A NOTE ON STOCHASTIC SEARCH METHODS FOR GLOBAL OPTIMIZATION

D. P. KENNEDY,* *University of Cambridge*

Abstract

Let $[A_n, B_n]$ be random subintervals of $[0, 1]$ defined recursively as follows. Let $A_1 = 0$, $B_1 = 1$ and take C_n, D_n to be the minimum and maximum of k, i.i.d. random points uniformly distributed on $[A_n, B_n]$. Choose $[A_{n+1}, B_{n+1}]$ to be $[C_n, B_n]$, $[A_n, D_n]$ or $[C_n, D_n]$ with probabilities p, q, r respectively, $p + q + r = 1$. It is shown that the limiting distribution of $[A_n, B_n]$ has the beta distribution on $[0, 1]$ with parameters $k(p + r)$ and $k(q + r)$. The result is used to consider a randomized version of Golden Section search.

RANDOM BINARY SEARCH; RANDOM GOLDEN SECTION SEARCH; INTERVAL-REDUCTION TECHNIQUES; RANDOMIZED GLOBAL OPTIMIZATION

1. Introduction

For a real-valued, differentiable, unimodal function, $f(x)$, defined on an interval, say $[0, 1]$, the method of binary search locates the value x^* maximizing f by successively reducing the interval in which x^* is known to lie; at the nth stage of the search if x^* is known to be in the subinterval $[a_n, b_n]$, and if c_n is the midpoint of $[a_n, b_n]$, the interval at the $(n + 1)$th stage is taken to be $[a_n, c_n]$ or $[c_n, b_n]$ according as $f'(c_n) < 0$ or $f'(c_n) > 0$. For functions f which are not necessarily unimodal, a randomized version of this algorithm for locating the local maxima (and hence the global maximum) of f had been investigated by Zemel [5]. Each application (or pass) of the algorithm locates a (random) point in X^*, the set of local maxima of f; successive passes sample points independently from X^*. Within each pass, at the nth stage, the search is reduced to a random subinterval $[A_n, B_n]$ $(A_0 = 0, B_0 = 1)$, C_n is (independently) uniformly distributed on $[A_n, B_n]$ and the interval at the $(n + 1)$th stage is taken to be $[A_n, C_n]$ or $[C_n, B_n]$ according as $f'(C_n) < 0$ or $f'(C_n) > 0$. As $n \to \infty$ the interval $[A_n, B_n]$ converges to a local maximum of f.

It is observed in [5] that for functions f for which the set X^* is evenly spread over $[0, 1]$, this randomized algorithm has a tendency to oversample those points in X^* which lie towards the edges of the interval $[0, 1]$. The reason for this may be seen by considering the situation where C_n is uniformly distributed over $[A_n, B_n]$ and the interval $[A_{n+1}, B_{n+1}]$ is taken to be $[A_n, C_n]$ or $[C_n, B_n]$ with equal probabilities $\frac{1}{2}$. Then, as has been noted by Chen et al. [2], $[A_n, B_n]$ converges (almost surely) to a point which has the arcsine distribution with probability density function $1/(\pi \sqrt{x(1-x)})$ on $[0, 1]$; this density is maximized at the edges of $[0, 1]$, since it is unbounded as $x \to 0$ or 1.

It might be asked whether randomized versions of other interval reduction techniques such as Golden Section search (or Fibonacci search), cf. [1], might sample points from X^* more uniformly. In Golden Section search, if f is unimodal (but not necessarily differentiable), at the nth stage the maximum is confined to the subinterval $[a_n, b_n]$ and f is evaluated at two points c_n, d_n from this interval with $c_n < d_n$. The interval at the next stage is taken to be $[a_n, d_n]$ or $[c_n, b_n]$ according to whether $f(c_n) > f(d_n)$ or $f(c_n) < f(d_n)$. In the next section we

consider a natural randomized version of this algorithm to locate local maxima for an arbitrary function f and derive a result which suggests how this algorithm might sample points from X^*.

2. The algorithm

Suppose that at the nth stage the search is reduced to the random subinterval $[A_n, B_n]$. Take k points $U_1^n, \ldots, U_k^n$, independently and uniformly distributed over $[A_n, B_n]$ and let $C_n = U_{(1)}^n$, $D_n = U_{(k)}^n$, where $U_{(1)}^n \leqq \cdots \leqq U_{(k)}^n$ are the order statistics of $U_1^n, \cdots, U_k^n$. Then the interval $[A_{n+1}, B_{n+1}]$ is taken to be $[A_n, D_n]$ or $[C_n, B_n]$ according as $f(C_n) > f(D_n)$ or $f(C_n) < f(D_n)$. It may be noted that, after the initial stage, to implement this algorithm only one additional point uniformly distributed over the remaining interval need be added at each stage. This is because conditional on C_n, B_n the points $U_{(2)}^n, \cdots, U_{(k)}^n$ are distributed as the order statistics of $k-1$ uniformly distributed points on $[C_n, B_n]$ (and similarly for $U_{(1)}^n, \cdots, U_{(k-1)}^n$ conditional on A_n, D_n). Furthermore, as $n \to \infty$ the algorithm will converge (almost surely) to a local maximum of f.

To analyse how the algorithm might sample points from X^*, we establish a limit result for the following scheme. Suppose that at the nth stage C_n, D_n are chosen in $[A_n, B_n]$ as described above, but now the interval $[A_{n+1}, B_{n+1}]$ is taken to be $[C_n, B_n]$, $[A_n, D_n]$ or $[C_n, D_n]$ independently at each stage with probabilities p, q, r, respectively, with $p + q + r = 1$. Then it is trivial to show that the interval $[A_n, B_n]$ converges (almost surely) to a point Z in $[0, 1]$. Denote by Beta (α, β) the beta distribution on $[0, 1]$ with parameters α, β having probability density function $(\Gamma(\alpha + \beta)/(\Gamma(\alpha)\Gamma(\beta))x^{\alpha-1}(1-x)^{\beta-1}$, $0 \leqq x \leqq 1$.

Proposition. The point Z has the Beta $(k(p + r), k(q + r))$ distribution.

Proof. Let $g(t) = E \exp(tZ)$ be the moment generating function of Z. By considering which of the three intervals is chosen at the first stage we obtain the following integral equation for $g(t)$:

$$g(t) = pk \int_0^1 (1-x)^{k-1} \exp(tx)g(t(1-x)) \, dx + qk \int_0^1 x^{k-1} g(t-x) \, dx$$

$$+ rk(k-1) \iint_{0 \leqq x \leqq y \leqq 1} (y-x)^{k-2} \exp(tx)g(t(y-x)) \, dx \, dy.$$

By making the change of variable $u = t(y - x)$, $v = t(1 - x)$ in the double integral it may be seen that

$$t^k g(t)/k = pe^t \int_0^t e^{-u} u^{k-1} g(u) \, du + q \int_0^t u^{k-1} g(u) \, du$$

$$+ r(k-1) \int_{0 \leqq u \leqq v \leqq t} u^{k-2} \exp(t-v)g(u) \, du \, dv$$

$$= \int_0^t (p \exp(t-u) + q)u^{k-1} g(u) \, du + r(k-1)$$

$$\times \int_0^t (\exp(t-u) - 1)u^{k-2} g(u) \, du.$$

Differentiating twice with respect to t, it follows that

(2.1) $$tg''(t) + \{(r+1)k - t\}g'(t) - (r+p)kg(t) = 0.$$

But, since $k \geqq 1$, from ([3], p. 1059) the solution of (2.1) satisfying $g(t) \to 1$ as $t \to 0$ is $g(t) = \Phi(k(r+q), k(1+r), t)$ where Φ is a degenerate hypergeometric function, and this solution in turn is the moment generating function of the Beta $(k(q+r), k(1-p))$ distribution, giving the result.

The proposition now suggests that the version of the randomized search algorithm with $k = 2$ will tend to sample the local maxima of f in X^* uniformly; for with $k = 2$ and $p = q = \frac{1}{2}$, $(r = 0)$, the limiting distribution in the proposition will be Beta $(1, 1)$ which is the uniform distribution on $[0, 1]$. Likewise, for $k > 2$ the algorithm will tend to oversample points towards the centre of the interval. Notice that, as described above, the case $k = 1$, $p = q = \frac{1}{2}$ gives as the limit the Beta $(\frac{1}{2}, \frac{1}{2})$ distribution which is the arcsine distribution.

It might also be pointed out that the proposition gives an example of an explicit limiting distribution for products of i.i.d. 2×2 stochastic matrices. Suppose that $\{\lambda_i, \mu_i\}$, $i = 1, 2, \ldots\}$ are i.i.d. random 2-vectors taking values in the unit square $[0, 1] \times [0, 1]$, with joint distribution $F(x, y) = P\{\lambda_1 \leqq x, \mu_1 < y\}$. If we define (A_n, B_n) by

$$\binom{A_{n+1}}{B_{n+1}} = \begin{pmatrix} 1 - \lambda_n & \lambda_n \\ 1 - \mu_n & \mu_n \end{pmatrix} \binom{A_n}{B_n}, \qquad n \geqq 1$$

with $A_1 = 0$, $B_1 = 1$, then $\begin{pmatrix} A_{n+1} & 1 - A_{n+1} \\ B_{n+1} & 1 - B_{n+1} \end{pmatrix}$ is the product of the 2×2 stochastic matrices

$$\begin{pmatrix} 1 - \lambda_t & \lambda_t \\ 1 - \mu_t & \mu_i \end{pmatrix}, \qquad i = 1, \cdots, n.$$

It is known that (cf. [4]) provided F is not concentrated on $\{(0, 1), (1, 0)\}$ then the pair (A_n, B_n) converges in distribution to a point (Z, Z), $0 \leqq Z \leqq 1$. However, the distribution of Z is known only for a limited number of distributions F (cf. [4] and references therein). The proposition gives the distribution of Z in the case where $F(x, y) = (r + q)y^k - r(y - x)^k$, $0 \leqq x \leqq y < 1$, $F(x, 1) - F(x, 1-) = p(1 - (1 - x)^k)$, $0 \leqq x \leqq 1$ and $F(x, y) = F(y, y)$ for $x \geqq y$.

Acknowledgement

References

[1] BEIGHTLER, C. S., PHILLIPS, D. T. AND WILDE, D. J. (1979) *Foundations of Optimization.* Prentice-Hall, Englewood Cliffs, New Jersey.

[2] CHEN, R., LIN, E. AND ZAME, A. (1981) Another arc sine law. *Sankhyā A* **43,** 371–373.

[3] GRADSHTEYN, I. S. AND RYZHIK, I. M. (1980) *Tables of Integrals, Series and Products.* Academic Press, New York.

[4] VAN ASSCHE, W. (1986) Products of 2×2 stochastic matrices with random entries. *J. Appl. Prob.* **23,** 1019–1024.

[5] ZEMEL, E. (1986) Random binary search: a randomizing algorithm for global optimization in $\mathbb{R}^1$. *Math. Operat. Res.* **11,** 651–662.