Ivan Kramosil

On extremum-searching approximate probabilistic algorithms

Kybernetika, Vol. 19 (1983), No. 5, 365--373

Persistent URL: http://dml.cz/dmlcz/125786

KYBERNETIKA — VOLUME 19 (1983), NUMBER 5

ON EXTREMUM-SEARCHING APPROXIMATE PROBABILISTIC ALGORITHMS

IVAN KRAMOSIL

In this paper a simple probabilistic algorithm is proposed the aim of which is to find an approximation for the maximum value of a recursive function defined on a finite set. It is shown that if this function is "regular" or "continuous" in a sense, then its maximum value can be relatively approximated within a computational complexity which is substantially smaller than that of corresponding deterministic procedures. The basic idea of the proposed algorithm is very close to that of the well-known Monte-Carlo methods.

1. THE LIMITS OF POWERS OF CLASSICAL PROBABILISTIC ALGORITHMS

Instead of abstract theoretical considerations let us start this paper with a simple example the generality of which seems to be sufficient in order to cover a large class of computational and decision problems. Let f be a function defined on a finite set with n elements and taking values in the set $\mathcal{N} = \{0, 1, 2, \ldots\}$ of naturals, f is supposed to be recursive and its computational complexity is supposed to be uniformly majorized by a constant; this constant may be taken as the unit without a loss of generality. The nature of the set on which f is defined is irrelevant, so we may suppose that $Dom(f) = \{1, 2, \ldots, n\} = \mathcal{N}(n) \subset \mathcal{N}$.

Now, we are looking for the maximal value $\max(f, n) = \max\{f(i) : i \leqq n\}$ taken by f on its domain. Clearly, computational complexity of this problem, i.e. computational complexity of the function $\max(f, n)$ is majorized by n and, in general, no better upper bound can be given, as the necessity to compute and compare all the values $f(i)$, $i \leqq n$, cannot be avoided. For the sake of simplicity we do not take into consideration, here or in what follows, the computational complexity connected with the necessity to remember the temporary maximum value, to compare it with the new value and to actualize the temporary maximum if necessary.

Hence, the computational complexity of $\max(f, n)$ is a linear function of the

input size supposing that this size is defined by the cardinality of set on which f is defined. If the input size is measured by the number of bits, necessary to inscribe the input argument of f, then the input size of the function $\max(f, n)$ is majorized by the value $Int(\log_2(n) + 1)$, where $Int(a)$ denotes the integer part of a real a Then the computational complexity of the function $\max(f, n)$ will be an exponential function of the input size, as $n = 2^{\log_2 n}$. The same situation may occur, when searching for the maximum of a function defined on the system of all subsets of a finite set when the cardinality of this set serves as the input size. For example, let us mention the problem to find one or all combination(s) of truth-values which verifies (verify) or falsifies (falsify) given propositional formula the complexity of which is measured by the number of different propositional indeterminates occuring in this formula. As a rule, the exponential dependence between the input size and computational complexity is considered not as the acceptable one from the computer implementation point in view. This is why we shall study the possibilities how to reduce the computational complexity of $\max(f, n)$ below n even if we were forced to a resignation to the precision or reliability of the results obtained by a modified algorithm. Let α, β be two functions defined on $\mathcal{N} - \{0\}$ and taking it into itself, we write $\alpha \in O(\beta)$, if $0 < \lim_{n\to\infty} (\alpha(n))(\beta(n))^{-1} < \infty$, we write $\alpha \in o(\beta)$, if $\lim_{n\to\infty} (\alpha(n))(\beta(n))^{-1} = 0$. Hence, we would like to find such an "approximate" algorithm for $\max(f, n)$, the computational complexity of which belongs to $o(n)$ and, particularly, to $O(\log_2 n) \subset o(n)$.

What about the possibilities of probabilistic algorithms (p.a.) in this domain? They are rather limited when considering p.a. in their classical sense, i.e. measuring the quality of a p.a. by its probability of error, hence, by its probability to obtain no matter which other result but the correct one. Let $X_1, X_2, \ldots$ be independent random variables defined on a probability space $\langle \Omega, \mathcal{N}, P \rangle$, taking Ω into $\mathcal{N}(n)$, and such that

$$P(\{X_i(\omega) = j\}) = n^{-1}, \quad i = 1, 2, \ldots, j = 1, 2, \ldots, n . \tag{1}$$

(We write $P(\{X(\omega) \in B\})$ for $P(\{\omega : \omega \in \Omega,\ X(\omega) \in B\})$, in general, $B \subset \mathcal{N}$) Given $N \in \mathcal{N}$, we sample $X_1(\omega), X_2(\omega), \ldots, X_N(\omega)$ and compute

$$M(f, n, N, \omega) = \max\{f(X_i(\omega)) : i \leqq N\}, \quad \max \emptyset = 0 . \tag{2}$$

If there is just one $n_0 \leqq n$ such that $f(n_0) = \max(f, n)$, then we arrive at the error iff no of $X_i(\omega)$, $i \leqq N$, equals n_0, so the probability of error is $(1 - n^{-1})^N$, as it is nothing else than

$$P(\{M(f, n, N, \omega) \neq \max(f, n)\}) . \tag{3}$$

But

$$\lim_{n\to\infty} (1 - n^{-1})^n = e^{-1}, \quad e = 2{\cdot}718\ldots, \tag{4}$$

hence, given an $\varepsilon > 0$,

$$\lim_{n\to\infty} (1 - n^{-1})^{(\ln\varepsilon^{-1})n} = (\lim_{n\to\infty} (1 - n^{-1})^n)^{\ln(\varepsilon^{-1})} = (e^{-1})^{-\ln\varepsilon} = e^{\ln\varepsilon} = \varepsilon . \tag{5}$$

So the number of samples necessary in order to keep the probability of error below ε increases linearly with the computational complexity in the deterministic case and the ratio is even greater than one when $\varepsilon < e^{-1}$. Hence, the randomization does not reduce, in this case, the computational complexity and we have to look for another solution.

2. APPROXIMATE PROBABILISTIC ALGORITHMS

The probability of error admits an uncertainty as far as the correectness of the result is considered, but does not admit the acceptability of an unprecise or approximate result. In the terms of loss functions we can say that the contemporary probabilistic algorithms use the zero-one loss function, i.e., the loss suffered when taking the result of the probabilistic algorithm for the desired value, equals zero if p.a. gives the desired value and equals one otherwise, not taking into account which the difference between the desired and the obtained value may be. Then the expected loss is nothing else than the probability of error given the argument; the p.a. is usually taken as acceptable, if this probability of error may be kept under an a priori given threshold value uniformly for all values of the argument. Now, this demand will be weakened in the sense that we want only that the obtained value differ from the desired one relatively little with a high probability.

Definition 1. Let g be a function which takes $\mathcal{N}$ into $\mathcal{N}$. *Probabilistic algorithm* (in the most general sense) *which computes* g is a pair $\langle g, \mathscr{X}\rangle$, where $\mathscr{X} = \langle Y_1, Y_2, \ldots\rangle$, is a sequence of random variables defined on the probability space $\langle \Omega, \mathscr{S}, P\rangle$ and taking their values in $\mathcal{N}$. *Probability of error* connected with $\langle g, \mathscr{X}\rangle$ and corresponding to $n \in \mathcal{N}$ will be denoted by $pe(g, \mathscr{X}, n)$ and defined as

$$pe(g, \mathscr{X}, n) = P(\{Y_n(\omega) \neq g(n)\}), \tag{6}$$

if $g(n)$ is defined. The pair $\langle g, \mathscr{X}\rangle$ is called *ε-classical p.a. which computes g*, if $pe(g, \mathscr{X}, n) \leqq \varepsilon$ for all $n \in Dom(g)$. The pair $\langle g, \mathscr{X}\rangle$ is called *(ε, δ)-approximate p.a.* (a.p.a.) *wich computes g*, if for all $n \in Dom(g)$,

$$P(\{|Y_n(\omega)(g(n))^{-1} - 1| \geqq \delta\}) \leqq \varepsilon, \tag{7}$$

setting $0/0 = 1$, $a/0 = \pm\infty$ for $a \neq 0$. Clearly, each ε-classical p.a. is also an $(\varepsilon, 0)$-a.p.a. which computes the same function. □

Definition 2. Let $\langle g, \mathscr{X}\rangle$ be a p.a., let $n \in Dom(g)$. The *risk* $r(g, \mathscr{X}, n)$ connected with $\langle g, \mathscr{X}\rangle$ and corresponding to $n \in \mathcal{N}$ is defined by

$$r(g, \mathscr{X}, n) = E_P(|Y_n(\omega)(g(n))^{-1} - 1|) = \int_\Omega |Y_n(\omega)(g(n))^{-1} - 1| \, P(d\omega). \tag{8}$$

Theorem 1. Let $\langle g, \mathscr{X}\rangle$ be such a p.a., that $r(g, \mathscr{X}, n) \leqq \varepsilon$ for all $n \in Dom(g)$. Then $\langle g, \mathscr{X}\rangle$ is also a $\langle\sqrt{\varepsilon}, \sqrt{\varepsilon}\rangle$ – a.p.a.

Proof. Denote

$$A_n = \{\omega : \omega \in \Omega, \left|Y_n(\omega)\,(g(n))^{-1} - 1\right| \geqq \sqrt{\varepsilon}\}\,. \tag{9}$$

If $P(A_n) > \sqrt{\varepsilon}$, then

$$r(g, \mathscr{X}, n) = \int_{\Omega} \left|Y_n(\omega)\,(g(n))^{-1} - 1\right| \mathrm{d}P \geqq \int_{A_n} \left|Y_n(\omega)\,(g(n))^{-1} - 1\right| \mathrm{d}P \geqq \tag{10}$$

$$\geqq (\sqrt{\varepsilon})\,.\,P(A_n) > \varepsilon$$

and it is a contradiction, so $P(A_n) \leqq \sqrt{\varepsilon}$ and the assertion holds. □

Theorem 2. Let $\langle g, \mathscr{X}\rangle$ be such an $\langle\varepsilon, \delta\rangle$ – a.p.a., that $0 \leqq Y_n(\omega)\,(g(n))^{-1} \leqq M < \infty$ holds for all $n \in Dom(g)$, $\omega \in \Omega$. Then $r(g, \mathscr{X}, n) \leqq M\varepsilon + \delta$ for all $n \in Dom(g)$.

Proof. Denote

$$A_n(\delta) = \{\omega : \omega \in \Omega, \left|Y_n(\omega)\,(g(n))^{-1} - 1\right| \geqq \delta\}\,, \tag{11}$$

so $P(A_n(\delta)) \leqq \varepsilon$ and, for all $n \in Dom(g)$,

$$r(g, \mathscr{X}, n) = \int_{A_n(\delta)} \left|Y_n(\omega)\,(g(n))^{-1}\right| \mathrm{d}P + \tag{12}$$

$$+ \int_{\Omega - A_n(\delta)} \left|Y_n(\omega)\,(g(n))^{-1} - 1\right| \mathrm{d}P \leqq M\,.\,P(A_n(\delta)) + \delta(1 - P(A_n(\delta))) \leqq$$

$$\leqq M\varepsilon + \delta\,.$$ □

It would be clearly possible to strengthen the definition of a.p.a. replacing (7) by the demand $r(g, \mathscr{X}, n) \leqq \varepsilon$ for all $n \in Dom(g)$. Theorems 1 and 2 prove this modification not to be substantial when the ratio $Y_n(\omega) : g(n)$ is bounded; in the case of searching for the maximum value of a function the ratio $M(f, n, N, \omega) : \max(f, n)$ is always majorized by 1. The transformation from ε to $\langle\sqrt{\varepsilon}, \sqrt{\varepsilon}\rangle$ or from $\langle\varepsilon, \delta\rangle$ to $M\varepsilon + \delta$, as far as the values of parameters are concerned, is not substantial, as we always study, whether there exists, for *all* $\varepsilon < 0$, $\delta < 0$, a.p.a. or an a.p.a. with the corresponding risk, which solves the given problem within computational complexity substantially reduced (e.g., to a constant, linear or polynomial complexity) when compared with corresponding deterministic algorithms, if any.

3. APPROXIMATION OF EXTREMAL VALUES

Even when we make our demands more weakened and if we use an appropriate a.p.a. when searching for $\max(f, n)$, it is not possible, in general, to reduce the computational complexity of this a.p.a. (i.e., keeping in mind the simplifying assump-

tions mentioned in the beginning of this paper, the size of the necessary random sample) in such a way that it would belong to $o(n)$. Let $Dom(f) = \mathcal{N}(n) = \{1, 2, \ldots, n\}$ $f(j) = 0$, $j \leqq n$, $j \neq j_0 \leqq n$, $f(j_0) = 1$. So $\max(f, n) = 1$, but $M(f, n, N, .)$, i.e. Y_n in the sense of Definition 1, equals 0 if j_0 is not sampled by a random variable X_i, $i \leqq N$, which satisfies (1). Hence, for $\delta < 1$,

$$(13)\quad P(\{|M(f, n, N, \omega)(\max(f, n))^{-1} - 1| \geqq \delta)\} = P(\{M(f, n, N, \omega) = 0\}) =$$
$$= P(\{j_0 \in \mathcal{N} - \bigcup_{i=1}^{N} \{X_i(\omega)\}\}) = (1 - n^{-1})^N,$$

and it is the same expression as in the case of classical p.a. with the same negative consequences as far as the necessary speed of increasing of N is concerned. The following theorem shows simple sufficient conditions under which there exists a.p.a. which computes $\max(f, n)$ and such that $N(n) \in o(n)$.

Definition 3. Let $f : \mathcal{N}(n) \to \mathcal{N}$ be a function, let π_n be a permutation of $\langle 1, 2, \ldots, n\rangle$, denote by $\pi_n f$ the function taking $\mathcal{N}(n)$ into $\mathcal{N}$ which is defined, for all $i \leqq n$, by $(\pi_n f)(i) = f(\pi_n i)$. Let π_n^0 be the permutation of $\{1, 2, \ldots, n\}$, for which $(\pi_n^0 f)(i) \leqq$ $\leqq (\pi_n^0 f)(j)$ for all $i \leqq j \leqq n$; if there are more such permutations, i.e., if there are $i, j \leqq n$, $i \neq j$, such that $f(i) = f(j)$, then π_n^0 is fixed arbitrarily among those permutations. Set $\tilde{f} = \pi_n^0 f$.

Theorem 3. Let $\mathcal{F} = \{f_1, f_2, \ldots\}$ be a sequence of recursive functions such that each f_n takes $\mathcal{N}(n)$ into $\mathcal{N}$. Let there exist positive $K_1, K_2, K \geqq 1$, c, such that, for all $n \in \mathcal{N}$

$$(14)\quad K_1 n \leqq \tilde{f}_n(n),$$

$$(15)\quad \tilde{f}_n(n) - \tilde{f}_n(x) \leqq K_2(n - x)^K \quad \text{for all} \quad x \in \mathcal{N}(n), \quad (n - x) \leqq cn.$$

Then, for all $\varepsilon > 0$, $\delta > 0$, the function $N(\delta, \varepsilon) : \mathcal{N} \to \mathcal{N}$

$$(16)\quad N(\delta, \varepsilon)(n) = Int[(K_1 K_2^{-1}\delta)^{-1} \ln(2\varepsilon^{-1})]\, n^{1-1/K} + const,$$

which belongs to $o(n)$, possesses the property that $\langle \max(f_n, n), \{M(f_n, n, N(\delta, \varepsilon)(n), .)\}\rangle$ is an $\langle \varepsilon, \delta\}$ – a.p.a. which computes (estimates) the maximum value of f_n on $\mathcal{N}(n)$.

Proof. Let $\delta > 0$, $\varepsilon > 0$ be fixed, write f instead of $\tilde{f}_n$. As $f(n) = \tilde{f}_n(n) =$ $= \max\{f_n(i) : i \leqq n\}$, (15) yields $f(n) - f(n - j) \leqq K_2 j^K$, if $j \leqq cn$, hence, $f(n) -$ $- f(n - j) \leqq K_2 m^K$ for all $j \leqq m \leqq cn$. So, for $m \leqq cn$

$$(17)\quad 1 - f(n - j)(f(n))^{-1} \leqq K_2 m^K (f(n))^{-1} \leqq K_2 m^K (K_1 n)^{-1},$$

using (14). Set

$$(18)\quad m(n) = Int((K_1 K_2^{-1}\delta n)^{1/K}) \leqq (K_1 K_2^{-1}\delta n)^{1/K},$$

then $m^K \leqq K_1 K_2^{-1}\delta n$, so $K_2 m^K (K_1 n)^{-1} \leqq \delta$, hence, $1 - f(n - j)(f(n))^{-1} \leqq \delta$.

If $K < 1$, then $m(n) \in o(n)$, so there exists $n_1 = n_1(c) \in \mathcal{N}$ such that $n - m(n) \geqq$ $\geqq n(1-c)$ for $n \geqq n_1$. Let $K < 1$, $n < n_1$, denote by $\mathcal{U}(n, m)$ the set $\{n, n-1, \ldots$ $\ldots, n-m\}$ of naturals. The relation

$$|1 - M(f, n, N, \omega)(f(n))^{-1}| < \delta \tag{19}$$

is assured, for $K < 1$ and $n \geqq n_1$, if at least one among the random variables $X_1, X_2, \ldots$, satisfying (1), takes its value in $\mathcal{U}(n, m(n))$. The probability of the opposite event is, due to the supposed statistical independence and equiprobable distribution, given by

$$\begin{aligned} &P(\{\omega : \omega \in \Omega, |1 - M(f, n, N, \omega)(f(n))^{-1}| \geqq \delta\}) = \\ &= (1 - (card\, \mathcal{U}(n, m(n)))\, n^{-1})^N \leqq (1 - 1 + m(n))\, n^{-1})^N \leqq \\ &\leqq (1 - (K_3(\delta)^K \sqrt{(n)})\, n^{-1})^N = (1 - K_3(\delta)\, n^{(1-1/K)})^N , \end{aligned} \tag{20}$$

where $K_3(\delta) = (K_1 K_2^{-1} \delta)^{1/K}$. If $K = 1$ or $n < n_1$, then the same probability is majorized by either 0, if $\mathcal{U}(n, m(n)) = \mathcal{N}(n)$, or $(1-c)^N$, if $m(n) < cn$. So there exists $c_1 \in \mathcal{N}$ such that the probability of non-validity of (19) is smaller than ε for $n < n_1$, if $N \geqq c_1$, namely $c_1 \geqq (\log_2 (1-c))^{-1} \log_2 \varepsilon$ will do. The following assertion holds:

$$\begin{aligned} &\lim_{n \to \infty} ((1 - K_3(\delta)\, n^{-(1-1/K)})^{n^{1-1/K}})^{(K_3(\delta))^{-1} \ln(2\varepsilon^{-1})} = \\ &= (\lim_{n \to \infty} (1 - K_3(\delta)\, n^{-(1-1/K)})^{n^{1-1/K}})^{(K_3(\delta))^{-1} \ln(2\varepsilon^{-1})} = \\ &= (e^{-K_3(\delta)})^{(K_3(\delta))^{-1} \ln(2\varepsilon^{-1})} = e^{-\ln 2\varepsilon^{-1}} = e^{\ln(\varepsilon/2)} = \varepsilon/2 , \end{aligned} \tag{21}$$

as $(1 - K_3(\delta)\, n^{-(1-1/K)})^{n^{1-1/K}}$ is a subsequence of $(1 - K_3(\delta)\, x^{-1})^x$ with the same limit value $e^{-K_3(\delta)}$. Followingly, there exists $n_2 \in \mathcal{N}$ such that

$$P(\{\omega : \omega \in \Omega, |1 - M(f, n, N_1(n), \omega)(f(n))^{-1}| < \delta\}) < \varepsilon \tag{22}$$

holds, for all $n \geqq n_2$, with $N_1(n)$ being defined by

$$N_1(n) = N_1(\delta, \varepsilon)(n) = Int((K_3(\delta))^{-1} \ln 2\varepsilon^{-1})\, n^{1-1/K}) + 1 , \tag{23}$$

clearly $N_1(n) \in o(n)$, as $\lim_{n \to \infty} N_1(n)\, n^{-1} = const\, (\sqrt[K]{(n)})^{-1} = 0$. There exists $c_2 \in \mathcal{N}$ such that (22) holds for $n \leqq n_2$, if $N_1(n)$ is replaced by $N_1(n) + c_2$, clearly,

$$c_2 = Int(\max \{(\log_2 \varepsilon)(\log_2 (1 - m(n)\, n^{-1}))^{-1} : n \leqq n_2\}) + 1 \tag{24}$$

will do. Hence, $N(n) = N_1(n) + \max(c_1, c_2)$ assures, that $\langle \max(f, n), \{M(f, n, N(n), .)\}\rangle$ is an $\langle \varepsilon, \delta \rangle$ – a.p.a. for $\max(f, n)$ on $\mathcal{N}(n)$. However, the same rests valid if $f = \tilde{f}_n$ is replaced by the original function f_n, as the proved property of our a.p.a. depends only on the number of argument values, for which the corresponding function value approximates $\max(f_n, n)$ in the sense of the adopted criterion. This number is $card\, \mathcal{U}(n, m(n)) = m(n) + 1$ for all $\pi_n \tilde{f}$, particularly for f_n itself. Hence, (22) holds for f_n as well with $N(n)$ defined by (16) and with $const = \max(c_1, c_2)$. □

The complexity reduction following when $N(n) \in O(n^{1-1/K}) \subset o(n)$ is not sufficient, in general to avoid the exponential complexity if the input size is measured by $m = \log_2 n$. Or, then $n = 2^m$, $n^{1-1/K} = (2^m)^{1-1/K} = c^m$, $1 < c < 2$, and this is, again, an exponential function of m. We would like to make the conditions stronger to assure the validity of the assertion of Theorem 3 even for an appropriate $N(n) \in$ $\in O(\log_2 n)$, e.g., for $N(n) = c_1 \log n + c_2$.

Theorem 4. Let the notations and conditions of Theorrm 3 hold with (15) replaced by

$$\tilde{f}_n(n) - \tilde{f}_n(x) \leqq K_2(n-x)\log_2^*(n-x), \quad x \in \mathcal{N}(n), \quad (n-x) \leqq cn, \tag{25}$$

where $\log_2^* x = \max(\log_2 x, 1)$. Then, for all $\varepsilon > 0$, $\delta > 0$, the function $N(\delta, \varepsilon)$: $\mathcal{N} \to \mathcal{N}$,

$$N(\delta, \varepsilon)(n) = Int[K_2 K_1^{-1} \ln(2\varepsilon^{-1})\delta^{-1}] \log_2 n + const, \tag{26}$$

which belongs to $O(\log_2 n)$, satisfies the assertion of Theorem 3.

Proof. The proof is similar to that of Theorem 3, it is why we introduce here just some of its key points. (26) yields that $1 - f(n-j)(f(n))^{-1} \leqq K_2 m \log_2^* m K_1^{-1} n^{-1}$. Take

$$m(n) = Int(n(\log_2^* n)^{-1} \delta K_1 K_2^{-1}), \tag{27}$$

then, for $n \geqq 2$,

$$K_2\, m(n) \log_2 m(n) K_1^{-1} n^{-1} \leqq$$

$$\leqq \delta + \delta(\log_2 n)^{-1}(\log_2(\delta K_1 K_2^{-1}) - \log_2 \log_2 n) \leqq \delta$$

for $n < 2^{\delta K_1 K_2^{-1}}$, but this may be always assured taking K_2 large enough. Hence,

$$|1 - M(f, n, N, \omega)(f(n))^{-1}| \leqq \delta \tag{28}$$

holds, if at least one value among $X_1(\omega), \ldots, X_N(\omega)$ does not differ from n by more than $m(n)$. The probability that (28) does not holds is majorized by

$$(1 - (m(n)+1)n^{-1})^N \leqq (1 - (\delta K_1 K_2^{-1})(\log_2 n)^{-1})^{(\ln 2\varepsilon^{-1})\delta^{-1} K_2 K_1^{-1} \log_2 n},$$

and the limit value of the last expression can be proved to be $\varepsilon/2$ using the same argumentation as in the proof of Theorem 3. Hence, the probability that (28) does not hold is smaller than ε for n large enough; enlarging $N(n)$ by an appropriate additive constant we extend the validity of the last statement to all $n \in \mathcal{N}$. Hence, there exists $c \in \mathcal{N}$ such that, for

$$N(n) = Int[\ln(2\varepsilon^{-1})\delta^{-1} K_2 K_1^{-1}] \log_2 n + c \in O(\log_2 n), \tag{29}$$

$\langle \max(f, n), \{M(f, n, N(n), .)\}\rangle$ is an a.p.a. which estimates $\max(f, n)$. □

4. CONCLUSIVE REMARKS

Let us close this paper by four remarks which comment the preceding considerations and results.

(1) Theorem 3 holds also in the case $K = 1$. Then it is sufficient that $N(n) \in$ $\in O(n^{1-1/K}) = O(1)$, i.e., that $M(n)$ be a large enough constant. Considerations analogous to those in the proofs of Theorems 3 and 4 give

$$N = N(\varepsilon, \delta) \geqq (\log_2 \varepsilon)\,(\log_2 (1 - \delta K_1 K_2^{-1}))^{-1}\,. \tag{30}$$

Verbally said, a finite number of samples suffice, for all n and independently of n, to discover with the precision of 100δ %, and probability $1 - \varepsilon$, which is the maximum value of the function f for arguments smaller than n and under the condition that the variation of f lies in an angle limited by two lines (intersecting themselves in the beginning of coordinates, but this is not substantial).

(2) When considering some other ways how the values taken by a random variable Y_n may approximate the desired value $g(n)$, we may try to replace the demand $|Y_n(\omega)\,(g(n))^{-1} - 1| \leqq \delta$, $0 \leqq \delta \leqq 1$, by the demand $|Y_n(\omega) - g(n)| \leqq \delta$, $0 \leqq \delta < \infty$ Under such a more strict condition imposed to a.p.a. Theorems 3 and 4 can be proved to fail. E.g., even if (15) holds, we must take $m(n) = m(\delta, \varepsilon)$ independently of n and the corresponding probability of error would be $(1 - m(\delta, \varepsilon)\, n^{-1})^N$, hence, N would increase as a linear function of n in order to assure the possibility to majorize this probability by an $\varepsilon > 0$. This is why a weaker and practically justifiable demand of a relatively good approximation of $g(n)$, by $Y_n(\omega)$ has been applied throughout this paper.

(3) In this paper we always suppose, that the limitations to the precision of the obtained results (given by δ) and to their reliability (given by ε) must hold uniformly for all possible values of arguments. Instead of this assumption we may suppose to have at our disposal an a priori probability distribution on the set of argument values (e.g., the equiprobable one in the finite case) and to classify the quality of p.a. or a.p.a. with respect to the expected probability of error or expected risk under this a priori distribution. The weak point of this approach consists in the necessity to justify why such and such a priori distribution gas been used, by reasons staying outside the mathematical theory itself. This situation is difficult namely if we cannot apply the equiprobable distribution and the Laplace principle according to which the absence of any reason supporting other than equiprobable distribution may be taken as a sufficient reason for this equiprobable distribution. This is a more general problem of the so called Bayes statistics, in [3] this approach is studied in more details as far as p.a. which test propositional formulas are concerned.

(4) The similar forms of Theorems 3 and 4 lead to the idea that they both are particular cases of a more general assertion proclaiming some dependences between lower and upper bound for the variation of the function f, and the number $N(n)$

of samples. Searching for such an assertion will be a matter of further investigation, however, we believe that even the particular assertions presented here are of certain illustrative importance as far as the possibilities and limits of powers of p.a. are concerned.

The approach to the problem of approximations of extremal values as presented here is far from being the only possible, however, it is qualitatively different from the other ones. The deterministic methods are based, as a rule, on a global knowledge of the investigated function (e.g., when the gradient method is to be used, we suppose that the function in question has no local extremum, that it is differentiable, etc.) In the approach presented here, on the other hand, the only needed is that an algorithm produces the value of the function given the value of its argument. From the other side, the stochastic approximation methods suppose also the investigated function itself to be of stochastic character and they aim to estimate the extremum of its expected value. The results are of asymptotic character and are conditioned by rather complicated assumptions of probabilistic nature. Moreover, it is rather difficult to define and estimate, in this case, the computational complexity of the statistical decision procedure in a way reasonable enough and comparable with the complexity measures used in this paper. Hence, the approach presented here seems to be an acceptable outcome in the situation when our possibilities to handle with the tested function is limited to the algorithm computing the corresponding values and the results presented above demonstrate our possibilities and limits of our powers in such a situation.

In spite of the self-explanatory character of this paper we refer below the monography [1] which may serve as an informal introduction into algorithms and their complexities from the implementational point of view; [2] presents the basic ideas of probabilistic algorithms.

(Received November 26, 1982.)

REFERENCES

[1] A. V. Aho, J. E. Hopcroft and J. D. Ullman: The Design and Analysis of Computer Algorithms. Reading, Addison-Wesley, London 1974 (Russian translation: Mir, Moscow 1978).

[2] M. O. Rabin: Probabilistic algorithms. In: Proceedings of the Symposium 1976, Carnegie-Mellon Univ., Pittsburgh, Academic Press, New York 1976, pp. 21—39.

[3] I. Kramosil: Statistical verification procedures for propositional calculus. Computers and Artificial Intelligence *2* (1983), 3, 235—258.

RNDr. Ivan Kramosil, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.