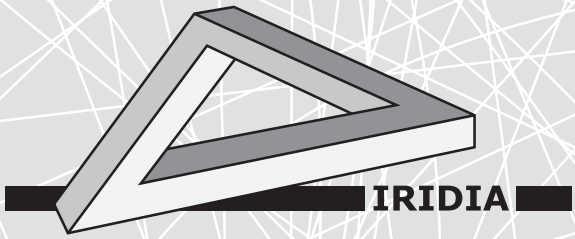

Université Libre de Bruxelles

Institut de Recherches Interdisciplinaires
et de Développements en Intelligence Artificielle

On the Estimation of the Expected Performance of a Metaheuristic on a Class of Instances

How many instances, how many runs?

Mauro Birattari

IRIDIA – Technical Report Series

Technical Report No.
TR/IRIDIA/2004-001

April 2004
Last revision: May 2005

IRIDIA – Technical Report Series
ISSN 1781-3794

Published by:

IRIDIA, *Institut de Recherches Interdisciplinaires et de Développements en Intelligence Artificielle*
Université Libre de Bruxelles
Av F. D. Roosevelt 50, CP 194/6
1050 Bruxelles, Belgium

Technical report number TR/IRIDIA/2004-001

Revision history:

TR/IRIDIA/2004-001.004	May 2005
TR/IRIDIA/2004-001.003	April 2005
TR/IRIDIA/2004-001.002	October 2004
TR/IRIDIA/2004-001.001	April 2004

On the Estimation of the Expected Performance of a Metaheuristic on a Class of Instances

How many instances, how many runs?

Mauro BIRATTARI mbiro@ulb.ac.be
IRIDIA, Université Libre de Bruxelles, Brussels, Belgium

Last revision: May 2005

Abstract

This paper discusses the problem of estimating, on the basis of a given number of say N experiments, the expected performance of a metaheuristic on a class I of benchmark problem instances. The problem of the empirical estimation of the expected behavior of a stochastic optimization algorithm has great relevance both in academic studies and in practical applications. This is particularly true for metaheuristics, a class of stochastic optimization algorithms for which gaining an analytical insight appears rather problematic.

In the paper, the estimation problem is formally posed in a probabilistic framework and the main theorems on the empirical estimation of the expected performance of a metaheuristic are enunciated and proved. In particular, the paper proves that the widely adopted methodology consisting in considering K instances and running the metaheuristic n times on each (with $K \times n = N$) is a suboptimal choice. Indeed, contrary to popular belief, *one single run* on each of N independently selected instances provides the most reliable estimation.

1 Introduction

Metaheuristics (Glover, 1986) are a class of promising algorithms for tackling hard combinatorial optimization problems. First introduced by Glover (1986) on the basis of early ideas formulated a decade before (Glover, 1977), *tabu search* is nowadays among the most cited metaheuristics. Other prominent examples of metaheuristics are *simulated annealing* (Kirkpatrick *et al.*, 1983; Cerny, 1985), *genetic algorithms* (Fogel, 1962; Fogel *et al.*, 1966; Rechenberg, 1973; Holland, 1975; Goldberg, 1989), *iterated local search* (Lourenço *et al.*, 2002), and *ant colony optimization* (Dorigo, 1992; Dorigo *et al.*, 1996, 1999; Dorigo & Stützle, 2004). Much research has been devoted to metaheuristics in the last two decades and a number of successful algorithms belonging to this class have been proposed. The following, is just a sample of research works that have been recently published by IEEE Transactions on *Systems, Man, and Cybernetics—Part B:* Pierre & Houéto (2002) discuss an application of *tabu search*, Blum & Dorigo (2004) propose an original problem-independent improvement of *ant colony optimization*, while Kato & Sakawa (2003), Guanqi & Shouyi (2003), Jin *et al.* (2003) Balakrishnan *et al.* (2004), and Tsai *et al.* (2004) profitably adopt various flavors of *genetic algorithms*, and more in general of *evolutionary algorithms*, for tackling different optimization problems. For recent surveys on metaheuristics, we refer the reader to Blum & Roli (2003) or Dréo *et al.* (2003).

Metaheuristics are particularly complex algorithms and, so far, relatively little insight could be gained on their behavior through analytical tools. For this reason, much of the research on metaheuristics is of an empirical nature.

A number of good articles have been published which prescribe some useful experimental guidelines and discuss the problem of defining an appropriate experimental methodology for studying metaheuristics—see, for example, Barr *et al.* (1995), Hooker (1995), and Rardin & Uzsoy (2001).

Nonetheless, a number of issues are still open and the need for an improvement of the quality standards of empirical analyses is still felt by a large share of the metaheuristics community. With this paper, we wish to contribute to the definition of a correct practice in the design of an experimental setting.

A typical problem faced by practitioners when conducting an empirical analysis of a metaheuristic consists in estimating its expected behavior on some given class I of instances, in terms of the quality of the best solution found in some given amount of time.[1] The basic ability of estimating the expected behavior of a metaheuristic, or more in general of any stochastic algorithm, is apparently of fundamental importance. Beside being extremely relevant in the empirical evaluation, assessment, and comparison of algorithms, it is the key issue in fine-tuning metaheuristics, when the task is selecting, among a set of candidate configurations of the parameters of the metaheuristic, the one which is expected to yield the best performance over a class of problem instances.

In order to give a precise meaning to the very concept of *expected* performance, we need to refer to a *probabilistic model* of the class I. Such idea, although already present in Wolpert & Macready (1997), does not seem to be widely accepted and exploited in the combinatorial optimization community. To the best of our knowledge, Birattari *et al.* (2002) is the only work that adopted a probabilistic model of the class of instances in the context of the definition and solution of the problem of tuning metaheuristics. The idea, which is further discussed and developed in Zlochin *et al.* (2004) and Birattari (2004), proves to be extremely convenient both in the theoretical analysis and in practical applications.

When it comes to defining an experimental setting for estimating the expected performance of a metaheuristic given that we can perform a maximum number N of runs, practitioners are often embarrassed. The question is invariably: *How many instances, how many runs?* The answers typically cover a wide range. Often it is believed that some sort of trade-off is involved in the choice: If, given an instance, the metaheuristic tends to be quite *erratic* and produces rather different results on subsequent runs, practitioners typically feel the need to perform more runs on each instance and are therefore inclined to *trade* instances for runs. On the other hand, if the metaheuristic shows a quite stable behavior on each instance, they feel like considering less runs and more instances. In any case, most of them would be quite unhappy at the idea of performing only one single run on each instance: They would argue that a metaheuristic is in any case a stochastic algorithm and that if you want your conclusions to be somehow meaningful, you need to average the results of more runs. They would probably add that it is pointless to average across more instances if the results you average, each concerning a different instance, are spoiled by a large margin of uncertainty ... The theorems presented in this paper prove them wrong on the issue. Indeed, contrary to popular belief, there is no trade-off involved in the definition of the experimental setting when the total number of runs N is fixed: The setting "N instances, one *single* run per instance" dominates all other choices.

The rest of the paper is structured as follows: Section 2 formally poses the estimation problem in a probabilistic framework and defines a class of estimators of the expected behavior of a metaheuristic. The estimators belonging to the class differ in the *experimental setting* they implement, that is, in the number of instances and number of runs. Section 3 proposes a first order analysis of the given estimators. In particular, we prove that, irrespectively of the ratio instances/runs, all estimators in the considered class are unbiased. On the other hand, Section 4 proposes a second order analysis which shows that the possible estimator are **not** equivalent with respect to their variance. Section 5 analyzes another classes of estimators that are slightly different from those introduced in Section 2. Section 6 concludes the paper.

[1] Alternatively, one could measure the performance of a metaheuristic in terms of the time needed to find a solution whose cost is below a given threshold. The results presented in the paper can be easily reformulated for the *time-to-threshold* case.

2 Formal position of the estimation problem

The framework we consider is based on the concept of a *stream* of instances: Every T seconds an instance i is selected from a class I of instances. The given metaheuristic m is supposed to run for T seconds on i and return the best solution s found during the run; that is, the solution whose cost c is not larger than the cost associated to any other solution s' visited during the run. The process is then iterated *ad libitum*. In this paper, we consider the case in which instances belonging to I are *a priori* indistinguishable: if we are given 3 instances—say i_a, i_b, and i_c—prior to running the metaheuristic on them we are not able to predict if the cost of the best solution we will obtain for i_a will be closer to the one we will obtain for i_b or the one we will obtain for i_c.[2] Further, we consider the case in which the instances are selected *independently* from a *fixed but unknown* probability measure P_I defined on I. Instance i is therefore selected with probability $P_I(i)$. Similarly, given the instance i, the cost c of the best solution found by the metaheuristic in the given amount of time, is a stochastic quantity arising with probability $P_C(c|i)$.[3]

Definition 1. We call a *scenario* for the estimation of the expected behavior of a metaheuristic, the joint probability measure $P(c, i) = P_C(c|i)P_I(i)$.

The Estimation Problem. *Estimate on the basis of N runs of a given metaheuristic, its average behavior on the class I, that is, the expected value of the cost c with respect to the scenario $P(c, i) = P_C(c|i)P_I(i)$:*

$$\mu = E[c] = \int c \, \mathrm{d}P_C(c|i) \, \mathrm{d}P_I(i), \tag{1}$$

where the operator E denotes the expectation taken with respect to the joint probability $P(c, i)$.

To this aim, we run a set J of experiments, with $|J| = N$. For each experiment $j \in \{1, 2, \ldots, N\}$ we observe a cost c_j. The quantity μ can be estimated by the estimator $\hat{\mu}$:

$$\hat{\mu} = \frac{1}{N} \sum_{j=1}^{N} c_j.$$

To be more precise, let us suppose we sample K distinct instances $i_1, i_2, \ldots, i_K$, with $K \leq N$, and we run the given metaheuristic for n_1 times on instance i_1, for n_2 times on instance i_2, and so on. This amounts to considering a set of experiments J which is partitioned in subsets $J_1, J_2, \ldots, J_K$, where $|J_k| = n_k$ and $\sum_k n_k = N$: each experiment j in the subset J_k consisting in running once the metaheuristic on instance i_k for observing the cost c_{kj}.

Definition 2. We call an *experimental setting*, or more simply a *setting*, the sequence of natural numbers $\mathcal{S}_N = (K, n_1, n_2, \ldots, n_K)$, that is, the specification of how many instances have to be considered, together with the number of runs to perform on each of them.

For convenience, we also introduce the following notation:

Definition 3. If K divides N, we denote with $\mathcal{H}_{K|N/K}$ the *homogeneous* setting, that is, $\mathcal{S}_N = (K, n_1, n_2, \ldots, n_K)$, where $n_k = N/K$ for all k. In particular, $\mathcal{H}_{N|1} = (N, n_1, n_2, \ldots, n_N)$, with $n_k = 1$ for all k, is the setting "N instances, one run per instance." Similarly, $\mathcal{H}_{1|N} = (1, N)$ is the setting "one instance, N runs."

[2] Such hypothesis is not too restrictive: In the case we are able to *a priori* distinguish among instances, we can consider a partition of I into disjoint subsets $I_1, I_2, \ldots, I_D$. The discussion we present in the paper holds within each of the sets $I_1, I_2, \ldots, I_D$. The partition of the original set and the decomposition of the original estimation problem into subproblems is connected to the notion of *stratified sampling* which is a well known *variance reduction technique* adopted in Monte Carlo estimation—See for example Rubinstein (1981). A discussion of this issue goes beyond the aims of this paper.

[3] In the following, integration is taken in the sense of *Lebesgue* in order to transparently account for both discrete and continuous quantities (Billingsley, 1986; Khuri, 2003). We adopt the notation $P_I(i)$ as a shorthand for the more correct $P_I(\{i\})$, that is, the measure of the singleton $\{i\}$. This remark is rather technical and does not have any major impact on the following discussion. It can be safely skipped if the reader is not particularly familiar with integration theory.

Definition 4. In a given scenario $P(c,i) = P_C(c|i)P_I(i)$, and for a given experimental setting $\mathcal{S}_N = (K, n_1, n_2, \ldots, n_K)$, the estimator $\hat{\mu}_{\mathcal{S}_N}$ of the expected value μ of the cost c is given by:

$$\hat{\mu}_{\mathcal{S}_N} = \frac{1}{N} \sum_{k=1}^{K} \sum_{j=1}^{n_k} c_{kj}, \tag{2}$$

where $N = \sum_{k=1}^{K} |J_k|$, $n_k = |J_k|$, and instances i_k and costs c_{kj} are extracted according to $P(c,i)$.

The following quantities are used in the following:

Definition 5. The expected value of the cost c within instance i is given by:

$$\mu_i = E[c|i] = \int c \, \mathrm{d}P_C(c|i).$$

Definition 6. The variance of the cost c within instance i is given by:

$$\sigma_i^2 = E\big[(c-\mu_i)^2|i\big] = \int (c-\mu_i)^2 \, \mathrm{d}P_C(c|i).$$

Definition 7. The expected *within-instance* variance is:

$$\bar{\sigma}^2_{\mathrm{WI}} = \int \sigma_i^2 \, \mathrm{d}P_I(i),$$

that is, the expected value with respect to the distribution of the instances of the variance of c within a same instance.

Definition 8. The *across-instance* variance is:

$$\sigma^2_{\mathrm{AI}} = \int (\mu_i - \mu)^2 \, \mathrm{d}P_I(i),$$

that is, the variance across the instances of the expected value of the cost for each instance.

3 First order analysis of the estimator $\hat{\mu}_{\mathcal{S}_N}$

Lemma 1. *In a given scenario $P(c,i) = P_C(c|,i)P_I(i)$, and for a given experimental setting $\mathcal{S}_N = (K, n_1, n_2, \ldots, n_K)$, the probability of obtaining the specific instances $i_1, i_2, \ldots, i_K$ and the specific results $c_{11}, c_{12}, \ldots, c_{1n_1}, c_{21}, c_{22}, \ldots, c_{2n_2}, \ldots, c_{K1}, c_{K2}, \ldots, c_{Kn_K}$, on which $\hat{\mu}_{\mathcal{S}_N}$ is based, is given by:*

$$P(i_1, i_2, \ldots, i_K, c_{11}, \ldots, c_{1n_1}, c_{21}, \ldots, c_{2n_2}, \ldots, c_{K1}, \ldots, c_{Kn_K}) = \prod_{k=1}^{K} P_I(i_k) \prod_{j=1}^{n_k} P_C(c_{kj}|i_k),$$

where $P_I(i_k)$ is the probability of sampling instance i_k, and $P_C(c_{kj}|i_k)$ is the probability of obtaining the cost c_{kj} as best result in a run of the metaheuristic on instance i_k.

Proof. The K instances are sampled independently according to the probability measure $P_I(i)$. Similarly, the costs c obtained on a given instance i are sampled independently according to $P_C(c|i)$. The joint probability is therefore the product of the terms. □

Theorem 1. *In all scenarios, irrespectively of the setting $\mathcal{S}_N$, that is, of how K and n_k with $k = 1 \ldots K$ are selected, $\hat{\mu}_{\mathcal{S}_N}$ is an unbiased estimator of μ.*

Proof. The proof is immediate and is given only for the sake of completeness:[4]

$$\int \hat{\mu}_{\mathcal{S}_N} \, \mathrm{d}P(\hat{\mu}_{\mathcal{S}_N}) = \int \frac{1}{N} \sum_{k=1}^{K} \sum_{j=1}^{n_k} c_{kj} \bigodot_{k=1}^{K} \mathrm{d}P_I(i_k) \bigodot_{j=1}^{n_k} \mathrm{d}P_C(c_{kj}|i_k)$$

$$= \frac{1}{N} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \int c_{kj} \, \mathrm{d}P_C(c_{kj}|i_k) \, \mathrm{d}P_I(i_k) = \mu.$$

□

In particular, $\hat{\mu}_{\mathcal{H}_{1|1}}$, based on a single run on a single instance, is an unbiased estimator of μ, irrespectively of which instance is considered, provided it is selected from I according to the *unknown* probability $P_I(i)$.[5] Similarly, the estimator $\hat{\mu}_{\mathcal{H}_{1|N}}$ based on N runs on one single instance is unbiased as well as $\hat{\mu}_{\mathcal{H}_{N/10|10}}$ which considers $N/10$ instances, 10 runs per instance.

4 Second order analysis of the estimator $\hat{\mu}_{\mathcal{S}_N}$

All possible estimators that can be written in the form given in Equation 2 are therefore equivalent for what concerns their expected behavior. Nonetheless, they differ for what concerns second order statistics. We are therefore interested here in finding the best *minimum-variance* estimator when the total number N of experiments is fixed. In simple words, we want to answer the question:

> *If I can run $N = 100$ experiments, should I consider (i) 1 instance and 100 runs; (ii) 10 instances and 10 runs on each; (iii) 100 instance and 1 single run on each; or what else?*

Lemma 2. *In a given scenario $P(c,i) = P_C(c|i)P_I(i)$, and for a given experimental setting $\mathcal{S}_N = (K, n_1, n_2, \ldots, n_K)$, the variance of the estimator $\hat{\mu}_{\mathcal{S}_N}$ is given by:*

$$\int (\hat{\mu}_{\mathcal{S}_N} - \mu)^2 \, \mathrm{d}P(\hat{\mu}_{\mathcal{S}_N}) = \frac{1}{N} \, \bar{\sigma}^2_{\mathrm{WI}} + \frac{\sum_{k=1}^{K} n_k^2}{N^2} \, \sigma^2_{\mathrm{AI}}.$$

Proof. It results:

$$\int (\hat{\mu}_{\mathcal{S}_N} - \mu)^2 \, \mathrm{d}P(\hat{\mu}_{\mathcal{S}_N}) = \int \left(\frac{1}{N} \sum_{k=1}^{K} \sum_{j=1}^{n_k} c_{kj} - \mu \right)^2 \bigodot_{k=1}^{K} \mathrm{d}P_I(i_k) \bigodot_{j=1}^{n_k} \mathrm{d}P_C(c_{kj}|i_k) =$$

$$= \int \left(\frac{1}{N} \sum_{k=1}^{K} \sum_{j=1}^{n_k} (c_{kj} - \mu_{i_k} + \mu_{i_k} - \mu) \right)^2 \bigodot_{k=1}^{K} \mathrm{d}P_I(i_k) \bigodot_{j=1}^{n_k} \mathrm{d}P_C(c_{kj}|i_k),$$

[4] In the following, with the notation: $\int f(x_1, x_2, \ldots, x_L) \bigodot_{l=1}^{L} \mathrm{d}P(x_l)$, we denote the sequence of nested integrals $\iint \cdots \int f(x_1, x_2, \ldots, x_L) \, \mathrm{d}P(x_1) \, \mathrm{d}P(x_2) \ldots \mathrm{d}P(x_L)$.

[5] The fact that $P_I(i)$ is unknown does not pose here any problem: in order to obtain an instance i extracted from I according to the unknown $P_I(i)$ it is sufficient to take randomly any of the instances that appear in the above described stream of instances: let's say the next one!

It follows that:

$$\int (\hat{\mu}_{\mathcal{S}_N} - \mu)^2 \, \mathrm{d}P(\hat{\mu}_{\mathcal{S}_N}) =$$

$$= \frac{1}{N^2} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \sum_{k'=1}^{K} \sum_{j'=1}^{n_{k'}} \int (c_{kj} - \mu_{i_k})(c_{k'j'} - \mu_{i_{k'}}) \, \mathrm{d}P_C(c_{kj}|i_k) \, \mathrm{d}P_I(i_k) \, \mathrm{d}P_C(c_{k'j'}|i_{k'}) \, \mathrm{d}P_I(i_{k'}) + \quad (3a)$$

$$+ \frac{1}{N^2} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \sum_{k'=1}^{K} \sum_{j'=1}^{n_{k'}} \int (\mu_{i_k} - \mu)(\mu_{i_{k'}} - \mu) \, \mathrm{d}P_I(i_k) \, \mathrm{d}P_I(i_{k'}) + \quad (3b)$$

$$+ \frac{1}{N^2} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \sum_{k'=1}^{K} \sum_{j'=1}^{n_{k'}} 2 \int (c_{kj} - \mu_{i_k})(\mu_{i_{k'}} - \mu) \, \mathrm{d}P_C(c_{kj}|i_k) \, \mathrm{d}P_I(i_k) \, \mathrm{d}P_I(i_{k'}). \quad (3c)$$

Let us now consider one by one the three addends given in 3a, 3b, and 3c.

Addend 3a: Concerning the terms for which $k \neq k'$, it results:[6]

$$\int (c_{kj} - \mu_{i_k})(c_{k'j'} - \mu_{i_{k'}}) \, \mathrm{d}P_C(c_{kj}|i_k) \, \mathrm{d}P_I(i_k) \, \mathrm{d}P_C(c_{k'j'}|i_{k'}) \, \mathrm{d}P_I(i_{k'}) =$$

$$= \int (c_{kj} - \mu_{i_k}) \, \mathrm{d}\cancelto{0}{P_C(c_{kj}|i_k)} \, \mathrm{d}P_I(i_k) \int (c_{k'j'} - \mu_{i_{k'}}) \, \mathrm{d}\cancelto{0}{P_C(c_{k'j'}|i_{k'})} \, \mathrm{d}P_I(i_{k'}) = 0.$$

Similarly, if $k = k'$ but $j \neq j'$, it results:

$$\int (c_{kj} - \mu_{i_k})(c_{k'j'} - \mu_{i_{k'}}) \, \mathrm{d}P_C(c_{kj}|i_k) \, \mathrm{d}P_I(i_k) \, \mathrm{d}P_C(c_{k'j'}|i_{k'}) \, \mathrm{d}P_I(i_{k'}) =$$

$$= \int \left(\int (c_{kj} - \mu_{i_k}) \, \mathrm{d}\cancelto{0}{P_C(c_{kj}|i_k)} \int (c_{kj'} - \mu_{i_k}) \, \mathrm{d}\cancelto{0}{P_C(c_{kj'}|i_k)} \right) \mathrm{d}P(i_k) = 0.$$

On the other hand, if $k = k'$ and $j = j'$, it results:

$$\int (c_{kj} - \mu_{i_k})(c_{k'j'} - \mu_{i_{k'}}) \, \mathrm{d}P_C(c_{kj}|i_k) \, \mathrm{d}P_I(i_k) \, \mathrm{d}P_C(c_{k'j'}|i_{k'}) \, \mathrm{d}P_I(i_{k'}) =$$

$$= \int (c_{kj} - \mu_{i_k})^2 \, \mathrm{d}P_C(c_{kj}|i_k) \, \mathrm{d}P_I(i_k).$$

Thus, addend 3a amounts to:

$$\frac{1}{N^2} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \int (c_{kj} - \mu_{i_k})^2 \, \mathrm{d}P_C(c_{kj}|i_k) \, \mathrm{d}P_I(i_k).$$

Addend 3b: since the integrand is independent from j and j', it results:

$$\frac{1}{N^2} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \sum_{k'=1}^{K} \sum_{j'=1}^{n_{k'}} \int (\mu_{i_k} - \mu)(\mu_{i_{k'}} - \mu) \, \mathrm{d}P_I(i_k) \, \mathrm{d}P_I(i_{k'}) =$$

$$= \frac{1}{N^2} \sum_{k=1}^{K} \sum_{k'=1}^{K} n_k n_{k'} \int (\mu_{i_k} - \mu)(\mu_{i_{k'}} - \mu) \, \mathrm{d}P_I(i_k) \, \mathrm{d}P_I(i_{k'}).$$

[6] In the following, with the notation

$$\int g(y) \, \mathrm{d}\cancelto{0}{y},$$

we graphically indicate that the integral is null. This will help visualizing which terms of an equations are null. For example, in

$$\int \left(f(x) \int g(y) \, \mathrm{d}\cancelto{0}{y} \right) \mathrm{d}x = 0,$$

the adopted notation helps showing that the reason why the overall double integral is null is because $\int g(y) \, \mathrm{d}y = 0$.

If $k \neq k'$, it results:

$$\int (\mu_{i_k} - \mu)(\mu_{i_{k'}} - \mu)\, \mathrm{d}P_I(i_k)\, \mathrm{d}P_I(i_{k'}) = \int (\mu_{i_k} - \mu)\, \mathrm{d}P_I(i_k) \overset{0}{\cancel{\int}} (\mu_{i_{k'}} - \mu)\, \mathrm{d}P_I(i_{k'}) = 0.$$

Otherwise, if $k = k'$, it results:

$$\int (\mu_{i_k} - \mu)(\mu_{i_{k'}} - \mu)\, \mathrm{d}P_I(i_k)\, \mathrm{d}P_I(i_{k'}) = \int (\mu_{i_k} - \mu)^2\, \mathrm{d}P_I(i_k).$$

Thus, addend 3b amounts to:

$$\frac{1}{N^2} \sum_{k=1}^{K} n_k^2 \int (\mu_{i_k} - \mu)^2\, \mathrm{d}P_I(i_k).$$

Addend 3c: it results:

$$\int (c_{kj} - \mu_{i_k})(\mu_{i_{k'}} - \mu)\, \mathrm{d}P_C(c_{kj}|i_k)\, \mathrm{d}P_I(i_k)\, \mathrm{d}P_I(i_{k'}) =$$
$$= \int \left((\mu_{i_{k'}} - \mu) \overset{0}{\cancel{\int}} (c_{kj} - \mu_{i_k})\, \mathrm{d}P_C(c_{kj}|i_k) \right) \mathrm{d}P_I(i_k)\, \mathrm{d}P_I(i_{k'}) = 0,$$

Thus, addend 3c is identically null.

It results therefore:

$$\int (\hat{\mu}_{\mathcal{S}_N} - \mu)^2\, \mathrm{d}P(\hat{\mu}_{\mathcal{S}_N}) =$$
$$= \frac{1}{N^2} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \int (c_{kj} - \mu_{i_k})^2\, \mathrm{d}P_C(c_{kj}|i_k)\, \mathrm{d}P_I(i_k) + \frac{1}{N^2} \sum_{k=1}^{K} n_k^2 \int (\mu_{i_k} - \mu)^2\, \mathrm{d}P_I(i_k).$$

On the basis of Definitions 7 and 8 we can write:

$$\int (\hat{\mu}_{\mathcal{S}_N} - \mu)^2\, \mathrm{d}P(\hat{\mu}_{\mathcal{S}_N}) = \frac{1}{N^2} \sum_{k=1}^{K} n_k\, \bar{\sigma}^2_{\mathrm{WI}} + \frac{1}{N^2} \sum_{k=1}^{K} n_k^2\, \sigma^2_{\mathrm{AI}}.$$

Remembering that $\sum_{k=1}^{K} n_k = N$, it results:

$$\int (\hat{\mu}_{\mathcal{S}_N} - \mu)^2\, \mathrm{d}P(\hat{\mu}_{\mathcal{S}_N}) = \frac{1}{N}\, \bar{\sigma}^2_{\mathrm{WI}} + \frac{\sum_{k=1}^{K} n_k^2}{N^2}\, \sigma^2_{\mathrm{AI}}.$$

□

Let us go back to our original question: With the constraint that the total number of runs must be N, what is the optimal number of instances to consider and how many runs to perform on each?

Theorem 2. *The variance of $\hat{\mu}_{\mathcal{S}_N}$ is minimized by the experimental setting $\bar{\mathcal{S}}_N = \mathcal{H}_{N|1}$, that is, by the setting "N instances, one run per instance."*

Proof. According to Lemma 2, the variance of $\hat{\mu}_{\mathcal{S}_N}$ is:

$$\int (\hat{\mu}_{\mathcal{S}_N} - \mu)^2\, \mathrm{d}P(\hat{\mu}_{\mathcal{S}_N}) = \frac{1}{N}\, \bar{\sigma}^2_{\mathrm{WI}} + \frac{\sum_{k=1}^{K} n_k^2}{N^2}\, \sigma^2_{\mathrm{AI}}.$$

Since the first addend does not depend on $\mathcal{S}_N = (K, n_1, \ldots, n_K)$, we can focus on the minimization of the second. Moreover, since N is fixed and σ^2_{AI} is out of our control, we focus on the minimization of:

$$\mathcal{C}(\mathcal{S}_N) = \sum_{k=1}^{K} n_k^2, \qquad \text{under the constraint:} \quad \sum_{k=1}^{K} n_k = N.$$

Let us assume now, by way of contradiction, that an experimental setting $\mathcal{S}_N = (K, n_1, \ldots, n_K)$ exists which is different form $\bar{\mathcal{S}}_N = \mathcal{H}_{N|1}$, satisfies the constraint, and which minimizes $\mathcal{C}$; that is, $\mathcal{S}_N$ is such that $\mathcal{C}(\mathcal{S}_N) \leq \mathcal{C}(\mathcal{S}'_N)$, for all $\mathcal{S}'_N$. Clearly, it must be $K \leq N$—otherwise the constraint would not be satisfied. Indeed, more precisely, it must be $K < N$ because otherwise it would be mandatory, in order to satisfy the constraint, to set $n_k = 1$ for all k and, in this case, we would fall back to the original statement to be proved. If $K < N$, in order to satisfy the constraint, there must exist at least an index q for which $n_q > 1$.

On the basis of the setting $\mathcal{S}_N$, we can construct another setting $\mathcal{S}'_N = (K', n'_1, \ldots, n'_{K'})$ where $K' = K + 1$, $n'_q = n_q - 1$, $n'_{K'} = 1$, and $n'_j = n_j$ otherwise. It is immediate to check that this second sequence satisfies the constraint if $\mathcal{S}_N$ does. Moreover, it results:

$$\mathcal{C}(\mathcal{S}'_N) = \sum_{k=1}^{K'} n_k'^2 = \sum_{k=1}^{K} n_k^2 - n_q^2 + n_q'^2 + 1 = \mathcal{C}(\mathcal{S}_N) - n_q^2 + (n_q - 1)^2 + 1 = \mathcal{C}(\mathcal{S}_N) - 2(n_q - 1).$$

Since $n_q > 1$, the term $2(n_q - 1)$ is strictly positive and the experimental setting $\mathcal{S}'_N$ is thus better that $\mathcal{S}_N$, which is a contradiction. □

Corollary 1. *The variance of the best estimator $\hat{\mu}_{\mathcal{H}_{N|1}}$ is:*

$$E\big[(\hat{\mu}_{\mathcal{H}_{N|1}} - \mu)^2\big] = \frac{1}{N}\Big(\bar{\sigma}^2_{\text{WI}} + \sigma^2_{\text{AI}}\Big).$$

Proof. It follows trivially from Lemma 2. □

Corollary 2. *$\hat{\mu}_{\mathcal{H}_{N|1}}$ is a consistent estimator of μ, that is, it converges in probability to μ:*

$$\lim_{N\to\infty} \mathit{Prob}\left\{|\hat{\mu}_{\mathcal{H}_{N|1}} - \mu| > \epsilon\right\} = 0, \quad \forall \epsilon > 0$$

Proof. The proof descends directly from Corollary 1. Indeed, $\hat{\mu}_{\mathcal{H}_{N|1}}$ converges to μ in the *mean square* sense: Provided that $\bar{\sigma}^2_{\text{WI}}$ and σ^2_{AI} are finite, as N tends to infinity, $E\big[(\hat{\mu}_{N|1} - \mu)^2\big]$ converges to zero. The statement follows, since convergence in *mean square* implies convergence in *probability* (Papoulis, 1991). □

It is interesting to consider here a numerical example that compares the best estimator $\hat{\mu}_{\mathcal{H}_{N|1}}$ with other possible estimators of μ. For definiteness, let us assume in this example that the total number of runs is fixed to $N = 100$, and let us study the variance of the estimators that are obtained under the following three different experimental settings: (i) N instances, one run per instance, (ii) 10 instances, 10 runs per instance and, finally, (iii) one single instance, 100 runs. From Lemma 2 it results:

Setting 1: 100 instances, 1 run per instance—best estimator according to Theorem 2.

$$E\big[(\hat{\mu}_{\mathcal{H}_{100|1}} - \mu)^2\big] = \frac{1}{100}\,\bar{\sigma}^2_{\text{WI}} + \frac{1}{100}\,\sigma^2_{\text{AI}}.$$

Setting 2: 10 instances, 10 runs per instance.

$$E\big[(\hat{\mu}_{\mathcal{H}_{10|10}} - \mu)^2\big] = \frac{1}{100}\,\bar{\sigma}^2_{\text{WI}} + \frac{1}{10}\,\sigma^2_{\text{AI}}.$$

Setting 3: 1 instance, 100 runs.

$$E\big[(\hat{\mu}_{\mathcal{H}_{1|100}} - \mu)^2\big] = \frac{1}{100}\,\bar{\sigma}^2_{\text{WI}} + \,\sigma^2_{\text{AI}}.$$

While the three settings act in the same way on the coefficient of the first term, a difference emerges for what concerns the coefficient of the second term: Settings 2 and 3 fail to efficiently reduce the contribution of the *across-instance* variance. The variance yielded by the three settings is equal only in the trivial case in which the *across-instance* variance is null, that is, when all instances share the same expected cost $\mu = \mu_i$, for all $i \in I$.

Remark 1. Although the estimator $\hat{\mu}_{\mathcal{H}_{10|10}}$ considered in Setting 2 is less *data-efficient* than the best $\hat{\mu}_{\mathcal{H}_{100|1}}$ considered in Setting 1, it is nonetheless consistent. On the other hand, the estimator $\hat{\mu}_{\mathcal{H}_{1|100}}$ given in Setting 3 is not consistent—apart for the trivial case in which $\sigma^2_{\mathrm{AI}} = 0$.

Remark 2. It should be noticed that no scenario exists in which the estimator $\hat{\mu}_{\mathcal{H}_{N|1}}$ yields a higher variance than any other estimator $\hat{\mu}_{\mathcal{S}_N}$. That is, no better setting exists than "N instances, one run per instance," irrespectively of the measures P_I and P_C.

Corollary 3. *The variance of the cost c obtained by the given metaheuristic on the whole class I of instances can be decomposed in two terms, the expected* within-instance *variance and the* across-instance *variance:*

$$\sigma^2 = E\big[(c-\mu)^2\big] = \bar{\sigma}^2_{\mathrm{WI}} + \sigma^2_{\mathrm{AI}}.$$

Proof. The result follows immediately Corollary 1 if we notice that the variance of the cost c is equal to the variance of an estimator $\hat{\mu}_{\mathcal{H}_{1|1}} = c$ based on a single sample. □

The expected *within-instance* variance $\bar{\sigma}^2_{\mathrm{WI}}$ measures how different can be the costs c obtained by the metaheuristic in different runs on the same instance; this quantity is averaged over all instance in I. On the other hand, the *across-instance* variance σ^2_{AI} measures how different the instances are one from the other for what concerns the expected value of the cost obtained by the given metaheuristic.

Remark 3. Taken together, Corollaries 1 and 3 are just the statement, in a multivariate setting, of a basic and well known property of the variance of empirical estimates of univariate quantities: Given an univariate stochastic variable x with $E[x] = \mu$ and $E\big[(x-\mu)^2\big] = \sigma^2$, the variance of $\hat{\mu}_N = 1/N \sum_{j=1}^{N} x_j$, where x_j are independently realizations of x, is given by $E\big[(\hat{\mu}_N - \mu)^2\big] = \sigma^2/N$.

5 Yet another possible estimator

Somebody might wish to consider the *average* across different instances of the *averages* of the results obtained for each of the considered instances. Formally, this estimator is:

Definition 9. The estimator $\tilde{\mu}_{\mathcal{S}_N}$ is given by:

$$\tilde{\mu}_{\mathcal{S}_N} = \frac{1}{K} \sum_{k=1}^{K} \left(\frac{1}{n_k} \sum_{j=1}^{n_k} c_{kj} \right).$$

Remark 4. It can be immediately verified that if the experimental setting is homogeneous, that is, if $\mathcal{S}_N = \mathcal{H}_{K|N/K}$, then $\tilde{\mu}_{\mathcal{H}_{K|N/K}} = \hat{\mu}_{\mathcal{H}_{K|N/K}}$.

Theorem 3. *$\tilde{\mu}_{\mathcal{S}_N}$ is an unbiased estimator of μ.*

Proof. The proof is immediate and is given only for the sake of completeness:

$$\begin{aligned}
\int \tilde{\mu}_{\mathcal{S}_N} \, \mathrm{d}P(\tilde{\mu}_{\mathcal{S}_N}) &= \int \sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{c_{kj}}{n_k K} \bigodot_{k=1}^{K} \mathrm{d}P_I(i_k) \bigodot_{j=1}^{n_k} \mathrm{d}P_C(c_{kj}|i_k) \\
&= \sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{1}{n_k K} \int c_{kj} \, \mathrm{d}P_C(c_{kj}|i_k) \, \mathrm{d}P_I(i_k) = \mu.
\end{aligned}$$

□

Lemma 3. *In a given scenario $P(c,i) = P_C(c|i)P_I(i)$, and for a given experimental setting $\mathcal{S}_N = (K, n_1, n_2, \ldots, n_K)$, the variance of the estimator $\tilde{\mu}_{\mathcal{S}_N}$ is given by:*

$$\int (\tilde{\mu}_{\mathcal{S}_N} - \mu)^2 \, \mathrm{d}P(\hat{\mu}_{\mathcal{S}_N}) = \frac{1}{K^2} \sum_{k=1}^{K} \frac{1}{n_k} \, \bar{\sigma}^2_{\mathrm{WI}} + \frac{1}{K} \, \sigma^2_{\mathrm{AI}}.$$

Proof. It results:

$$\int (\tilde{\mu}_{\mathcal{S}_N} - \mu)^2 \, \mathrm{d}P(\tilde{\mu}_{\mathcal{S}_N}) = \int \left(\sum_{k=1}^{K} \left(\frac{1}{n_k K} \sum_{j=1}^{n_k} c_{kj} \right) - \mu \right)^2 \bigodot_{k=1}^{K} \mathrm{d}P_I(i_k) \bigodot_{j=1}^{n_k} \mathrm{d}P_C(c_{kj}|i_k) =$$

$$= \int \left(\sum_{k=1}^{K} \sum_{j=1}^{n_k} \frac{c_{kj} - \mu_{i_k} + \mu_{i_k} - \mu}{n_k K} \right)^2 \bigodot_{k=1}^{K} \mathrm{d}P_I(i_k) \bigodot_{j=1}^{n_k} \mathrm{d}P_C(c_{kj}|i_k).$$

It follows that:

$$\int (\tilde{\mu}_{\mathcal{S}_N} - \mu)^2 \, \mathrm{d}P(\tilde{\mu}_{\mathcal{S}_N}) =$$

$$= \sum_{k=1}^{K} \sum_{j=1}^{n_k} \sum_{k'=1}^{K} \sum_{j'=1}^{n_{k'}} \int \frac{c_{kj} - \mu_{i_k}}{n_k K} \frac{c_{k'j'} - \mu_{i_{k'}}}{n_{k'} K} \, \mathrm{d}P_C(c_{kj}|i_k) \, \mathrm{d}P_I(i_k) \, \mathrm{d}P_C(c_{k'j'}|i_{k'}) \, \mathrm{d}P_I(i_{k'}) + \tag{4a}$$

$$+ \sum_{k=1}^{K} \sum_{j=1}^{n_k} \sum_{k'=1}^{K} \sum_{j'=1}^{n_{k'}} \int \frac{\mu_{i_k} - \mu}{n_k K} \frac{\mu_{i_{k'}} - \mu}{n_{k'} K} \, \mathrm{d}P_I(i_k) \, \mathrm{d}P_I(i_{k'}) + \tag{4b}$$

$$+ \sum_{k=1}^{K} \sum_{j=1}^{n_k} \sum_{k'=1}^{K} \sum_{j'=1}^{n_{k'}} 2 \int \frac{c_{kj} - \mu_{i_k}}{n_k K} \frac{\mu_{i_{k'}} - \mu}{n_{k'} K} \, \mathrm{d}P_C(c_{kj}|i_k) \, \mathrm{d}P_I(i_k) \, \mathrm{d}P_I(i_{k'}). \tag{4c}$$

Let us now consider one by one the three addends given in 4a, 4b, and 4c.

Addend 4a: Concerning the terms for which $k \neq k'$, it results:

$$\int \frac{c_{kj} - \mu_{i_k}}{n_k K} \frac{c_{k'j'} - \mu_{i_{k'}}}{n_{k'} K} \, \mathrm{d}P_C(c_{kj}|i_k) \, \mathrm{d}P_I(i_k) \, \mathrm{d}P_C(c_{k'j'}|i_{k'}) \, \mathrm{d}P_I(i_{k'}) =$$

$$= \int \frac{c_{kj} - \mu_{i_k}}{n_k K} \cancelto{0}{\mathrm{d}P_C(c_{kj}|i_k)} \, \mathrm{d}P_I(i_k) \int \frac{c_{k'j'} - \mu_{i_{k'}}}{n_{k'} K} \cancelto{0}{\mathrm{d}P_C(c_{k'j'}|i_{k'})} \, \mathrm{d}P_I(i_{k'}) = 0.$$

Similarly, if $k = k'$ but $j \neq j'$, it results:

$$\int \frac{c_{kj} - \mu_{i_k}}{n_k K} \frac{c_{k'j'} - \mu_{i_{k'}}}{n_{k'} K} \, \mathrm{d}P_C(c_{kj}|i_k) \, \mathrm{d}P_I(i_k) \, \mathrm{d}P_C(c_{k'j'}|i_{k'}) \, \mathrm{d}P_I(i_{k'}) =$$

$$= \int \left(\int \frac{c_{kj} - \mu_{i_k}}{n_k K} \cancelto{0}{\mathrm{d}P_C(c_{kj}|i_k)} \int \frac{c_{kj'} - \mu_{i_k}}{n_k K} \cancelto{0}{\mathrm{d}P_C(c_{kj'}|i_k)} \right) \mathrm{d}P(i_k) = 0.$$

On the other hand, if $k = k'$ and $j = j'$, it results:

$$\int \frac{c_{kj} - \mu_{i_k}}{n_k K} \frac{c_{k'j'} - \mu_{i_{k'}}}{n_{k'} K} \, \mathrm{d}P_C(c_{kj}|i_k) \, \mathrm{d}P_I(i_k) \, \mathrm{d}P_C(c_{k'j'}|i_{k'}) \, \mathrm{d}P_I(i_{k'}) =$$

$$= \int \frac{(c_{kj} - \mu_{i_k})^2}{(n_k K)^2} \, \mathrm{d}P_C(c_{kj}|i_k) \, \mathrm{d}P_I(i_k).$$

Thus, addend 4a amounts to:

$$\frac{1}{K^2} \sum_{k=1}^{K} \left(\frac{1}{n_k^2} \sum_{j=1}^{n_k} \int (c_{kj} - \mu_{i_k})^2 \, \mathrm{d}P_C(c_{kj}|i_k) \, \mathrm{d}P_I(i_k) \right).$$

Addend 4b: since the integrand is independent from j and j', it results:

$$\sum_{k=1}^{K}\sum_{j=1}^{n_k}\sum_{k'=1}^{K}\sum_{j'=1}^{n_{k'}}\int \frac{\mu_{i_k}-\mu}{n_k K}\frac{\mu_{i_{k'}}-\mu}{n_{k'}K}\,\mathrm{d}P_I(i_k)\,\mathrm{d}P_I(i_{k'}) =$$

$$= \frac{1}{K^2}\sum_{k=1}^{K}\sum_{k'=1}^{K}\int(\mu_{i_k}-\mu)(\mu_{i_{k'}}-\mu)\,\mathrm{d}P_I(i_k)\,\mathrm{d}P_I(i_{k'}).$$

If $k \neq k'$, it results:

$$\int(\mu_{i_k}-\mu)(\mu_{i_{k'}}-\mu)\,\mathrm{d}P_I(i_k)\,\mathrm{d}P_I(i_{k'}) = \int(\mu_{i_k}-\mu)\,\mathrm{d}P_I(i_k)\overset{0}{\cancel{\int}}(\mu_{i_{k'}}-\mu)\,\mathrm{d}P_I(i_{k'}) = 0.$$

Otherwise, if $k = k'$, it results:

$$\int(\mu_{i_k}-\mu)(\mu_{i_{k'}}-\mu)\,\mathrm{d}P_I(i_k)\,\mathrm{d}P_I(i_{k'}) = \int(\mu_{i_k}-\mu)^2\,\mathrm{d}P_I(i_k).$$

Thus, addend 4b amounts to:

$$\frac{1}{K^2}\sum_{k=1}^{K}\int(\mu_{i_k}-\mu)^2\,\mathrm{d}P_I(i_k).$$

Addend 4c: it results:

$$\int\frac{c_{kj}-\mu_{i_k}}{n_k K}\frac{\mu_{i_{k'}}-\mu}{n_{k'}K}\,\mathrm{d}P_C(c_{kj}|i_k)\,\mathrm{d}P_I(i_k)\,\mathrm{d}P_I(i_{k'}) =$$

$$= \int\left(\frac{\mu_{i_{k'}}-\mu}{n_{k'}K}\overset{0}{\cancel{\int}}\frac{c_{kj}-\mu_{i_k}}{n_k K}\,\mathrm{d}P_C(c_{kj}|i_k)\right)\mathrm{d}P_I(i_k)\,\mathrm{d}P_I(i_{k'}) = 0.$$

Thus, addend 4c is identically null.

It results therefore:

$$\int(\tilde{\mu}_{\mathcal{S}_N}-\mu)^2\,\mathrm{d}P(\tilde{\mu}_{\mathcal{S}_N}) =$$

$$= \frac{1}{K^2}\sum_{k=1}^{K}\left(\frac{1}{n_k^2}\sum_{j=1}^{n_k}\int(c_{kj}-\mu_{i_k})^2\,\mathrm{d}P_C(c_{kj}|i_k)\,\mathrm{d}P_I(i_k)\right) + \frac{1}{K^2}\sum_{k=1}^{K}\int(\mu_{i_k}-\mu)^2\,\mathrm{d}P_I(i_k).$$

On the basis of Definitions 7 and 8 we can write:

$$\int(\tilde{\mu}_{\mathcal{S}_N}-\mu)^2\,\mathrm{d}P(\tilde{\mu}_{\mathcal{S}_N}) = \frac{1}{K^2}\sum_{k=1}^{K}\frac{1}{n_k}\,\bar{\sigma}^2_{\mathrm{WI}} + \frac{1}{K}\,\sigma^2_{\mathrm{AI}}.$$

□

Let us consider the coefficient of the first term.

Lemma 4. *For a given $K < N$ and under the additive constraint $\sum_{k=1}^{K} n_k = N$, the quantity $\mathcal{C}(\mathcal{S}_N) = \sum_{k=1}^{K}(1/n_k)$ is minimized if and only if $\max_k n_k - \min_k n_k \leq 1$, that is, $n_k = N/K\ \forall x$ when K divides N, or $n_v = n$ for $K-r$ distinct n_v and $n_w = n+1$ for other r distinct n_w, where n and r are quotient and rest of the integer division of N by K, respectively. For the given K, the optimal value of the function $\mathcal{C}$ is therefore*

$$\mathcal{C}(\bar{\mathcal{S}}_N|K) = \frac{K-r}{n} + \frac{r}{n+1}.$$

Proof. In order to prove the statement, let us assume by way of contradiction, that the setting $\mathcal{S}_N = \{K, n_1, \ldots, n_K\}$ with $\max_k n_k - \min_k n_k > 1$ minimizes $\mathcal{C}(\mathcal{S}_N)$ while satisfying the additive constraint. Let further $M = \arg\max_k n_k$ and $m = \arg\min_k n_k$, we have thus $n_M - n_m - 1 > 0$. We can generate a setting $\mathcal{S}'_N = \{K, n'_1, \ldots, n'_K\}$ with $n'_m = n_m + 1$, $n'_M = n_M - 1$, and $n'_k = n_k$, otherwise. Clearly $\mathcal{S}_N$ satisfies the additive constraint. Moreover, it results:

$$\begin{aligned}\mathcal{C}(\mathcal{S}'_N) = \sum_{k=1}^{K} \frac{1}{n'_k} = \mathcal{C}(\mathcal{S}_N) - \left(\frac{1}{n_m} + \frac{1}{n_M}\right) + \left(\frac{1}{n'_m} + \frac{1}{n'_M}\right) = \\ = \mathcal{C}(\mathcal{S}_N) - \frac{n_m + n_M}{n_m n_M} + \frac{n'_m + n'_M}{n'_m n'_M} = \mathcal{C}(\mathcal{S}_N) - \frac{n_m + n_M}{n_m n_M} + \frac{n_m + 1 + n_M - 1}{(n_m + 1)(n_M - 1)} = \\ = \mathcal{C}(\mathcal{S}_N) - \frac{n_m + n_M}{n_m n_M} + \frac{n_m + n_M}{n_m n_M + n_M - n_m - 1}.\end{aligned}$$

Since the $n_M - n_m - 1 > 0$, the second fraction is smaller than the first and thus $\mathcal{C}(\mathcal{S}'_N) < \mathcal{C}(\mathcal{S}_N)$, which is a contradiction. To conclude the prof we need to observe that for a given K, all possible settings for which $\max_k n_k - \min_k n_k \leq 1$ is satisfied are just permutations of the same set $\{n_1, n_2, \ldots, n_K\}$, where the first r elements have value $n + 1$ and the other $K - r$ elements have value n. All such settings share therefore the same value of the function $\mathcal{C}$, which is clearly invariant under permutation of the addends, and thus all of them minimize it. The condition is therefore necessary and sufficient. The rest of the statement follows trivially. □

Lemma 5. *For any possible setting $\mathcal{S}_N = \{K, n_1, \ldots, n_K\}$ that satisfies the additive constraint $\sum_k n_k = N$, it results:*

$$\frac{1}{K^2} \sum_{k=1}^{K} \frac{1}{n_k} \geq \frac{1}{Kn},$$

where n is the result of the integer division of N by K.

Proof. From Lemma 4 it follows that for any generic setting $\mathcal{S}_N$,

$$\frac{1}{K^2} \sum_{k=1}^{K} \frac{1}{n_k} \geq \frac{1}{K^2} \left(\frac{K - r}{n} + \frac{r}{n + 1}\right) = \frac{1}{K} \left(\frac{K - r}{K} \frac{1}{n} + \frac{r}{K} \frac{1}{n + 1}\right).$$

The two addends in parenthesis in the last term represent a weighted average of $1/n$ and $1/(n+1)$, with weights $(K - r)/K$ and r/K, respectively. The value in parenthesis is therefore constrained to stay between $1/n$ and $1/(n + 1)$, thus:

$$\frac{1}{Kn} \leq \frac{1}{K} \left(\frac{K - r}{K} \frac{1}{n} + \frac{r}{K} \frac{1}{n + 1}\right) \leq \frac{1}{K(n + 1)}.$$

The statement follows. □

Theorem 4. *In any scenario $P(c, i) = P_C(c|i)P_I(i)$, for any given number of total runs N, the estimator $\tilde{\mu}_{\mathcal{S}_N}$ is not better than the estimator $\hat{\mu}_{\mathcal{H}_{N|1}}$ for what concerns the variance of the estimate.*

Proof. Let us recall that according to Corollary 1, the variance of $\hat{\mu}_{\mathcal{H}_{N|1}}$ is:

$$E\left[(\hat{\mu}_{\mathcal{H}_{N|1}} - \mu)^2\right] = \frac{1}{N} \bar{\sigma}^2_{\mathrm{WI}} + \frac{1}{N} \sigma^2_{\mathrm{AI}}.$$

According to Lemma 3, the variance of $\tilde{\mu}_{\mathcal{S}_N}$ is:

$$E\left[(\tilde{\mu}_{\mathcal{S}_N} - \mu)^2\right] = \frac{1}{K^2} \sum_{k=1}^{K} \frac{1}{n_k} \bar{\sigma}^2_{\mathrm{WI}} + \frac{1}{K} \sigma^2_{\mathrm{AI}}.$$

Let us compare the coefficients of σ^2_{WI} and σ^2_{AI} in the two equations. According to Lemma 5, $1/K^2 \sum_{k=1}^{K} 1/n_k \geq 1/Kn$, where n is the result of the integer division of N by K. Therefore $Kn \leq N$ and

$$\frac{1}{K^2} \sum_{k=1}^{K} \frac{1}{n_k} \geq \frac{1}{N}. \tag{5}$$

On the other hand, $K \leq N$ and therefore

$$\frac{1}{K} \geq \frac{1}{N}. \tag{6}$$

In both inequalities 5 and 6, the equal sign holds if and only if $\tilde{\mu}_{\mathcal{S}_N} = \hat{\mu}_{\mathcal{H}_{N|1}}$. □

6 Conclusions

The paper considers a class of *linear unbiased* estimators to empirically evaluate, on the basis of N runs the expected behavior of the given metaheuristic on a class of instances. In particular, we formally show that performing one single run on N different instances guarantees that the variance of the estimate is minimized. Any other experimental setting fails being efficient for what concerns the reduction of the variance. Moreover, we show that the total variance can be decomposed in two terms: the expected *within-instance* variance and the *across-instance* variance. A suboptimal experimental setting fails to act on the latter.

Contrary to popular belief, there is no trade-off involved in the definition of the experimental setting when the total number of runs is fixed. The setting "N instances, one run per instance" is shown to be *uniformly* the best across all possible scenarios, that is, irrespectively of the ratio between expected *within-instance* variance and *across-instance* variance.

Acknowledgments. This work was supported by the *Metaheuristics Network* and by the *ANTS* project. The *Metaheuristics Network* is a Research Training Network funded by the Improving Human Potential Programme of the Commission of the European Communities, grant HPRN-CT-1999-00106. The *ANTS* project is an *Action de Recherche Concertée* funded by the Scientific Research Directorate of the French Community of Belgium. The information provided is the sole responsibility of the author and does not reflect the opinion of the sponsors. The European Community and the French Community of Belgium are not responsible for any use that might be made of data appearing in this publication.

References

Balakrishnan, P. V., Gupta, R., & Jacob, V. S. (2004). Development of hybrid genetic algorithms for product line design. *IEEE Transactions on Systems, Man, and Cybernetics—Part B*, *34*(1):468–483.

Barr, R. S., Golden, B. L., Kelly, J. P., Resende, M. G. C., & Stewart, W. R. (1995). Designing and reporting computational experiments with heuristic methods. *Journal of Heuristics*, *1*(1):9–32.

Billingsley, P. (1986). *Probability and Measure*. Second edition. John Wiley & Sons. New York, NY, USA.

Birattari, M. (2004). *The Problem of Tuning Metaheuristics as Seen from a Machine Learning Perspective*. PhD thesis, Université Libre de Bruxelles, Brussels, Belgium.

Birattari, M., Stützle, T., Paquete, L., & Varrentrapp, K. (2002). A racing algorithm for configuring metaheuristics. In Langdon, W. B., Cantú-Paz, E., Mathias, K., Roy, R., Davis, D.,

Poli, R., Balakrishnan, K., Honavar, V., Rudolph, G., Wegener, J., Bull, L., Potter, M. A., Schultz, A. C., Miller, J. F., Burke, E., & Jonoska, N. (Eds.), *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 11–18, Morgan Kaufmann. San Francisco, CA, USA.

Blum, C. & Dorigo, M. (2004). The hyper-cube framework for ant colony optimization. *IEEE Transactions on Systems, Man, and Cybernetics—Part B*, *34*(2):1161–1172.

Blum, C. & Roli, A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys*, *35*(3):268–308.

Cerny, V. (1985). A thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm. *Journal of Optimization Theory and Applications*, *45*:41–51.

Dorigo, M. (1992). *Ottimizzazione, apprendimento automatico, ed algoritmi basati su metafora naturale*. PhD thesis, Politecnico di Milano, Milano, Italy. In Italian.

Dorigo, M., Di Caro, G., & Gambardella, L. M. (1999). Ant algorithms for distributed discrete optimization. *Artificial Life*, *5*(2):137–172.

Dorigo, M., Maniezzo, V., & A.Colorni (1996). Ant System: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics—Part B*, *26*(1):29–41.

Dorigo, M. & Stützle, T. (2004). *Ant Colony Optimization*. MIT Press. Cambridge, MA, USA.

Dréo, J., Pétrowski, A., Siarry, P., & Taillard, E. (2003). *Métahauristiques pour l'optimisation difficile*. Éditions Eyrolles. Paris, France. In French.

Fogel, L. J. (1962). Toward inductive inference automata. In *Proceedings of the International Federation for Information Processing Congress*, pp. 395–399. Munich, Germany.

Fogel, L. J., Owens, A. J., & Walsh, M. J. (1966). *Artificial Intelligence through Simulated Evolution*. John Wiley & Sons. New York, NY, USA.

Glover, F. (1977). Heuristics for integer programming using surrogate constraints. *Decision Sciences*, *8*:156–166.

Glover, F. (1986). Future paths for integer programming and links to artificial intelligence. *Computers & Operations Research*, *13*(5):533–549.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Publishing Company. Reading, MA, USA.

Guanqi, G. & Shouyi, Y. (2003). Evolutionary parallel local search for function optimization. *IEEE Transactions on Systems, Man, and Cybernetics—Part B*, *33*(6):864–876.

Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press. Ann Harbor, MI, USA.

Hooker, J. N. (1995). Testing heuristics: We have it all wrong. *Journal of Heuristics*, *1*(1):33–42.

Jin, H.-D., Leung, K.-S., Wong, M.-L., & Xu, Z.-B. (2003). An efficient self-organizing map designed by genetic algorithms for the traveling salesman problem. *IEEE Transactions on Systems, Man, and Cybernetics—Part B*, *33*(6):877–887.

Kato, K. & Sakawa, M. (2003). Genetic algorithms with decomposition procedures for multidimensional 0-1 knapsac problems with block angular structures. *IEEE Transactions on Systems, Man, and Cybernetics—Part B*, *33*(3):410–419.

Khuri, A. I. (2003). *Advanced Calculus with Applications in Statistics*. Second edition. John Wiley & Sons. Hoboken, NJ, USA.

Kirkpatrick, S., Gelatt Jr., C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, *220*:671–680.

Lourenço, H. R., Martin, O., & Stützle, T. (2002). Iterated local search. In Glover, F. & Kochenberger, G. (Eds.), *Handbook of Metaheuristics*, pp. 321–353. Kluwer Academic Publisher. Norwell, MA, USA.

Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes.* Third edition. McGraw-Hill. New York, NY, USA.

Pierre, S. & Houéto, F. (2002). Assigning cells to switches using taboo search. *IEEE Transactions on Systems, Man, and Cybernetics—Part B*, *32*(3):351–355.

Rardin, R. R. & Uzsoy, R. (2001). Experimental evaluation of heuristic optimization algorithms: A tutorial. *Journal of Heuristics*, *7*(2):261–304.

Rechenberg, I. (1973). *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Frommann-Holzboog. Stuttgart, Germany.

Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo Method*. John Wiley & Sons. New York, NY, USA.

Tsai, H.-K., Yang, J.-M., Tsai, Y.-F., & Kao, C.-Y. (2004). An evolutionary algorithm for large traveling salesman problems. *IEEE Transactions on Systems, Man, and Cybernetics—Part B*, *34*(4):1718–1742.

Wolpert, D. H. & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*(1):67–82.

Zlochin, M., Birattari, M., & Dorigo, M. (2004). Towards a theory of practice in metaheuristics design. A machine learning perspective. Technical Report MCS04-01, Computer Science and Applied Mathematics, The Weizmann Institute of Science, Rehovot, Israel. Submitted for journal publication.