |\prod_{i=1}^{6} a_i^{x_i} - \prod_{i=1}^{6} a_i^{(1-x_i)}|$$

, with $(a_1, a_2, a_3, a_4, a_5, a_6) = (1.3, 0.2, 6.1, 4.9, 2.5, 8.3)$ and $X_i \in \{0, 1\}$

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	1	0	0	0	1	0	46.367
2	0	1	0	0	0	1	95.483
3	0	0	0	0	1	1	12.979
4	0	1	1	1	1	1	122.744
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
7	0	0	0	1	1	1	100.089
8	1	0	0	1	1	1	130.958
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
11	0	1	0	0	0	0	806.083
12	1	1	1	0	0	1	100.089
13	0	0	0	0	0	1	11.129
14	1	1	0	1	0	0	125.301
15	0	0	1	1	0	1	247.437
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
18	0	1	0	0	0	1	95.483
19	0	0	0	0	0	0	160.257
20	0	0	1	1	0	0	24.495

Example1 - estimate a product of univariate marginals

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

Example1 - estimate a product of univariate marginals

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6) = p(x_1) \cdot p(x_2) \cdot p(x_3) \cdot p(x_4) \cdot p(x_5) \cdot p(x_6), \quad X \sim Bernoulli$$

Example1 - estimate a product of univariate marginals

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\mathbf{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\mathbf{x}) = p(x_1, \ldots, x_6) = p(x_1) \cdot p(x_2) \cdot p(x_3) \cdot p(x_4) \cdot p(x_5) \cdot p(x_6), \quad X \sim Bernoulli$$

$$p(X_1 = 1) =$$

Example1 - estimate a product of univariate marginals

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\mathbf{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\mathbf{x}) = p(x_1, \ldots, x_6) = p(x_1) \cdot p(x_2) \cdot p(x_3) \cdot p(x_4) \cdot p(x_5) \cdot p(x_6), \quad X \sim Bernoulli$$

$$p(X_1 = 1) =$$

Example1 - estimate a product of univariate marginals

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\mathbf{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\mathbf{x}) = p(x_1, \ldots, x_6) = p(x_1) \cdot p(x_2) \cdot p(x_3) \cdot p(x_4) \cdot p(x_5) \cdot p(x_6), \quad X \sim Bernoulli$$

$$p(X_1 = 1) = \frac{4}{10}$$

Example1 - estimate a product of univariate marginals

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\mathbf{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\mathbf{x}) = p(x_1, \dots, x_6) = p(x_1) \cdot p(x_2) \cdot p(x_3) \cdot p(x_4) \cdot p(x_5) \cdot p(x_6), \quad X \sim Bernoulli$$

$$p(X_1 = 1) = \frac{4}{10} \quad p(X_2 = 1) = \frac{2}{10} \quad p(X_3 = 1) = \frac{4}{10}$$

$$p(X_4 = 1) = \frac{4}{10} \quad p(X_5 = 1) = \frac{4}{10} \quad p(X_6 = 1) = \frac{5}{10}$$

Example1 - sample solutions from a product of univariate marginals

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6) = p(x_1) \cdot p(x_2) \cdot p(x_3) \cdot p(x_4) \cdot p(x_5) \cdot p(x_6)$$

$$p(X_1 = 1) = \frac{4}{10}$$

Example1 - sample solutions from a product of univariate marginals

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6) = p(x_1) \cdot p(x_2) \cdot p(x_3) \cdot p(x_4) \cdot p(x_5) \cdot p(x_6)$$

$$p(X_1 = 1) = \frac{4}{10} \qquad 0.65$$

Example1 - sample solutions from a product of univariate marginals

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6) = p(x_1) \cdot p(x_2) \cdot p(x_3) \cdot p(x_4) \cdot p(x_5) \cdot p(x_6)$$

$$p(X_1 = 1) = \frac{4}{10} \quad < \quad 0.65$$

Example1 - sample solutions from a product of univariate marginals

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6) = p(x_1) \cdot p(x_2) \cdot p(x_3) \cdot p(x_4) \cdot p(x_5) \cdot p(x_6)$$

$$p(X_1 = 1) = \frac{4}{10} \quad < \quad 0.65 \quad \longrightarrow \quad 0$$

Example1 - sample solutions from a product of univariate marginals

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6) = p(x_1) \cdot p(x_2) \cdot p(x_3) \cdot p(x_4) \cdot p(x_5) \cdot p(x_6)$$

$$p(X_1 = 1) = \frac{4}{10} \quad < \quad 0.65 \quad \longrightarrow \quad 0$$

$$p(X_2 = 1) = \frac{2}{10} \quad < \quad 0.43 \quad \longrightarrow \quad 0$$

$$p(X_3 = 1) = \frac{4}{10} \quad < \quad 0.71 \quad \longrightarrow \quad 0$$

$$p(X_4 = 1) = \frac{4}{10} \quad > \quad 0.25 \quad \longrightarrow \quad 1$$

$$p(X_5 = 1) = \frac{4}{10} \quad > \quad 0.38 \quad \longrightarrow \quad 1$$

$$p(X_6 = 1) = \frac{5}{10} \quad < \quad 0.59 \quad \longrightarrow \quad 0$$

Example1 - the new population

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	0	1	1	1	0	0	20.997
2	1	1	0	0	1	1	24.495
3	1	0	0	0	0	1	4.155
4	0	0	0	1	0	1	36.705
5	0	0	0	1	1	0	0.914
6	0	0	0	0	0	1	11.129
7	0	0	1	0	0	0	20.336
8	0	0	0	0	1	1	12.979
9	1	0	0	1	1	0	5.799
10	0	1	1	0	1	1	18.945
11	0	0	0	0	0	1	11.129
12	0	0	1	1	0	0	24.495
13	0	0	1	0	0	0	20.336
14	1	0	0	0	0	1	4.155
15	1	0	0	0	1	1	20.997
16	0	0	0	0	1	1	12.979
17	0	1	1	0	1	1	18.945
18	1	0	0	0	1	0	46.367
19	0	0	0	0	0	1	11.129
20	0	0	1	1	0	0	24.495

Example2 - estimate a product of bivariate marginals

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

Example2 - estimate a product of bivariate marginals

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6) = p(x_1, x_2) \cdot p(x_3, x_4) \cdot p(x_5, x_6), \quad (X, Y) \sim \textit{bivariate Bernoulli}$$

Example2 - estimate a product of bivariate marginals

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6) = p(x_1, x_2) \cdot p(x_3, x_4) \cdot p(x_5, x_6), \quad (X, Y) \sim \textit{bivariate Bernoulli}$$

$$p(X_1 = 0, X_2 = 0) =$$

Example2 - estimate a product of bivariate marginals

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\mathbf{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\mathbf{x}) = p(x_1, \ldots, x_6) = p(x_1, x_2) \cdot p(x_3, x_4) \cdot p(x_5, x_6), \quad (X, Y) \sim \textit{bivariate Bernoulli}$$

$$p(X_1 = 0, X_2 = 0) =$$

Example2 - estimate a product of bivariate marginals

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6) = p(x_1, x_2) \cdot p(x_3, x_4) \cdot p(x_5, x_6), \quad (X, Y) \sim \textit{bivariate Bernoulli}$$

$$p(X_1 = 0, X_2 = 0) = \frac{5}{10}$$

Example2 - estimate a product of bivariate marginals

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6) = p(x_1, x_2) \cdot p(x_3, x_4) \cdot p(x_5, x_6), \quad (X, Y) \sim \textit{bivariate Bernoulli}$$

$$p(X_1 = 0, X_2 = 0) = \frac{5}{10} \qquad p(X_1 = 0, X_2 = 1) = \frac{1}{10} \qquad p(X_1 = 1, X_2 = 0) = \frac{3}{10}$$

Example2 - estimate a product of bivariate marginals

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6) = p(x_1, x_2) \cdot p(x_3, x_4) \cdot p(x_5, x_6), \quad (X, Y) \sim \textit{bivariate Bernoulli}$$

$$p(X_1 = 0, X_2 = 0) = \frac{5}{10} \quad p(X_1 = 0, X_2 = 1) = \frac{1}{10} \quad p(X_1 = 1, X_2 = 0) = \frac{3}{10}$$

$$p(X_3 = 0, X_4 = 0) = \frac{5}{10} \quad p(X_3 = 0, X_4 = 1) = \frac{1}{10} \quad p(X_3 = 1, X_4 = 0) = \frac{1}{10}$$

$$p(X_5 = 0, X_6 = 0) = \frac{3}{10} \quad p(X_5 = 0, X_6 = 1) = \frac{2}{10} \quad p(X_5 = 1, X_6 = 0) = \frac{2}{10}$$

Example2 - sample solutions from a product of bivariate marginals

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6) = p(x_1, x_2) \cdot p(x_3, x_4) \cdot p(x_5, x_6)$$

$$p(X_1 = 0, X_2 = 0) = \tfrac{5}{10}$$
$$p(X_1 = 0, X_2 = 1) = \tfrac{1}{10}$$
$$p(X_1 = 1, X_2 = 0) = \tfrac{3}{10}$$
$$p(X_1 = 1, X_2 = 1) = \tfrac{1}{10}$$

Example2 - sample solutions from a product of bivariate marginals

$$p(\mathbf{x}) = p(x_1, \ldots, x_6) = p(x_1, x_2) \cdot p(x_3, x_4) \cdot p(x_5, x_6)$$

$$p(X_1 = 0, X_2 = 0) = \tfrac{5}{10} \qquad [0, \tfrac{5}{10}]$$

$$p(X_1 = 0, X_2 = 1) = \tfrac{1}{10}$$

$$p(X_1 = 1, X_2 = 0) = \tfrac{3}{10}$$

$$p(X_1 = 1, X_2 = 1) = \tfrac{1}{10}$$

Example2 - sample solutions from a product of bivariate marginals

$$p(\mathbf{x}) = p(x_1, \ldots, x_6) = p(x_1, x_2) \cdot p(x_3, x_4) \cdot p(x_5, x_6)$$

$$p(X_1 = 0, X_2 = 0) = \tfrac{5}{10} \qquad [0, \tfrac{5}{10}]$$

$$p(X_1 = 0, X_2 = 1) = \tfrac{1}{10} \qquad (\tfrac{5}{10}, \tfrac{5}{10} + \tfrac{1}{10}]$$

$$p(X_1 = 1, X_2 = 0) = \tfrac{3}{10}$$

$$p(X_1 = 1, X_2 = 1) = \tfrac{1}{10}$$

Example2 - sample solutions from a product of bivariate marginals

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6) = p(x_1, x_2) \cdot p(x_3, x_4) \cdot p(x_5, x_6)$$

$$p(X_1 = 0, X_2 = 0) = \tfrac{5}{10} \qquad [0, \tfrac{5}{10}]$$
$$p(X_1 = 0, X_2 = 1) = \tfrac{1}{10} \qquad (\tfrac{5}{10}, \tfrac{6}{10}]$$
$$p(X_1 = 1, X_2 = 0) = \tfrac{3}{10}$$
$$p(X_1 = 1, X_2 = 1) = \tfrac{1}{10}$$

Example2 - sample solutions from a product of bivariate marginals

$$p(\boldsymbol{x}) = p(x_1, \dots, x_6) = p(x_1, x_2) \cdot p(x_3, x_4) \cdot p(x_5, x_6)$$

$p(X_1 = 0, X_2 = 0) = \frac{5}{10}$	$[0, \frac{5}{10}]$
$p(X_1 = 0, X_2 = 1) = \frac{1}{10}$	$(\frac{5}{10}, \frac{6}{10}]$
$p(X_1 = 1, X_2 = 0) = \frac{3}{10}$	$(\frac{6}{10}, \frac{9}{10}]$
$p(X_1 = 1, X_2 = 1) = \frac{1}{10}$	$(\frac{9}{10}, 1]$

Example2 - sample solutions from a product of bivariate marginals

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6) = p(x_1, x_2) \cdot p(x_3, x_4) \cdot p(x_5, x_6)$$

$p(X_1 = 0, X_2 = 0) = \frac{5}{10}$	0.35	$[0, \frac{5}{10}]$
$p(X_1 = 0, X_2 = 1) = \frac{1}{10}$		$(\frac{5}{10}, \frac{6}{10}]$
$p(X_1 = 1, X_2 = 0) = \frac{3}{10}$		$(\frac{6}{10}, \frac{9}{10}]$
$p(X_1 = 1, X_2 = 1) = \frac{1}{10}$		$(\frac{9}{10}, 1]$

Example2 - sample solutions from a product of bivariate marginals

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6) = p(x_1, x_2) \cdot p(x_3, x_4) \cdot p(x_5, x_6)$$

$$\begin{array}{lllll}
p(X_1 = 0, X_2 = 0) = \frac{5}{10} & & 0.35 & \in & [0, \frac{5}{10}] \\
p(X_1 = 0, X_2 = 1) = \frac{1}{10} & & & & (\frac{5}{10}, \frac{6}{10}] \\
p(X_1 = 1, X_2 = 0) = \frac{3}{10} & & & & (\frac{6}{10}, \frac{9}{10}] \\
p(X_1 = 1, X_2 = 1) = \frac{1}{10} & & & & (\frac{9}{10}, 1]
\end{array}$$

Example2 - sample solutions from a product of bivariate marginals

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6) = p(x_1, x_2) \cdot p(x_3, x_4) \cdot p(x_5, x_6)$$

$p(X_1 = 0, X_2 = 0) = \frac{5}{10}$	0.35	$\in$	$[0, \frac{5}{10}]$	$\longrightarrow$	0 0
$p(X_1 = 0, X_2 = 1) = \frac{1}{10}$			$(\frac{5}{10}, \frac{6}{10}]$		
$p(X_1 = 1, X_2 = 0) = \frac{3}{10}$			$(\frac{6}{10}, \frac{9}{10}]$		
$p(X_1 = 1, X_2 = 1) = \frac{1}{10}$			$(\frac{9}{10}, 1]$		

Example2 - sample solutions from a product of bivariate marginals

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6) = p(x_1, x_2) \cdot p(x_3, x_4) \cdot p(x_5, x_6)$$

$p(X_1 = 0, X_2 = 0) = \frac{5}{10}$	0.35	$\in$	$[0, \frac{5}{10}]$	$\longrightarrow$	0 0
$p(X_1 = 0, X_2 = 1) = \frac{1}{10}$			$(\frac{5}{10}, \frac{6}{10}]$		
$p(X_1 = 1, X_2 = 0) = \frac{3}{10}$			$(\frac{6}{10}, \frac{9}{10}]$		
$p(X_1 = 1, X_2 = 1) = \frac{1}{10}$			$(\frac{9}{10}, 1]$		
$p(X_3 = 0, X_4 = 0) = \frac{5}{10}$	0.17	$\in$	$[0, \frac{5}{10}]$	$\longrightarrow$	0 0
$p(X_3 = 0, X_4 = 1) = \frac{1}{10}$			$(\frac{5}{10}, \frac{6}{10}]$		
$p(X_3 = 1, X_4 = 0) = \frac{1}{10}$			$(\frac{6}{10}, \frac{7}{10}]$		
$p(X_3 = 1, X_4 = 1) = \frac{3}{10}$			$(\frac{7}{10}, 1]$		
$p(X_5 = 0, X_6 = 0) = \frac{3}{10}$			$[0, \frac{3}{10}]$		
$p(X_5 = 0, X_6 = 1) = \frac{2}{10}$			$(\frac{3}{10}, \frac{5}{10}]$		
$p(X_5 = 1, X_6 = 0) = \frac{2}{10}$			$(\frac{5}{10}, \frac{7}{10}]$		
$p(X_5 = 1, X_6 = 1) = \frac{3}{10}$	0.82	$\in$	$(\frac{7}{10}, 1]$	$\longrightarrow$	1 1

Example2 - the new population

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	0	0	1	1	0	0	24.495
2	0	0	0	0	1	1	12.979
3	0	0	0	0	0	1	11.129
4	1	0	0	0	0	1	4.155
5	0	1	1	1	1	1	122.744
6	1	0	0	0	1	1	20.997
7	0	0	0	1	0	1	36.705
8	0	0	1	0	0	0	20.336
9	1	0	1	0	1	0	11.691
0	0	0	0	1	1	0	0.914
11	1	1	0	0	1	1	24.495
12	0	0	0	0	0	0	160.257
13	0	0	0	1	0	1	36.705
14	1	0	0	0	0	1	4.155
15	1	1	0	0	0	0	12.979
16	0	0	0	0	0	1	11.129
17	0	0	0	0	1	0	62.003
18	1	0	0	1	1	0	5.799
19	0	0	0	1	0	0	28.01
20	1	0	1	1	0	0	34.707

Example3 - estimate a multivariate distribution

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

Example3 - estimate a multivariate distribution

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6), \quad \boldsymbol{X} \sim \textit{multivariate Bernoulli}$$

Example3 - estimate a multivariate distribution

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6), \quad \boldsymbol{X} \sim \textit{multivariate Bernoulli}$$

$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 0, X_6 = 0) =$

Example3 - estimate a multivariate distribution

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6), \quad \boldsymbol{X} \sim \textit{multivariate Bernoulli}$$

$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 0, X_6 = 0) =$

Example3 - estimate a multivariate distribution

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\mathbf{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\mathbf{x}) = p(x_1, \ldots, x_6), \quad \boldsymbol{X} \sim \textit{multivariate Bernoulli}$$

$$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 0, X_6 = 0) = \frac{0}{10}$$

Example3 - estimate a multivariate distribution

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\mathbf{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\mathbf{x}) = p(x_1, \ldots, x_6), \quad \mathbf{X} \sim \textit{multivariate Bernoulli}$$

$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 0, X_6 = 0) = \frac{0}{10}$
$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 0) = \frac{0}{10}$
$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 1, X_5 = 0, X_6 = 0) = \frac{0}{10}$
$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 1, X_5 = 1, X_6 = 0) = \frac{0}{10}$
$p(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 0, X_5 = 0, X_6 = 0) = \frac{1}{10}$
$p(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 0, X_5 = 1, X_6 = 0) = \frac{0}{10}$
$p(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 0) = \frac{1}{10}$
$p(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 1, X_6 = 0) = \frac{0}{10}$
$p(X_1 = 0, X_2 = 1, X_3 = 0, X_4 = 0, X_5 = 0, X_6 = 0) = \frac{0}{10}$
$p(X_1 = 0, X_2 = 1, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 0) = \frac{0}{10}$

$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 0, X_6 = 1) = \frac{1}{10}$
$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{1}{10}$
$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 1, X_5 = 0, X_6 = 1) = \frac{1}{10}$
$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 1, X_5 = 1, X_6 = 1) = \frac{0}{10}$
$p(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 0, X_5 = 0, X_6 = 1) = \frac{0}{10}$
$p(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{0}{10}$
$p(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 1) = \frac{0}{10}$
$p(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 1, X_6 = 1) = \frac{0}{10}$
$p(X_1 = 0, X_2 = 1, X_3 = 0, X_4 = 0, X_5 = 0, X_6 = 1) = \frac{0}{10}$
$p(X_1 = 0, X_2 = 1, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{0}{10}$

Example3 - estimate a multivariate distribution

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6), \quad \boldsymbol{X} \sim \textit{multivariate Bernoulli}$$

$$p(X_1 = 0, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 0, X_6 = 0) = \frac{0}{10}$$
$$p(X_1 = 0, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 1, X_6 = 0) = \frac{0}{10}$$
$$p(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 0, X_6 = 0) = \frac{0}{10}$$
$$p(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 1, X_6 = 0) = \frac{0}{10}$$
$$p(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 0) = \frac{1}{10}$$
$$p(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 1, X_6 = 0) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 0, X_6 = 1) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{1}{10}$$
$$p(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 1, X_5 = 0, X_6 = 1) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 1, X_5 = 1, X_6 = 1) = \frac{0}{10}$$

$$p(X_1 = 0, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 0, X_6 = 1) = \frac{0}{10}$$
$$p(X_1 = 0, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 1, X_6 = 1) = \frac{0}{10}$$
$$p(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 0, X_6 = 1) = \frac{0}{10}$$
$$p(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{0}{10}$$
$$p(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 1) = \frac{0}{10}$$
$$p(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 1, X_6 = 1) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 0) = \frac{1}{10}$$
$$p(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 1, X_5 = 0, X_6 = 0) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 1, X_5 = 1, X_6 = 0) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 0, X_5 = 0, X_6 = 0) = \frac{0}{10}$$

Example3 - estimate a multivariate distribution

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\mathbf{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\mathbf{x}) = p(x_1, \ldots, x_6), \quad \mathbf{X} \sim \textit{multivariate Bernoulli}$$

$$p(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 0, X_5 = 0, X_6 = 1) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 1) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 1, X_6 = 1) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0, X_5 = 0, X_6 = 1) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{1}{10}$$
$$p(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 0, X_6 = 1) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 1, X_6 = 1) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 0, X_6 = 1) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{0}{10}$$

$$p(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 0, X_5 = 1, X_6 = 0) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 0) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 1, X_6 = 0) = \frac{1}{10}$$
$$p(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0, X_5 = 0, X_6 = 0) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 0) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 0, X_6 = 0) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 1, X_6 = 0) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 0, X_6 = 0) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 1, X_6 = 0) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 0) = \frac{0}{10}$$

Example3 - estimate a multivariate distribution

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	1	0	0	0	1	0	46.367
3	0	0	0	0	1	1	12.979
5	1	0	0	0	1	1	20.997
6	1	0	1	1	1	0	95.483
9	0	0	1	0	0	0	20.336
10	1	1	0	0	1	1	24.495
13	0	0	0	0	0	1	11.129
16	0	1	1	1	0	0	20.997
17	0	0	0	1	0	1	36.705
20	0	0	1	1	0	0	24.495

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6), \quad \boldsymbol{X} \sim \textit{multivariate Bernoulli}$$

$$p(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 1) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 1, X_6 = 1) = \frac{0}{10}$$
$$p(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 1, X_6 = 0) = \frac{0}{10}$$

Example3 - sample solutions from a multivariate distribution

$$p(\mathbf{x}) = p(x_1, \ldots, x_6)$$

$$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 0, X_6 = 1) = \frac{1}{10}$$
$$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{1}{10}$$
$$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 1, X_5 = 0, X_6 = 1) = \frac{1}{10}$$
$$p(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 0, X_5 = 0, X_6 = 0) = \frac{1}{10}$$
$$p(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 0) = \frac{1}{10}$$
$$p(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 0) = \frac{1}{10}$$
$$p(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 0) = \frac{1}{10}$$
$$p(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{1}{10}$$
$$p(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 1, X_6 = 0) = \frac{1}{10}$$
$$p(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{1}{10}$$

Example3 - sample solutions from a multivariate distribution

$$p(\mathbf{x}) = p(x_1, \ldots, x_6)$$

$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 0, X_6 = 1) = \frac{1}{10}$	$[0, \frac{1}{10}]$
$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{1}{10}$	$(\frac{1}{10}, \frac{2}{10}]$
$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 1, X_5 = 0, X_6 = 1) = \frac{1}{10}$	$(\frac{2}{10}, \frac{3}{10}]$
$p(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 0, X_5 = 0, X_6 = 0) = \frac{1}{10}$	$(\frac{3}{10}, \frac{4}{10}]$
$p(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 0) = \frac{1}{10}$	$(\frac{4}{10}, \frac{5}{10}]$
$p(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 0) = \frac{1}{10}$	$(\frac{5}{10}, \frac{6}{10}]$
$p(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 0) = \frac{1}{10}$	$(\frac{6}{10}, \frac{7}{10}]$
$p(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{1}{10}$	$(\frac{7}{10}, \frac{8}{10}]$
$p(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 1, X_6 = 0) = \frac{1}{10}$	$(\frac{8}{10}, \frac{9}{10}]$
$p(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{1}{10}$	$(\frac{9}{10}, 1]$

Example3 - sample solutions from a multivariate distribution

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6)$$

$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 0, X_6 = 1) = \frac{1}{10}$		$[0, \frac{1}{10}]$
$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{1}{10}$		$(\frac{1}{10}, \frac{2}{10}]$
$p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 1, X_5 = 0, X_6 = 1) = \frac{1}{10}$	0.22	$(\frac{2}{10}, \frac{3}{10}]$
$p(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 0, X_5 = 0, X_6 = 0) = \frac{1}{10}$		$(\frac{3}{10}, \frac{4}{10}]$
$p(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 0) = \frac{1}{10}$		$(\frac{4}{10}, \frac{5}{10}]$
$p(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 0) = \frac{1}{10}$		$(\frac{5}{10}, \frac{6}{10}]$
$p(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 0) = \frac{1}{10}$		$(\frac{6}{10}, \frac{7}{10}]$
$p(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{1}{10}$		$(\frac{7}{10}, \frac{8}{10}]$
$p(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 1, X_6 = 0) = \frac{1}{10}$		$(\frac{8}{10}, \frac{9}{10}]$
$p(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{1}{10}$		$(\frac{9}{10}, 1]$

Example3 - sample solutions from a multivariate distribution

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6)$$

$$
\begin{array}{lcr}
p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 0, X_6 = 1) = \frac{1}{10} & & [0, \frac{1}{10}] \\
p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{1}{10} & & (\frac{1}{10}, \frac{2}{10}] \\
p(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 1, X_5 = 0, X_6 = 1) = \frac{1}{10} & 0.22 & \in (\frac{2}{10}, \frac{3}{10}] \\
p(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 0, X_5 = 0, X_6 = 0) = \frac{1}{10} & & (\frac{3}{10}, \frac{4}{10}] \\
p(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 0) = \frac{1}{10} & & (\frac{4}{10}, \frac{5}{10}] \\
p(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 0) = \frac{1}{10} & & (\frac{5}{10}, \frac{6}{10}] \\
p(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 0) = \frac{1}{10} & & (\frac{6}{10}, \frac{7}{10}] \\
p(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{1}{10} & & (\frac{7}{10}, \frac{8}{10}] \\
p(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 1, X_6 = 0) = \frac{1}{10} & & (\frac{8}{10}, \frac{9}{10}] \\
p(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0, X_5 = 1, X_6 = 1) = \frac{1}{10} & & (\frac{9}{10}, 1]
\end{array}
$$

Example3 - sample solutions from a multivariate distribution

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_6)$$

$$\begin{array}{ll}
p(X_1=0, X_2=0, X_3=0, X_4=0, X_5=0, X_6=1) = \frac{1}{10} & [0, \frac{1}{10}] \\
p(X_1=0, X_2=0, X_3=0, X_4=0, X_5=1, X_6=1) = \frac{1}{10} & (\frac{1}{10}, \frac{2}{10}] \\
p(X_1=0, X_2=0, X_3=0, X_4=1, X_5=0, X_6=1) = \frac{1}{10} \quad 0.22 \in & (\frac{2}{10}, \frac{3}{10}] \longrightarrow 0\ 0\ 0\ 1\ 0\ 1 \\
p(X_1=0, X_2=0, X_3=1, X_4=0, X_5=0, X_6=0) = \frac{1}{10} & (\frac{3}{10}, \frac{4}{10}] \\
p(X_1=0, X_2=0, X_3=1, X_4=1, X_5=0, X_6=0) = \frac{1}{10} & (\frac{4}{10}, \frac{5}{10}] \\
p(X_1=0, X_2=1, X_3=1, X_4=1, X_5=0, X_6=0) = \frac{1}{10} & (\frac{5}{10}, \frac{6}{10}] \\
p(X_1=1, X_2=0, X_3=0, X_4=0, X_5=1, X_6=0) = \frac{1}{10} & (\frac{6}{10}, \frac{7}{10}] \\
p(X_1=1, X_2=0, X_3=0, X_4=0, X_5=1, X_6=1) = \frac{1}{10} & (\frac{7}{10}, \frac{8}{10}] \\
p(X_1=1, X_2=0, X_3=1, X_4=1, X_5=1, X_6=0) = \frac{1}{10} & (\frac{8}{10}, \frac{9}{10}] \\
p(X_1=1, X_2=1, X_3=0, X_4=0, X_5=1, X_6=1) = \frac{1}{10} & (\frac{9}{10}, 1]
\end{array}$$

Example3 - the new population

	X_1	X_2	X_3	X_4	X_5	X_6	$f(\boldsymbol{x})$
1	0	0	1	0	0	0	20.336
2	1	1	0	0	1	1	24.495
3	0	0	0	1	0	1	36.705
4	1	0	0	0	1	1	20.997
5	0	0	0	0	1	1	12.979
6	0	1	1	1	0	0	20.997
7	1	1	0	0	1	1	24.495
8	1	0	1	1	1	0	95.483
9	1	0	0	0	1	0	46.367
10	0	0	0	0	0	1	11.129
11	0	0	1	1	0	0	24.495
12	0	0	0	0	1	1	12.979
13	0	1	1	1	0	0	20.997
14	1	0	1	1	1	0	95.483
15	1	0	0	0	1	0	46.367
16	0	0	0	1	0	1	36.705
17	0	0	1	0	0	0	20.336
18	0	0	0	0	0	1	11.129
19	0	0	1	1	0	0	24.495
20	0	0	0	0	1	1	12.979

Outline

The need for factorizing

$$X \text{ and } Y \text{ independent} \quad \Rightarrow \quad p(X, Y) = p(X) \cdot p(Y)$$

The need for factorizing

X and Y independent $\quad \Rightarrow \quad p(X, Y) = p(X) \cdot p(Y)$

Why not dealing with the joint probability distribution?

The need for factorizing

$$X \text{ and } Y \text{ independent} \quad \Rightarrow \quad p(X,Y) = p(X) \cdot p(Y)$$

Why not dealing with the joint probability distribution?

- Given $\boldsymbol{X} = (X_1, X_2, ..., X_n)$, $X_i \in \{0,1\}$, $p(\boldsymbol{x})$ has $2^n - 1$ parameters
- Example1: 1 parameter Example2: 3 parameters Example3: 63 parameters

The need for factorizing

$$X \text{ and } Y \text{ independent} \quad \Rightarrow \quad p(X,Y) = p(X) \cdot p(Y)$$

Why not dealing with the joint probability distribution?

- Given $\boldsymbol{X} = (X_1, X_2, ..., X_n)$, $X_i \in \{0,1\}$, $p(\boldsymbol{x})$ has $2^n - 1$ parameters
- Example1: 1 parameter Example2: 3 parameters Example3: 63 parameters

Do we need to calculate all the parameters?

The need for factorizing

$$X \text{ and } Y \text{ independent} \quad \Rightarrow \quad p(X,Y) = p(X) \cdot p(Y)$$

Why not dealing with the joint probability distribution?

- Given $\boldsymbol{X} = (X_1, X_2, ..., X_n)$, $X_i \in \{0, 1\}$, $p(\boldsymbol{x})$ has $2^n - 1$ parameters
- Example1: 1 parameter   Example2: 3 parameters   Example3: 63 parameters

Do we need to calculate all the parameters?

- We are estimating

The need for factorizing

$$X \text{ and } Y \text{ independent} \quad \Rightarrow \quad p(X,Y) = p(X) \cdot p(Y)$$

Why not dealing with the joint probability distribution?

- Given $\boldsymbol{X} = (X_1, X_2, ..., X_n)$, $X_i \in \{0,1\}$, $p(\boldsymbol{x})$ has $2^n - 1$ parameters
- Example1: 1 parameter Example2: 3 parameters Example3: 63 parameters

Do we need to calculate all the parameters?

- We are estimating
- $MSE(\hat{\Theta}) = Var(\hat{\Theta}) + [\Theta - E(\hat{\Theta})]^2$
- Desirable properties for an estimator, e.g. unbiasness, efficiency, consistency
- Often, properties are attained asymptotically, e.g. maximum likelihood estimators
- Number of required sample points is exponential in n, e.g. for Boltzmann distribution, about $20 \cdot (2^n - 1)$ [Vapnik, 98]

Factorization of the probability distribution

$$p(\boldsymbol{x}) = \prod_{i=1}^{k} \psi_i(\boldsymbol{x}_{S_i}), S_i \subseteq \{1, 2, \ldots, n\}$$

Factorization of the probability distribution

$$p(\boldsymbol{x}) = \prod_{i=1}^{k} \psi_i(\boldsymbol{x}_{S_i}), S_i \subseteq \{1, 2, \ldots, n\}$$

- Probabilistic graphical models: independence graph G (qualitative part) + potentials Ψ (quantitative part)
- E.g., G undirected and $\Psi = \{\psi_1(c_1), \cdots, \psi_k(c_k)\}$, C_i is a clique $\longrightarrow$ Markov network
- E.g., G directed acyclic graph and $\Psi = \{p(x_1|\pi_1), \cdots, p(x_n|\pi_n)\}$, Π_i is a set of *parents* of X_i in G $\longrightarrow$ Bayesian network

Factorization of the probability distribution

$$p(\boldsymbol{x}) = \prod_{i=1}^{k} \psi_i(\boldsymbol{x}_{S_i}), S_i \subseteq \{1, 2, \ldots, n\}$$

- Probabilistic graphical models: independence graph G (qualitative part) + potentials Ψ (quantitative part)
- E.g., G undirected and $\Psi = \{\psi_1(c_1), \cdots, \psi_k(c_k)\}$, C_i is a clique $\longrightarrow$ Markov network
- E.g., G directed acyclic graph and $\Psi = \{p(x_1|\pi_1), \cdots, p(x_n|\pi_n)\}$, Π_i is a set of *parents* of X_i in G $\longrightarrow$ Bayesian network

An extra benefit!

- We see the independence relations among variables

Factorization of the probability distribution

$$p(\boldsymbol{x}) = \prod_{i=1}^{k} \psi_i(\boldsymbol{x}_{S_i}), S_i \subseteq \{1, 2, \ldots, n\}$$

- Probabilistic graphical models: independence graph G (qualitative part) + potentials Ψ (quantitative part)
- E.g., G undirected and $\Psi = \{\psi_1(c_1), \cdots, \psi_k(c_k)\}$, C_i is a clique $\longrightarrow$ Markov network
- E.g., G directed acyclic graph and $\Psi = \{p(x_1|\pi_1), \cdots, p(x_n|\pi_n)\}$, Π_i is a set of *parents* of X_i in G $\longrightarrow$ Bayesian network

An extra benefit!

- We see the independence relations among variables

Building (learning) a probabilistic graphical model (from data)

Factorization of the probability distribution

$$p(\boldsymbol{x}) = \prod_{i=1}^{k} \psi_i(\boldsymbol{x}_{s_i}), S_i \subseteq \{1, 2, \ldots, n\}$$

- Probabilistic graphical models: independence graph G (qualitative part) + potentials Ψ (quantitative part)
- E.g., G undirected and $\Psi = \{\psi_1(c_1), \cdots, \psi_k(c_k)\}$, C_i is a clique $\longrightarrow$ Markov network
- E.g., G directed acyclic graph and $\Psi = \{p(x_1|\pi_1), \cdots, p(x_n|\pi_n)\}$, Π_i is a set of *parents* of X_i in G $\longrightarrow$ Bayesian network

An extra benefit!

- We see the independence relations among variables

Building (learning) a probabilistic graphical model (from data)

- Model fitting: G fixed somehow (a-priori knowledge, assumptions), estimate $\psi_i(\boldsymbol{x}_{s_i})$
- Model selection: find a G such that $\psi_i(\boldsymbol{x}_{s_i})$ estimates are *adequate*
  - With dense G, estimates fit data well
  - With sparse G, more reliable estimates

Outline

Classification of EDAs

- At each generation a probabilistic graphical model is learnt

Classification of EDAs

- At each generation a probabilistic graphical model is learnt
- Classification according to the order of interactions between variables which the model allows:

Classification of EDAs

- At each generation a probabilistic graphical model is learnt
- Classification according to the order of interactions between variables which the model allows:
  - Univariate EDAs: $p(\boldsymbol{x}) = \prod_{i=1}^{k} \psi_i(x_{j(i)})$
    all the variables are independent - UMDA, PBIL, CGA

Classification of EDAs

- At each generation a probabilistic graphical model is learnt
- Classification according to the order of interactions between variables which the model allows:
    - Univariate EDAs: $p(\boldsymbol{x}) = \prod_{i=1}^{k} \psi_i(x_{j(i)})$
      all the variables are independent - UMDA, PBIL, CGA
    - Bivariate EDAs: $p(\boldsymbol{x}) = \prod_{i=1}^{k} \psi_i(x_{j_1(i)}, x_{j_2(i)})$
      variables with order one dependencies - MIMIC, TREE, BMDA

Classification of EDAs

- At each generation a probabilistic graphical model is learnt
- Classification according to the order of interactions between variables which the model allows:
    - Univariate EDAs: $p(\boldsymbol{x}) = \prod_{i=1}^{k} \psi_i(x_{j(i)})$
      all the variables are independent - UMDA, PBIL, CGA
    - Bivariate EDAs: $p(\boldsymbol{x}) = \prod_{i=1}^{k} \psi_i(x_{j_1(i)}, x_{j_2(i)})$
      variables with order one dependencies - MIMIC, TREE, BMDA
    - Multivariate EDAs: $p(\boldsymbol{x}) = \prod_{i=1}^{k} \psi_i(\boldsymbol{x}_{\boldsymbol{S}_i})$
      no restriction on the number of dependencies - FDA, EBNA, BOA, LFDA, ECGA

Classification of EDAs

- At each generation a probabilistic graphical model is learnt
- Classification according to the order of interactions between variables which the model allows:
  - Univariate EDAs: $p(\boldsymbol{x}) = \prod_{i=1}^{k} \psi_i(x_{j(i)})$
    all the variables are independent - UMDA, PBIL, CGA
  - Bivariate EDAs: $p(\boldsymbol{x}) = \prod_{i=1}^{k} \psi_i(x_{j_1(i)}, x_{j_2(i)})$
    variables with order one dependencies - MIMIC, TREE, BMDA
  - Multivariate EDAs: $p(\boldsymbol{x}) = \prod_{i=1}^{k} \psi_i(\boldsymbol{x}_{\boldsymbol{S}_i})$
    no restriction on the number of dependencies - FDA, EBNA, BOA, LFDA, ECGA

UMDA

- Univariate Marginal Distribution Algorithm (UMDA) employs a model where variables are independent

UMDA

- Univariate Marginal Distribution Algorithm (UMDA) employs a model where variables are independent
- Distribution factorization:

$$p_l(\boldsymbol{x}) = \prod_{i=1}^{n} p_l(x_i)$$

- Parameters estimation through maximum likelihood (frequency counts in D_l^{Sel}):

$$p_l(x_i) = \frac{1}{N} \sum_{\boldsymbol{x^j} \in D_l^{Sel}} x_i^j$$

UMDA

- Univariate Marginal Distribution Algorithm (UMDA) employs a model where variables are independent
- Distribution factorization:

$$p_l(\boldsymbol{x}) = \prod_{i=1}^{n} p_l(x_i)$$

- Parameters estimation through maximum likelihood (frequency counts in D_l^{Sel}):

$$p_l(x_i) = \frac{1}{N} \sum_{\boldsymbol{x^j} \in D_l^{Sel}} x_i^j$$

- Obtain new solutions by sampling with replacement of each univariate probability distribution (as in Example1)

PBIL

- Population Based Incremental Learning (PBIL) employs a model where variables are independent

PBIL

- Population Based Incremental Learning (PBIL) employs a model where variables are independent
- Distribution factorization:

$$p_l(\boldsymbol{x}) = \prod_{i=1}^{n} p_l(x_i)$$

- Parameters estimation through maximum likelihood and *hebbian* rule:

$$p_l(x_i) = (1-\alpha)p_{l-1}(x_i) + \alpha \frac{1}{N} \sum_{\boldsymbol{x^j} \in D_l^{Sel}} x_i^j$$

PBIL

- Population Based Incremental Learning (PBIL) employs a model where variables are independent
- Distribution factorization:

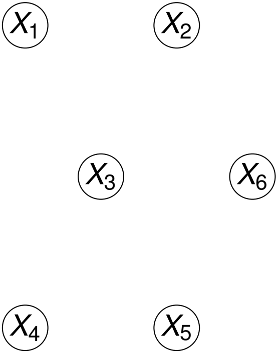
$$p_l(\boldsymbol{x}) = \prod_{i=1}^{n} p_l(x_i)$$

- Parameters estimation through maximum likelihood and *hebbian* rule:

$$p_l(x_i) = (1-\alpha)p_{l-1}(x_i) + \alpha \frac{1}{N} \sum_{\boldsymbol{x}^{\boldsymbol{j}} \in D_l^{Sel}} x_i^j$$

- Obtain new solutions by sampling with replacement of each univariate probability distribution

Factorization example for UMDA and PBIL



$$p_l(\boldsymbol{x}) = p_l(x_1) \cdot p_l(x_2) \cdot p_l(x_3) \cdot p_l(x_4) \cdot p_l(x_5) \cdot p_l(x_6)$$

MIMIC

- Mutual Information Maximizing Input Clustering (MIMIC) uses a first order dependencies model

MIMIC

- Mutual Information Maximizing Input Clustering (MIMIC) uses a first order dependencies model
- Distribution factorization:

$$p_l(\boldsymbol{x}) = \prod_{i=1}^{n-1} p_l(x_{i:n}|x_{(i+1):n}) \cdot p_l(x_{n:n})$$

, with $X_{i:n}$, $i \in \{1, \ldots, n\}$, being a permutation

MIMIC

- Mutual Information Maximizing Input Clustering (MIMIC) uses a first order dependencies model
- Distribution factorization:

$$p_l(\boldsymbol{x}) = \prod_{i=1}^{n-1} p_l(x_{i:n}|x_{(i+1):n}) \cdot p_l(x_{n:n})$$

, with $X_{i:n}$, $i \in \{1, \ldots, n\}$, being a permutation

- Graph selection:
  - Choose the permutation minimizing $KL(p|p_l)$, the Kullback-Leibler divergence from p_l to the empirical distribution p
  - Use a greedy algorithm to search over the $n!$ permutations
- Parameters estimation by frequency counts over the (conditional) probabilities

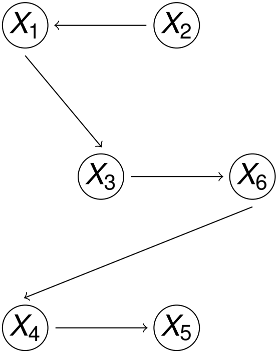
MIMIC

- Mutual Information Maximizing Input Clustering (MIMIC) uses a first order dependencies model
- Distribution factorization:

$$p_l(\boldsymbol{x}) = \prod_{i=1}^{n-1} p_l(x_{i:n}|x_{(i+1):n}) \cdot p_l(x_{n:n})$$

  , with $X_{i:n}$, $i \in \{1, \ldots, n\}$, being a permutation
- Graph selection:
  - Choose the permutation minimizing $KL(p|p_l)$, the Kullback-Leibler divergence from p_l to the empirical distribution p
  - Use a greedy algorithm to search over the $n!$ permutations
- Parameters estimation by frequency counts over the (conditional) probabilities
- Obtain new solutions by sampling following the order given by the permutation

Factorization example of MIMIC



$$p_l(\boldsymbol{x}) = p_l(x_1|x_2) \cdot p_l(x_2) \cdot p_l(x_3|x_1) \cdot p_l(x_4|x_6) \cdot p_l(x_5|x_4) \cdot p_l(x_6|x_3)$$

TREE

- TREE employs a first order dependencies probabilistic graphical model

TREE

- TREE employs a first order dependencies probabilistic graphical model
- Distribution factorization:

$$p_l(\boldsymbol{x}) = \prod_{i=1}^{n} p_l(x_i | x_{j(i)})$$

, with $X_{j(i)}$ the *parent* variable of X_i

TREE

- TREE employs a first order dependencies probabilistic graphical model
- Distribution factorization:

$$p_l(\boldsymbol{x}) = \prod_{i=1}^{n} p_l(x_i | x_{j(i)})$$

, with $X_{j(i)}$ the *parent* variable of X_i

- Graph selection:
  - Chow & Liu algorithm: find the tree structure minimizing $KL(p|p_l) = -\sum_{i=1}^{n} I(X_i, X_{j(i)}) + \sum_{i=1}^{n} H(X_i) - H(\boldsymbol{X})$
  - Exact algorithm for finding the structure maximizing $\sum_{i=1}^{n} I(X_i, X_{j(i)})$
- Parameters through maximum likelihood estimates

TREE

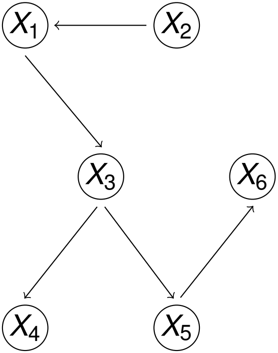
- TREE employs a first order dependencies probabilistic graphical model
- Distribution factorization:

$$p_l(\boldsymbol{x}) = \prod_{i=1}^{n} p_l(x_i|x_{j(i)})$$

, with $X_{j(i)}$ the *parent* variable of X_i

- Graph selection:
  - Chow & Liu algorithm: find the tree structure minimizing $KL(p|p_l) = -\sum_{i=1}^{n} I(X_i, X_{j(i)}) + \sum_{i=1}^{n} H(X_i) - H(\boldsymbol{X})$
  - Exact algorithm for finding the structure maximizing $\sum_{i=1}^{n} I(X_i, X_{j(i)})$
- Parameters through maximum likelihood estimates
- Obtain new solutions by Probabilistic Logic Sampling [Henrion, 88]

Factorization example of TREE



$$p_l(\boldsymbol{x}) = p_l(x_1|x_2) \cdot p_l(x_2) \cdot p_l(x_3|x_1) \cdot p_l(x_4|x_3) \cdot p_l(x_5|x_3) \cdot p_l(x_6|x_5)$$

FDA

- Factorized Distribution Algorithm (FDA) uses a model where multiple dependencies are allowed

FDA

- Factorized Distribution Algorithm (FDA) uses a model where multiple dependencies are allowed
- Designed for additively decomposed functions: $f(\boldsymbol{x}) = \sum_{i=1}^{m} f_i(\boldsymbol{x}_{s_i})$, with $s_i \subseteq \{1, \ldots, n\}$
- E.g., $f(x_1, x_2, x_3, x_4, x_5, x_6) = f_1(x_1, x_2, x_4) + f_2(x_3, x_4, x_6) + f_3(x_1, x_5, x_6)$

FDA

- Factorized Distribution Algorithm (FDA) uses a model where multiple dependencies are allowed
- Designed for additively decomposed functions: $f(\boldsymbol{x}) = \sum_{i=1}^{m} f_i(\boldsymbol{x}_{s_i})$, with $s_i \subseteq \{1, \ldots, n\}$
- E.g., $f(x_1, x_2, x_3, x_4, x_5, x_6) = f_1(x_1, x_2, x_4) + f_2(x_3, x_4, x_6) + f_3(x_1, x_5, x_6)$
- Distribution factorization:
  - Fixed in advance
  - Tries to mimic the structure of the objective function

FDA

- Factorized Distribution Algorithm (FDA) uses a model where multiple dependencies are allowed
- Designed for additively decomposed functions: $f(\boldsymbol{x}) = \sum_{i=1}^{m} f_i(\boldsymbol{x}_{s_i})$, with $s_i \subseteq \{1, \ldots, n\}$
- E.g., $f(x_1, x_2, x_3, x_4, x_5, x_6) = f_1(x_1, x_2, x_4) + f_2(x_3, x_4, x_6) + f_3(x_1, x_5, x_6)$
- Distribution factorization:
  - Fixed in advance
  - Tries to mimic the structure of the objective function
  - Let $d_j = \cup_{i=1}^{j} s_i$, $b_j = s_j \backslash d_{j-1}$, $c_j = s_j \cap d_{j-1}$, for $i = 1, \ldots, m$, with $d_m = \{1, \ldots, n\}$ and $d_0 = \emptyset$
  - E.g., $d_1 = \{1, 2, 4\}$, $d_2 = \{1, 2, 3, 4, 6\}$, $d_3 = \{1, 2, 3, 4, 5, 6\}$

    $b_1 = \{1, 2, 4\}$, $b_2 = \{3, 6\}$, $b_3 = \{5\}$

    $c_1 = \emptyset$, $c_2 = \{4\}$, $c_3 = \{1, 6\}$

FDA

- Factorized Distribution Algorithm (FDA) uses a model where multiple dependencies are allowed
- Designed for additively decomposed functions: $f(\boldsymbol{x}) = \sum_{i=1}^{m} f_i(\boldsymbol{x}_{s_i})$, with $s_i \subseteq \{1, \ldots, n\}$
- E.g., $f(x_1, x_2, x_3, x_4, x_5, x_6) = f_1(x_1, x_2, x_4) + f_2(x_3, x_4, x_6) + f_3(x_1, x_5, x_6)$
- Distribution factorization:
  - Fixed in advance
  - Tries to mimic the structure of the objective function
  - Let $d_j = \cup_{i=1}^{j} s_i$, $b_j = s_j \backslash d_{j-1}$, $c_j = s_j \cap d_{j-1}$, for $i = 1, \ldots, m$, with $d_m = \{1, \ldots, n\}$ and $d_0 = \emptyset$
  - E.g., $d_1 = \{1, 2, 4\}$, $d_2 = \{1, 2, 3, 4, 6\}$, $d_3 = \{1, 2, 3, 4, 5, 6\}$

    $b_1 = \{1, 2, 4\}$, $b_2 = \{3, 6\}$, $b_3 = \{5\}$

    $c_1 = \emptyset$, $c_2 = \{4\}$, $c_3 = \{1, 6\}$
  - If $b_i \neq \emptyset$,

$$p_l(\boldsymbol{x}) = \prod_{i=1}^{m} p_l(\boldsymbol{x}_{b_i} | \boldsymbol{x}_{c_i})$$

FDA

- Factorized Distribution Algorithm (FDA) uses a model where multiple dependencies are allowed
- Designed for additively decomposed functions: $f(\boldsymbol{x}) = \sum_{i=1}^{m} f_i(\boldsymbol{x}_{s_i})$, with $s_i \subseteq \{1, \ldots, n\}$
- E.g., $f(x_1, x_2, x_3, x_4, x_5, x_6) = f_1(x_1, x_2, x_4) + f_2(x_3, x_4, x_6) + f_3(x_1, x_5, x_6)$
- Distribution factorization:
    - Fixed in advance
    - Tries to mimic the structure of the objective function
    - Let $d_j = \cup_{i=1}^{j} s_i$, $b_j = s_j \backslash d_{j-1}$, $c_j = s_j \cap d_{j-1}$, for $i = 1, \ldots, m$, with $d_m = \{1, \ldots, n\}$ and $d_0 = \emptyset$
    - E.g., $d_1 = \{1, 2, 4\}$, $d_2 = \{1, 2, 3, 4, 6\}$, $d_3 = \{1, 2, 3, 4, 5, 6\}$

      $b_1 = \{1, 2, 4\}$, $b_2 = \{3, 6\}$, $b_3 = \{5\}$

      $c_1 = \emptyset$, $c_2 = \{4\}$, $c_3 = \{1, 6\}$
    - If $b_i \neq \emptyset$,

$$p_l(\boldsymbol{x}) = \prod_{i=1}^{m} p_l(\boldsymbol{x}_{b_i} | \boldsymbol{x}_{c_i})$$

- Parameters estimated through maximum likelihood

FDA

- Factorized Distribution Algorithm (FDA) uses a model where multiple dependencies are allowed
- Designed for additively decomposed functions: $f(\boldsymbol{x}) = \sum_{i=1}^{m} f_i(\boldsymbol{x}_{s_i})$, with $s_i \subseteq \{1, \ldots, n\}$
- E.g., $f(x_1, x_2, x_3, x_4, x_5, x_6) = f_1(x_1, x_2, x_4) + f_2(x_3, x_4, x_6) + f_3(x_1, x_5, x_6)$
- Distribution factorization:
  - Fixed in advance
  - Tries to mimic the structure of the objective function
  - Let $d_j = \cup_{i=1}^{j} s_i$, $b_j = s_j \backslash d_{j-1}$, $c_j = s_j \cap d_{j-1}$, for $i = 1, \ldots, m$, with $d_m = \{1, \ldots, n\}$ and $d_0 = \emptyset$
  - E.g., $d_1 = \{1, 2, 4\}$, $d_2 = \{1, 2, 3, 4, 6\}$, $d_3 = \{1, 2, 3, 4, 5, 6\}$

    $b_1 = \{1, 2, 4\}$, $b_2 = \{3, 6\}$, $b_3 = \{5\}$

    $c_1 = \emptyset$, $c_2 = \{4\}$, $c_3 = \{1, 6\}$
  - If $b_i \neq \emptyset$,

$$p_l(\boldsymbol{x}) = \prod_{i=1}^{m} p_l(\boldsymbol{x}_{b_i} | \boldsymbol{x}_{c_i})$$

- Parameters estimated through maximum likelihood
- Obtain new solutions by Probabilistic Logic Sampling

FDA - some ideas

Is this a good factorization?

FDA - some ideas

Is this a good factorization?

- Under the factorization conditions, if data follows Boltzmann distribution p_β and $\forall\, i \geq 2\ \exists\, j < i$ such that $c_i \subseteq s_j$ (RIP) then

$$p_\beta(\boldsymbol{x}) = \prod_{i=1}^{m} p_\beta(\boldsymbol{x}_{b_i} | \boldsymbol{x}_{c_i})$$

- E.g., $c_3 \not\subseteq s_1$ and $c_3 \not\subseteq s_2$

FDA - some ideas

Is this a good factorization?

- Under the factorization conditions, if data follows Boltzmann distribution p_β and $\forall\ i \geq 2\ \exists\ j < i$ such that $c_i \subseteq s_j$ (RIP) then

$$p_\beta(\boldsymbol{x}) = \prod_{i=1}^{m} p_\beta(\boldsymbol{x}_{b_i} | \boldsymbol{x}_{c_i})$$

- E.g., $c_3 \not\subseteq s_1$ and $c_3 \not\subseteq s_2$
- If RIP holds, FDA with Boltzmann selection converges to a global optimum for infinite populations

FDA - some ideas

Is this a good factorization?

- Under the factorization conditions, if data follows Boltzmann distribution p_β and $\forall\ i \geq 2\ \exists\ j < i$ such that $c_i \subseteq s_j$ (RIP) then

$$p_\beta(\boldsymbol{x}) = \prod_{i=1}^{m} p_\beta(\boldsymbol{x}_{b_i}|\boldsymbol{x}_{c_i})$$

- E.g., $c_3 \not\subseteq s_1$ and $c_3 \not\subseteq s_2$
- If RIP holds, FDA with Boltzmann selection converges to a global optimum for infinite populations

The FDA in practice

- Try to fulfil RIP
- If not possible, too difficult, or c_i,b_i not k-bounded then use the *approximate* factorization

EBNA

- Estimation of Bayesian Network Algorithm (EBNA) uses a model where multiple dependencies are allowed

EBNA

- Estimation of Bayesian Network Algorithm (EBNA) uses a model where multiple dependencies are allowed
- Distribution factorization:

$$p_l(\boldsymbol{x}) = \prod_{i=1}^{n} p_l(x_i|\boldsymbol{\pi_i})$$

, $\boldsymbol{\Pi_i}$ the set of *parent* variables of X_i

EBNA

- Estimation of Bayesian Network Algorithm (EBNA) uses a model where multiple dependencies are allowed
- Distribution factorization:

$$p_l(\boldsymbol{x}) = \prod_{i=1}^{n} p_l(x_i|\boldsymbol{\pi_i})$$

  , $\boldsymbol{\Pi_i}$ the set of *parent* variables of X_i
- Induce a Bayesian network from dataset D_l^{Sel}
  - A Bayesian network is a pair $(S, \boldsymbol{\theta})$
  - S is a directed acyclic graph representing the (in)dependencies between the variables
  - $\boldsymbol{\theta}$ is a set of conditional probability values
- Parameters of p_l given by $\boldsymbol{\theta}$

EBNA

- Estimation of Bayesian Network Algorithm (EBNA) uses a model where multiple dependencies are allowed
- Distribution factorization:

$$p_l(\boldsymbol{x}) = \prod_{i=1}^{n} p_l(x_i|\boldsymbol{\pi_i})$$

, $\boldsymbol{\Pi_i}$ the set of *parent* variables of X_i

- Induce a Bayesian network from dataset D_l^{Sel}
  - A Bayesian network is a pair $(S, \boldsymbol{\theta})$
  - S is a directed acyclic graph representing the (in)dependencies between the variables
  - $\boldsymbol{\theta}$ is a set of conditional probability values
- Parameters of p_l given by $\boldsymbol{\theta}$
- Obtain new solutions by Probabilistic Logic Sampling

EBNA - Bayesian network learning

How to learn a Bayesian network from data?

EBNA - Bayesian network learning

How to learn a Bayesian network from data?

- Finding the best network is NP-hard

EBNA - Bayesian network learning

How to learn a Bayesian network from data?

- Finding the best network is NP-hard
- Two steps: structure S induction and parameters θ induction

EBNA - Bayesian network learning

How to learn a Bayesian network from data?

- Finding the best network is NP-hard
- Two steps: structure S induction and parameters $\boldsymbol{\theta}$ induction
- Given S, $\boldsymbol{\theta}$ can be elicited through maximum likelihood estimates

EBNA - Bayesian network learning

How to learn a Bayesian network from data?

- Finding the best network is NP-hard
- Two steps: structure S induction and parameters $\boldsymbol{\theta}$ induction
- Given S, $\boldsymbol{\theta}$ can be elicited through maximum likelihood estimates
- Structure induction: “detect conditional (in)dependencies” or “score + search”

EBNA - Bayesian network learning

How to learn a Bayesian network from data?

- Finding the best network is NP-hard
- Two steps: structure S induction and parameters $\boldsymbol{\theta}$ induction
- Given S, $\boldsymbol{\theta}$ can be elicited through maximum likelihood estimates
- Structure induction: “detect conditional (in)dependencies” or “score + search”
- “score + search”:
  - A score measures the goodness of a structure given the data
  - A search method is used to find the best structure

EBNA - Bayesian network learning

How to learn a Bayesian network from data?

- Finding the best network is NP-hard
- Two steps: structure S induction and parameters $\boldsymbol{\theta}$ induction
- Given S, $\boldsymbol{\theta}$ can be elicited through maximum likelihood estimates
- Structure induction: "detect conditional (in)dependencies" or "score + search"
- "score + search":
  - A score measures the goodness of a structure given the data
  - A search method is used to find the best structure

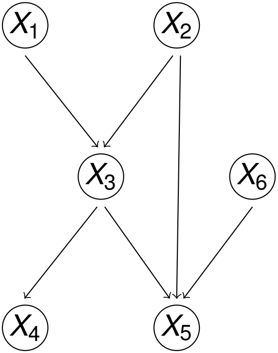
EBNA_{BIC}

- Penalized maximum likelihood score: $\sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{ijk} \log \frac{N_{ijk}}{N_{ij}} - \frac{1}{2} \log N \, dim(S)$
- Hill climbing search

EBNA_{K2+pen}

- Penalized bayesian score: $\log[\prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{(r_i-1)!}{(N_{ij}+r_i-1)!} \prod_{k=1}^{r_i} N_{ijk}!] - \frac{1}{2} \log N \, dim(S)$
- Greedy search

Factorization example of EBNA



$$p_l(\boldsymbol{x}) = p_l(x_1) \cdot p_l(x_2) \cdot p_l(x_3|x_1, x_2) \cdot p_l(x_4|x_3) \cdot p_l(x_5|x_2, x_3, x_6) \cdot p_l(x_6)$$

More on EDAS

Other discrete EDAs

Continuous EDAs

Permutation representations

Multiobjective EDAs

Parallel EDAs

Hybrid EDAs

...

Outline

Motivation from evolutionary computation

EDAs to improve GAs

- Poor results of GA in some functions, e.g. deceptive
- *Linkage* learning is believed to be useful
- Many parameters to be tuned in a GA

Motivation from evolutionary computation

EDAs to improve GAs

- Poor results of GA in some functions, e.g. deceptive
- *Linkage* learning is believed to be useful
- Many parameters to be tuned in a GA

From GAs to EDAs

- Uniform Crossover $\longrightarrow$ get the gene value of a parent chosen randomly
- Bit-based Simulated Crossover $\longrightarrow$ choose the gene value of a parent with probability proportional to its fitness

Motivation from statistical physics

Boltzmann distribution: $p(\boldsymbol{x}) = \frac{e^{\beta f(\boldsymbol{x})}}{\sum_{\boldsymbol{y}} e^{\beta f(\boldsymbol{y})}}$, with $\beta > 0$

Motivation from statistical physics

$$\text{Boltzmann distribution: } p(\boldsymbol{x}) = \frac{e^{\beta f(\boldsymbol{x})}}{\sum_{\boldsymbol{y}} e^{\beta f(\boldsymbol{y})}}, \text{ with } \beta > 0$$

- If $f(\mathbf{x}) > f(\mathbf{y})$ then $p(\mathbf{x}) > p(\mathbf{y})$
- When $\beta \to \infty$, only the optima will have positive probability
- If it was possible to sample the Boltzmann distribution then optimization would be easy
- Problem? There is a exponential number of terms in the denominator

Motivation from statistical physics

Boltzmann distribution: $p(\boldsymbol{x}) = \frac{e^{\beta f(\boldsymbol{x})}}{\sum_{\boldsymbol{y}} e^{\beta f(\boldsymbol{y})}}$, with $\beta > 0$

- If $f(\mathbf{x}) > f(\mathbf{y})$ then $p(\mathbf{x}) > p(\mathbf{y})$
- When $\beta \to \infty$, only the optima will have positive probability
- If it was possible to sample the Boltzmann distribution then optimization would be easy
- Problem? There is a exponential number of terms in the denominator

Boltzmann selection

- Guarantees selected solutions follow a Boltzmann distribution

Motivation from statistical physics

$$\text{Boltzmann distribution: } p(\boldsymbol{x}) = \frac{e^{\beta f(\boldsymbol{x})}}{\sum_{\boldsymbol{y}} e^{\beta f(\boldsymbol{y})}}, \text{ with } \beta > 0$$

- If $f(\mathbf{x}) > f(\mathbf{y})$ then $p(\mathbf{x}) > p(\mathbf{y})$
- When $\beta \rightarrow \infty$, only the optima will have positive probability
- If it was possible to sample the Boltzmann distribution then optimization would be easy
- Problem? There is a exponential number of terms in the denominator

Boltzmann selection

- Guarantees selected solutions follow a Boltzmann distribution

Boltzmann EDA

- At each step, solutions are chosen through Boltzmann selection
- Create new solutions by sampling the obtained Boltzmann distribution
- It has a convergence proof
- Just a conceptional algorithm

Advantages and disadvantages

Advantages

- Able to deal with linkage information
- Learn the current structure of the search space
- Suitable for black-box optimization
- Explicit expression of the relations between variables
- Suitable for the injection of a-priori knowledge
- Less parameters, in general
- Rely on solid theoretical concepts

Advantages and disadvantages

Advantages

- Able to deal with linkage information
- Learn the current structure of the search space
- Suitable for black-box optimization
- Explicit expression of the relations between variables
- Suitable for the injection of a-priori knowledge
- Less parameters, in general
- Rely on solid theoretical concepts

Disadvantages

- Require some degree of expertise
- Not very friendly
- Sometimes harder problems than the one at hand arise
- Lack of flexibility for certain problems, e.g. management of variable length solutions

Advantages and disadvantages

Advantages

- Able to deal with linkage information
- Learn the current structure of the search space
- Suitable for black-box optimization
- Explicit expression of the relations between variables
- Suitable for the injection of a-priori knowledge
- Less parameters, in general
- Rely on solid theoretical concepts

Disadvantages

- Require some degree of expertise
- Not very friendly
- Sometimes harder problems than the one at hand arise
- Lack of flexibility for certain problems, e.g. management of variable length solutions

EDAS are one more method!

Outline

Still many things to do

Relation between problem and probability model

- How to relate the objective function structure and the probabilistic graphical model
- Which is the best model for a given problem?

Diversity

- Is it good to make very accurate estimations?
- Which is the proper balance?

Theoretical work

- Run-time analysis
- Convergence in probability

New EDAs

- Continuous EDAs, variable length solutions
- Island models in EDAs, constraint handling with EDAs

Applications

- Application of EDAs to the real world

Outline

Main points

Three main ideas

- At each step of the search, estimate the probability distribution of promising solutions
- Factorization is a key point $\longrightarrow$ probabilistic graphical models
- Different types of models are possible

Main points

Three main ideas

- At each step of the search, estimate the probability distribution of promising solutions
- Factorization is a key point $\longrightarrow$ probabilistic graphical models
- Different types of models are possible

More generally ...

- EDAs is a highly interdisciplinary field
- Connect evolutionary computation, statistical inference and sampling, information theory, machine learning, optimization ...

Where to learn more from ...

Books

- Pelikan, M., K. Sastry, E. Cantú-Paz (eds.) *Scalable Optimization via Probabilistic Mode ling* Springer-Verlag. 2006.
- Lozano, J.A., P. Larrañaga, I. Inza, E. Bengoetxea (eds.) *Towards a New Evolutionary Computation. Advances on Estimation of Distribution Algorithms* Springer-Verlag. 2006.
- Pelikan, M. *Hierarchical Bayesian Optimization Algorithm: Toward a New Generation of Evolutionary Algorithms* Springer-Verlag. 2005.
- Larrañaga, P., J.A. Lozano (eds.) *Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation*. Kluwer. 2002.

Journal special issues

- Lozano, J.A., Q. Zhang, P. Larrañaga (eds.) *Evolutionary Algorithms based on Probabilistic Models*. IEEE Transactions on Evolutionary Computation.
- Larrañaga, P., J.A. Lozano (eds.) *Estimation of Distribution Algorithms*. Evolutionary Computation 13(1). 2005.
- Larrañaga, P., J.A. Lozano (eds.) *Synergies between Probabilistic Graphical Models and Evolutionary Computation*. International Journal of Approximate Reasoning 31. 2002.

Conference special sessions and workshops

- *Estimation of Distribution Algorithms*. Workshops and track at GECCO.
- *Evolutionary Algorithms based on Probabilistic Models*. Special session at CEC.

Browse the web ...

Estimation of Distribution Algorithms, an introduction

Ramón Sagarna

Department of Computer Science
University of Birmingham

SEBASE Summer School 2007