Inferential performance assessment of stochastic optimisers and the attainment function*

Viviane Grunert da Fonseca[1], Carlos M. Fonseca[1,2], and Andreia O. Hall[3]

[1] ADEEC, UCE
Universidade do Algarve
Faro, Portugal
[2] Instituto de Sistemas e Robótica
Coimbra, Portugal
[3] Departamento de Matemática
Universidade de Aveiro
Aveiro, Portugal
vgrunert@ualg.pt, cmfonsec@ualg.pt, andreia@mat.ua.pt

Abstract The performance of stochastic optimisers can be assessed experimentally on given problems by performing multiple optimisation runs, and analysing the results. Since an optimiser may be viewed as an estimator for the (Pareto) minimum of a (vector) function, stochastic optimiser performance is discussed in the light of the criteria applicable to more usual statistical estimators. Multiobjective optimisers are shown to deviate considerably from standard point estimators, and to require special statistical methodology. The attainment function is formulated, and related results from random closed-set theory are presented, which cast the attainment function as a mean-like measure for the outcomes of multiobjective optimisers. Finally, a covariance-measure is defined, which should bring additional insight into the stochastic behaviour of multiobjective optimisers. Computational issues and directions for further work are discussed at the end of the paper.

1 Introduction

Stochastic optimisers, such as evolutionary algorithms, simulated annealing and tabu search, have found many successful applications in a broad range of scientific domains. However, only limited theoretical results concerning their performance are available. Typically, simple versions of the algorithms and/or objective functions must be considered in order to make the theoretical analysis possible, which limits their practical applicability. As an alternative, the performance of

* in E. Zitzler, K. Deb, L. Thiele, C. A. Coello Coello and D. Corne, eds., *Evolutionary Multi-Criterion Optimization, First International Conference, EMO 2001*, vol. 1993 of Lecture Notes in Computer Science, pp. 213–225, Berlin: Springer-Verlag, 2001.

stochastic optimisers may be assessed experimentally on given problems by performing multiple, independent optimisation runs, and statistically analysing the results.

Two main issues are raised by such an inferential approach. Firstly, the very notion of optimiser performance must take into account the stochastic nature of the optimisers considered, as well as any other relevant optimiser characteristics, such as scale-independence, for example. As the same considerations apply to statistical estimators, optimiser performance will be discussed in that light in Section 2.

Secondly, specific statistical methodology may be needed, depending on the notion of performance adopted, in order to analyse the data produced by the optimisation runs. In particular, multiobjective optimisers such as multiobjective genetic algorithms (Fonseca and Fleming, 1995) produce sets of non-dominated objective vectors, instead of a single optimal objective value per run. Dealing with random sets introduces additional difficulties into the analysis. In Section 3, the attainment function is formally defined, and shown to relate closely to established results in random closed set theory. In particular, it is shown to be a measure analogous to the common mean, which considerably strengthens its role as a measure of multiobjective optimiser performance. Based on the same theory, variance-like and covariance-like measures are introduced which should provide additional insight into multiobjective optimiser performance.

Finally, computational issues are discussed. The paper concludes with a summary of the results, and a discussion of their implications for future work.

2 Inferential performance assessment

Optimiser performance can ultimately be understood in terms of the trade-off between the quality of the solutions produced and the computational effort required to produce those solutions, for a given class of optimisation problems. Experimentally, optimiser performance may be assessed in terms of:

1. The time taken to produce a solution with a given level of quality (run time),
2. The quality of the solutions produced within a given time,

where time may be measured in terms of number of iterations, number of function evaluations, CPU time, elapsed time, etc., and solution quality is defined by the problem's objective function(s). When considering *stochastic* optimisers, or deterministic optimisers under *random* initial conditions, both run time, in the first case, and solution quality, in the second case, are random, and the study of optimiser performance is reduced to the study of the corresponding distributions.

Hoos and Stützle (1998) propose the estimation and analysis of run-time distributions. It is worth noting that such time-to-event data may originate from improper distributions, since an optimiser may fail to find a solution with the desired quality in some runs. Also, the data may be subject to censoring whenever the actual run-time of the optimiser exceeds the practical time-limits of the experiment. Thus, the data may require special statistical treatment, of the kind usually encountered in statistical survival analysis. Run-time distributions are univariate distributions by definition, even if the problem considered involves multiple objectives.

Fonseca and Fleming (1996) suggested the study of solution-quality distributions. The outcome of a multiobjective optimisation run was considered to be the set of non-dominated objective vectors evaluated during that run. In the single-objective case, this reduces to a single objective value per run, corresponding to the quality of the best solution(s) found, and leads to the study of univariate distributions. In the multiple objective case, however, solution-quality distributions are either multivariate distributions, in the case where optimisers produce a single non-dominated vector per run, or set distributions, in the general case.

In this context, optimisers may be seen as estimators for the global (Pareto) optimum of a (vector) function. Therefore, optimiser performance can be viewed in the light of the performance criteria usually considered for classical statistical estimators. However, it must be noted that optimisers are actually more than simple estimators, as they must also provide the actual *solutions* corresponding to their estimates of the function's optimum.

2.1 The single-objective case

As discussed above, the outcomes of single-objective optimisers consist of a single value per optimisation run, which is the objective value corresponding to the best solution(s) found. Therefore, one is interested in the stochastic behaviour of random variables X in $\mathbb{R}$, and the performance of optimisers and that of point estimators may be seen in parallel.

Good estimators should produce estimates which are *close* to the unknown estimand, both in terms of location and spread. The same applies to the outcomes of single-objective optimisers. Closeness in terms of location may be measured by the difference between the mean or the median of the corresponding distributions and the unknown estimand. This is known as the mean-bias and the median-bias, respectively. Ideally, both should be zero. Possible measures of spread are the variance and the interquartile-range, both of which should be small. Alternatively, location and spread may be combined in terms of the mean-squared-error, which should also be small.

Mean and variance are the first moment and the second centred moment of a distribution. They are efficiently estimated by the arithmetic mean $\bar{X}$ and the empirical variance s^2, respectively, when the underlying distribution is close to normal. This is the case with many statistical estimators, at least for sufficiently large sample sizes. The solution-quality distributions of optimisers, on the other hand, can (and should) be very asymmetric. Moreover, objective-scale information is ignored by some optimisers, which rely solely on order information. Thus, estimating the median and the inter-quartile range through their empirical counterparts might be preferred here, since quantiles are scale-invariant, i.e. $\tau[\gamma(X)] = \gamma[\tau(X)]$ for any quantile γ and any strictly monotonic transformation τ (Witting, 1985, p. 23).

In addition to closeness considerations, point estimates and optimisation outcomes should follow a type of distribution easy to deal with. In the case of estimators, this is usually the normal distribution. Optimisation outcomes, however, must follow a distribution which is bounded below (considering minimisation problems). Its left end-point should be as close to the unknown minimum as possible, and it should be right skewed, so that outcomes are likely to be close to the minimum. Given that the outcome of a single-objective optimisation run is the minimum of all objective values computed in the course of the run, ideal solution-quality distributions would be extreme-value distributions, the estimation of which has been vastly studied in the literature (see for instance Smith (1987), Lockhart and Stephens (1994), and Embrechts et al. (1997)), both in a parametric and in a semi/non-parametric setting.

The shape of a distribution can be assessed directly by estimating the cumulative distribution function, $F_X(\cdot)$, which completely characterises the underlying distribution. One may also wish to study specific aspects of the distribution, such as skewness (e.g. through the kurtosis) and tail behaviour (through end-point and tail-index estimation, for example). For minimisation problems, left and right-tail behaviour is related to best and worst-case performance, respectively.

2.2 The multiobjective case

When the optimisation problem is multiobjective, a whole front of Pareto-optimal solutions in $\mathbb{R}^d$ is to be approximated, and the outcome of an optimisation run may be a set of non-dominated objective vectors. For simplicity, the situation where the outcome of a run consists of a single objective vector shall be considered first.

Single objective vectors The most common multivariate measure of location is possibly the arithmetic mean, which is now a vector in $\mathbb{R}^d$. If the unknown estimand is also a vector, as is the case with multivariate point estimators, this is clearly appropriate. The mean-bias of a point estimator, for example, can be written as the difference between the mean of corresponding distribution and the unknown estimand. Common measures of spread are the covariance matrix and other measures related to it (Mood *et al.*, 1974, p. 351ff). All formulate spread in terms of deviation from the mean, which is a point.

In a multiobjective optimisation context, however, both bias and spread should be understood in terms of Pareto fronts. Note that the mean-vector of a number of non-dominated vectors could be located beyond a concave Pareto-optimal front to be approximated, outside the collection of all possible outcomes! Useful, alternative measures of bias and spread shall be given later in Section 3.

The shape of a multivariate distribution can be assessed through estimation of the cumulative multivariate distribution function, even though this is more challenging computationally than the corresponding univariate case. Again, solution-quality distributions should be skewed in the sense that outcomes should be likely to be close to the unknown Pareto front. Note that the Pareto front imposes a bound on the support of solution-quality distributions. Multivariate extreme-value theory is currently an active, but very specialised, area of research.

Multiple non-dominated objective vectors Outcomes are represented by the random (point) sets $\mathcal{X} = \{X_j \in \mathbb{R}^d,\ j = 1, \ldots, M\}$ where the elements X_j are non-dominated within the set and random, and the number M of elements is random. Performance assessment oriented towards solution quality, as discussed so far, must take into account the particular set-character of the distributions involved.

Statistical estimators which produce a set of non-dominated vectors in $\mathbb{R}^d$ when applied to a data-set are not known to the authors, but curve estimators, seen as (continuous) random curve sets in $\mathbb{R}^2$, could be related. Bias measures for curve estimators $\hat{g}(\cdot)$, such as the average sum of squares

$$\frac{1}{k}\sum_{i=1}^{k}[\hat{g}(Z_i) - g(Z_i)]^2$$

or the supremum-norm

$$\sup_{i=1,\ldots,k} |\hat{g}(Z_i) - g(Z_i)|,$$

where the $Z_i \in \mathbb{R}$ are either random or deterministic, might suggest suitable analogues for the performance assessment of multiobjective optimisers. If the difference is replaced by the minimum Euclidean-distance between the random-outcomes X_j and the Pareto-optimal front to be approximated, one obtains measures similar in spirit to the *generational distance*, proposed by Van Veldhuizen and Lamont (2000).

Unlike curve-estimators, the performance of multiobjective optimisers is additionally affected by the variability of the outcomes within a set and by how uniformly the outcomes are distributed along the final trade-off surface (Zitzler, 1999; Zitzler *et al.*, 1999; Van Veldhuizen and Lamont, 2000). Hence, taking into account the *overall* point-set character of the outcomes promises to be much more informative than just relying on summary measures such as the above. *Random closed set theory* (Matheron, 1975; Kendall, 1974) addresses precisely this issue. Note that the outcome-set $\mathcal{X}$ is closed.

The mean of a random-set distribution has been defined in various set-valued ways. One of the most popular is the *Aumann-mean*, which is defined as "the set of expected selections, where a selection is any random vector that almost surely belongs to the random set" (Cressie, 1993, p. 751). A possible estimator for this mean of some (general) random closed set $\mathcal{W}$ is formulated as

$$\bar{\mathcal{W}_n} = \frac{1}{n}(\mathcal{W}_1 \oplus \mathcal{W}_2 \oplus \ldots \oplus \mathcal{W}_n),$$

which is the Minkowski average of n independent copies $\mathcal{W}_1, \ldots, \mathcal{W}_n$ of $\mathcal{W}$ (Cressie, 1993, p. 751). Note that the Minkowski addition of two sets $\mathcal{A}_1$ and $\mathcal{A}_2$ is defined as

$$\mathcal{A}_1 \oplus \mathcal{A}_2 = \{a_1 + a_2 \mid a_1 \in \mathcal{A}_1, a_2 \in \mathcal{A}_2\}.$$

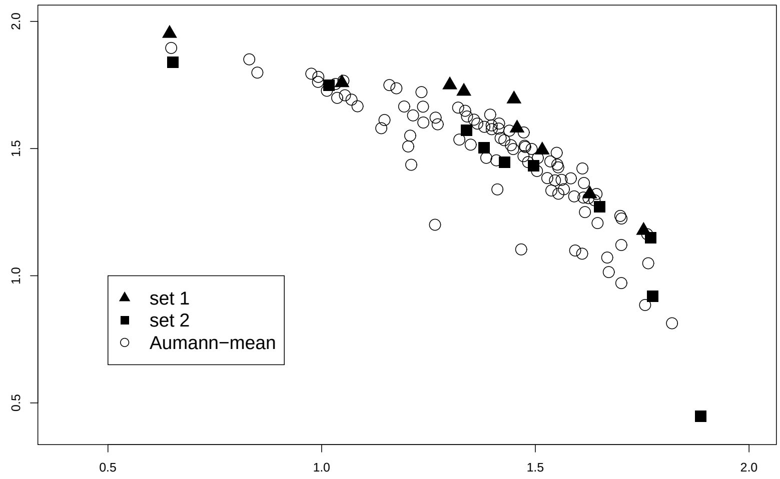

Figure 1. *The estimated Aumann-mean for two sets of non-dominated points in* $\mathbb{R}^2$.

Clearly, the estimated Aumann-mean of the outcome set $\mathcal{X}$ of a multiobjective optimiser contains many more elements than the observed sets themselves (see Figure 1). In addition, the theoretical mean is typically a convex set, and does *not* contain exclusively non-dominated elements. As for the vector-mean, some elements might even be located beyond Pareto-optimal front to be approximated, if it is concave. The Aumann-mean is therefore unsuitable as a measure of location in an optimisation context.

An alternative (empirical) mean-formula appears to be more useful. It is the *empirical covering function*, which is defined for a (general) random-set $\mathcal{W}$ as

$$p_n(z) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{I}\{z \in \mathcal{W}_i\}, \quad z \in \mathbb{R}^d. \tag{1}$$

The random sets $\mathcal{W}_1, \ldots, \mathcal{W}_n$ are independently and identically distributed like $\mathcal{W}$, and $\boldsymbol{I}\{\cdot\}$ denotes the indicator function. The empirical covering function has been applied in the area of “Particle Statistics” to describe the average of possibly non-convex particles. Note that particles must be transformed into sets first, by choosing “reasonable locations and orientations” for them (Stoyan, 1998).

The attainment function and its empirical estimator (Shaw *et al.*, 1999; Fonseca and Fleming, 1996) turn out to be equivalent to the theoretical covering function $p(z) = P(z \in \mathcal{W})$ and its empirical counterpart. The definition of the attainment function and additional theoretical results are given in the following section.

3 The attainment function

3.1 Definition, interpretation, and estimation

The attainment function provides a description of the distribution of an outcome set $\mathcal{X} = \{X_j \in \mathbb{R}^d,\ j = 1, \ldots, M\}$ in a simple and elegant way, using the notion of goal-attainment. It is defined by the function $\alpha_{\mathcal{X}}(\cdot) : \mathbb{R}^d \longrightarrow [0, 1]$ with

$$\begin{aligned}\alpha_{\mathcal{X}}(z) &= P(X_1 \leq z \ \vee \ X_2 \leq z \ \vee \ldots \vee \ X_M \leq z) \\ &= P(\mathcal{X} \trianglelefteq z).\end{aligned}$$

The symbol "$\vee$" denotes the logical "or". The expression $\alpha_{\mathcal{X}}(z)$ corresponds to the probability of at least one element of $\mathcal{X}$ being smaller than or equal to $z \in \mathbb{R}^d$, that is, the probability of an optimiser finding at least one solution which attains the goal-vector z in a single run. Clearly, the attainment function is a *generalisation of the multivariate cumulative distribution function* $F_X(z) = P(X \leq z)$. It reduces to the latter when $M = 1$, i.e. when the optimiser produces only one random objective vector per optimisation run.

The attainment function simultaneously addresses the three criteria of solution quality in the multiobjective context pointed out by Zitzler and colleagues (Zitzler, 1999; Zitzler *et al.*, 1999), although not separately: a long tail (in the multidimensional sense) away from the true Pareto front may be due to the location of individual outcome elements in some runs (first criterion), to the lack of uniformity of the elements within runs (second criterion), or to the small extent of the outcome non-dominated sets (third criterion).

The attainment function can be estimated via its empirical counterpart

$$\alpha_n(z) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{I}\{\mathcal{X}_i \trianglelefteq z\},$$

the *empirical attainment function*, where the random sets $\mathcal{X}_1, \ldots, \mathcal{X}_n$ correspond to the outcomes of n independent runs of the optimiser. Note the similarity to the empirical covering function (1).

3.2 The link to random closed set theory

The attainment function can be written in terms of so called "hit-or-miss probabilities", which are of fundamental importance in random closed set theory. For this, an alternative representation of the outcome set $\mathcal{X} = \{X_j \in \mathbb{R}^d,\ j = 1, \ldots, M\}$ with equivalent stochastic behaviour is chosen. It is the random (closed) set

$$\begin{aligned}\mathcal{Y} &= \{y \in \mathbb{R}^d \mid X_1 \leq y \ \vee \ X_2 \leq y \ \vee \ldots \vee \ X_M \leq y\} \\ &= \{y \in \mathbb{R}^d \mid \mathcal{X} \trianglelefteq y\}\end{aligned} \tag{2}$$

describing the region in $\mathbb{R}^d$ which is attained by $\mathcal{X}$ (see Figure 2). Using this alternative representation of $\mathcal{X}$, the attainment function may be expressed as

$$\alpha_{\mathcal{X}}(z) = P(z \in \mathcal{Y}), \quad z \in \mathbb{R}^d.$$

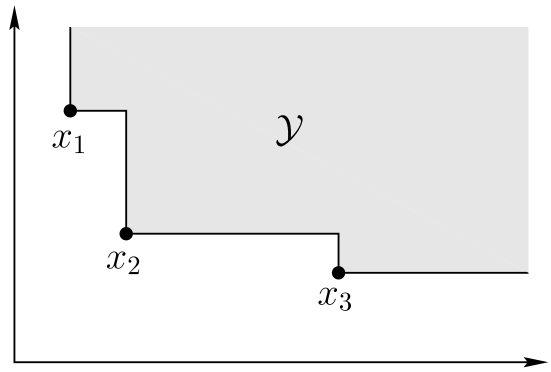

Figure 2. *Outcome set $\mathcal{X}$ with non-dominated realizations x_1, x_2, and x_3 and the set $\mathcal{Y}$ (here as a realization).*

Hence, the attainment function of the outcome-set $\mathcal{X}$ is identical to the *covering function* of the associated random set $\mathcal{Y}$. Denoting n independent copies of the random set $\mathcal{Y}$ as $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$ one can rewrite the empirical attainment function as

$$\alpha_n(z) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{I}\{z \in \mathcal{Y}_i\},$$

which shows the identity between the empirical attainment function of $\mathcal{X}$ and the empirical covering function of $\mathcal{Y}$ (compare with (1)).

Finally, the (theoretical) attainment function of $\mathcal{X}$ is identical to the *hitting function* or capacity functional (see e.g. Cressie (1993), Goutsias (1998)) of $\mathcal{Y}$ with support restricted to the collection of all one-point sets $\{z\}$ in $\mathbb{R}^d$. Hence, it can be expressed via hit-or-miss probabilities as

$$\alpha_{\mathcal{X}}(z) = P\big(\mathcal{Y} \cap \{z\} \neq \emptyset\big). \tag{3}$$

In general, the hitting function is defined over *all* compact subsets K in $\mathbb{R}^d$ (a definition for spaces more general than $\mathbb{R}^d$ is not of interest here). It fully characterises the stochastic behaviour of a random closed set in $\mathbb{R}^d$, and is of essential importance in random closed set theory. Note that the attainment function does *not* contain enough information to uniquely describe the stochastic behaviour of $\mathcal{X}$ and of $\mathcal{Y}$.

3.3 First-order moment concepts

The hitting function of a (general) random closed set $\mathcal{W}$ defined over all compact subsets K in $\mathbb{R}^d$ is identical to the general first-order moment measure $C^{(1)}_{\mathcal{W}}(\cdot)$ of the same set, i.e.

$$C^{(1)}_{\mathcal{W}}(K) = P\big(\mathcal{W} \cap K \neq \emptyset\big).$$

The above definition generalises the notion of first-order moment of a binary random field $\{b(z) \mid z \in \mathbb{R}^d\}$, which is a collection of random function values $b(z)$ where $b(z)$ can be 0 or 1. Here, the first-order moment (measure) is defined as

$$P\big(b(z) = 1\big) = P\big(\mathcal{W} \cap \{z\} \neq \emptyset\big)$$

where the random closed set $\mathcal{W}$ is related to the binary random field according to $\mathcal{W} = \{z \in \mathbb{R}^d \mid b(z) = 1\}$. See Goutsias (1998).

As its formulation in (3) shows, the attainment function $\alpha_{\mathcal{X}}(\cdot)$ is the *first-order moment measure of the binary random field derived from the random set* $\mathcal{Y}$ in (2) so that

$$\mathcal{Y} = \{z \in \mathbb{R}^d \mid b(z) = 1\}.$$

In other words, the attainment function $\alpha_{\mathcal{X}}(z)$ is the expected value of the binary random variable $\boldsymbol{I}\{\mathcal{Y} \cap \{z\} \neq \emptyset\} = \boldsymbol{I}\{\mathcal{X} \trianglelefteq z\}$ for all goals $z \in \mathbb{R}^d$. Hence, it makes sense to see the attainment function as a mean-measure for the set-distribution of $\mathcal{Y}$ and, in the wider sense, also of the outcome-set $\mathcal{X}$. Note again that the empirical covering function is used as a mean-measure in particle statistics.

As remarked above, the attainment function reduces to the (multivariate) distribution function $F_X(\cdot)$ for singular sets $\mathcal{X} = \{X\}$. The distribution function is a mean-measure for the distribution of the random set $\mathcal{Y} = \{y \in \mathbb{R}^d \mid X \leq y\}$ and, in a wider sense, also of $\mathcal{X} = \{X\}$. Thus, $F_X(\cdot)$ is a suitable alternative for the mean-vector of single objective vectors as a measure of location, when the optimisation problem is multiobjective and a whole Pareto-front is to be approximated (see the discussion in 2.2).

A notion of bias may be constructed in terms of the difference between the attainment function $\alpha_{\mathcal{X}}(z)$ associated with the optimisation outcomes and the ideal attainment function $\alpha_I(z) = \boldsymbol{I}\{z \in \mathcal{Y}^*\}$, where $\mathcal{Y}^*$ denotes the deterministic region attained by the true Pareto-optimal set of the problem. The bias, in this sense, is a function of a goal z, and indicates how far from ideal the optimiser is regarding the attainment of that goal.

3.4 Higher-order moment concepts

When the first-order moment does not fully characterise a distribution, higher-order moments can contribute with additional information about the stochastic behaviour of a random variable/vector/set. Depending on the actual distribution, a finite number of higher-order moments may, or may not, be enough to fully characterise it. In statistics, this problem is known as the *problem of moments* (Mood *et al.*, 1974, p. 81).

The attainment function, as mentioned before, does not uniquely determine the underlying set-distribution of $\mathcal{Y}$ (and of $\mathcal{X}$). In fact, it just addresses one aspect of optimiser performance, which is location-closeness. Closeness of the approximations in terms of spread (variability across runs) could be described by the variance (second centred moment). The second-order moment (measure) of $\mathcal{Y}$

$$C^{(2)}_{\mathcal{Y}}(\{z_1\},\{z_2\}) = P\Big[\big(\mathcal{Y}\cap\{z_1\}\neq\emptyset\big) \;\wedge\; \big(\mathcal{Y}\cap\{z_2\}\neq\emptyset\big)\Big]$$
$$= P\big[\big(\mathcal{X}\trianglelefteq z_1\big) \;\wedge\; \big(\mathcal{X}\trianglelefteq z_2\big)\big]$$

(originally defined for the binary random field associated with $\mathcal{Y}$, see above) describes the probability of hitting two goals z_1 and z_2 simultaneously. Together with the first-order moment $C^{(1)}_{\mathcal{Y}}(\{z\})$, the attainment function, it can be used to explain the dependence structure between the two binary random variables $\boldsymbol{I}\{\mathcal{Y}\cap\{z_1\}\neq\emptyset\}$ and $\boldsymbol{I}\{\mathcal{Y}\cap\{z_2\}\neq\emptyset\}$. The difference

$$C^{(2)}_{\mathcal{Y}}(\{z_1\},\{z_2\}) - C^{(1)}_{\mathcal{Y}}(\{z_1\})\cdot C^{(1)}_{\mathcal{Y}}(\{z_2\})$$
$$= P\big[\big(\mathcal{X}\trianglelefteq z_1\big) \;\wedge\; \big(\mathcal{X}\trianglelefteq z_2\big)\big] - \alpha_{\mathcal{X}}(z_1)\cdot\alpha_{\mathcal{X}}(z_2)$$

can be seen as a form of covariance. If it equals zero, the two random variables are uncorrelated. On the other hand, if the event of attaining a goal z_1 is independent from the event of attaining the goal z_2 then the difference is zero (compare with Goutsias (1998)). Dependencies between more than two goals can be explored through higher-order moments of $\mathcal{Y}$. Eventually one can hope to completely characterise the distribution of the outcome-set $\mathcal{X}$ (through $\mathcal{Y}$).

Setting $z_1 = z_2 = z$, one obtains

$$C^{(2)}_{\mathcal{Y}}(\{z\},\{z\}) - C^{(1)}_{\mathcal{Y}}(\{z\})\cdot C^{(1)}_{\mathcal{Y}}(\{z\}) = \alpha_{\mathcal{X}}(z) - \alpha^2_{\mathcal{X}}(z) = \beta_{\mathcal{X}}(z)$$

which is simply the variance of the binary random variable $\boldsymbol{I}\{\mathcal{Y}\cap\{z\}\neq\emptyset\} = \boldsymbol{I}\{\mathcal{X}\trianglelefteq z\}$ for all $z\in\mathbb{R}^d$. The corresponding empirical estimator would be

$$\beta_n(z) = \frac{1}{n}\sum_{i=1}^{n}\big(\alpha_n(z) - \boldsymbol{I}\{z\in\mathcal{Y}_i\}\big)^2,$$

which is rather similar to the variance estimator defined by Stoyan (1998) for particle data.

4 Computational issues

The practical usefulness of the attainment function as a measure of multiobjective optimiser performance is tied to the ability to estimate it from experimental data. The computation of the empirical attainment function (EAF) in arbitrary dimensions (i.e., number of objectives) is related to the computation of the multivariate empirical cumulative distribution function (ECDF), but computing the multivariate ECDF efficiently is not considered an easy task, either (see Justel et al. (1997)). In fact, whereas the univariate ECDF exhibits discontinuities at the data points only, the multivariate ECDF exhibits discontinuities at the data points *and* at other points, the coordinates of which are combinations of the coordinates of the data points. As the number of dimensions increases, the number of points needed to describe the ECDF (and the EAF) may easily become too large to store. Storing all relevant points may not always be necessary, however. The maximum difference between two EAFs, for example, can be computed without that requirement.

Similar considerations apply, to an even greater extent, to the estimation of the second-order moments. Work in this area is currently in progress.

5 Conclusions and future perspectives

The performance assessment of stochastic (multiobjective) optimisers was discussed in the light of existing criteria for the performance of classical statistical estimators, and theoretical foundations for the attainment function were established within the field known as random closed set theory.

The outcomes of multiobjective optimisers are random point sets in $\mathbb{R}^d$ denoted by $\mathcal{X}$. Alternatively, they can be represented by (continuous) random closed sets $\mathcal{Y}$ of a particular type with equivalent stochastic behaviour. Considering minimisation problems, the sets $\mathcal{Y}$ are unbounded towards $+\infty$ in every dimension, and are bounded below by the elements of $\mathcal{X}$.

The attainment function of an outcome set $\mathcal{X}$ is a first-order moment measure of the corresponding set $\mathcal{Y}$, defined over all possible one-point sets in $\mathbb{R}^d$ (the general moment measure is defined over all compact subsets in $\mathbb{R}^d$). Comparing the performance assessment of optimisers with that of statistical estimators showed that the attainment function is a kind of mean measure of the outcome-set $\mathcal{X}$. As such, it does indeed address a very sensible aspect of the stochastic behaviour of the optimiser, i.e. the location of the approximation. A suitable definition of bias was also suggested, which allows the location of the approximation to be seen with respect to the unknown Pareto-front.

The attainment function is a generalisation of the (multivariate) cumulative distribution function to the case of random non-dominated point sets. Thus, also the cumulative distribution function can be seen as a mean-measure for the set $\mathcal{Y}$ describing the region in $\mathbb{R}^d$ which is attained by a single objective vector X. In a wider sense, the cumulative distribution function can be seen as a mean-measure of $\{X\}$ itself. Regarding the empirical attainment function, it is hoped that it preserves some of the good properties of the empirical cumulative distribution function. Also, the attainment function makes it possible to compare the performance of multiobjective optimisers regardless of whether they produce one or many objective vectors per run!

The attainment function does not *fully* characterise the distribution of the random sets $\mathcal{X}$ or $\mathcal{Y}$. Extensions of the attainment function based on higher-order moment concepts were introduced which could contribute with additional information. They might eventually lead to the full characterisation of the distributions considered. This perspective gives the attainment function an advantage over performance measures such as the volume measure of the attained region $\mathcal{Y}$, which is related, for example, to the "size of the dominated space" in Zitzler

(1999, p. 43f). In Matheron's (1975) theory, the distribution of a random closed set is characterised by hit-or-miss events (on which the attainment function is based) and "not by measures or contents" (Stoyan, 1998).

The results presented here are mainly of probabilistic nature. They are needed to support inferential methodology such as the test procedure for the comparison of optimiser performance used in Shaw et al. (1999), which is based on the maximum difference between two empirical attainment functions and on a permutation argument (see Good (2000)). Unlike the methodology proposed by Knowles and Corne (2000), such a test does not depend on auxiliary lines or suffer from multiple testing issues (see also Fonseca and Fleming (1996)). To a great extent, inferential methodology which truly exploits the attainment function and related concepts has yet to be developed.

Finally, the solution-quality view of optimiser performance could be combined with the run-time perspective by considering time an additional objective to be minimised. The outcome of an optimisation run would then be the set of non-dominated objective-vectors, augmented with time, evaluated during the run.

Acknowledgement

This work was funded by the Fundação para a Ciência e a Tecnologia under the PRAXIS XXI programme (Project PRAXIS-P-MAT-10135-1998), with the support of the European Social Fund, FEDER, and the Portuguese State. The authors wish to acknowledge the anonymous reviewers for their valuable comments on the original manuscript.

References

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, revised edition.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events*. Springer-Verlag, Berlin.

Fonseca, C. M. and Fleming, P. J. (1995). An overview of evolutionary algorithms in multiobjective optimization. *Evolutionary Computation*, 3(1):1–16.

Fonseca, C. M. and Fleming, P. J. (1996). On the performance assessment and comparison of stochastic multiobjective optimizers. In Voigt, H.-M., Ebeling, W., Rechenberg, I., and Schwefel, H.-P., editors, *Parallel Problem Solving from Nature – PPSN IV*, number 1141 in Lecture Notes in Computer Science, pages 584–593. Springer Verlag, Berlin.

Good, P. I. (2000). *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer Series in Statistics. Springer Verlag, New York, 2nd edition.

Goutsias, J. (1998). Modeling random shapes: An introduction to random closed set theory. Technical Report JHU/ECE 90-12, Department of Electrical and Computer Engineering, Image Analysis and Comunnications Laboratory, The John Hopkins University, Baltimore, MD 21218.

Hoos, H. and Stutzle, T. (1998). Evaluating Las Vegas algorithms — pitfalls and remedies. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 238–245.

Justel, A., Peña, D., and Zamar, R. (1997). A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statistics and Probability Letters*, 35:251–259.

Kendall, D. G. (1974). Foundations of a theory of random sets. In Harding, E. F. and Kendall, D. G., editors, *Stochastic Geometry. A Tribute to the Memory of Rollo Davidson*, pages 322–376. John Wiley & Sons, New York.

Knowles, J. D. and Corne, D. W. (2000). Approximating the nondominated front using the Pareto Archived Evolution Strategy. *IEEE Transactions on Evolutionary Computation*, 8(2):149–172.

Lockhart, R. and Stephens, M. (1994). Estimation and tests of fit for the three-parameter Weibull distribution. *Journal of the Royal Statistical Society, Series B*, 56(3):491–500.

Matheron, G. (1975). *Random Sets and Integral Geometry.* John Wiley & Sons, New York.

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the Theory of Statistics.* McGraw-Hill Series in Probability and Statistics. McGraw-Hill Book Company, Singapore, 3rd edition.

Shaw, K. J., Fonseca, C. M., Nortcliffe, A. L., Thompson, M., Love, J., and Fleming, P. J. (1999). Assessing the performance of multiobjective genetic algorithms for optimization of a batch process scheduling problem. In *Proceedings of the Congress on Evolutionary Computation (CEC99)*, volume 1, pages 37–45, Washington DC.

Smith, R. (1987). Estimating tails of probability distributions. *The Annals of Statistics*, 15(3):1174–1207.

Stoyan, D. (1998). Random sets: Models and statistics. *International Statistical Review*, 66:1–27.

Van Veldhuizen, D. and Lamont, G. B. (2000). On measuring multiobjective evolutionary algorithm performance. In *Proceedings of the 2000 Congress on Evolutionary Computation*, pages 204–211.

Witting, H. (1985). *Mathematische Statistik I.* B. G. Teubner, Stuttgart.

Zitzler, E. (1999). *Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications.* PhD thesis, Computer Engineering and Networks Laboratory, Swiss Federal Institute of Technology Zurich.

Zitzler, E., Deb, K., and Thiele, L. (1999). Comparison of multiobjective evolutionary algorithms: Empirical results (Revised version). Technical Report 70, Computer Engineering and Networks Laboratory, Swiss Federal Institute of Technology Zurich.