ON VARIABILITY ANALYSIS OF EVOLUTIONARY ALGORITHM-BASED ESTIMATION

Manuel Rizzo [1]

[1] Dipartimento di Scienze Statistiche, Universita' degli studi di Roma "La Sapienza", (e-mail: manuel.rizzo@uniroma1.it)

ABSTRACT: A theoretical framework to analyze variability of parametric estimates obtained via Evolutionary Algorithms (EAs) is proposed. The nature of EAs, in fact, introduces a further source of variability, due to stochastic elements of the procedure. A simulation study employing Genetic Algorithms and Differential Evolution is also conducted in order to make comments on the effect of these stochastic elements on variability.

KEYWORDS: Genetic algorithms, Differential evolution, Least absolute deviation regression, Convergence rate

1 Introduction

Evolutionary Algorithms (EAs) are nature inspired methods, introduced in the 1960s in the field of Artificial Intelligence as a tool to analyze complex systems and problems (De Jong, 2006). In the last decades many researchers have employed such methodologies in statistics, where complex optimization problems often arise (Baragona *et al.*, 2011). In the present paper we focus on parameter estimation problems tackled by EAs for which an analytical solution can not be found, or complexity of optimization problem may cause failure of standard methods. In this case an additional source of variability due to the stochastic elements in EAs must be considered in the analysis. As far as this kind of variability has not been addressed yet in statistical literature, we shall propose a framework to analyze it as a component which is independent from sampling variability.

Paper is organized as follows: in Section 2 EAs employed in the study are introduced; in Section 3 we present the framework of analysis of variability; Section 4 outlines a simulation study; in Section 5 we resume comments and conclusions.

2 Evolutionary Algorithms

EAs fall in the category of *population based* methods, for which at each iteration (or *generation*) a population of N individuals, coded to represent solutions, is evaluated. The objective function, called *fitness*, measures solutions goodness and biases search process towards modal zones. Stochastic operators which modify and create solutions at each generation characterize specific EAs. We shall now shortly describe the two algorithms implemented in our parametric estimation framework, valid for consistent estimators and convergent EAs.

2.1 Genetic Algorithm

Genetic Algorithms (GAs; Holland, 1975) are among the most important EAs. In our problem a binary vector $\underline{\psi}$ (called *chromosome*) encodes a real parameter vector $\underline{\theta} = (\theta_1, ..., \theta_k)$ by standard binary coding rule; each fixed length bit interval of $\underline{\psi}$ encodes a component θ_j of $\underline{\theta}$. At each generation N individuals are selected proportionally with respect to fitness, and are subject to operators of *crossover* and *mutation*: according to crossover, with fixed probability pC, pairs of chromosomes are selected to exchange a section of their bits, specified by a common cutting point; *mutation* step, on the other side, allows each bit of each chromosome to flip its value from 0 to 1 or viceversa, with fixed probability pM. GA convergence is assured (Rudolph, 1997) if $pM > 0$ and if an *elitist* strategy is employed, for which the best individual in the population at each generation is always maintained.

2.2 Differential Evolution

Differential Evolution (DE) algorithm has been introduced in the 1990s and became soon a powerful EA for continuous optimization (Price *et al.*, 2006). In this case parameter vectors are directly represented in the population. At each generation, for each vector x_i in the population a so-called *mutant* v_i is built by *differential mutation*, using three randomly chosen vectors x_{r0}, x_{r1}, x_{r2}, by: $v_i = x_{r0} + F(x_{r1} - x_{r2})$, where $F > 0$ is a scale factor. Afterwards x_i and v_i are recombined by use of *binomial crossover*, for which each component of new vector u_i can be inherited either from x_i or v_i, with fixed probability CR. If resulting u_i has a better fitness value than x_i it is accepted in the next generation, otherwise x_i is retained (*selection* operator); this latter step implies elitism. In order to guarantee global convergence of the procedure (Hu *et al.*, 2013;

Knobloch *et al.*, 2017) we modify standard DE by forcing at least one vector in the population at each generation to be uniformly regenerated at random within its boundaries, before the selection step.

3 Variability decomposition

Let us consider a sample of n observations generated from a model indexed by a parameter vector $\underline{\theta}$. In a frequentist inference framework we denote $\hat{\underline{\theta}}$ as the best theoretical value of a consistent estimator (for example a Maximum Likelihood estimator), not available in practice. If $\underline{\theta}^*$ is the approximation of $\hat{\underline{\theta}}$, obtained by a convergent EA with V fitness function evaluations, then we must account for extra variability introduced in the analysis, due to the stochastic elements of EA. Following Rizzo & Battaglia, 2017 we assume independence between data-generating model and process generating EA random seeds, introduce the tjwo related expectations $\mathbb{E}_S$ and $\mathbb{E}_{EA}$, and decompose trace of total covariance matrix Σ_{TOT} (whose generic element refers to total error $[\theta^* - \theta]$) by:

$$tr(\Sigma_{TOT}) = tr(\Sigma_S) + tr(\Sigma_{EA}) = tr(W_S)\frac{1}{f(n)} + tr(W_{EA})\frac{1}{h(V)},$$

where Σ_S is sampling covariance matrix (generic element is $\sigma^S_{ij} = \mathbb{E}_S[(\widehat{\theta}_i - \theta_i)(\widehat{\theta}_j - \theta_j)]$), Σ_{EA} is EA covariance matrix ($\sigma^*_{ij} = \mathbb{E}_{EA}[(\theta^*_i - \widehat{\theta}_i)(\theta^*_j - \widehat{\theta}_j)]$), $f(n)$ and $h(V)$ are, respectively, estimator consistency rate and EA convergence rate, W_S and W_{EA} are matrices composed by elements that depend, respectively, from statistical model and from EA but are constant with respect to n and V.

Whilst sampling variability is generally well understood in consistency theory (for example, in the case of asymptotically efficient estimators we have $f(n) = n$), EAs convergence rate, which corresponds to algorithmic variability in our framework, is not an easy task to analyze theoretically.

In the next section we shall consider an application of GA and DE based estimation, for which we will evaluate algorithmic variability. We shall also make some comments on the effect of algorithm configurations on $h(V)$.

4 Applications

Application considered is a Least Absolute Deviation estimation of Linear Regression model: $y_i = \beta_0 - \beta_1 x_{i,1} - \beta_2 x_{i,2} + \varepsilon_i,\ i = 1, ..., n$. This estimator is the

function that minimizes the sum of absolute value of errors, and it is not differentiable. We employed both GA and DE in this estimation problem, using data simulated from a model with $\underline{\beta} = (0.5, 0.5, -0.5)$ and $\varepsilon_i \sim t_5$. We chose fitness function f, to be maximized, as:

$$f = \exp\{-\sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_{i,1} - \beta_2 x_{i,2}| \ / \ n\}.$$

In both algorithm applications we estimated EA variability using $J = 500$ algorithm runs at each generation g by:

$$\sigma_{ii}^{*(g)} = \frac{1}{J} \sum_{j=1}^{J} [\beta_{j,i}^{*(g)} - \widehat{\beta}_i]^2, \ i = 1, ..., 3,$$

where $\beta_{j,i}^{*(g)}$ is the best solution found up to generation g in run j, and $\widehat{\beta}_i$ is the best theoretical value, available in simulation. Then a linear EA convergence rate was estimated by regression:

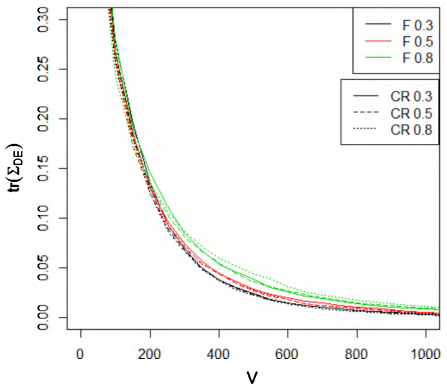
$$tr(\Sigma_{EA}^{(g)}) = tr(W_{EA}) \frac{1}{[V^{(g)}]} + e_g, \quad g = 1, ..., G, \tag{1}$$

where $tr(\Sigma_{EA}^{(g)})$ includes elements $\sigma_{ii}^{*(g)}$, $tr(W_{EA})$ is treated as regression parameter, $[V^{(g)}]$ is the number of fitness evaluations up to generation g, e_g is the error. By considering $[V^{(g)}]$ instead of g we were allowed to compare algorithms with different population sizes. In fact we considered population sizes $N = 50, 70$ (with related maximum number of generations, respectively, $G = 2000, 1450$) and following configurations for GAs and DE:

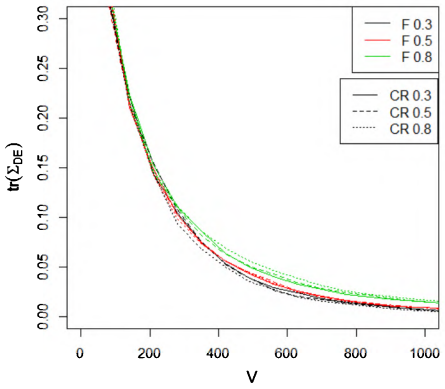
- GA: $pM = 0.01, 0.05, 0.10$; $pC = 0.5, 0.7, 0.9$
- DE: $F = 0.3, 0.5, 0.8$; $CR = 0.3, 0.5, 0.8$,

so that 18 algorithms have been employed in the study.

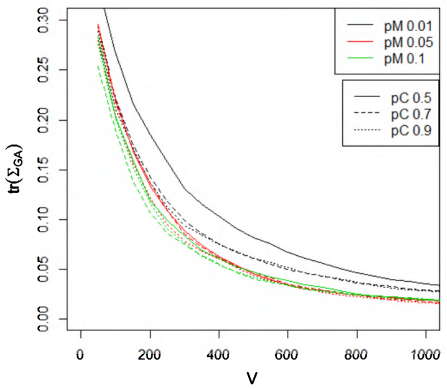
Figure 1 shows the curves of $tr(\Sigma_{EA}^{(g)})$ estimates for all scenarios. DE experiments show a more homogeneous behaviour with respect to GAs (in particular CR seems to have a very low effect), and in both algorithms as N increases differences between experiments in each panel tend to reduce. However DE estimation seems to improve as F decreases, as the best behaviour is registered at 0.3. On GA side the same happens for low mutation rate pM (with a worsening for low pC), possibly because if an elitist strategy is adopted then effect of exploration (task assigned to mutation operator) become crucial in the analysis.



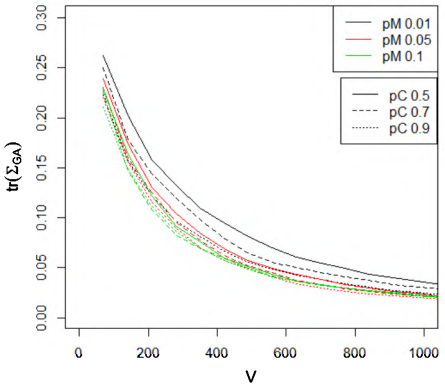
(a) DE, population size $N = 50$



(b) DE, population size $N = 70$



(c) GA, population size $N = 50$



(d) GA, population size $N = 70$

Figure 1: Estimates of EA covariance matrix trace

Table 1 shows goodness of fit results of regression analysis (1), in term of R^2, for all scenarios. These results clearly indicate that a linear convergence rate is reasonable for all different EAs configurations in this application, and this extends part of results in Rizzo & Battaglia, 2017, which considers only a single configuration of GA.

5 Conclusions

In this paper we analyzed variability of EA-based estimates, considering the effect of stochastic elements of these methods and estimating their convergence

Table 1: R^2 coefficient values for EAs linear convergence rate estimates

GA	N=50	N=70	**DE**	N=50	N=70
pM=0.01 pC=0.5	0.87	0.89	CR=0.3 F=0.3	0.93	0.93
pM=0.01 pC=0.7	0.90	0.91	CR=0.3 F=0.5	0.95	0.95
pM=0.01 pC=0.9	0.89	0.92	CR=0.3 F=0.8	0.95	0.95
pM=0.05 pC=0.5	0.92	0.93	CR=0.5 F=0.3	0.94	0.93
pM=0.05 pC=0.7	0.92	0.93	CR=0.5 F=0.5	0.95	0.94
pM=0.05 pC=0.9	0.93	0.93	CR=0.5 F=0.8	0.96	0.95
pM=0.10 pC=0.5	0.93	0.94	CR=0.8 F=0.3	0.94	0.93
pM=0.10 pC=0.7	0.92	0.95	CR=0.8 F=0.5	0.95	0.94
pM=0.10 pC=0.9	0.93	0.93	CR=0.8 F=0.8	0.96	0.96

rates in a simulation study. Further studies are needed to better understand theoretically the effect of algorithm configurations on variability, that could support applications of EAs to statistical problems.

References

BARAGONA, R., BATTAGLIA, F., & POLI, I. 2011. *Evolutionary statistical procedures*. Berlin: Springer.

DE JONG, K. A. 2006. *Evolutionary computation: A unified approach*. MIT press.

HOLLAND, J. H. 1975. *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.

HU, Z., XIONG, S., SU, Q, & ZHANG, X. 2013. Sufficient conditions for global convergence of differential evolution algorithm. *Journal of Applied Mathematics*, **2013**.

KNOBLOCH, R., MLỲNEK, J., & SRB, R. 2017. The classic differential evolution algorithm and its convergence properties. *Applications of Mathematics*, **62**(2), 197–208.

PRICE, K. V., STORN, R. M., & LAMPINEN, J. A. 2006. *Differential evolution: A practical approach to global optimization*. Springer.

RIZZO, M., & BATTAGLIA, F. 2017. Statistical and Computational Tradeoff in Genetic Algorithm-Based Estimation. arXiv:1703.08676.

RUDOLPH, G. 1997. *Convergence properties of evolutionary algorithms*. Kovac.