Drift Analysis

GECCO 2012 Tutorial

Per Kristian Lehre

ASAP Research Group
School of Computer Science
University of Nottingham, UK
PerKristian.Lehre@nottingham.ac.uk

http://www.sigevo.org/gecco-2012/

The latest version of these slides are available from:
http://www.cs.nott.ac.uk/~pkl/drift/

GECCO'12 Companion, July 7-11, 2012, Philadelphia, PA, USA.
ACM 978-1-4503-1178-6/12/07.

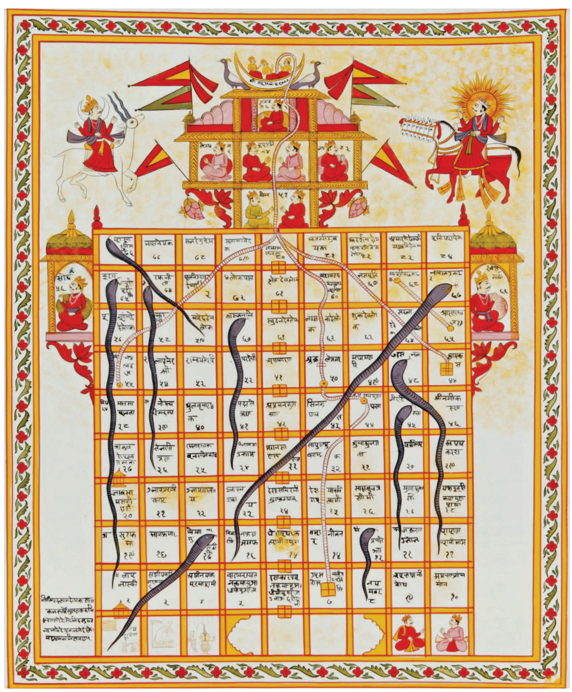
About the speaker

Per Kristian Lehre

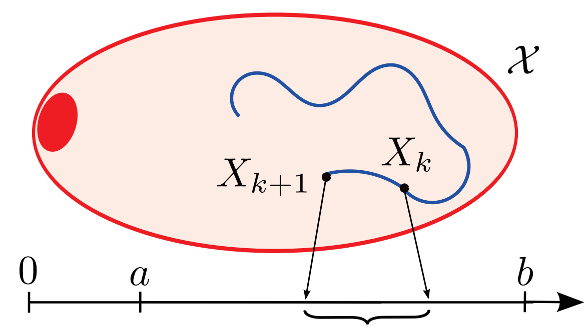
- Lecturer at University of Nottingham, UK
- PhD in CS from NTNU (Norway) 2006
- Working in EC since 2003

Research Interests

- Runtime analysis of evolutionary algorithms, especially population-based EAs
- Black-box complexity
- ...



What is Drift Analysis?

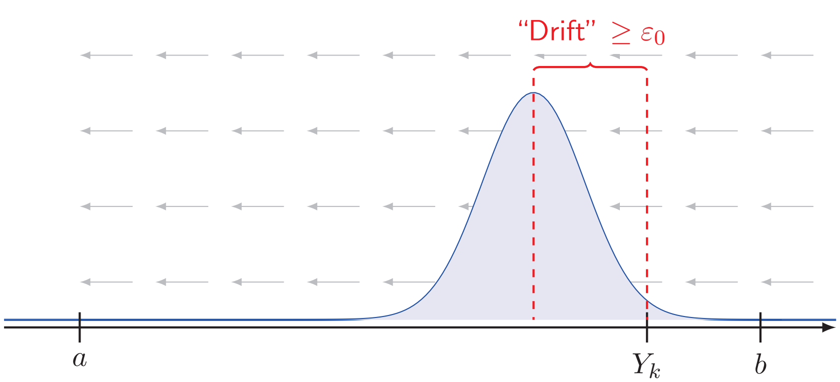


$$\Delta_k := g(X_{k+1}) - g(X_k)$$

- Prediction of the long term behaviour of a process X
  - hitting time, stability, occupancy time etc.

  from properties of Δ.

What is Drift Analysis[1]?



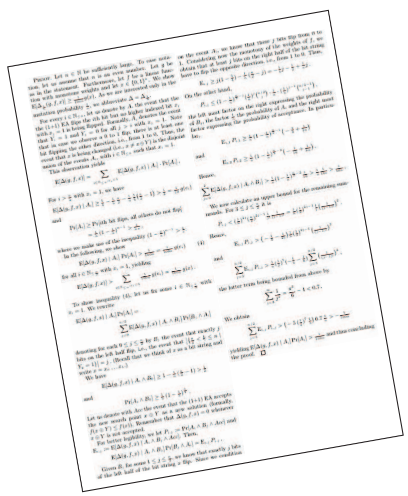
- A powerful tool in time-complexity analysis of EAs.

[1] NB! Drift is a different concept than *genetic drift* in evolutionary biology.

Runtime of (1+1) EA on Linear Functions [3]

Droste, Jansen & Wegener (2002)

Runtime of (1+1) EA on Linear Functions [2]



Doerr, Johannsen, and Winzen (GECCO 2010)

Some history

Origins

- Stability of equilbria in ODEs (Lyapunov, 1892)
- Stability of Markov Chains (see eg [14])
- 1982 paper by Hajek [6]
  - Simulated annealing [19]

Drift Analysis of Evolutionary Algorithms

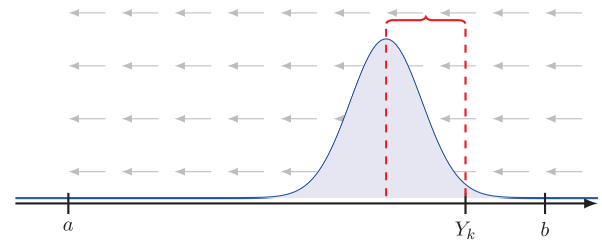
- Introduced to EC in 2001 by He and Yao [7, 8]
  - (1+1) EA on linear functions: $O(n \ln n)$ [7]
  - (1+1) EA on maximum matching by Giel and Wegener [5]
- Simplified drift in 2008 by Oliveto and Witt [18]
- Multiplicative drift by Doerr et al [2]
  - (1+1) EA on linear functions: $en \ln(n) + O(n)$ [22]
- Variable drift by Johannsen [11] and Mitavskiy et al. [15]
- Population drift [12]

About this tutorial...

- Assumes no or little background in probability theory
- Main focus will be on drift theorems and their proofs
  - Some theorems are presented in a simplified form
    full details are available in the references
- A few simple applications will be shown

- Please feel free to interrupt me with questions!

General Assumptions



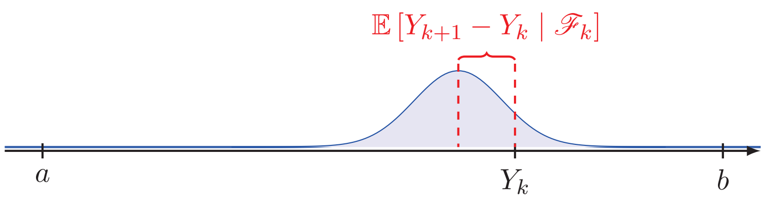
- X_k is a stochastic process[2] in some general state space $\mathcal{X}$ representing the state of the evolutionary algorithm
  - eg. X_k is the current search point of (1+1) EA in $\{0,1\}^n$
- $Y_k := g(X_k)$ were $g : \mathcal{X} \to \mathbb{R}$ is a “distance function”
- Our goal is to say something about these two stopping times

$$\tau_a := \min\{k \geq 0 \mid Y_k \leq a\} \quad \tau_b := \min\{k \geq 0 \mid Y_k \geq b\}$$

where we assume $-\infty \leq a < b < \infty$ and $Y_0 \in (a, b)$.

[2]We do not require X_k to be a Markov process.

Overview of Tutorial

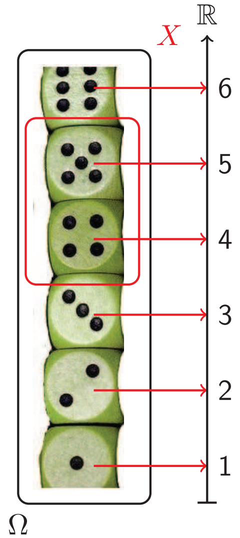


Drift Condition[3]	Statement	Note
$\mathbb{E}\left[Y_{k+1} \mid \mathscr{F}_k\right] \leq Y_k - \varepsilon_0$	$\mathbb{E}\left[\tau_a\right] \leq$	Additive drift [7, 10]
	$\Pr\left(\tau_a > B\right) \leq$	[6]
	$\Pr\left(\tau_b < B\right) \leq$	Simplified drift [6, 17]
$\mathbb{E}\left[Y_{k+1} \mid \mathscr{F}_k\right] \geq Y_k - \varepsilon_0$	$\leq \mathbb{E}\left[\tau_a\right]$	Additive drift (lower b.) [7, 9]
$\mathbb{E}\left[Y_{k+1} \mid \mathscr{F}_k\right] \leq Y_k$	$\mathbb{E}\left[\tau_a\right] \leq$	Supermartingale [16]
$\mathbb{E}\left[Y_{k+1} \mid \mathscr{F}_k\right] \leq (1-\delta)Y_k$	$\mathbb{E}\left[\tau_a\right] \leq$	Multiplicative drift [2, 4]
	$\Pr\left(\tau_a > B\right) \leq$	[1]
$\mathbb{E}\left[Y_{k+1} \mid \mathscr{F}_k\right] \geq (1-\delta)Y_k$	$\leq \mathbb{E}\left[\tau_a\right]$	Multipl. drift (lower b.) [13]
$\mathbb{E}\left[Y_{k+1} \mid \mathscr{F}_k\right] \leq Y_k - h(Y_k)$	$\mathbb{E}\left[\tau_a\right] \leq$	Variable drift [11]
$\mathbb{E}\left[e^{\lambda Y_{k+1}} \mid \mathscr{F}_k\right] \leq \frac{e^{\lambda Y_k}}{\alpha_0}$	$\Pr\left(\tau_b < B\right) \leq$	Population drift [12]

[3]Some drift theorems need additional conditions.

Part 1 - Basic Probability Theory

Basic Probability Theory



Probability Triple $(\Omega, \mathscr{F}, \Pr)$

- Ω : Sample space
- $\mathscr{F}$: σ-algebra on Ω (family of events)
- $\Pr : \mathscr{F} \to \mathbb{R}$ probability function (satisfying probability axioms)

Events

- $\mathcal{E} \in \mathscr{F}$

Random Variable

- $X : \Omega \to \mathbb{R}$ and $X^{-1} : \mathcal{B} \to \mathscr{F}$
- $X = y \Longleftrightarrow \{\omega \in \Omega \mid X(\omega) = y\}$

Stochastic Processes and Filtration

Definition

- A *stochastic process* is a sequence of random variables $Y_1, Y_2, \ldots$ on the same probability space.
- A *filtration* is an increasing family of sub σ-algebras[4] of $\mathscr{F}$

$$\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \cdots \subseteq \mathscr{F}$$

- A stochastic process Y_k is *adapted to a filtration* $\mathscr{F}_k$ if Y_k is $\mathscr{F}_k$-measurable for all k: $\forall A \in \mathcal{B}\{\omega \in \Omega \mid Y_k(\omega) \in A\} \in \mathscr{F}_k$

$\Longrightarrow$ Informally, $\mathscr{F}_k$ represents the information that has been revealed about the process during the first k steps, including the value of Y_k.

[4] An event family closed under any countable collection of set operations [21].

Stopping Time

Definition

A rv. $\tau : \Omega \to \mathbb{N}$ is called *a stopping time* if for all $k \geq 0$

$$\{\tau \leq k\} \in \mathscr{F}_k$$

- The information obtained until step k is sufficient to decided whether the event $\{\tau \leq k\}$ is true or not.

Example

- The smallest t such that $Y_t < a$ in a stochastic process.
- The runtime of an evolutionary algorithm

Expectation

The **expectation** of a discrete random variable X is

$$\mathbb{E}[X] := \sum_x x \Pr(X = x)$$

Independence
Two random variables X and Y are **independent** if

$$\forall x, y \in \mathbb{R} \quad \Pr(\{X \leq x\} \cap \{Y \leq y\}) = \Pr(X \leq x) \Pr(Y \leq y)$$

If X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$
Linearity of expectation
For any random variables X and Y, and constants $a, b \in \mathbb{R}$

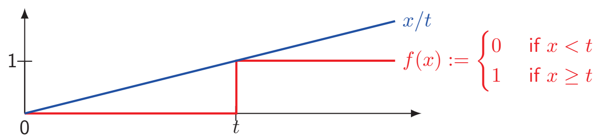
$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$
$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

$\Longrightarrow$ Often easier to find $\mathbb{E}[X]$ than $\Pr(X \geq k)$.

Markov's Inequality

Theorem (Markov's Inequality)

Let X be a random variable assuming only non-negative values. Then $\Pr(X \geq t) \leq \mathbb{E}[X]/t$ for all $t \in \mathbb{R}^+$.



$$\Pr(X \geq t) = 0 \cdot \Pr(X < t) + 1 \cdot \Pr(X \geq t) = \mathbb{E}[f(X)] \leq \mathbb{E}[X/t].$$

Jensen's Inequality

Theorem

If $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function on (a, b), and X a random variable taking values in (a, b), then

$$\mathbb{E}[\varphi(X)] \geq \varphi(\mathbb{E}[X])$$

- If $\varphi''(x) \geq 0$ for all $x \in (a, b)$, then φ is convex on (a, b).

Jensen's Inequality

Theorem

If $\varphi : \mathbb{R} \to \mathbb{R}$ is a concave function on (a, b), and X a random variable taking values in (a, b), then

$$\mathbb{E}[\varphi(X)] \leq \varphi(\mathbb{E}[X])$$

- If $\varphi''(x) < 0$ for all $x \in (a, b)$, then φ is concave on (a, b).

Conditional Probability and Expectation

Given two events $\mathcal{A}$ and $\mathcal{E}$, where $\Pr(\mathcal{E}) > 0$

$$\Pr(\mathcal{A} \mid \mathcal{E}) := \frac{\Pr(\mathcal{A} \cap \mathcal{E})}{\Pr(\mathcal{E})}$$

Conditional Expectation

$$\mathbb{E}[X \mid \mathcal{E}] := \sum_x x \cdot \Pr(X = x \mid \mathcal{E})$$

Law of total probability, $0 < \Pr(\mathcal{E}) < 1$

$$\Pr(\mathcal{A}) = \;\geq \Pr(\mathcal{E}) \cdot \Pr(\mathcal{A} \mid \mathcal{E}) + \Pr(\overline{\mathcal{E}}) \cdot \Pr(\mathcal{A} \mid \overline{\mathcal{E}})$$
$$\mathbb{E}[X] = \;\geq \Pr(\mathcal{E}) \cdot \mathbb{E}[X \mid \mathcal{E}] + \Pr(\overline{\mathcal{E}}) \cdot \mathbb{E}[X \mid \overline{\mathcal{E}}] \text{ if } X \geq 0$$

Tower Property, if $\mathscr{H}$ is a sub-σ-algebra of $\mathscr{G}$, then

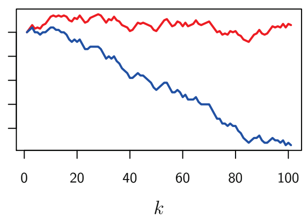
$$\mathbb{E}[\mathbb{E}[X \mid \mathscr{G}] \mid \mathscr{H}] = \mathbb{E}[X \mid \mathscr{H}]$$

Martingales

Definition (Supermartingale)

Any process Y_k st $\forall k$

1. Y_k is adapted to $\mathscr{F}_k$
2. $\mathbb{E}[|X_k|] < \infty$
3. $\mathbb{E}[Y_{k+1} \mid \mathscr{F}_k] \leq Y_k$



Example

Let $\Delta_1, \Delta_2, \ldots$ be rvs with $-\infty < \mathbb{E}[\Delta_{k+1} \mid \mathscr{F}_k] \leq -\varepsilon_0$ for $k \geq 0$
Then the following sequence is a super-martingale

$$Y_k := \Delta_1 + \cdots + \Delta_k \qquad Z_k := Y_k + k\varepsilon_0$$

$$\begin{aligned}\mathbb{E}[Y_{k+1} Z_{k+1} \mid \mathscr{F}_k] &= \Delta_1 + \cdots + \Delta_k + \mathbb{E}[\Delta_{k+1} \mid \mathscr{F}_k] + (k+1)\varepsilon_0 \\ &\leq \Delta_1 + \cdots + \Delta_k - \varepsilon_0 + (k+1)\varepsilon_0 < Y_k = Z_k\end{aligned}$$

Part 2 - Additive Drift

Martingales

Lemma

If Y is a supermartingale, then $\mathbb{E}[Y_k \mid \mathscr{F}_0] \leq Y_0$ for all fixed $k \geq 0$.

Proof.

$$\begin{aligned}\mathbb{E}[Y_k \mid \mathscr{F}_0] &= \mathbb{E}[\mathbb{E}[Y_k \mid \mathscr{F}_k] \mid \mathscr{F}_0] \\ &\leq \mathbb{E}[Y_{k-1} \mid \mathscr{F}_0] \underbrace{\leq \cdots \leq \mathbb{E}[Y_0 \mid \mathscr{F}_0]}_{\text{by induction on } k} = Y_0\end{aligned}$$

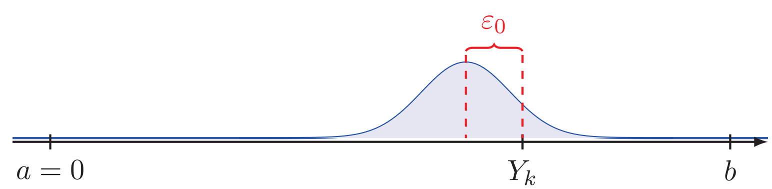
Example

Where is the process Y in the previous example after k steps?

$$Y_0 = Z_0 \geq \mathbb{E}[Z_k \mid \mathscr{F}_0] = \mathbb{E}[Y_k + k\varepsilon_0 \mid \mathscr{F}_0]$$

Hence, $\mathbb{E}[Y_k \mid \mathscr{F}_0] \leq Y_0 - \varepsilon_0 k$, which is not surprising...

Additive Drift



(C1+) $\forall k \quad \mathbb{E}[Y_{k+1} - Y_k \mid Y_k > 0 \wedge \mathscr{F}_k] \leq -\varepsilon_0$
(C1−) $\forall k \quad \mathbb{E}[Y_{k+1} - Y_k \mid Y_k > 0 \wedge \mathscr{F}_k] \geq -\varepsilon_0$

Theorem ([7, 9, 10])

Given a sequence $(Y_k, \mathscr{F}_k)$ over an interval $[0, b] \subset \mathbb{R}$.
Define $\tau := \min\{k \geq 0 \mid Y_k = 0\}$, and assume $\mathbb{E}[\tau \mid \mathscr{F}_0] < \infty$.

- *If (C1+) holds for an $\varepsilon_0 > 0$, then $\mathbb{E}[\tau \mid \mathscr{F}_0] \leq Y_0/\varepsilon_0 \leq b/\varepsilon_0$.*
- *If (C1−) holds for an $\varepsilon_0 > 0$, then $\mathbb{E}[\tau \mid \mathscr{F}_0] \geq Y_0/\varepsilon_0$.*

Obtaining Supermartingales from Drift Conditions

Definition (Stopped Process)

Let Y be a stochastic process and τ a stopping time.

$$Y_{k\wedge\tau} := \begin{cases} Y_k & \text{if } k < \tau \\ Y_\tau & \text{if } k \geq \tau \end{cases}$$

Let $\tau_a := \min\{k \geq 0 \mid Y_k \leq a\}$, and assume

(C1) $\mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > a \wedge \mathscr{F}_k\right] \leq -\varepsilon_0$

- Condition (C1) only required to hold when $Y_k > a$,
- Hence, Y_k is not necessarily a supermartingale...
- But the "stopped process" $Y_{k\wedge\tau_a}$ is a supermartingale, so

$$\mathbb{E}\left[Y_{k\wedge\tau_a} \mid \mathscr{F}_0\right] \leq Y_0 \quad \forall k$$

Dominated Convergence Theorem

Theorem

Suppose X_k is a sequence of random variables such that for each outcome in the sample space

$$\lim_{k\to\infty} X_k = X.$$

Let $Y \geq 0$ be a random variable with $\mathbb{E}[Y] < \infty$ such that for each outcome in the sample space, and for each k

$$|X_k| \leq Y.$$

Then it holds

$$\lim_{k\to\infty} \mathbb{E}[X_k] = \mathbb{E}\left[\lim_{k\to\infty} X_k\right] = \mathbb{E}[X]$$

Proof of Additive Drift Theorem

(C1+) $\forall k \quad \mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > 0 \wedge \mathscr{F}_k\right] \leq -\varepsilon_0$

(C1−) $\forall k \quad \mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > 0 \wedge \mathscr{F}_k\right] \geq -\varepsilon_0$

Theorem

Given a sequence $(Y_k, \mathscr{F}_k)$ over an interval $[0, b] \subset \mathbb{R}$
Define $\tau := \min\{k \geq 0 \mid Y_k = 0\}$, and assume $\mathbb{E}[\tau \mid \mathscr{F}_0] < \infty$.

- *If (C1+) holds for an $\varepsilon_0 > 0$, then $\mathbb{E}[\tau \mid \mathscr{F}_0] \leq Y_0/\varepsilon_0$.*
- *If (C1−) holds for an $\varepsilon_0 > 0$, then $\mathbb{E}[\tau \mid \mathscr{F}_0] \geq Y_0/\varepsilon_0$.*

Proof.

By (C1+), $Z_k := Y_{k\wedge\tau} + \varepsilon_0(k \wedge \tau)$ is a super-martingale, so

$$Y_0 = \mathbb{E}[Z_0 \mid \mathscr{F}_0] \geq \mathbb{E}[Z_k \mid \mathscr{F}_0] \quad \forall k.$$

Since Y_k is bounded to $[0, b]$, and τ has finite expectation, the dominated convergence theorem applies and

$$Y_0 \geq \lim_{k\to\infty} \mathbb{E}[Z_k \mid \mathscr{F}_0] = \mathbb{E}[Y_\tau + \varepsilon_0\tau \mid \mathscr{F}_0] = \varepsilon_0\mathbb{E}[\tau \mid \mathscr{F}_0].$$

Proof of Additive Drift Theorem

(C1+) $\forall k \quad \mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > 0 \wedge \mathscr{F}_k\right] \leq -\varepsilon_0$

(C1−) $\forall k \quad \mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > 0 \wedge \mathscr{F}_k\right] \geq -\varepsilon_0$

Theorem

Given a sequence $(Y_k, \mathscr{F}_k)$ over an interval $[0, b] \subset \mathbb{R}$
Define $\tau := \min\{k \geq 0 \mid Y_k = 0\}$, and assume $\mathbb{E}[\tau \mid \mathscr{F}_0] < \infty$.

- *If (C1+) holds for an $\varepsilon_0 > 0$, then $\mathbb{E}[\tau \mid \mathscr{F}_0] \leq Y_0/\varepsilon_0$.*
- *If (C1−) holds for an $\varepsilon_0 > 0$, then $\mathbb{E}[\tau \mid \mathscr{F}_0] \geq Y_0/\varepsilon_0$.*

Proof.

By (C1−), $Z_k := Y_{k\wedge\tau} + \varepsilon_0(k \wedge \tau)$ is a sub-martingale, so

$$Y_0 = \mathbb{E}[Z_0 \mid \mathscr{F}_0] \leq \mathbb{E}[Z_k \mid \mathscr{F}_0] \quad \forall k.$$

Since Y_k is bounded to $[0, b]$, and τ has finite expectation, the dominated convergence theorem applies and

$$Y_0 \leq \lim_{k\to\infty} \mathbb{E}[Z_k \mid \mathscr{F}_0] = \mathbb{E}[Y_\tau + \varepsilon_0\tau \mid \mathscr{F}_0] = \varepsilon_0\mathbb{E}[\tau \mid \mathscr{F}_0].$$

Example: (1+1) EA on LeadingOnes

1 (1+1) EA

Sample $x^{(0)}$ uniformly at random from $\{0,1\}^n$.
for $k = 0, 1, 2, \ldots$ **do**
  Set $y := x^{(k)}$, and flip each bit of y with probability $1/n$.

$$x^{(k+1)} := \begin{cases} y & \text{if } f(y) \geq f(x^{(k)}) \\ x^{(k)} & \text{otherwise.} \end{cases}$$

end for

Example 1: (1+1) EA on LeadingOnes

$$\textsc{Lo}(x) := \sum_{i=1}^{n} \prod_{j=1}^{i} x_j$$

$$x = \overbrace{1111111111111111}^{\text{Leading 1-bits}}\overbrace{0{*}{*}{*}{*}{*}{*}{*}{*}{*}{*}{*}{*}{*}{*}{*}{*}}^{\text{Remaining bits}}.$$

Left-most 0-bit

- Let $Y_k := n - \textsc{Lo}(x^{(k)})$ be the "remaining" bits in step $k \geq 0$.
- Let $\mathcal{E}$ be the event that only the left-most 0-bit flipped in y.
- The sequence Y_k is non-increasing, so

$$\begin{aligned} &\mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > 0 \wedge \mathscr{F}_k\right] \\ &\quad\leq (-1) \Pr\left(\mathcal{E} \mid Y_k > 0 \wedge \mathscr{F}_k\right) \\ &\quad= (-1)(1/n)(1-1/n)^{n-1} \leq -1/en. \end{aligned}$$

- By the additive drift theorem, $\mathbb{E}\left[\tau \mid \mathscr{F}_0\right] \leq enY_0 \leq en^2$.

Example 2: (1+1) EA on Linear Functions

- Given some constants $w_1, \ldots, w_n \in [w_{\min}, w_{\max}]$, define

$$f(x) := w_1 x_1 + w_2 x_2 + \cdots + w_n x_n$$

- Let Y_k be the function value that "remains" at time k, ie

$$Y_k := \left(\sum_{i=1}^{n} w_i\right) - \left(\sum_{i=1}^{n} w_i x_i^{(k)}\right) = \sum_{i=1}^{n} w_i \left(1 - x_i^{(k)}\right).$$

- Let $\mathcal{E}_i$ be the event that only bit i flipped in y, then

$$\mathbb{E}\left[Y_{k+1} - Y_k \mid \mathscr{F}_k\right] \leq \sum_{i=1}^{n} \Pr\left(\mathcal{E}_i \mid \mathscr{F}_k\right) \mathbb{E}\left[Y_{k+1} - Y_k \mid \mathcal{E}_i \wedge \mathscr{F}_k\right]$$

$$\leq \left(\frac{1}{n}\right)\left(1 - \frac{1}{n}\right)^{n-1} \sum_{i=1}^{n} w_i \left(x_i^{(k)} - 1\right) \leq -\frac{Y_k}{en} \leq -\frac{w_{\min}}{en}$$

- By the additive drift theorem, $\mathbb{E}\left[\tau \mid \mathscr{F}_0\right] \leq en^2 (w_{\max}/w_{\min})$.

Remarks on Example Applications

Example 1: (1+1) EA on LeadingOnes

- The upper bound en^2 is very accurate.
- The exact expression is $c(n)n^2$, where $c(n) \to (e-1)/2$ [20].

Example 2: (1+1) EA on Linear Functions

- The upper bound $en^2(w_{\max}/w_{\min})$ is correct, but very loose.
- The linear function BinVal has $(w_{\max}/w_{\min}) = 2^{n-1}$.
- The tightest known bound is $en\log(n) + O(n)$ [22].

$\Longrightarrow$ A poor choice of distance function gives an imprecise bound!

What is a good distance function?

Theorem ([8])

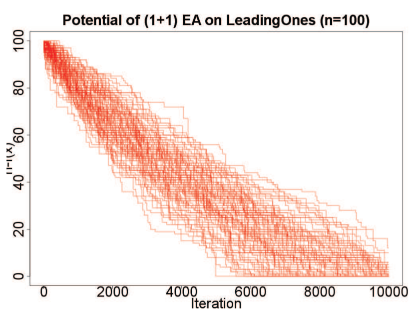
Assume Y is a homogeneous Markov chain, and τ the time to absorption. Then the function $g(x) := \mathbb{E}[\tau \mid Y_0 = x]$, satisfies

$$\begin{cases} g(x) = 0 & \text{if } x \text{ is an absorbing state} \\ \mathbb{E}[g(Y_{k+1}) - g(Y_k) \mid \mathscr{F}_k] = -1 & \text{otherwise.} \end{cases}$$

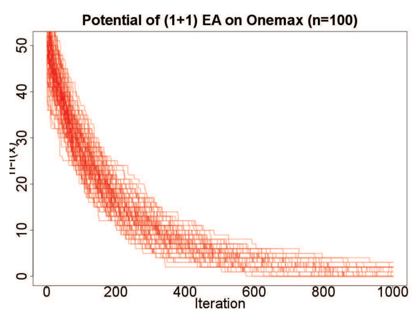
- Distance function g gives exact expected runtime!
- But g requires complete knowledge of the expected runtime!
- Still provides insight into what is a good distance function:
  - a good approximation (or guess) for the remaining runtime

Part 3 - Variable Drift

Drift may be Position-Dependant



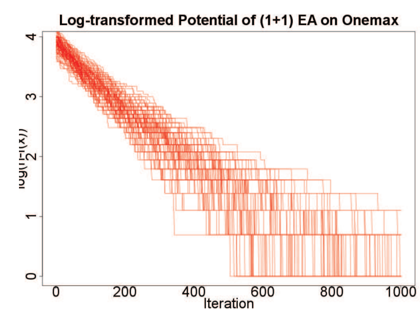
Constant Drift



Variable Drift

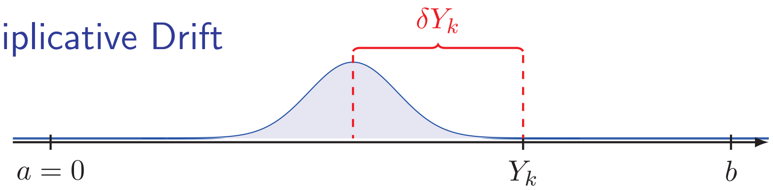
Drift may be Position-Dependant

Idea: Find a function $g : \mathbb{R} \to \mathbb{R}$ st. the transformed stochastic process $g(X_1), g(X_2), g(X_3), \ldots$ has constant drift.



Constant Drift

Multiplicative Drift



(M) $\forall k \quad \mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > a \wedge \mathscr{F}_k\right] \le -\delta Y_k$

Theorem ([2])

Given a sequence $(Y_k, \mathscr{F}_k)$ *over an interval* $[a, b] \subset \mathbb{R}$, $a > 0$
Define $\tau_a := \min\{k \ge 0 \mid Y_k = a\}$, *and assume* $\mathbb{E}\left[\tau_a \mid \mathscr{F}_0\right] < \infty$.

- *If (M) holds for a* $\delta > 0$, *then* $\mathbb{E}\left[\tau_a \mid \mathscr{F}_0\right] \le \ln(Y_0/a)/\delta$.

Proof.

$g(s) := \ln(s/a)$ is concave, so by Jensen's inequality

$$\begin{aligned}&\mathbb{E}\left[g(Y_{k+1}) - g(Y_k) \mid Y_k > a \wedge \mathscr{F}_k\right]\\ &\quad \le \ln(\mathbb{E}\left[Y_{k+1} \mid Y_k > a \wedge \mathscr{F}_k\right]) - \ln(Y_k) \le \ln(1-\delta) \le -\delta.\end{aligned}$$

$\square$

Example: Linear Functions Revisited

- For any $c \in (0,1)$, define the distance at time k as

$$Y_k := c w_{\min} + \sum_{i=1}^{n} w_i \left(1 - x_i^{(k)}\right)$$

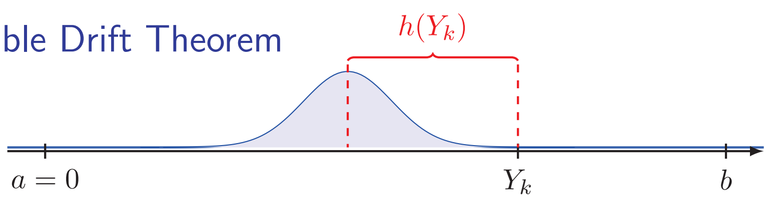
- We have already seen that

$$\begin{aligned}\mathbb{E}\left[Y_{k+1} - Y_k \mid \mathscr{F}_k\right] &\le \frac{1}{en} \sum_{i=1}^{n} w_i \left(x_i^{(k)} - 1\right)\\ &= -\frac{Y_k - c w_{\min}}{en} \le -\frac{Y_k(1-c)}{en}\end{aligned}$$

- By the multiplicative drift theorem ($a := c w_{\min}$ and $\delta := \frac{1-c}{en}$)

$$\mathbb{E}\left[\tau_a \mid \mathscr{F}_0\right] \le \left(\frac{en}{1-c}\right) \ln\left(1 + \frac{n w_{\max}}{c w_{\min}}\right)$$

Variable Drift Theorem



(V) $\forall k \quad \mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > 0 \wedge \mathscr{F}_k\right] \le -h(Y_k)$

Theorem ([15, 11])

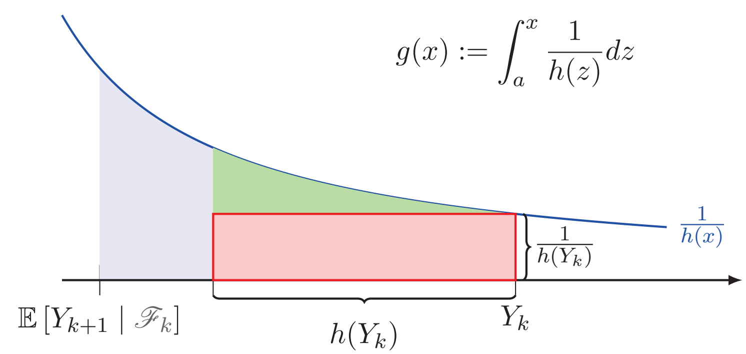
Given a sequence $(Y_k, \mathscr{F}_k)$ *over an interval* $[a, b] \subset \mathbb{R}, a > 0$.
Define $\tau_a := \min\{k \ge 0 \mid Y_k = a\}$, *and assume* $\mathbb{E}\left[\tau_a \mid \mathscr{F}_0\right] < \infty$.
If there exists a function $h : \mathbb{R} \to \mathbb{R}$ *such that*

- $h(x) > 0$ *and* $h'(x) > 0$ *for all* $x \in [a, b]$, *and*
- *drift condition (V) holds, then*

$$\mathbb{E}\left[\tau_a \mid \mathscr{F}_0\right] \le \int_a^{Y_0} \frac{1}{h(z)} dz$$

$\Longrightarrow$ The multiplicative drift theorem is the special case $h(x) = \delta x$.

Variable Drift Theorem: Proof



Proof.

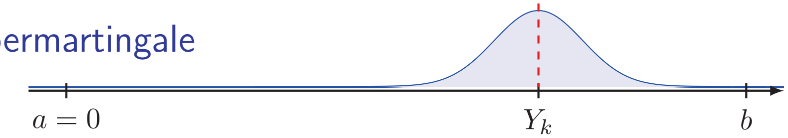
The function g is concave ($g'' < 0$), so by Jensen's inequality

$$\begin{aligned}\mathbb{E}\left[g(Y_k) - g(Y_{k+1}) \mid \mathscr{F}_k\right] &\ge g(Y_k) - g(\mathbb{E}\left[g(Y_{k+1}) \mid \mathscr{F}_k\right])\\ &\ge \int_{Y_k - h(Y_k)}^{Y_k} \frac{1}{h(z)} dz \ge 1\end{aligned}$$

$\square$

Part 4 - Supermartingale

Supermartingale



(S1) $\forall k \quad \mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > 0 \wedge \mathscr{F}_k\right] \leq 0$

(S2) $\forall k \quad \mathbf{Var}\left[Y_{k+1} \mid Y_k > 0 \wedge \mathscr{F}_k\right] \geq \sigma^2$

Theorem (See eg. [16])

Given a sequence $(Y_k, \mathscr{F}_k)$ *over an interval* $[0, b] \subset \mathbb{R}$.
Define $\tau := \min\{k \geq 0 \mid Y_k = 0\}$, *and assume* $\mathbb{E}\left[\tau \mid \mathscr{F}_0\right] < \infty$.

- *If (S1) and (S2) hold for* $\sigma > 0$, *then* $\mathbb{E}\left[\tau \mid \mathscr{F}_0\right] \leq \frac{Y_0(2b - Y_0)}{\sigma^2}$

Proof.

Let $Z_k := b^2 - (b - Y_k)^2$, and note that $b - Y_k \leq \mathbb{E}\left[b - Y_{k+1} \mid \mathscr{F}_k\right]$.

$$
\begin{aligned}
\mathbb{E}\left[Z_{k+1} - Z_k \mid \mathscr{F}_k\right] &= -\mathbb{E}\left[(b - Y_{k+1})^2 \mid \mathscr{F}_k\right] + (b - Y_k)^2 \\
&\leq -\mathbb{E}\left[(b - Y_{k+1})^2 \mid \mathscr{F}_k\right] + \mathbb{E}\left[b - Y_{k+1} \mid \mathscr{F}_k\right]^2 \\
&= -\mathbf{Var}\left[b - Y_{k+1}\right] = -\mathbf{Var}\left[Y_{k+1} \mid \mathscr{F}_k\right] \leq -\sigma^2 \qquad \square
\end{aligned}
$$

Part 5 - Hajek's Theorem

Hajek's Theorem[5]

Theorem

If there exist $\lambda, \varepsilon_0 > 0$ *and* $D < \infty$ *such that for all* $k \geq 0$

(C1) $\mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > a \wedge \mathscr{F}_k\right] \leq -\varepsilon_0$

(C2) $(|Y_{k+1} - Y_k| \mid \mathscr{F}_k) \prec Z$ *and* $\mathbb{E}\left[e^{\lambda Z}\right] = D$

then for any $\delta \in (0, 1)$

(2.9) $\Pr\left(\tau_a > B \mid \mathscr{F}_0\right) \leq e^{\eta(Y_0 - a - B(1-\delta)\varepsilon_0)}$

(*) $\Pr\left(\tau_b < B \mid \mathscr{F}_0\right) \leq \frac{BD}{(1-\delta)\eta\varepsilon_0} \cdot e^{\eta(a-b)}$

for some $\eta \geq \min\{\lambda, \delta\varepsilon_0\lambda^2/D\} > 0$.

- If $\lambda, \varepsilon_0, D \in O(1)$ and $b - a \in \Omega(n)$,
  then there exists a constant $c > 0$ such that

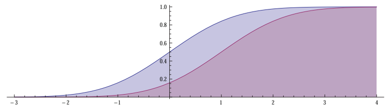
$$\Pr\left(\tau_b \leq e^{cn} \mid \mathscr{F}_0\right) \leq e^{-\Omega(n)}$$

[5] The theorem presented here is a corollary to Theorem 2.3 in [6].

Stochastic Dominance - $(|Y_{k+1} - Y_k| \mid \mathscr{F}_k) \prec Z$

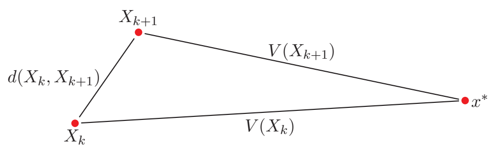
Definition

$Y \prec Z$ if $\Pr(Z \le c) \le \Pr(Y \le c)$ for all $c \in \mathbb{R}$



Example

1. If $Y \le Z$, then $Y \prec Z$.
2. Let (Ω, d) be a metric space, and $V(x) := d(x, x^*)$. Then $|V(X_{k+1}) - V(X_k)| \prec d(X_{k+1}, X_k)$



Condition (C2) implies that "long jumps" must be rare

Assume that

(C2) $(|Y_{k+1} - Y_k| \mid \mathscr{F}_k) \prec Z$ and $\mathbb{E}\left[e^{\lambda Z}\right] = D$

Then for any $j \ge 0$,

$$\begin{aligned}
\Pr(|Y_{k+1} - Y_k| \ge j) &= \Pr\left(e^{\lambda |Y_{k+1} - Y_k|} \ge e^{\lambda j}\right) \\
&\le \mathbb{E}\left[e^{\lambda |Y_{k+1} - Y_k|}\right] e^{-\lambda j} \\
&\le \mathbb{E}\left[e^{\lambda Z}\right] e^{-\lambda j} \\
&= D e^{-\lambda j}.
\end{aligned}$$

Moment Generating Function (mgf) $\mathbb{E}\left[e^{\lambda Z}\right]$

Definition

The mgf of a rv X is $M_X(\lambda) := \mathbb{E}\left[e^{\lambda X}\right]$ for all $\lambda \in \mathbb{R}$.

- The n-th derivative at $t = 0$ is $M_X^{(n)}(0) = \mathbb{E}[X^n]$, hence M_X provides all moments of X, thus the name.
- If X and Y are independent rv. and $a, b \in \mathbb{R}$, then

$$M_{aX+bY}(t) = \mathbb{E}\left[e^{t(aX+bX)}\right] = \mathbb{E}\left[e^{taX}\right]\mathbb{E}\left[e^{tbX}\right] = M_X(at) M_Y(bt)$$

Example

- Let $X := \sum_{i=1}^n X_i$ where X_i are independent rvs with $\Pr(X_i = 1) = p$ and $\Pr(X_i = 0) = 1 - p$. Then

$$M_{X_i}(\lambda) = (1-p)e^{\lambda \cdot 0} + p e^{\lambda \cdot 1}$$
$$M_X(\lambda) = M_{X_1}(\lambda) M_{X_2}(\lambda) \cdots M_{X_n}(\lambda) = (1 - p + pe^{\lambda})^n.$$

Moment Generating Functions

Distribution		mgf
Bernoulli	$\Pr(X = 1) = p$	$1 - p + pe^t$
Binomial	$X \sim \mathrm{Bin}(n, p)$	$(1 - p + pe^t)^n$
Geometric	$\Pr(X = k) = (1-p)^{k-1} p$	$\frac{pe^t}{1-(1-p)e^t}$
Uniform	$X \sim U(a, b)$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$
Normal	$X \sim N(\mu, \sigma^2)$	$\exp(t\mu + \frac{1}{2}\sigma^2 t^2)$

The mgf. of $X \sim \mathrm{Bin}(n, p)$ at $t = \ln(2)$ is

$$(1 - p + pe^t)^n = (1 + p)^n \le e^{pn}.$$

Condition (C2) often holds trivially

Example ((1+1) EA)

Choose x uniformly from $\{0,1\}^n$
for $k = 0, 1, 2, ...$
Set $x' := x^{(k)}$, and flip each bit of x' with probability p.
If $f(x') \geq f(x^{(k)})$, **then** $x^{(k+1)} := x'$ **else** $x^{(k+1)} := x^{(k)}$

Assume
- Fitness function f has unique maximum $x^* \in \{0,1\}^n$.
- Distance function is $g(x) = H(x, x^*)$

Then
- $|g(x^{(k+1)}) - g(x^{(k)})| \prec Z$ where $Z := H(x^{(k)}, x')$
- $Z \sim \mathrm{Bin}(n,p)$ so $\mathbb{E}\left[e^{\lambda Z}\right] \leq e^{np}$ for $\lambda = \ln(2)$

Proof overview

Theorem (2.3 in [6])

Assume that there exists $0 < \rho < 1$ and $D \geq 1$ such that

(D1) $\mathbb{E}\left[e^{\eta Y_{k+1}} \mid Y_k > a \wedge \mathscr{F}_k\right] \leq \rho e^{\eta Y_k}$

(D2) $\mathbb{E}\left[e^{\eta Y_{k+1}} \mid Y_k \leq a \wedge \mathscr{F}_k\right] \leq D e^{\eta a}$

Then

(2.6) $\mathbb{E}\left[e^{\eta Y_{k+1}} \mid \mathscr{F}_0\right] \leq \rho^k e^{\eta Y_0} + D e^{\eta a}(1-\rho^k)/(1-\rho)$.

(2.8) $\Pr\left(Y_k \geq b \mid \mathscr{F}_0\right) \leq \rho^k e^{\eta(Y_0 - b)} + D e^{\eta(a-b)}(1-\rho^k)/(1-\rho)$.

(*) $\Pr\left(\tau_b < B\right) \leq e^{\eta(a-b)} BD/(1-\rho)$

(2.9) $\Pr\left(\tau_a > k \mid \mathscr{F}_0\right) \leq e^{\eta(Y_0 - a)} \rho^k$

Lemma

Assume that there exists a $\varepsilon_0 > 0$ such that

(C1) $\mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > a \wedge \mathscr{F}_k\right] \leq -\varepsilon_0$

(C2) $(|Y_{k+1} - Y_k| \mid \mathscr{F}_k) \prec Z$ *and* $\mathbb{E}\left[e^{\lambda Z}\right] = D < \infty$ *for a* $\lambda > 0$.

then (D1) and (D2) hold for some η and $\rho < 1$

Simple Application (1+1) EA on Needle

(1+1) EA with mutation rate $p = 1/n$ on

$$\textsc{Needle}(x) := \prod_{i=1}^{n} x_i \qquad \begin{aligned} Y_k &:= H(x^{(k)}, 0^n) \\ a &:= (3/4)n \\ b &:= n \end{aligned}$$

Condition (C2) satisfied[6] with $D = \mathbb{E}\left[e^{\lambda Z}\right] \leq e$ where $\lambda = \ln(2)$.
Condition (C1) satisfied for $\varepsilon_0 := 1/2$ because

$$\mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > a \wedge \mathscr{F}_k\right] \leq (n-a)p - ap = -\varepsilon_0.$$

Thus, $\eta \geq \min\{\lambda, \delta\varepsilon_0\lambda^2/D\} > 1/25$ when $\delta = 1/2$ and

$$\Pr\left(\tau_a > n + k \mid \mathscr{F}_0\right) \leq e^{(1/25)(Y_0 - a - (n+k)(1-\delta)\varepsilon_0)} \leq e^{-k/100}$$
$$\Pr\left(\tau_b < e^{n/200} \mid \mathscr{F}_0\right) = e^{-\Omega(n)}$$

[6] See previous slide.

Theorem

(D1) $\mathbb{E}\left[e^{\eta(Y_{k+1} - Y_k)} \mid Y_k > a \wedge \mathscr{F}_k\right] \leq \rho$

(D2) $\mathbb{E}\left[e^{\eta(Y_{k+1} - a)} \mid Y_k \leq a \wedge \mathscr{F}_k\right] \leq D$

Assume that (D1) and (D2) hold. Then

(2.6) $\mathbb{E}\left[e^{\eta Y_{k+1}} \mid \mathscr{F}_0\right] \leq \rho^k e^{\eta Y_0} + D e^{\eta a}(1-\rho^k)/(1-\rho)$.

Proof.

By the law of total probability, and the conditions (D1) and (D2)

$$\mathbb{E}\left[e^{\eta Y_{k+1}} \mid \mathscr{F}_k\right] \leq \rho e^{\eta Y_k} + D e^{\eta a} \tag{1}$$

By the law of total expectation, Ineq. (1), and induction on k

$$\begin{aligned} \mathbb{E}\left[e^{\eta Y_{k+1}} \mid \mathscr{F}_0\right] &= \mathbb{E}\left[\mathbb{E}\left[e^{\eta Y_{k+1}} \mid \mathscr{F}_k\right] \mid \mathscr{F}_0\right] \\ &\leq \rho \mathbb{E}\left[e^{\eta Y_k} \mid \mathscr{F}_0\right] + D e^{\eta a} \\ &\leq \rho^k e^{\eta Y_0} + (1 + \rho + \rho^2 + \cdots + \rho^{k-1}) D e^{\eta a}. \end{aligned}$$

□

Proof of (2.8)

Theorem

(D1) $\mathbb{E}\left[e^{\eta(Y_{k+1}-Y_k)} \mid Y_k > a \wedge \mathscr{F}_k\right] \leq \rho$

(D2) $\mathbb{E}\left[e^{\eta(Y_{k+1}-a)} \mid Y_k \leq a \wedge \mathscr{F}_k\right] \leq D$

Assume that (D1) and (D2) hold. Then

(2.6) $\mathbb{E}\left[e^{\eta Y_{k+1}} \mid \mathscr{F}_0\right] \leq \rho^k e^{\eta Y_0} + De^{\eta a}(1-\rho^k)/(1-\rho)$.

(2.8) $\Pr(Y_k \geq b \mid \mathscr{F}_0) \leq \rho^k e^{\eta(Y_0-b)} + De^{\eta(a-b)}(1-\rho^k)/(1-\rho)$.

Proof.

(2.8) follows from Markov's inequality and (2.6)

$$\begin{aligned}\Pr(Y_{k+1} \geq b \mid \mathscr{F}_0) &= \Pr\left(e^{\eta Y_{k+1}} \geq e^{\eta b} \mid \mathscr{F}_0\right)\\ &\leq \mathbb{E}\left[e^{\eta Y_{k+1}} \mid \mathscr{F}_0\right] e^{-\eta b}\end{aligned}$$

□

Proof of (2.9)

Theorem

(D1) $\mathbb{E}\left[e^{\eta(Y_{k+1}-Y_k)} \mid Y_k > a \wedge \mathscr{F}_k\right] \leq \rho$

Assume that (D1) hold. Then

(2.9) $\Pr(\tau_a > k \mid \mathscr{F}_0) \leq e^{\eta(Y_0-a)}\rho^k$

Proof.

By (D1) $Z_k := e^{\eta Y_{k\wedge\tau}}\rho^{-k\wedge\tau}$ is a supermartingale, so

$$e^{\eta Y_0} = Z_0 \geq \mathbb{E}[Z_k \mid \mathscr{F}_0] = \mathbb{E}\left[e^{\eta Y_{k\wedge\tau}}\rho^{-k\wedge\tau} \mid \mathscr{F}_0\right] \qquad (2)$$

By (2) and the law of total probability

$$\begin{aligned}e^{\eta Y_0} &\geq \Pr(\tau_a > k \mid \mathscr{F}_0)\,\mathbb{E}\left[e^{\eta Y_{k\wedge\tau}}\rho^{-k\wedge\tau} \mid \tau_a > k \wedge \mathscr{F}_0\right]\\ &= \Pr(\tau_a > k \mid \mathscr{F}_0)\,\mathbb{E}\left[e^{\eta Y_k}\rho^{-k} \mid \tau_a > k \wedge \mathscr{F}_0\right]\\ &\geq \Pr(\tau_a > k \mid \mathscr{F}_0)\, e^{\eta a}\rho^{-k}\end{aligned}$$

□

Proof of (*)

Theorem

(D1) $\mathbb{E}\left[e^{\eta(Y_{k+1}-Y_k)} \mid Y_k > a \wedge \mathscr{F}_k\right] \leq \rho$

(D2) $\mathbb{E}\left[e^{\eta(Y_{k+1}-a)} \mid Y_k \leq a \wedge \mathscr{F}_k\right] \leq D$

Assume that (D1) and (D2) hold for $D \geq 1$. Then

(2.8) $\Pr(Y_k \geq b \mid \mathscr{F}_0) \leq \rho^k e^{\eta(Y_0-b)} + De^{\eta(a-b)}(1-\rho^k)/(1-\rho)$.

(*) $\Pr(\tau_b < B) \leq e^{\eta(a-b)}BD/(1-\rho)$

Proof.

By the union bound and (2.8)

$$\begin{aligned}\Pr(\tau_b < B \mid Y_0 < a \wedge \mathscr{F}_0) &\leq \sum_{k=1}^{B} \Pr(Y_k \geq b \mid Y_0 < a \wedge \mathscr{F}_0)\\ &\leq \sum_{k=1}^{B} De^{\eta(a-b)}\left(\rho^k + \frac{1-\rho^k}{1-\rho}\right) \leq \frac{BDe^{\eta(a-b)}}{1-\rho}\end{aligned}$$

□

(C1) and (C2) $\Longrightarrow$ (D1)

(C1) $\mathbb{E}[Y_{k+1} - Y_k \mid Y_k > a \wedge \mathscr{F}_k] \leq -\varepsilon_0$

(C2) $(|Y_{k+1} - Y_k| \mid \mathscr{F}_k) \prec Z$ and $\mathbb{E}\left[e^{\lambda Z}\right] = D < \infty$ for a $\lambda > 0$.

(D1) $\mathbb{E}\left[e^{\eta(Y_{k+1}-Y_k)} \mid Y_k > a \wedge \mathscr{F}_k\right] \leq \rho$

Lemma

Assume (C1) and (C2). Then (D1) holds when $\rho \geq 1 - \eta\varepsilon_0 + \eta^2 c$, and $0 < \eta \leq \min\{\lambda, \varepsilon_0/c\}$ where $c := \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{k!}\mathbb{E}\left[Z^k\right]$.

Proof.

Let $X := (Y_{k+1} - Y_k \mid Y_k > a \wedge \mathscr{F}_k)$.
By (C2) it holds, $|X| \prec Z$, so $\mathbb{E}\left[X^k\right] \leq \mathbb{E}\left[|X|^k\right] \leq \mathbb{E}\left[Z^k\right]$.
From $e^x = \sum_{k=0}^{\infty} x^k/(k!)$ and linearity of expectation

$$0 < \mathbb{E}\left[e^{\eta X}\right] = 1 + \eta\mathbb{E}[X] + \sum_{k=2}^{\infty}\frac{\eta^k}{k!}\mathbb{E}\left[X^k\right] \leq \rho.$$

□

(C2) $\implies$ (D2)

(C2) $(|Y_{k+1} - Y_k| \mid \mathscr{F}_k) \prec Z$ and $\mathbb{E}\left[e^{\lambda Z}\right] = D < \infty$ for a $\lambda > 0$.

(D2) $\mathbb{E}\left[e^{\eta(Y_{k+1}-a)} \mid Y_k \le a \wedge \mathscr{F}_k\right] \le D$

Theorem

Assume (C2) and $0 < \eta \le \lambda$. Then (D2) holds.

Proof.

If $Y_k \le a$ then $Y_{k+1} - a \le Y_{k+1} - Y_k \le |Y_{k+1} - Y_k|$, so

$$\mathbb{E}\left[e^{\eta(Y_{k+1}-a)} \mid Y_k \le a \wedge \mathscr{F}_k\right] \le \mathbb{E}\left[e^{\lambda|Y_{k+1}-Y_k|} \mid Y_k \le a \wedge \mathscr{F}_k\right]$$

Furthermore, by (C2)

$$\mathbb{E}\left[e^{\lambda|Y_{k+1}-Y_k|} \mid Y_k \le a \wedge \mathscr{F}_k\right] \le \mathbb{E}\left[e^{\lambda Z}\right] = D.$$

$\square$

(C1) and (C2) $\implies$ (D1) and (D2)

(C1) $\mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > a \wedge \mathscr{F}_k\right] \le -\varepsilon_0$

(C2) $(|Y_{k+1} - Y_k| \mid \mathscr{F}_k) \prec Z$ and $\mathbb{E}\left[e^{\lambda Z}\right] = D < \infty$ for a $\lambda > 0$.

(D1) $\mathbb{E}\left[e^{\eta(Y_{k+1}-Y_k)} \mid Y_k > a \wedge \mathscr{F}_k\right] \le \rho$

(D2) $\mathbb{E}\left[e^{\eta(Y_{k+1}-a)} \mid Y_k \le a \wedge \mathscr{F}_k\right] \le D$

Lemma

Assume (C1) and (C2). Then (D1) and (D2) hold when

$$\rho \ge 1 - \eta\varepsilon_0 + \eta^2 c \quad \textit{and} \quad 0 < \eta \le \min\{\lambda, \varepsilon_0/c\}$$

where $c := \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{k!}\mathbb{E}\left[Z^k\right] = (D - 1 - \lambda\mathbb{E}[Z])\lambda^{-2} > 0.$

Corollary

Assume (C1), (C2) and $0 < \delta < 1$. Then (D1) and (D2) hold for

$$\eta := \min\{\lambda, \delta\varepsilon_0/c\} \implies \delta\varepsilon_0 \ge \eta c$$
$$\rho := 1 - (1-\delta)\eta\varepsilon_0 = 1 - \eta\varepsilon_0 + \eta\delta\varepsilon_0 \ge 1 - \eta\varepsilon + \eta^2 c$$

Reformulation of Hajek's Theorem

Theorem

If there exist $\lambda, \varepsilon > 0$ and $1 < D < \infty$ such that for all $k \ge 0$

(C1) $\mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > a \wedge \mathscr{F}_k\right] \le -\varepsilon_0$

(C2) $(|Y_{k+1} - Y_k| \mid \mathscr{F}_k) \prec Z$ *and* $\mathbb{E}\left[e^{\lambda Z}\right] = D$

then for any $\delta \in (0,1)$

(2.9) $\Pr\left(\tau_a > B \mid \mathscr{F}_0\right) \le e^{\eta(Y_0-a)}\rho^B$

(*) $\Pr\left(\tau_b < B \mid \mathscr{F}_0\right) \le \frac{BD}{(1-\rho)} \cdot e^{\eta(a-b)}$

where $\eta := \min\{\lambda, \delta\varepsilon_0/c\}$ and $\rho := 1 - (1-\delta)\eta\varepsilon_0$

1. Note that $\ln(\rho) \le \rho - 1$ for all $\rho \ge 0$, so

$$\begin{aligned}\Pr\left(\tau_a > B \mid \mathscr{F}_0\right) &\le e^{\eta(Y_0-a)}\rho^B = e^{\eta(Y_0-a)}e^{B\ln(\rho)} \\ &\le e^{\eta(Y_0-a-B(1-\delta)\varepsilon_0)}.\end{aligned}$$

2. $c = (D - 1 - \lambda\mathbb{E}[Z])\lambda^{-2} < D/\lambda^2$ so $\eta \ge \min\{\lambda, \delta\varepsilon_0\lambda^2/D\}$.

Reformulation of Hajek's Theorem

Theorem

If there exist $\lambda, \varepsilon > 0$ and $1 < D < \infty$ such that for all $k \ge 0$

(C1) $\mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > a \wedge \mathscr{F}_k\right] \le -\varepsilon_0$

(C2) $(|Y_{k+1} - Y_k| \mid \mathscr{F}_k) \prec Z$ *and* $\mathbb{E}\left[e^{\lambda Z}\right] = D$

then for any $\delta \in (0,1)$

(2.9) $\Pr\left(\tau_a > B \mid \mathscr{F}_0\right) \le e^{\eta(Y_0-a-B(1-\delta)\varepsilon_0)}$

(*) $\Pr\left(\tau_b < B \mid \mathscr{F}_0\right) \le \frac{BD}{(1-\delta)\eta\varepsilon_0} \cdot e^{\eta(a-b)}$

for some $\eta \ge \min\{\lambda, \delta\varepsilon_0\lambda^2/D\}$.

Simplified Drift Theorem [17]

We have already seen that

(C2) $(|Y_{k+1} - Y_k| \mid \mathscr{F}_k) \prec Z$ and $\mathbb{E}\left[e^{\lambda Z}\right] = D$

implies $\Pr\left(|Y_{k+1} - Y_k| \geq j\right) \leq De^{-\lambda j}$ for all $j \in \mathbb{N}_0$.

The simplified drift theorem replaces (C2) with

(S) $\Pr\left(Y_{k+1} - Y_k \geq j \mid Y_k < b\right) \leq r(n)(1+\delta)^{-j}$ for all $j \in \mathbb{N}_0$.

and with some additional assumptions, provides a bound of type[7]

$$\Pr\left(\tau_b < 2^{c(b-a)}\right) \leq 2^{-\Omega(b-a)}. \qquad (3)$$

- Until 2008, conditions (D1) and (D2) were used in EC.
- (D1) and (D2) can lead to highly tedious calculations.
- Oliveto and Witt were the first to point out that the much simpler to verify (C1), along with (S) is sufficient.

[7] See [17] for the exact statement.

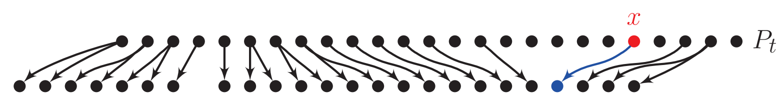
Part 6 - Population Drift

Drift Analysis of Population-based Evolutionary Algorithms

- Evolutionary algorithms generally use populations.
- So far, we have analysed the drift of the (1+1) EA, ie an evolutionary algorithm with population size one.
- The state aggregation problem makes analysis of population-based EAs with classical drift theorems difficult: How to define an appropriate distance function?
  - Should reflect the progress of the algorithm
  - Often hard to define for single-individual algorithms
  - Highly non-trivial for population-based algorithms

$\Longrightarrow$ This part of the tutorial focuses on a drift theorem for populations which alleviates the state aggregation problem.

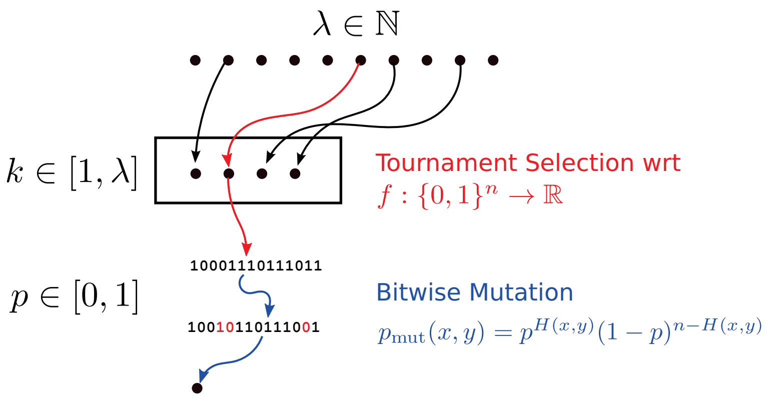
Population-based Evolutionary Algorithms



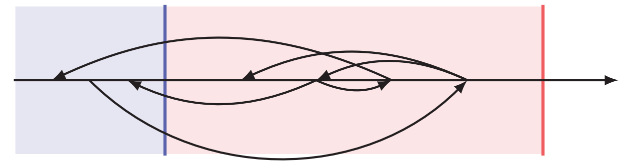
Require: ,
  Finite set $\mathcal{X}$, and initial population $P_0 \in \mathcal{X}^\lambda$
  Selection mechanism $p_{\mathsf{sel}} : \mathcal{X}^\lambda \times \mathcal{X} \to [0,1]$
  Variation operator $p_{\mathsf{mut}} : \mathcal{X} \times \mathcal{X} \to [0,1]$

  for $t = 0, 1, 2, \ldots$ until termination condition **do**
    for $i = 1$ to λ **do**
      Sample i-th parent x according to $p_{\mathsf{sel}}(P_t, \cdot)$
      Sample i-th offspring $P_{t+1}(i)$ according to $p_{\mathsf{mut}}(x, \cdot)$
    end for
  end for

Selection and Variation - Example



Population Drift



Central Parameters

- Reproductive rate of selection mechanism p_{sel}

$$\alpha_0 = \max_{1 \leq j \leq \lambda} \mathbb{E}\left[\#\text{offspring from parent } j\right],$$

- Random walk process corresponding to variation operator p_{mut}

$$X_{k+1} \sim p_{\mathsf{mut}}(X_k)$$

Population Drift [12]

(C1P) $\forall k \quad \mathbb{E}\left[e^{\kappa(g(X_{k+1})-g(X_k))} \mid a < g(X_k) < b\right] < 1/\alpha_0$

Theorem

Define $\tau_b := \min\{k \geq 0 \mid g(P_k(i)) > b$ *for some* $i \in [\lambda]\}$.

If there exists constants $\alpha_0 \geq 1$ *and* $\kappa > 0$ *such that*

- p_{sel} *has reproductive rate less than* α_0
- *the random walk process corresponding to* p_{mut} *satisfies (C1P)*

and some other conditions hold,[8] *then for some constants* $c, c' > 0$

$$\Pr\left(\tau_b \leq e^{c(b-a)}\right) = e^{-c'(b-a)}$$

[8] Some details are omitted. See Theorem 1 in [12] for all details.

Population Drift: Decoupling Selection & Variation

Population drift
If there exists a $\kappa > 0$ such that

$$M_{\Delta_{\mathsf{mut}}}(\kappa) < 1/\alpha_0$$

where

$$\Delta_{\mathsf{mut}} = g(X_{k+1}) - g(X_k)$$
$$X_{k+1} \sim p_{\mathsf{mut}}(X_k)$$

and

$$\alpha_0 = \max_j \mathbb{E}\left[\#\text{offspring from parent } j\right],$$

then the runtime is exponential.

Classical drift [6]
If there exists a $\kappa > 0$ such that

$$M_{\Delta}(\kappa) < 1$$

where

$$\Delta = h(P_{k+1}) - h(P_k),$$

then the runtime is exponential.

Conclusion

- Drift analysis is a powerful tool for analysis of EAs
  - Mainly used in EC to bound the expected runtime of EAs
  - Useful when the EA has non-monotonic progress, eg. when the fitness value is a poor indicator of progress
- The "art" consists in finding a good distance function
  - No simple receipe
- A large number of drift theorems are available
  - Additive, multiplicative, variable, population drift...
  - Significant related literature from other fields than EC
- Not the only tool in the toolbox, also
  - Artificial fitness levels, Markov Chain theory, Concentration of measure, Branching processes, Martingale theory, Probability generating functions, ...

Acknowledgements

Thanks to

- David Hodge
- Carsten Witt, and
- Daniel Johannsen

for insightful discussions.

References I

[1] Benjamin Doerr and Leslie Ann Goldberg.
Drift analysis with tail bounds.
In *Proceedings of the 11th international conference on Parallel problem solving from nature: Part I*, PPSN'10, pages 174–183, Berlin, Heidelberg, 2010. Springer-Verlag.

[2] Benjamin Doerr, Daniel Johannsen, and Carola Winzen.
Multiplicative drift analysis.
In *GECCO '10: Proceedings of the 12th annual conference on Genetic and evolutionary computation*, pages 1449–1456, New York, NY, USA, 2010. ACM.

[3] Stefan Droste, Thomas Jansen, and Ingo Wegener.
On the analysis of the (1+1) Evolutionary Algorithm.
Theoretical Computer Science, 276:51–81, 2002.

[4] Simon Fischer, Lars Olbrich, and Berthold Vöcking.
Approximating wardrop equilibria with finitely many agents.
Distributed Computing, 21(2):129–139, 2008.

[5] Oliver Giel and Ingo Wegener.
Evolutionary algorithms and the maximum matching problem.
In *Proceedings of the 20th Annual Symposium on Theoretical Aspects of Computer Science (STACS 2003)*, pages 415–426, 2003.

References II

[6] Bruce Hajek.
Hitting-time and occupation-time bounds implied by drift analysis with applications.
Advances in Applied Probability, 14(3):502–525, 1982.

[7] Jun He and Xin Yao.
Drift analysis and average time complexity of evolutionary algorithms.
Artificial Intelligence, 127(1):57–85, March 2001.

[8] Jun He and Xin Yao.
A study of drift analysis for estimating computation time of evolutionary algorithms.
Natural Computing, 3(1):21–35, 2004.

[9] Jens Jägersküpper.
Algorithmic analysis of a basic evolutionary algorithm for continuous optimization.
Theoretical Computer Science, 379(3):329–347, 2007.

[10] Jens Jägersküpper.
A Blend of Markov-Chain and Drift Analysis.
In *Proceedings of the 10th International Conference on Parallel Problem Solving from Nature (PPSN 2008)*, 2008.

References III

[11] Daniel Johannsen.
Random combinatorial structures and randomized search heuristics.
PhD thesis, Universität des Saarlandes, 2010.

[12] Per Kristian Lehre.
Negative drift in populations.
In *Proceedings of Parallel Problem Solving from Nature - (PPSN XI)*, volume 6238 of *LNCS*, pages 244–253. Springer Berlin / Heidelberg, 2011.

[13] Per Kristian Lehre and Carsten Witt.
Black-box search by unbiased variation.
Algorithmica, pages 1–20, 2012.

[14] Sean P. Meyn and Richard L. Tweedie.
Markov Chains and Stochastic Stability.
Springer-Verlag, 1993.

[15] B. Mitavskiy, J. E. Rowe, and C. Cannings.
Theoretical analysis of local search strategies to optimize network communication subject to preserving the total number of links.
International Journal of Intelligent Computing and Cybernetics, 2(2):243–284, 2009.

[16] Frank Neumann, Dirk Sudholt, and Carsten Witt.
Analysis of different mmas aco algorithms on unimodal functions and plateaus.
Swarm Intelligence, 3(1):35–68, 2009.

References IV

[17] Pietro Oliveto and Carsten Witt.
Simplified drift analysis for proving lower bounds inevolutionary computation.
Algorithmica, pages 1–18, 2010.
10.1007/s00453-010-9387-z.

[18] Pietro S. Oliveto and Carsten Witt.
Simplified drift analysis for proving lower bounds in evolutionary computation.
Technical Report Reihe CI, No. CI-247/08, SFB 531, Technische Universität Dortmund, Germany, 2008.

[19] Galen H. Sasaki and Bruce Hajek.
The time complexity of maximum matching by simulated annealing.
Journal of the ACM, 35(2):387–403, 1988.

[20] Dirk Sudholt.
General lower bounds for the running time of evolutionary algorithms.
In *Proceedings of Parallel Problem Solving from Nature - (PPSN XI)*, volume 6238 of *LNCS*, pages 124–133. Springer Berlin / Heidelberg, 2010.

[21] David Williams.
Probability with Martingales.
Cambridge University Press, 1991.

References V

[22] Carsten Witt.
Optimizing linear functions with randomized search heuristics - the robustness of mutation.
In Christoph Dürr and Thomas Wilke, editors, *29th International Symposium on Theoretical Aspects of Computer Science (STACS 2012)*, volume 14 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 420–431, Dagstuhl, Germany, 2012. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.