An Application of Erlang Mixture Distributions to modeling the reproduction mechanism in estimation of distribution algorithms.

Joanna Kołodziej

University of Bielsko - Biała, Department of Mathematics and Computer Science, Bielsko-Biała, Poland, email: jkolodziej@ath.bielsko.pl

Abstract. We present the idea of an application of the mixtures of Erlang distributions in the construction of the recombination mechanism in estimation of distribution algorithms. We analyze main properties of Erlang mixtures and define a new Erlang Mixture Estimation of Distribution Algorithm. We try to compare the efficiencies of ErM-EDA and evolutionary strategy in case of large populations. Some experimental results are presented after simple theoretical studies.

1 Introduction

The modeling of the population distribution in evolutionary algorithms can be a key to analyzing the populations dynamics. The mathematical analysis of the population distributions is not easy because of an unknown analytical form of the empirical density functions.

The empirical distributions of the evolving populations could be estimated by mixtures of continuous distributions. The most convenient estimator is the mixture of Gaussian distributions. The parameters of this mixture can be calibrated by a quasi-Bayesian estimation technique [3] which is rather hard to implement and is time consuming. We proposed in [7] to replace the mixture of Gaussian distributions estimator by the mixture of Erlang distributions estimator.

Good approximation properties of the finite mixture of Erlang distributions are examined and theoretically proved in many papers [5]. It has been successfully used in engineering and business. The applications include queuing systems, reliability assessment, inventory control, computer evaluations and biological studies [13], [2]. The gamma and Erlang distributions are most frequently used to model the lifetime data. This is due to their flexibility in the choice of the shape and scale parameters [5].

In this paper we try to verify the efficiency of the Erlang mixture estimator in case of multiple dimension problems. We define an *Erlang Mixture Estimation of Distribution Algorithm (ErM-EDA)* as a new Estimated of Distribution Algorithm (EDA) [6]. The main goal of EDAs is to extract some global statistical information from a set of selected sample points (population) and build a posterior probability distribution model of promising solutions based on the extracted information. The new solutions are sampled from the model and they replace the current population.

We performed some simple experiments to compare efficiency of our method with the efficiency of the evolutionary algorithm in case of large finite populations.

The remainder of the paper is organized as follows. In section 2 we define a probability model for our algorithm and we introduce an *Erlang Mixture Population Generator* specially designed for this model. In section 3 we define the ErM-EDA algorithm. Results of performed numerical experiments are presented in Section 4. The paper ends with some final remarks.

2 Mixtures of Erlang distributions

The gamma distribution and its special case the Erlang distribution play a crucial role in mathematical statistic and many applied areas.

Definition 2.1. A random variable X is said to have a gamma distribution with parameters (b, λ), if X has the probability density function (pdf)

$$\gamma_{(b)}^{(\lambda)}(x) = \begin{cases} 0 & \text{, if } x \leq 0 \\ \frac{\lambda^b x^{b-1}}{\Gamma(b) e^{\lambda x}} & \text{, otherwise} \end{cases} \tag{1}$$

where $\lambda > 0, b > 0$ are the scale and shape parameters respectively and Γ is the Euler's gamma function.

The Erlang distribution is a special case of the gamma distribution, thus we can define it in the following way:

Definition 2.2. When the gamma distribution has an integral shape parameter m, it is called the Erlang distribution with the probability density function:

$$f_{(m)}^{(\lambda)}(x) = \begin{cases} 0 & \text{, if } x \leq 0 \\ \frac{\lambda^m x^{m-1}}{(m-1)!} e^{-\lambda x} & \text{, otherwise} \end{cases} \tag{2}$$

where $m \in \{1, 2, 3, 4, ...\}$ and λ is a scale parameter.

We denote the random variable with the Erlang distribution by $X \sim Er(m, \lambda)$.

The following are interesting and important properties of the Erlang distribution, $X \sim Er(m, \lambda)$:

1. The mean and variance of x are
$$E(X) = \frac{m}{\lambda}; V(X) = \frac{m}{\lambda^2}.$$
2. The Laplace transform of X is
$$E(exp(sX)) = (1 - s/\lambda)^{-m}.$$
3. X can be defined as a sum of m identically distributed exponential random variables with rate λ.
4. The Erlang distribution has reproductive and infinitely divisible property.

These properties provide very useful theoretical tools when using gamma and Erlang distributions in real applications. For a given mean value of the Erlang random variable varying the value of the shape parameter m is equivalent to varying the variance.

To solve a multimodal optimization problem we can apply a probabilistic model based on the mixtures of the continuous distributions. A finite mixture of univariate Erlang distributions can be defined in the following way:

Definition 2.3. Let X be an a random variable with the distribution which is a mixture of the univariate Erlang distributions. The probability density function of this mixture can be defined by the following formula:

$$g(x) = \begin{cases} 0 & , \quad x \leq 0 \\ \sum_{i=1}^{k} v_i \frac{\lambda_i^{m_i+1} \cdot x_i^{m_i}}{m_i!} \cdot e^{(-\lambda_i \cdot x_i)} & , \quad x > 0 \end{cases} ; \tag{3}$$

where:

k - is the number of parcels, $\lambda_i > 0, (i = 1, ..., k)$ - are the scale parameters of the mixture, $m_i > 0, (i = 1, ..., k)$ - are the shape parameters of the mixture, $v_i > 0$, $(i = 1, ..., k)$ - are mixture weights, such that: $\sum_{i=1}^{k} v_i = 1$.

The concept of mixture can be extended to multiple dimensions by means of the weighted sum of multivariate distributions.

Definition 2.4. The finite mixture of multivariate Erlangs is the probability distribution whose density function is given by the following formula:

$$G(x_1, ..., x_n) = \sum_{i=1}^{k} v_i \cdot f_{\widehat{m^i}}^{\widehat{\lambda^i}}(x_1, ..., x_n) \tag{4}$$

where $f_{\widehat{m^i}}^{\widehat{\lambda^i}}(x_1, ..., x_n)$ is the joint Erlang density function, $\widehat{m^i} = (m_1^i, ..., m_n^i)$ and $\widehat{\lambda^i} = (\lambda_1^i, ..., \lambda_n^i), (i = 1, ..., k)$ are shape and scale parameters for this function.

A particular case of multivariate mixture is the tensor product [10], which is a multivariate distribution built from the weighted sum of the product of n univariate distributions, where n is the number of random variables. The tensor product is nothing other than the mixture of multivariate distributions with independent coordinates. If $n = 1$, then the tensor product reduces to a mixture.

3 Erlang Mixture Population Generator

The Erlang mixture distribution can be very useful in the estimation of the empirical distributions of samples.

Let us define a set $\{x^1, ...x^N\}$ of sample points, $x^l = (x_1^l, ..., x_n^l)$, $x_j^l \in [a_j, b_j]$; $a_j, b_j \in \mathbf{R}; a_j < b_j, l = 1, ..., N$. To construct a multidimensional frequency histogram for the set of samples we have to divide the admissible domain $[a_1, b_1] \times ... \times [a_n, b_n]$ into $k = K \times ... \times K$(n times) classes. The class frequencies in the multidimensional histogram $W(i)$ are calculated in the following way $W(i) = W(1, i) \cdot ... \cdot W(n, i)$ and the class marks for this histogram are the points $c_i = (c_i(1), ..., c_i(n))$, where $W(j, t)$ and $c_t(j)$ are the class frequencies and the class marks for the coordinates respectively. Analyzing this histogram we can generate the new set of sample points with the mixture of Erlangs distribution. Assuming that the variances in each frequency histogram class are equal we can define *an Erlang Mixture Population Generator* in the following way:

1. For given $c_i = (c_i(1), ..., c_i(n))$ and fixed shape parameters $\widehat{m^i} = (m_1^i, ..., m_n^i)$ we calculate the scale parameters $\widehat{\lambda^i} = (\lambda_1^i, ..., \lambda_n^i)$ of the mixture according to the following formula

$$\lambda_j^i = \frac{m_j^i}{c_i(j)}; d_j = \frac{m_j^1}{\lambda_j^1} = ... = \frac{m_j^k}{\lambda_j^k}, j = 1, ..., n; (i = 1, ..., k). \tag{5}$$

where d_j is the variance of the variable X_j. The variance d_j should not be greater than $\frac{(b_j - a_j)^2}{4K^2}$.

2. We calculate the weight coefficients of the mixture $v_i; (i = 1, ..k)$ in the following way:

$$v_i = \frac{W(i)}{\sum_{i=1}^{k} W(i)}; \tag{6}$$

where $W(i)$ is the multidimensional class frequency

3. We define the probability density function of the mixture $G(x)$ according to the formula(4).

4. We generate the random variable X with the mixture of Erlang probability distributions.

In the step **4.** of the generator we have to use a random number generator for the mixtures. In case of mixtures of univariate distributions we can use the inverse transform method [16]. This method is very accurate but time-consuming. In case of mixtures of multivariate distributions we propose a simple mixture generator defined below:

Step1: Generate an index $I \in \{1, .., k\}$ according to the distribution $Prob(I = i) = v_i$

Step2: For given $I = i$ generate a random variable X with the Erlang distribution with parameters $\widehat{\lambda^i}$ and $\widehat{m^i}$.

Step3: Repeat steps 1 and 2 until the new set of samples is full.

The generation of the multidimensional random variable X in the **Step2** can be very complicated. A few general techniques for simulating non-normal multivariate distributions can be applied for this problem, for example the transformation approach and the rejection approach or the conditional distribution method [1].

In the simplest case of independent coordinates of X we can generate each univariate coordinate using one of the efficient gamma generators [12]. The joint Erlang density function is calculated as a product of the univariate ones. In this paper we assume that the multivariate Erlang mixture is a tensor product.

We applied the Erlang Mixture Population Generator as the recombination mechanism in evolutionary strategies. We defined a new estimation of distribution algorithm called *Erlang Mixture EDA (ErM-EDA)*. A framework for this algorithm is presented the next section.

4 Erlang Mixture Estimation of Distribution Algorithm

4.1 Main idea of EDAs

Estimation of Distribution Algorithms (EDAs) [6] are the stochastical methods designed for solving global optimization problems. The main goal of EDAs is to extract some global statistic information from a set of selected sample points (population) and build a posterior probability distribution model of promising solutions based on the extracted information. The new solutions are sampled from the model and they replace the current population.

Probability models of the sample sets in the existing estimation of distribution algorithms are based on a Gaussian distributions model [13], a Gaussian mixtures model [9] or a histogram model [15]. Each of these methods has many disadvantages. A Gaussian model can be built with very low computational cost but cannot be used to cope with multimodal objective function. A Gaussian mixture model is the most useful method, but the cost of building such a model is prohibitively high, particularly for large scale problems. The marginal histogram model is the simplest to implement but the independence of the variables must be assumed. We should also remember that usually a lot of sample points are needed to build a good probability model. Thus EDAs are often time-consuming.

One way to improve the performance of these algorithms is to hybridize a local method with EDA. We can refer to two recent propositions of hybrid EDAs: Iterated Estimation Evolutionary Algorithm (IEDA) with conjugate gradient algorithm developed by Bosman and Thierens [8] and Estimation of Distribution Algorithm with Local Search (EDA/L) defined by Zhang et al [11].

The other way to improve the performance of EDAs could be an application of another probability model. We defined a new density estimation algorithm using Erlang mixture generator as a method of reproduction in EDA. We tried to give an answer for the question about possibility.

4.2 ErM-EDA framework

The main idea of ErM-EDA is the replacement of the reproduction mechanism in the evolutionary strategies with the reproduction based on the generator of the mixture of Erlang distributions. Parameters of the empirical density function for this generator are determined from the frequency histogram of the parental population obtained as a result of the applied selection method.

Let us define some global optimization problem

$$\hat{x} = \arg\min_{x \in D} f(x), \tag{7}$$

where $x = (x_1, ...x_n) \in \mathbb{R}^n, f : \mathbb{R}^n \to \mathbb{R}$ is the objective function and $D = \{x : a_i \leq x \leq b_i; i = 1, ..., n\}$ is the feasible region. To apply the ErM-EDA to solve this problem we have first to make a linear transformation the objective function because of the assumption of nonnegative arguments for the Erlang Mixture Population Generator.

The framework of the proposed algorithm is as follows:

1. **Parameters setting:** Population sizes - N, $\tilde{N}$; the number of classes in a frequency histogram k

2. **Initialization:** Generate N points from D using uniform design technique.
3. **Selection of the gene pool:** Select $\tilde{N}$ to the parental population (gene pool) using fitness proportional selection mechanism.
4. **Reproduction:** Built a frequency histogram of the gene pool and calculate the parameters of the Erlang multivariate mixture empirical density function. Use the Erlang mixture generator to generate $\tilde{N}$ offsprings to the offspring population.
5. **Population updating:** Select N individuals from the offspring population to constitute a new base population.
6. **Stopping criterion:** Repeat steps **3. - 5.** until the stopping criterion is satisfied.

In **Reproduction** step we first built a frequency histogram of the gene pool created in step **3.** by using the roulette wheel selection technique [4]. The frequency classes defined for this histogram are the same in each iteration of ErM-EDA. Thus the scale and the shape parameters could be also fixed. Only the weights coefficients of the mixture should be calculated in each iteration.

In step **5.** we select individuals to the new base population according to the same selection scheme as in step **3.**. As a stopping criterion we accept the maximal number of iterations.

The accuracy of the ErM-EDA algorithm depends on the number of classes k in the frequency histogram.

5 Experiments

In this section we report on the main results of simple experiments which we have performed for ErM-EDA. The following test functions are used in our experimental studies:

- $T_1 = \sum_{i=1}^{2} -(x_i - 5.12)^2 - 10cos(2\Pi(x_i - 5.12) + 10); x_i \in [0; 10, 24]$
- $T_2 = \sum_{i=1}^{10} -(x_i - 5.12)^2 - 10cos(2\Pi(x_i - 5.12) + 10); x_i \in [0; 10, 24]$
- $T_3 = \exp\left(-5\sum_{i=1}^{2}(x_i - 0,5)^2\right) + \frac{1}{2}\exp\left(-5((1,5 - x_1)^2 + \sum_{i=1}^{2}(x_i - 0,5)^2)\right);$ $x_i \in [0, 1]$
- $T_4 = \exp\left(-5\sum_{i=1}^{10}(x_i - 0,5)^2\right) + \frac{1}{2}\exp\left(-5((1,5 - x_1)^2 + \sum_{i=1}^{10}(x_i - 0,5)^2)\right);$ $x_i \in [0, 1]$

The modified Rastrigin functions T_1 and T_2 were minimized while T_3 and T_4 functions (the sums of two Gaussian peaks) where maximized.

For the ErM-EDA algorithm we set the following parameters:

- The size of the populations

$$N = \begin{cases} 31 & , \quad n = 2 \\ 1000 & , \quad n = 10 \end{cases}; \tag{8}$$

  where n means the dimension of the search space.
- The variances of the univariate Erlang distributions are $d(1) = 1,050$ for T_1 and T_2, $d(2) = 0,0004$ for T_3 and T_4.
- The number of multidimensional frequency classes was $k = 5^n$.
- Stopping condition: The maximal number of executed iterations was 15.

We used the roulette wheel selection mechanism and gamma random number generator. We performed 30 independent runs for ErM-EDA and we compared its efficiency

Table 1. The ErM-EDA and EA performance

Test function	**nr (ErM-EDA)**	**nr (EA)**
T_1	30	22
T_2	4	14
T_3	30	26
T_4	5	19

with the efficiency of (μ, λ)-Evolutionary Strategy (ES) with Gaussian mutation on he same set of the test functions. We set $\lambda = 7\mu$ and the standard deviation of mutation $\sigma = 1.0$ for ES. The values of μ were the same as for the ErM-EDA populations. The experimental results are presented in Table 1. The parameter **nr** denotes the number of runs in which the global optimum was found.

The obtained results show that for two-dimensional functions T_1 and T_3 the performances of both applied algorithms are very similar. In cases of the large populations and ten-dimensional functions the performance of ErM-EDA was four times worse than ES.

6 Conclusions

- The modeling of the population distribution in the evolutionary algorithms can be a key to analyzing the populations dynamics.
- We applied the Erlang mixture model for the estimation of the empirical distributions of individuals selected for a recombination in an evolutionary process. We defined a new estimation of distribution algorithm called *Erlang Mixture Estimation of Distribution Algorithm*. We introduce the Erlang Mixture Population Generator as a reproduction mechanism in this algorithm.
- Simple experimental results show that our algorithm could be effective only in the low-dimensional cases.The performance of ErM-EDA was very disappointed for us. We think that the main reason of this bad performance is the problem with the generation of the uniformly distributed sample points in high dimensional spaces. Also the analysis of the frequency histograms is not an optimal method of calibration of the parameters of the cumulative probability distribution of the populations.
- In our current research we also try to apply another estimators, such as kernel estimators [14]. We also extended the estimator proposed [7] to the multidimensional case. We obtained a new model in which populations are represented by the matrixes. The columns of the matrix are vectors of the individuals coordinates and the distributions of the variables in rows are the mixtures of the univariate Erlang distributions. The results of the simple experiments are much better than for ErM-EDA and ES algorithms, but we need to perform more experiments to verify the efficiency of this method.

Bibliography

[1] L. Devroye. *Non-Uniform Random Variate Generation.* Springer Vlg., 1986.

[2] Y. Fang. Teletraffic analysis and mobility modeling of pcs networks. *EEE Trnasactions on Communication*, 47:1062–1073, 1999.

[3] J. D. Hamilton. A quasi-bayesian approach to estimating parameters for mixtures of normal distributions. *Journal of Business and Economic Statistics*, 9:27–39, 1991.

[4] J.Arabas. *Lectures on the Evolutionary Algorithms (in Polish).* WNT, 2001.

[5] F. P. Kelly. *Reversibility and Stochastic Networks.* Wiley & Sons, 1979.

[6] P. Larra naga and J. A. Lozano. *Estimation of Distribution Algorithms : A new Tool for Evolutionary Computation.* Kluwer Academic Publishers, 2001.

[7] J. Kołodziej and L. Ogiński. The estimation of the population distribution by the mixture of erlang distributions in evolutionary algorithms. 2005.

[8] D. Thierens P.A. Bosman. Expanding from discrete to continuous edas: The ieda. In *Proc. of Parallel Problems Solving from Nature, PPSN-VI*, pages 767–776, 2000.

[9] D. Thierens P.A. Bosman. Exploiting gradient information in continuous iterated density estimation evolutionary algorithms. In *Working report, UU-CS-2001-53, Universiteit Utrecht*, 2001.

[10] M.J.D. Powell. *Approximation Theory and Methods.* Cambridge University Press, 1981.

[11] Q. Zhang Q, J. Sun, E. Tsang, and J. Ford. Hybrid estimation of distribution algorithm for global optimization. *Engineering Computations*, 21, No.1:91–107, 2004.

[12] P. Wieczorkowski R and R. Zieliński. *Computer Random Numbers generators (in Polish).* WNT, 1997.

[13] S. Kotz S, N. Balakrishnan, and N.L. Johnson. *Continous Multivariate Distributions, volume 1: Models an Applications, 2nd ed.* ohn Wiley Interscience Publication, 2001.

[14] D.M. Titterington, A.F.M. Smith, and U.E. Makov. *Statistical Analysis of finite Mixture Distributions.* Wiley & Sons, 1985.

[15] S. Tsutsui, M. Pelikan, and D.E. Goldberg. Evolutionary algorithm using marginal histogram model in continuous domain. 2005.

[16] D.C. Montgomery W.W. Hines and, D.M. Goldsman, and C.M. Borror. *Probability and Statistics in Engineering, 4th ed.* Wiley & Sons, 2001.