On When to Stop Sampling for the Maximum

C. G. E. BOENDER and A. H. G. RINNOOY KAN
Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands

(Accepted: 16 October 1991)

Abstract. Suppose a sequential sample is taken from an unknown discrete probability distribution on an unknown range of integers, in an effort to sample its maximum. A crucial issue is an appropriate stopping rule determining when to terminate the sampling process. We approach this problem from a Bayesian perspective, and derive stopping rules that minimize loss functions which assign a loss to the sample size and to the deviation between the maximum in the sample and the true (unknown) maximum. We will show that our rules offer an extremely simple approximate solution to the well-known problem to terminate the Multistart method for continuous global optimization.

Key words. Bayesian stopping rules, Multistart.

1. Introduction

In this paper we describe a *Bayesian estimation procedure* for an *arbitrary discrete probability distribution* on $[y^-, y^+] \cap \mathbf{N}$ for which the *probabilities* $\Theta_{y^-}, \ldots, \Theta_{y^+}$ *and* the *minimum value* y^- and the *maximum value* y^+ are all unknown. In the Bayesian approach the user is asked to express his beliefs about the unknowns in the form of a *prior probability distribution*. The *sampling information* is then used to convert these initial beliefs into the *posterior distribution* of the unknowns through *Bayes's Theorem*.

A relevant extension of the above statistical framework is the development of *stopping rules* for sampling procedures which aim to detect the minimum value or maximum value at reasonable cost: Given the posterior distribution of the unknowns, a decision whether or not to stop sampling has to be taken which weighs the *cost* of further sampling against the *posterior expected revenues*.

In this paper we consider two *loss functions*, i.e., $c^E.n + c^T.[y^+ - y^+(n)]$ and $c^E.n + c^T.[y^+ - y^+(n)]/[y^+ - y^-]$, where n denotes the sample size, $y^+(n)$ is the maximum in the sample, and c^E and c^T are constants to be chosen by the user. Thus, the loss is determined by the sample size and by the deviation between the maximum in the sample and the true maximum. We show that, given *uniform prior distributions* for the unknowns, the *Bayesian one-step look ahead stopping rule* for the first loss function is to stop the sampling process when $[c^E/c^T].n.[n-3] \geq [y^+(n) - y^-(n) + 1]$, where $y^-(n)$ denotes the minimum in the sample. Under the same assumptions the *optimal rule* for the second loss function is to terminate sampling when $[c^E/c^T].n.[n+1] \geq 1$. Thus, for the second loss function the optimal sample size can be determined *prior to the sampling procedure*.

Such stopping rules apply, for example, to successful optimization methods for *global* and *combinatorial* optimization problems, where a *local search procedure* is started from a set of randomly sampled starting points. In the *Multistart* methods for global optimization, a local search procedure is started from every point in the sample or from an appropriate subset of the sample (Rinnooy Kan and Timmer 1987). Sampling in combination with local search techniques are also successfully applied to combinatorial problems (Lin 1965). In a proper implementation of these methods, the best observed value converges to the true optimum value with probability 1 when the sample size grows to infinity. However, since one is unable to continue the search forever, there is a need for stopping rules to determine the sample size which yields the optimal trade-off between *reliability* and *computational effort*. Since at each trial each local optimum has a fixed (unknown) probability of being sampled and since the true optimum value is unknown, our rules can be applied to these algorithms.

With respect to the application of Bayesian stopping rules to (stochastic) global optimization methods, we refer to Zielinski (1981), who originated the research in this field. However, the stopping rules in this reference are only based on the frequencies of occurrence of the sampled local optima: the rules prescribe to terminate a sequence of local searches on the basis of the posterior distribution of the number of unobserved local optima, *independently of the function values* of these optima, which may of course be worse than the ones which have already been observed (see also Boender and Rinnooy Kan 1987). To cope with this problem, Betro and Schoen (1987) introduced an alternative Bayesian framework, which processes the observed function values. Finally, in Piccioni and Ramponi (1991) a class of Bayesian stopping rules is described which involves both the frequencies of occurrence of the local optima, *and* their function values.

The rules in this paper are also based on the frequencies of occurrence of the local optima and their function values. However, we realize that, due to the integrality assumption, our rules only provide an *approximation* for terminating random sampling methods for continuous global optimization. However, opposite to the rules in Piccioni and Ramponi (1991), our rules are given by *extremely simple closed form* formulae, and from the numerical Section 5 it will appear that they work well, so that (in the opinion of the authors) they offer an appealing simple approximating alternative for the existing exact rules.

In Section 2 we describe the statistical model which is studied in this paper. Section 3 contains the prior and posterior distribution of the unknowns. This information is used in Section 4 to develop stopping rules for a sequential sample. Section 5 contains numerical experiments about the application of the rule to terminate the Multistart procedure on standard test problems for global optimization.

2. The Statistical Model

Our statistical model is an *arbitrary discrete* probability distribution

$$\Pr\{i\} = \Theta_i \qquad i \in [y^-, y^+] \cap \mathbf{N}, \tag{1}$$

where Θ_i is the probability of sampling the value i, and y^- and y^+ are the minimum and maximum possible value respectively. Prior to the sampling process the parameters Θ_i, y^- and y^+ are all unknown.

For a sample of n independent observations, define the random variables

N_i = Number of times that the value i occurs $(\Sigma N_i = n)$
$Y^-(n)$ = Minimum value in the sample
$Y^+(n)$ = Maximum value in the sample.

Then it is well known that the *sampling distribution* of the frequencies of occurrence N_i is *multinomial*, i.e.,

$$\Pr\{(N_{Y^-(n)}, \ldots, N_{Y^+(n)}) = (n_{y^-(n)}, \ldots, n_{y^+(n)})\} = n!\,.\prod_{i=y^-(n)}^{y^+(n)} \Theta_i^{n_i}/n_i!\,. \tag{2}$$

We note that the distribution (2) is actually conditional on the minimum value y^-, the maximum value y^+ and the probabilities $\Theta_{y^-}, \ldots, \Theta_{y^+}$, but in our notation these dependencies will henceforth be omitted.

3. The Prior and Posterior Distribution

In the Bayesian approach the unknowns y^-, y^+, $\Theta_{y^-}, \ldots, \Theta_{y^+}$ are themselves assumed to be random variables Y^-, Y^+, $\Theta_{y^-}, \ldots, \Theta_{y^+}$ for which a prior distribution is provided by the user. For a given and known minimum and maximum value, the standard Bayesian approach is to assume a *natural conjugate Dirichlet prior distribution* for the unknown probabilities. In our extension, the minimum and maximum value are also unknown parameters for which we assume a priori that the probabilities that $Y^- = m1$ and $Y^+ = m2$ $(m1 \leq m2)$ are equal to arbitrary constants τ_{m1} and τ_{m2}. Conditionally on $Y^- = m1$ and $Y^+ = m2$, the probabilities $\Theta_{m1}, \ldots, \Theta_{m2}$ are assumed to follow a Dirichlet distribution with parameters $\alpha_{m1}, \ldots, \alpha_{m2}$ $(\Sigma\alpha_i = \alpha)$. Note that this assumption implies that the prior expected value and variance of the unknown probabilities Θ_i are equal to α_i/α and $(\alpha_i/\alpha).[(\alpha_i + 1)/(\alpha + 1) - \alpha_i/\alpha]$ respectively (see, e.g., Wilks 1962). Thus, the prior expected value of Θ_i is proportional to α_i, and the corresponding prior variance decreases inversely with the total value of the α_i's.

Given the above sampling distribution, and the prior distributions for the unknown minimum and maximum and probabilities, it follows easily from *Bayes Theorem* that the *posterior distribution* of the unknowns is equal to:

$$\begin{aligned}
&\Pr\{Y^- = m1, Y^+ = m2, \Theta_{m1} \in d\phi_{m1}, \ldots, \Theta_{m2} \in d\phi_{m2} | \\
&(N_{Y^-(n)}, \ldots, N_{Y^+(n)}) = (n_{y^-(n)}, \ldots, n_{y^+(n)})\} = \\
&= \Pr\{(N_{Y^-(n)}, \ldots, N_{Y^+(n)}) = (n_{y^-(n)}, \ldots, n_{y^+(n)})\}\,. \\
&.\Pr\{Y^- = m1\}.\Pr\{Y^+ = m2\}.
\end{aligned}$$

$$\Pr\{\Theta_{m1} \in d\phi_{m1}, \ldots, \Theta_{m2} \in d\phi_{m2} | Y^- = m1, Y^+ = m2\}$$
$$\propto \tau_{m1} \cdot \tau_{m2} \cdot (\alpha - 1)! . \prod_{i=y^-(n)}^{y^+(n)} \phi_i^{n_i} . \prod_{j=m1}^{m2} \phi_j^{\alpha_j - 1}/(\alpha_j - 1)! , \tag{3}$$

where $\propto$ denotes proportionality. Note that the probability mass of the distribution 3 is concentrated on the domain $\{-\infty < m1 \leq m2 < \infty\}$. Secondly, given each pair $(m1, m2)$ with $m1 <= m2$, the probability distribution of the probabilities is a degenerate Dirichlet distribution on the unit simplex $\{\Theta_i >= 0, i = m1, \ldots, m2; \Theta_{m1} + \cdots + \Theta_{m2} = 1\}$. Note that in the right hand sides of (3), as well as henceforth, for reasons of abbreviation we omitted the terms $d\phi_i$.

As a special case we can choose the prior to be *non-informative*, i.e., for the minimum and maximum value each non-negative integer is regarded *equiprobable* ($\tau_{m1} = \tau_{m2} = 1$ for all $m1 <= m2$), and given $Y^- = m1$ and $Y^+ = m2$ the probabilities $\Theta_{m1}, \ldots, \Theta_{m2}$ are *uniformly* distributed on the unit simplex in $\mathbf{R}^{m2-m1+1}$ ($\alpha_i = 1$ for all i). Then, applying the equality,

$$\sum_{r=x}^{\infty} \frac{r!}{(n+r)!} = \frac{x!}{(n-1).(n+x-1)!} , \tag{4}$$

the following *posterior results* can be proved:

$$\Pr\{Y^- = m1, Y^+ = m2, \Theta_{m1} \in d\phi_{m1}, \ldots, \Theta_{m2} \in d\phi_{m2} |$$
$$(N_{Y^-(n)}, \ldots, N_{Y^+(n)}) = (n_{y^-(n)}, \ldots, n_{y^+(n)})\} =$$
$$= \frac{[(n-1).(n-2).(n+y^+(n)-y^-(n)-2)!.(m2-m1)!]}{[y^+(n)-y^-(n)]!} . \prod_{i=y^-(n)}^{y^+(n)} \phi_i^{n_i}/n_i! \tag{5}$$

The marginal posterior distribution of the maximum and minimum value:

$$\Pr\{Y^+ = m2 | (N_{Y^-(n)}, \ldots, N_{Y^+(n)}) = (n_{y^-(n)}, \ldots, n_{y^+(n)})\} =$$
$$\frac{[n-2].[n+y^+(n)-y^-(n)-2]!.[m2-y^-(n)]!}{[y^+(n)-y^-(n)]!.[n+m2-y^-(n)-1]!} , \tag{6}$$
$$\Pr\{Y^- = m1 | (N_{Y^-(n)}, \ldots, N_{Y^+(n)}) = (n_{y^-(n)}, \ldots, n_{y^+(n)})\} =$$
$$\frac{[n-2].[n+y^+(n)-y^-(n)-2]!.[y^+(n)-m1]!}{[y^+(n)-y^-(n)]!.[n+y^+(n)-m1-1]!} . \tag{7}$$

The posterior expectation of the maximum and minimum value:

$$E\{Y^+ | (N_{Y^-(n)}, \ldots, N_{Y^+(n)}) = (n_{y^-(n)}, \ldots, n_{y^+(n)})\}$$
$$= \frac{n-2}{n-3} \cdot [y^+(n) - y^-(n) + 1] + y^-(n) - 1 , \tag{8}$$
$$E\{Y^- | (N_{Y^-(n)}, \ldots, N_{Y^+(n)}) = (n_{y^-(n)}, \ldots, n_{y^+(n)})\}$$
$$= \frac{n-2}{n-3} \cdot [y^-(n) - y^+(n) - 1] + y^+(n) + 1 . \tag{9}$$

The posterior expectations of the probabilities of occurrence:

$$E\{\Theta_i|(N_{Y^-(n)}, \ldots, N_{Y^+(n)}) = (n_{y^-(n)}, \ldots, n_{y^+(n)})\}$$

$$= \frac{n-2}{n} \cdot \frac{n_i + 1}{n + y^+(n) - y^-(n) - 1} \quad \text{for } y^-(n) \leq i \leq y^+(n)$$

$$= \frac{n-2}{n} \cdot \frac{[i - y^-(n)]!.[n + y^+(n) - y^-(n) - 2]!}{[y^+(n) - y^-(n)]!.[n + i - y^-(n) - 1]!} \quad \text{for } i > y^+(n) \qquad (10)$$

$$= \frac{n-2}{n} \cdot \frac{[y^+(n) - i]!.[n + y^+(n) - y^-(n) - 2]!}{[y^+(n) - y^-(n)]!.[n + y^+(n) - i - 1]!} \quad \text{for } i < y^-(n) .$$

Given a quadratic loss function, the above expected values are *optimal Bayesian estimates* for the unknown minimum and maximum values y^- and y^+ and for the probabilities $\Theta_{y^-}, \ldots, \Theta_{y^+}$ (Lindley 1978). (We observe from (10) that the sum of the posterior expected probabilities is precisely equal to 1). These Bayesian estimators can be applied, for example, to the computation of an optimal (s, Q) *inventory policy* in the case that not only the probability distribution of lead time demand, but also the maximum lead time demand is unknown (Boender and Rinnooy Kan 1990).

4. Stopping Rules

In this section we use the posterior information to construct stopping rules for the search of the maximum value y^+. In *non-sequential* or *zero step look ahead* stopping rules the posterior information is used only to determine if the current best value $y^+(n)$ is satisfactory. For example, one may stop the sampling procedure when the posterior probability $[y^+(n) - y^-(n) + 1]/[n + y^+(n) - y^-(n) - 1]$, computed through (6), that there exists a better value than $y^+(n)$, is less than a prescribed value, or one may stop sampling when the round-off of the posterior expected optimum, given by (8), is equal to the sampled maximum.

Sequential stopping rules take into account the *posterior expected revenue* as well as the *cost* of further sampling. We assume that the sampling cost is equal to a fixed quantity c^E for each sample point. The revenue will first be assumed to be equal to a fixed quantity c^T per unit improvement of the deviation between the sampled maximum $y^+(n)$ and the true maximum y^+. Hence, the so-called *loss function* is equal to

$$L_n = c^T.[y^+ - y^+(n)] + c^E.n . \qquad (11)$$

In the Bayesian approach the unknown quantity of interest y^+ of (11) follows a posterior distribution, so that L_n is also a random variable, whose so-called *posterior loss* is equal to:

$$E\{L_n|(N_{Y^-(n)}, \ldots, N_{Y^+(n)}) = (n_{y^-(n)}, \ldots, n_{y^+(n)})\}$$

$$= c^T.[E\{Y^+|(N_{Y^-(n)}, \ldots, N_{Y^+(n)}) = (n_{y^-(n)}, \ldots, n_{y^+(n)})\} - y^+(n)] + c^E.n$$
$$= c^T.[y^+(n) - y^-(n) + 1]/[n-3] + c^E.n\,. \tag{12}$$

The *sequential* Bayesian analysis is based on the following observation. Given a sample of size n, the posterior loss of a sample of size $n+1$ is a random variable depending on the outcome of the additional sample point. The posterior probabilities of these outcomes are given by formula (10). Thus, the stochastic sequence of posterior losses satisfies the recurrence relation:

$$E\{E\{L_{n+1}|(N_{Y^-(n+1)}, \ldots, N_{Y^+(n+1)})\}|(N_{Y^-(n)}, \ldots, N_{Y^+(n)})$$
$$= (n_{y^-(n)}, \ldots, n_{y^+(n)})\}$$
$$= \sum_{i=y^+(n)+1}^{\infty} E\{\Theta_i|(N_{Y^-(n)}, \ldots, N_{Y^+(n)}) = (n_{y^-(n)}, \ldots, n_{y^+(n)})\}$$
$$.\, E\{L_{n+1}|(N_{Y^-(n+1)}, \ldots, N_{Y^+(n+1)}) = (n_{y^-(n)}, \ldots, n_{y^+(n)})\}$$
$$+ \sum_{i=y^-(n)}^{y^+(n)} E\{\Theta_i|(N_{Y^-(n)}, \ldots, N_{Y^+(n)}) = (n_{y^-(n)}, \ldots, n_{y^+(n)})\}$$
$$.\, E\{L_{n+1}|(N_{Y^-(n+1)}, \ldots, N_{Y^+(n+1)}) = (n_{y^-(n)}, \ldots, n_{y^+(n)})\}$$
$$+ \sum_{i=-\infty}^{y^-(n)-1} E\{\Theta_i|(N_{Y^-(n)}, \ldots, N_{Y^+(n)}) = (n_{y^-(n)}, \ldots, n_{y^+(n)})\}$$
$$.\, E\{L_{n+1}|(N_{Y^-(n+1)}, \ldots, N_{Y^+(n+1)}) = (n_i, \ldots, n_{y^+(n)})\}\,.$$

Substitution of (10) and (12) in (13) yields:

$$E\{E\{L_{n+1}|(N_{Y^-(n+1)}, \ldots, N_{Y^+(n+1)})\}|(N_{Y^-(n)}, \ldots, N_{Y^+(n)})$$
$$= (n_{y^-(n)}, \ldots, n_{Y^+(n)})\} = c^E.(n+1)$$
$$+ \sum_{i=y^+(n)+1}^{\infty} c^T.\frac{n-2}{n}.\frac{[n+y^+(n)-y^-(n)-1]!}{[y^+(n)-y^-(n)]!}$$
$$.\frac{[i-y^-(n)]!}{[n+i-y^-(n)-1]!}.\frac{i-y^-(n)+1}{n-2}$$
$$+ \sum_{i=y^-(n)}^{y^+(n)} c^T.\frac{n-2}{n}.\frac{n_i+1}{n+y^+(n)-y^-(n)-1}.\frac{y^+(n)-y^-(n)+1}{n-2}$$
$$+ \sum_{i=-\infty}^{y^-(n)-1} c^T.\frac{n-2}{n}.\frac{[n+y^+(n)-y^-(n)-1]!}{[y^+(n)-y^-(n)]!}$$
$$.\frac{[y^+(n)-i]!}{[n+y^+(n)-i-1]!}.\frac{y^+(n)-i+1}{n-2}$$
$$= c^E.(n+1) + c^T.[n-1].[y^+(n)-y^-(n)+1]/\{n.[n-3]\}\,. \tag{14}$$

The Bayesian *one step look ahead* stopping rule now prescribes to stop sampling when the expected posterior loss (14) is greater than the current posterior loss (12) (see, e.g., De Groot 1978), because in that case the posterior expected revenue of an additional sample point is clearly insufficient to compensate for the corresponding sampling cost c^E. In our model this turns out to be the case when

$$[c^E/c^T].n.[n-3] \geqslant [y^+(n) - y^-(n) + 1]\,. \tag{15}$$

Next we consider the loss function, initiated by a remark of prof. Törn at the IIASA conference on Global Optimization in Sopron, Hungary, 1990,

$$L_n = c^T.[y^+ - y^+(n)]/[y^+ - y^-] + c^E.n\,, \tag{16}$$

i.e., the user assigns a termination loss to the deviation between the sampled maximum $y^+(n)$ and the true maximum y^+, as a fraction of the range $[y^+ - y^-]$. Then the posterior loss can be shown to be equal to

$$E\{L_n|(N_{Y^-(n)}, \ldots, N_{Y^+(n)}) = (n_{y^-(n)}, \ldots, n_{y^+(n)})\} = c^T/n + c^E.n\,. \tag{17}$$

Thus, for this choice of the loss function the *optimal stopping rule* is to terminate when the sample size n satisfies

$$[c^E/c^T].n.[n+1] \geqslant 1\,. \tag{18}$$

Hence, for the loss function (16) the *optimal* sample size can be determined *prior to the sampling procedure*!

5. Numerical Results

In this section the rules (15) and (18) are applied to the Multi-start algorithm for global optimization of continuous multi-modal objective functions. Pure Multi-start proceeds by randomly sampling feasible points, usually according to the uniform distribution over the feasible region, and starting a local search procedure from each of the sampled points. Clearly, there is a need for a stopping rule which optimally weighs the potential revenues of further sampling of optima against the required additional computational effort.

We applied the stoping rules to a number of test functions for global optimization (Dixon and Szegö 1975). In Table I we show the values of the global maxima y^+, and the values of the lowest local maximum y^-, as well as the probabilities

Table I. Characteristics of the test functions for global optimization

	Global maximum & sampling probability		Lowest local maximum & sampling probability	
Goldstein & Price	$y^+ = -3.00$	57.1%	$y^- = -840.00$	6.0%
Shekel 5	$y^+ = +10.15$	32.0%	$y^- = +2.63$	17.9%
Shekel 7	$y^+ = +10.40$	30.1%	$y^- = +1.83$	2.5%
Shekel 10	$y^+ = +10.53$	22.7%	$y^- = +1.67$	2.6%

Table II. Average number of local searches and the probability that the global maximum is sampled in 25 experiments for three specifications of the loss function $L_n = c^T.[y^+ - y^+(n)] + c^E.n$ (11)

	Rule (15): $[c^E/c^T].n.[n-3] \geq [y^+(n) - y^-(n) + 1]$					
	$c^E/c^T = 0.1$		$c^E/c^T = 0.5$		$c^E/c^T = 1.0$	
Goldstein & Price	80.36	100%	27.00	100%	17.76	100%
Shekel 5	10.88	96%	5.88	88%	4.88	88%
Shekel 7	11.16	92%	5.84	76%	4.76	72%
Shekel 10	11.88	96%	6.36	76%	4.92	76%

that these local maxima are sampled. As local optimization procedure we used the simplex method of Nelder and Mead as programmed in Numerical Recipes (Press *et al.* 1986).

In Table II we show the results which are obtained by the one step look ahead rule (15) which is based on the loss function (11). For 25 sequencies of local searches and three different versions of (11) we show the average number of local searches and the percentage of times that the global maximum y^+ is sampled at the moment that (15) prescribes to stop. In Table III we show the results which are obtained by the optimal rule (18), based on four specifications of the loss function (16). Clearly for this rule the required number of local searches can be determined in advance, given that the loss function (16) is completely specified.

Note that for the loss function (11) c^E/c^T may be interpreted as the number of local searches that one is willing to do to improve the sample maximum $y^+(n)$ by one unit. For the loss function (16) $100 * c^E/c^T$ may be interpreted as the number of local searches that one is willing to do to improve the deviation of the sample maximum $y^+(n)$ from the true maximum y^+ by 1 percent of the range $[y^+ - y^-]$.

We observe from Table II that the number of local searches that is required for the Goldstein & Price test function is much larger than for the Shekel family, although the probability to sample the global maximum of the Goldstein & Price test function is larger than for the Shekel family. This phenomenon is evidently due to the fact that the Goldstein & Price test function has a much larger domain $[y^-, y^+]$ than the Shekel family. For this test function the worst local maximum, and the worst but one are respectively equal to -840.00 (6% of occurrence) and

Table III. Optimal number of local searches and the corresponding probability that the global maximum is sampled for four specifications of the loss function $L_n = c^T.[y^+ - y^+(n)] / [y^+ - y^-] + c^E.n$ (16)

	Rule (18): $[c^E/c^T].n.[n+1] \geq 1$							
	$c^E/c^T = 0.001$		$c^E/c^T = 0.01$		$c^E/c^T = 0.1$		$c^E/c^T = 0.5$	
Goldstein & Price	32.0	100%	10.0	100%	3.0	92%	1.0	57%
Shekel 5	32.0	100%	10.0	98%	3.0	69%	1.0	32%
Shekel 7	32.0	100%	10.0	97%	3.0	66%	1.0	30%
Shekel 10	32.0	100%	10.0	92%	3.0	54%	1.0	23%

Table IV. Required sample size n^* to detect the global maximum y^+ with probability 95%

	n^*
Goldstein & Price	3.5
Shekel 5	7.8
Shekel 7	8.4
Shekel 10	11.7

−84.00 (13% of occurrence). Thus, if one is sampling for the global maximum value −3.00 of the Goldstein & Price test function, and one is "unlucky" enough to decrease the sample minimum $y^-(n)$ from −84.00 to −840.00, then the stopping rule $[c^E/c^T].n.[n-3] \geqslant [y^+(n) - y^-(n) + 1]$ will obviously prescribe to continue sampling much longer.

We will not attempt to use the above numerical results to compare the quality of the rules (15) and (18) with the quality of other rules. First of all, such a comparison is clearly a three-criteria-decision-problem, since one is interested in minimizing the required number of local searches, in maximizing the probability that the global maximum is sampled, and in minimizing the computational effort to compute the decision whether to continue or stop. Furthermore, such a comparison is hampered by the fact that the different rules require different parameters to be set, like in our case the number of local searches that one is willing to do to gain improvement of the sample maximum.

Instead, we computed the number of local searches to obtain the global maximum with 95%, if the probability of sampling the global maximum y^+ (cf. Table I) would be known. Denote the probability of occurrence of the global maximum y^+ as p. Then it is easy to show that the required number of local searches is equal to $\ln(1.0-0.95)/\ln(1.0-p)$. We depicted these values for the different test functions in Table IV, and we let the readers conclude for themselves.

References

Betro, B. and F. Schoen (1987), Sequential Stopping Rules for the Multistart Algorithm in Global Optimization, *Mathematical Programming* **38**.

Boender, C. G. E. and A. H. G. Rinnooy Kan (1987), Bayesian Stopping Rules for Multi-start Global Optimization Methods, *Mathematical Programming* **36**.

Boender, C. G. E. and A. H. G. Rinnooy Kan (1990), A Bayesian Learning Procedure for the (s, Q) Inventory Policy, *Statistica Neerlandica* **44**, 3.

De Groot, M. H. (1978), *Optimal Statistical Decisions*, McGraw-Hill, New-York.

Dixon, L. C. W. and G. P. Szegö (1975), *Towards Global Optimisation*, North-Holland, Amsterdam.

Lin, S. (1965), Computer Solutions for the Travelling Salesman Problem, *Bell. Syst. Tech. J.* **44**.

Lindley, D. V. (1978), *Bayesian Statistics: A Review*, Society for Industrial and Applied Mathematics, Philadelphia.

Piccioni, M. and A. Ramponi (1991), Stopping Rules for the Multistart Method When Different Local Minima Have Different Function Values, to appear in *Optimization*.

Press, H. P., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1986), *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, New-York.

Rinnooy Kan, A. H. G. and G. T. Timmer (1987), Stochastic Global Optimization Methods. Part I: Clustering Methods, Part II; Multi Level Methods, *Mathematical Programming* **39**.

Wilks, S. S. (1962), *Mathematical Statistics*, Wiley, New-York.

Zielinski, R. (1981), A Statistical Estimate of the Structure of Multi-Extremal Problems, *Mathematical Programming* **21**.