Locating the Maximum of a Simple Random Sequence by Sequential Search

BRUCE HAJEK, SENIOR MEMBER, IEEE

Abstract —**Consider a stationary Gaussian process with $EX_iX_j = a^{|i-j|}$ where $0 < a < 1$, and let $0 < r < 1$. It is shown that to locate the maximum of $X_1, X_2, \cdots, X_N$ for large N with probability r, roughly $-rN \log a / \log\log N$ observations at sequentially determined locations are both sufficient and necessary.**

I. Introduction

Let $(X_i:\ i \in Z)$ denote a stationary Gaussian random process with mean zero and $EX_iX_j = a^{|i-j|}$ where a is a constant with $0 < a < 1$. Let M and N be integers, with $1 \le M \le N$. Consider a sequential strategy U for attempting to find the maximum of $(X_i:\ 1 \le i \le N)$. Assume that $U = (U_1, U_2, \cdots, U_M)$ where each U_i takes values in $\{1, 2, \cdots, N\}$ and that U_i is a function of $(X_{U_1}, \cdots, X_{U_{i-1}})$ for each i. Define X^* to be the maximum of $X_1, \cdots, X_N$, and let i^* be the random place in $\{1, \cdots, N\}$ for which $X^* = X_{i^*}$. Define S to be the event that $i^* \in \{U_1, \cdots, U_M\}$.

We are interested in choosing U to maximize the success probability $P[S]$. One possible strategy is to choose $U_1, \cdots, U_M$ to be fixed distinct elements of $\{1, 2, \cdots, N\}$. If $a = 0$ so that the X_i are independent standard Gaussian random variables, then $P[S] = M/N$, and clearly no strategy can yield a larger value of $P[S]$. Henceforth we assume that $0 < a < 1$.

Fix r with $0 < r < 1$, and define ϵ_N by

$$\epsilon_N = \frac{-\log a}{\log\log N}, \qquad N \ge 1.$$

Let $U_{N,r}$ denote the set of strategies for finding the maximum of $(X_1, \cdots, X_N)$ using $m = \lfloor Nr\epsilon_N \rfloor$ observations. The purpose of this paper is to prove the following theorem, which is proved in the next two sections.

Theorem: The following holds:

$$\lim_{N \to \infty} \max_{U \in U_{N,r}} P[S] = r.$$

Manuscript received April 18, 1986; revised August 15, 1986. This work was supported in part by the Army Research Office under Grant DAAG29-84-K-0005 and the Office of Naval Research under Grant US Navy N00014-82-K-0359.

The author was with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA. He is now with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory of the University of Illinois at Champaign-Urbana, 1101 W. Springfield, Urbana, IL 61801, USA.

IEEE Log Number 8716446.

The theorem shows that a successful search requires looking at roughly a fraction ϵ_N of the places. Unless a is close to one, N must be extremely large for ϵ_N to be close to zero. For example, if $a = 0.5$, then N must exceed 10^{444} for ϵ_N to drop below one-tenth.

It would be interesting to see how many observations are necessary to find an i' such that with probability r, $X_{i'}$ is within a specified amount of the maximum. The techniques used in this paper may help.

We examined the foregoing random process for its simplicity, not because it arises in any application. However, results similar to the theorem (perhaps less sharp) probably can be proved for more general collections of random variables. In fact, our proof is based almost entirely on the theory of extremes of random processes, and that theory now covers many non-Gaussian stationary random processes, continuous-time random processes, and quite general collections of Gaussian random variables [1]. Generalizations of the Theorem may have implications for the problem of maximizing an objective function which is costly to compute (so it is worthwhile to ponder about where the function should be evaluated) and which has many local maxima. Such a problem arises when determining multilens geometry in optical design, placing struts in structural design, placing sensors in imaging systems, locating drilling sites in geological exploration, etc. If the objective function can reasonably be modeled as a sample path from a specific random process, it might be possible to evaluate the likely performance of various strategies or to give absolute performance limitations.

We close this section by discussing a continuous-parameter analog of the theorem. Let $(X_t:\ t \in R)$ be a sample continuous Gaussian random process with $EX_sX_t = \exp(-\alpha|s-t|)$. For fixed $f > 0$, the theorem applied to the discrete time process $(X_{nf}:\ n \in Z)$ yields the following result. For large T, to find the maximum of $(X_t:\ t \in [0, T] \cap \{nf:\ n \in Z\})$ with success probability r, roughly $\alpha rT/\log\log T$ observations are both sufficient and necessary. Since this number does not depend on f, the following *conjecture* is plausible. Fix $d > 0$, and let T be large. To estimate the location t^* of the maximum of $(X_t:\ 0 \le t \le T)$ to within distance d with success probability r, roughly $\alpha rT/\log\log T$ observations are both sufficient and necessary.

II. The Positive Half

We begin the proof of the theorem by showing that a simple two-phase strategy can yield success probabilities as close to r as we would like, for N sufficiently large. Fix δ with $0 < \delta \leq 1/9$. Let

$$K = \left\lfloor \frac{(1-4\delta)\log\log N}{-2\log a} \right\rfloor,$$

$$F = \{h(2K+1) - K : 1 \leq h \leq |F|\}$$

where $|F| = \lfloor N r \epsilon_N (1-\delta) \rfloor$,

$$G_i = \{j : 1 \leq |i-j| \leq K\}, \qquad \text{for } i \in F,$$

and

$$G = \bigcup_{i \in F} G_i.$$

Then $G \cup F = \{j : 1 \leq j \leq (2K+1)|F|\}$. Since

$$\lim_{N \to \infty} (2K+1)|F|/N = r(1-4\delta)(1-\delta),$$

we can suppose that N is so large that

$$N(r-5\delta) \leq (2K+1)|F| \leq N.$$

The first phase of the strategy consists of taking $\lfloor N r \epsilon_N (1-\delta) \rfloor$ observations precisely at the places in F. Let $R(i)$ denote the place in F where the ith largest observation was made in this first phase, for $1 \leq i \leq |F|$. Begin the second phase by taking observations at all the places in $G_{R(1)}$. Next take observations at all the places in $G_{R(2)}$, and so forth, until a total of $\lfloor N r \epsilon_N \rfloor$ observations is taken for the two phases combined. Clearly, this two-phase strategy is in $U_{N,r}$.

Lemma 1: For δ fixed and all N sufficiently large, the two-phase strategy yields $P[S] \geq r - 9\delta$.

Proof: Let $\gamma = (\log N)^{\delta}$, and suppose Γ_N is to be specified. Observe that the event that the algorithm *fails* is contained in the union of W_1, W_2, W_3, and W_4, where

$$W_1 = \{i^* \notin F \cup G\},$$

$$W_2 = \left\{ \begin{array}{l} X_i \geq \gamma \text{ and } X_j \text{ is } \textit{not} \text{ observed, for} \\ \text{some } (i,j) \text{ with } i \in F,\ j \in G_i \end{array} \right\},$$

$$W_3 = \{X^* < \Gamma_N\},$$

$$W_4 = \left\{ \begin{array}{l} X_j \geq \Gamma_N \text{ and } X_i < \gamma, \text{ for some } (i,j) \\ \text{with } i \in F,\ j \in G_i \end{array} \right\}.$$

Since i^* is asymptotically uniformly distributed over $\{1, \cdots, N\}$ [1, ch. 5],

$$\begin{aligned} P[W_1] &= P[i^* > (2K+1)|F|] \\ &\leq P[i^* > N(r-5\delta)], \qquad \text{for } N \text{ large} \\ &\leq 1 - r + 6\delta, \qquad \text{for } N \text{ large.} \end{aligned} \tag{2.1}$$

Let V denote the number of i in F such that $X_i \geq \gamma$. Then

$$\begin{aligned} P[W_2] &\leq P[V \geq N r \epsilon_N \delta / 2K] \\ &\leq 2K E(V)/N r \epsilon_N \delta \leq \delta, \qquad \text{for } N \text{ large} \end{aligned} \tag{2.2}$$

where we use the fact that (using $Q(\gamma) = P[X_i \geq \gamma]$)

$$E(V) = |F| Q(\gamma) \leq N \exp\left(-\frac{1}{2}\gamma^2\right) \Big/ \gamma.$$

Next, we choose Γ_N so that

$$P[W_3] = P[X^* \leq \Gamma_N] = \delta. \tag{2.3}$$

Finally, if $i \in F$, $j \in G_i$ and N is so large that $\gamma - a^K \Gamma_N < 0$ (see the following), then

$$\begin{aligned} P[X_j \geq \Gamma_N \text{ and } X_i < \gamma] &= \int_{\Gamma_N}^{\infty} Q\left(\frac{a^{|i-j|} u - \gamma}{\sqrt{1 - a^{2|i-j|}}} \right) \phi(u)\, du \\ &\leq \int_{\Gamma_N}^{\infty} Q(a^K \Gamma_N - \gamma) \phi(u)\, du \\ &= Q(\Gamma_N) Q(a^K \Gamma_N - \gamma). \end{aligned}$$

Hence since at most N such values of (i,j) exist,

$$P[W_4] \leq N Q(\Gamma_N) Q(a^K \Gamma_N - \gamma).$$

By [1, part (i), theorem 4.3.3], $NQ(\Gamma_N)$ converges to $-\ln\delta$ as N tends to infinity. Also, $\Gamma_N/(2\log N)^{1/2}$ tends to one so that

$$\begin{aligned} a^K \Gamma_N - \gamma &\sim (\log N)^{-(1/2)+2\delta} (2\log N)^{1/2} - (\log N)^{\delta} \\ &\to \infty, \qquad \text{as } N \to \infty. \end{aligned}$$

Thus

$$P[W_4] \leq \delta, \qquad \text{for } N \text{ large.} \tag{2.4}$$

Combining (2.1)–(2.4), we see that the probability the two-phase algorithm fails is at most $1 - r + 9\delta$ for large enough N. Lemma 1 is thus proved.

III. The Negative Half

The proof of the Theorem will be completed in this section, and the following three facts will play a key role: a) X^* is distributed roughly like the maximum of N independent standard Gaussian variables (see (3.2) to follow), b) X^* is no larger than the maximum of those variables observed and those variables at locations not close to those observed (i.e., at locations in G to follow), and c) given that the observed values are not unusually large ($\Theta \in A$ to follow), the conditional distribution of the roughly $N(1-r)$ variables not close to those observed is nearly the same as the maximum of $N(1-r)$ standard Gaussian variables (Lemma 3 to follow). Using these facts and integration by parts, we can show that the maximum of the observed variables is less than the maximum of the variables not close to those observed, with probability nearly $1-r$. That will complete the proof. Let us begin.

Fix δ with $0 < \delta < 1$. At several places below we require N to be sufficiently large while a, r, and δ remain fixed. Let U be an arbitrary strategy in $U_{N,r}$.

Let Θ denote the information gained after U has been applied. That is,

$$\Theta = \{O, (X_i : i \in O)\}$$

where O is the random set of places at which observations were made. Let

$$\gamma = (\log N)^{\delta}$$

$$L = \left[\frac{(1+4\delta)\log\log N}{-2\log a} \right],$$

and define a subset A of the set of possible outcomes of Θ by

$$A = \left\{ \theta:\ \max_{i \in O} |X_i| \le (\log N)^{(1/2)+\delta}, \right.$$
$$\left. |\{ i \in O:\ |X_i| \ge \gamma \}| \le \frac{\delta N}{4L} \right\}.$$

Since

$$P\left[\max_{1 \le i \le N} |X_i| \ge (\log N)^{(1/2)+\delta} \right] \underset{N\to\infty}{\to} 0$$

and

$$P\left[|\{ i:\ 1 \le i \le N, |X_i| \ge \gamma \}| \ge \frac{\delta N}{4L} \right] \le \frac{4LQ(\gamma)}{N\delta} \underset{N\to\infty}{\to} 0,$$

we have that

$$P[\Theta \in A] \ge 1-\delta, \qquad \text{for } N \text{ large.}$$

Define a random set G by

$$G = \{ j:\ 1 \le j \le N, |j-i| \ge L, \qquad \text{for all } i \in O \text{ and}$$
$$|j-i| \ge 2L \text{ for all } i \in O \text{ with } |X_i| \ge \gamma \}.$$

Note that G is determined by Θ, and if $\Theta \in A$, then

$$|G| \ge N - (2L-1)Nr\epsilon_N - 4L(\delta N/4L)$$
$$\ge N(1 - r(1+4\delta+\epsilon_N) - \delta) \ge N(1-r-6\delta) \quad (3.1)$$

where we assume that N is so large that $\delta \ge r\epsilon_N$.

Given a possible outcome θ, we let P_θ denote the induced conditional distribution of $(X_i:\ i \in G)$. This conditional distribution is Gaussian. We let

$$(m_i:\ i \in G) \qquad (\sigma_i^2 = S_{ii}:\ i \in G) \qquad (S_{ij}:\ i, j \in G)$$

denote the means, variances, and covariances of $(X_i:\ i \in G)$ under the distribution P_θ. Note that G, the m_i, σ_i^2, and S_{ij} are all determined by θ.

Lemma 2: If $\theta \in A$, then for all i, j in G,

$$|m_i| \le m$$

where

$$m \triangleq (\log N)^{-(1/2)-\delta}$$

$$0 \le S_{ij} \le a^{|i-j|}$$

$$\sigma_i^2 \in [\sigma^2, 1]$$

where

$$\sigma^{-2} \triangleq 1 - 2(\log n)^{-(1+\delta)}$$

Proof: The Markov property of X easily implies the following Markov random field or "reciprocal" property. If $F = \{ j:\ a \le j \le b \}$ for integers a and b, then $(X_j:\ j \in F)$ and $(X_j:\ j \in F^c)$ are conditionally independent given (X_a, X_b). Thus m_i depends on only $(i-h, k-i, X_h, X_k)$ where h and k are the elements in $O \cup \{+\infty, -\infty\}$ lying to the left and right of i and as close as possible to i. Similarly, S_{ij} depends only on O (and not on $(X_i:\ i \in O)$) and $S_{ij} = 0$ if $i < h < j$ for some h in 0. The properties claimed in the lemma are then easy to establish using the following facts, which themselves are easy to establish by explicit calculation: For integers h, i, j, k with $h \le i \le j \le k$,

$$E[X_i | X_h, X_k] = m(i:\ h,k)X_h + m(i:\ k,h)X_k$$

where

$$0 \le m(i:\ h,k) \le a^{|h-i|}$$

$$0 \le \operatorname{cov}(X_i, X_j | X_h, X_k) \le a^{|i-j|}$$

$$1 - a^{|h-i|} - a^{|k-i|} \le \operatorname{var}(X_i | X_h, X_k) \le 1. \qquad \square$$

Let $a_N = (2\log N)^{1/2}$ and

$$b_N = (2\log N)^{1/2} - \frac{1}{2}(2\log N)^{-(1/2)}(\log\log N + \log 4\pi).$$

Define

$$\tilde{X} = a_N\left(\max_{1 \le i \le N} X_i - b_N \right)$$

$$\tilde{Y} = a_N\left(\max_{i \in O} X_i - b_N \right)$$

$$\tilde{Z} = a_N\left(\max_{i \in G} X_i - b_N \right).$$

It is known [1] that

$$P\{ \tilde{X} \le x + O(1) \} \to e^{-e^{-x}}, \qquad \text{as } N \to \infty. \quad (3.2)$$

Lemma 3: If N is sufficiently large and $\theta \in A$

$$P_\theta[\tilde{Z} \le x] \le \exp(-(1-r-6\delta)e^{-x}) + \delta$$

for all x.

Proof: It suffices to prove that

$$P_\theta[\tilde{Z} \le x] \le \exp(-(1-r-6\delta)e^{-x}) + \frac{\delta}{2}$$

for x in a large finite interval $[x_{\min}, x_{\max}]$ not depending on N. We will make use of the change of variables

$$u = \frac{x}{a_N} + b_N.$$

Now

$$P_\theta[\tilde{Z} \le x] = P_\theta[X_i \le u \text{ for all } i \in G]$$

$$= P_\theta\left[\frac{X_i - m_i}{\sigma_i} \le \frac{u - m_i}{\sigma_i} \text{ for all } i \in G \right].$$

Choose ρ with $a < \rho < 1$. ρ will not depend on N. We assume that N is so large that $a/\sigma^2 \le \rho$. Then under P_θ,

$$\operatorname{cov}\left(\frac{X_i - m_i}{\sigma_i}, \frac{X_j - m_j}{\sigma_j} \right) \le \frac{a^{|i-j|}}{\sigma^2}$$

$$\le \rho \text{ if } i \ne j.$$

Let $(\xi_i: i \in G)$ be a vector of independent Gaussian random variables with mean zero and variance one. By the normal comparison lemma [1, theorem 4.2.1, eq. 4.2.2] there is a constant C depending only on ρ such that

$$\begin{aligned}\left|P_\theta[\tilde{Z} \le x] - P\left[\xi_i \le \frac{u - m_i}{\sigma_i} \text{ for all } i \in G\right]\right| \\ &\le C \sum_{\substack{i<j \\ i,j \in G}} \operatorname{cov}\left(\frac{X_i - m_i}{\sigma_i}, \frac{X_j - m_j}{\sigma_j}\right) \\ &\quad \cdot \exp - \frac{(u - m_i)^2/\sigma_i + (u - m_j)^2/\sigma_j}{2(1+\rho)} \\ &\le C \sum_{\substack{i<j \\ i,j \in G}} \frac{a^{|i-j|}}{\sigma^2} \exp - \frac{(u_{\min} - m)^2}{(1+\rho)\sigma^2} \\ &\le \frac{2CN}{\sigma^2(1-a)} \exp\left(-\left(\frac{2}{1+\rho}\right)^{1/2} \log N\right) (\text{for } N \text{ large}) \\ &\le \frac{\delta}{4} \text{ (for } N \text{ large)} \end{aligned} \tag{3.3}$$

where we used the fact that

$$\begin{aligned} u_{\min} - m &= (x_{\min}/a_N + b_N) - m \\ &\ge \left(\frac{1+\rho}{2}\right)^{1/4} (2\log N)^{1/2} \text{ (for } N \text{ large).} \end{aligned}$$

We also have

$$\begin{aligned} P\left[\xi_i \le \frac{u - m_i}{\sigma_i} \text{ for all } i \in G\right] &\le P\left[\xi_i \le \frac{u+m}{\sigma} \text{ for all } i \in G\right] \\ &= P\left[a_N(\xi_i - b_N) \le x + \alpha(N,x) \text{ for all } i \text{ in } G\right] \end{aligned}$$

where

$$\begin{aligned} \alpha(N,x) &= \left(\frac{1}{\sigma} - 1\right)(x + a_N b_N) + \frac{m a_N}{\sigma} \\ &= O\left((\log N)^{-\delta}\right) = o(1). \end{aligned}$$

This, (3.1), and the basic convergence theorem for the maximum of independent standard Gaussian random variables [1] implies that for N sufficiently large,

$$P\left[\xi_i \le \frac{u - m_i}{\sigma_i} \text{ for all } i \in G\right] \le \exp(-(1 - r - 6\delta)e^{-x}) + \frac{\delta}{4}$$

for $x \in [x_{\min}, x_{\max}]$. Combining this with (3.3) proves the lemma.

Equation (3.2) implies that for N sufficiently large

$$\begin{aligned} \exp(-e^{-x}) - \delta &\le P[\tilde{X} \le x] \\ &\le P[\Theta \in A] + P[\tilde{X} \le x, \Theta \in A] \\ &\le \delta + P[\tilde{Y} \le x, \tilde{Z} \le x, \Theta \in A] \\ &= \delta + P[\tilde{Y} \le x, \Theta \in A] \\ &\quad \cdot P[\tilde{Z} \le x | \tilde{Y} \le x, \Theta \in A] \\ &\le \delta + P[\tilde{Y} \le x, \Theta \in A] \\ &\quad \cdot (\exp(-(1 - r - 6\delta)e^{-x}) + \delta) \end{aligned}$$

where we applied Lemma 3 and the fact that $\tilde{Y}$ is a function of Θ to get the last inequality. Thus we have

$$P[\tilde{Y} \le x, \Theta \in A] \ge \left[\frac{\exp(-e^{-x}) - 2\delta}{\exp(-(1 - r - 6\delta)e^{-x}) + \delta}\right]^+. \tag{3.4}$$

Note that the right side increases to $\exp(-re^{-x})$ as δ decreases to zero.

Next,

$$\begin{aligned} P[\tilde{Y} < \tilde{Z}] &\ge P[\tilde{Y} < \tilde{Z}, \Theta \in A] \\ &= \int_{-\infty}^{\infty} P[\tilde{Y} \in dx, x < \tilde{Z}, \Theta \in A] \\ &= \int_{-\infty}^{\infty} P[\tilde{Y} \in dx, \Theta \in A] P[x < \tilde{Z} | \tilde{Y} = x, \Theta \in A] \\ &\ge \int_{-\infty}^{\infty} P[\tilde{Y} \in dx, \Theta \in A] \\ &\quad \cdot (1 - \exp(-(1 - r - 6\delta)e^{-x} - \delta)^+ \end{aligned}$$

where once again we applied Lemma 3 and the fact that $\tilde{Y}$ is a function of Θ to get the last inequality. Let x_0 denote the value of x for which the expression in parentheses in the integrand is zero. Then by integration by parts, the last integral is equal to

$$\int_{-\infty}^{x_0} P[\tilde{Y} \le x, \Theta \in A] \frac{d}{dx} \exp(-(1 - r - 6\delta)e^{-x})\, dx \tag{3.5}$$

which in turn is at least as large as $\gamma(\delta)$, where $\gamma(\delta)$ is defined by (3.5) with the right side of (3.4) substituted in for $P[\tilde{Y} \le x, \Theta \in A]$. Thus

$$P[S] \le 1 - P[\tilde{Y} < \tilde{Z}] \le 1 - \gamma(\delta) \text{ for } N \text{ large.}$$

By the monotone convergence theorem,

$$\lim_{y \to 0} \gamma(y) = \int_{-\infty}^{\infty} e^{-re^{-x}} \frac{d}{dx}[\exp(-(1-r)e^{-x})]\, dx = 1 - r.$$

Since δ was arbitrary with $0 < \delta < 1$, we have proved that

$$\limsup_{N \to \infty} \max_{U \in U_{N,r}} P[S] \le r.$$

Together with Lemma 1, this proves the theorem.

ACKNOWLEDGMENT

Lennart Ljung got me started on this problem. I am grateful both to him and to John Tsitsiklis for enlightening discussions.

REFERENCES

[1] M. R. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes.* New York: Springer-Verlag, 1983.