Bisection Search in the Presence of Noise
Rolf Waerber, Peter I. Frazier, Shane G. Henderson
School of Operations Research and Information Engineering, Cornell University, NY, USA

Stochastic Root-Finding Problem

- Consider a function \( g : [0, 1] \to \mathbb{R} \).
- Monotonicity assumption: There exists a unique \( X^* \in [0, 1] \) such that
  - \( g(x) > 0 \) for \( x < X^* \);
  - \( g(x) < 0 \) for \( x > X^* \).
- Evaluating \( g \) at a point \( x \in [0, 1] \) returns \( Y(x) = g(x) + \epsilon(x) \), where \( \epsilon(x) \) is an independent noise with zero mean (median).
- Goal: Find \( X^* \in [0, 1] \).
- Decisions:
  - Where to place samples \( X_n \) for \( n = 0, 1, 2, \ldots \)
  - How to estimate \( X^* \) after \( n \) iterations.

Applications

- **Stochastic Gradient Methods**: \( X_{n+1} = \pi_0 \{ X_n + a_n Y_n(X_n) \} \).
  For example, training a maximum likelihood estimator on a large dataset (Bottou, 2006).
- **Simulation Optimization**: \( g(x) \) is a gradient.
- **Active learning**: For example, detecting a boundary efficiently with an airborne range sensor (Castro and Nowak, 2008).

Setting

We reduce the observation to \( Z_n(X_n) := \text{sign}(Y_n(X_n)) \). Then
\[
Z_n(X_n) = \begin{cases} 
\text{sign}(g(X_n)) & \text{with probability } p(X_n), \\
-\text{sign}(g(X_n)) & \text{with probability } 1 - p(X_n).
\end{cases}
\]

The probability of a correct sign \( p(\cdot) \) depends on \( g(\cdot) \) and the noise \( \epsilon_n \).

**Stylized Setting**:
- \( p(\cdot) \) is constant!
- \( p(\cdot) \) is known!
- Notation: \( p(\cdot) = \rho \in \{1/2, 1\} \), \( q := 1 - \rho \).
- Bayes setting: \( X^* \sim f_0 \), where \( f_0 \) is a density with domain \([0, 1]\).

The Probabilistic Bisection Algorithm

1. Place a prior density \( f_0 \) on the root \( X^* \), \( f_0 \) has domain \([0, 1]\).
   (Example: \( U(0,1) \))
2. For \( n=0,1,2, \ldots \)
   (a) Measure at the median \( X_n := F_n^{-1}(1/2) \).
   (b) Update the posterior density:
   \[
   f_{n+1}(x) = \begin{cases} 
2p \cdot f_n(x) & \text{if } x > X_n, \\
2q \cdot f_n(x) & \text{if } x \leq X_n.
\end{cases}
\]
3. Estimate after \( n \) iterations: \( X_n = F_n^{-1}(1/2) \).

**Results**

- **Theorem: (Optimality)** The Probabilistic Bisection Algorithm is optimal in minimizing the expected posterior entropy \( \mathbb{E}[h(f_n)] \) for any \( N \in \mathbb{N} \).
- **Theorem: (Consistency)** On a set of probability 1 the posterior distribution \( P_n(\cdot) \) converges weakly to a point mass at \( X^* \).
- **Theorem: (Rate of Convergence)** There exists a constant \( c = c(p) > 1 \) such that \( \mathbb{E}[|X^* - X_n|] = o(c^{-n}) \).

**Under the stylized setting**, the asymptotic rate of convergence of the Probabilistic Bisection Algorithm is much faster than that of a Stochastic Gradient Method:
\[
o(c^{-n}) \text{ vs. } O(n^{-1/2}).
\]

Future Research

- In practice, the probability \( p(x) \) varies with \( x \) and is unknown.
- Can sample sequentially to achieve a probability \( p(X_n) \) bounded from below by a constant, say \( p_c \).
- Given any \( 1/2 < p_c < 1 \), a test could be constructed with \( p(X_n) \geq p_c \) whenever \( P(Z_n > 0) \neq 1/2 \).

**Conclusion**

Probabilistic Bisection-based algorithms could be very powerful in solving Stochastic Root-finding problems.
- Very fast rate of convergence (geometric).
- No tuning parameters.
- Robust.
- \( p(x) \) needs to be estimated (could slow down the rate of convergence).
- No ready-to-use algorithms yet.

- This work was supported in part by NSF grant CMMI-0800688 and AFOSR YIP FA9550-11-1-0083.
- References are available upon request.

rw339@cornell.edu, pf98@cornell.edu, sgh9@cornell.edu
http://www.orie.cornell.edu