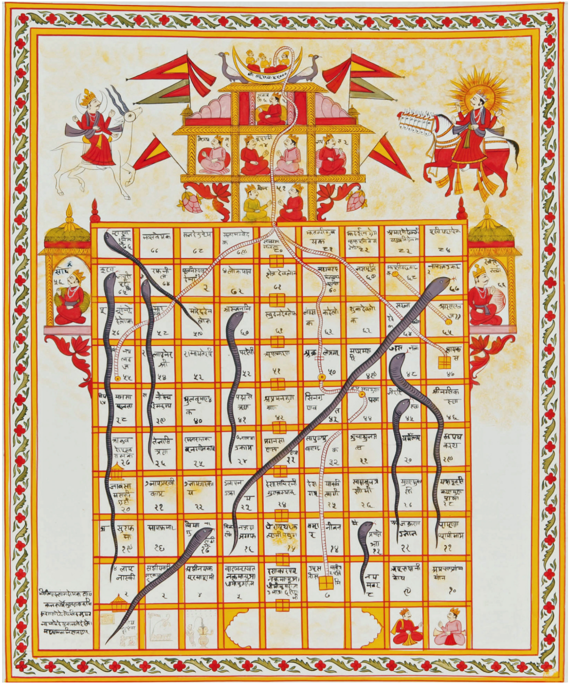
Drift Analysis - A Tutorial

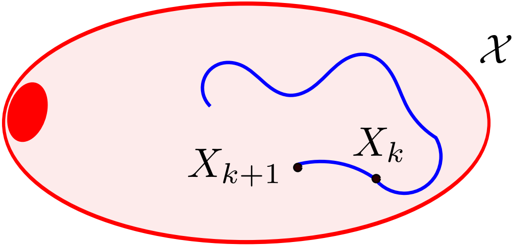
Per Kristian Lehre

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

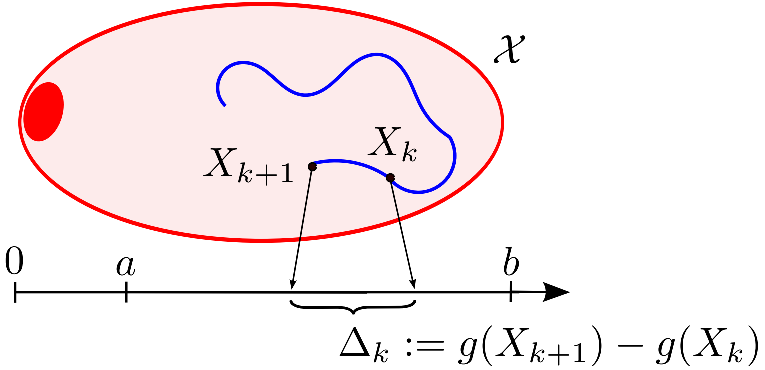
July 6, 2012



What is Drift Analysis?



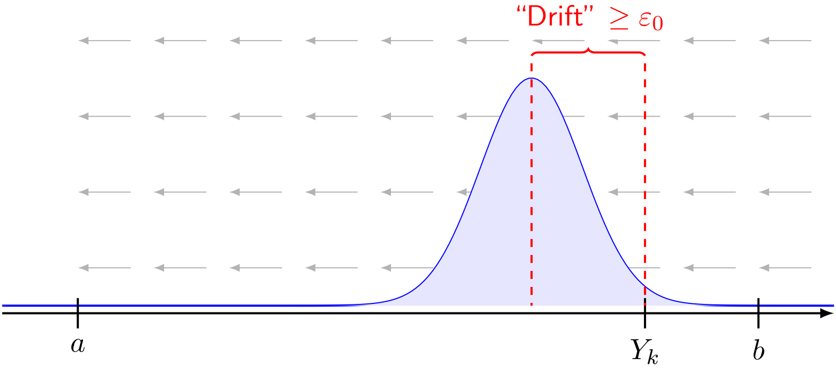
What is Drift Analysis?



- 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]?

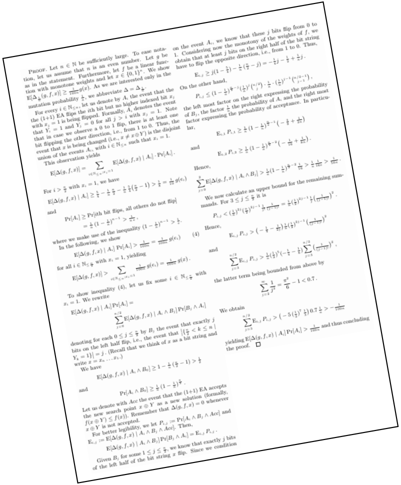


[1] NB! Drift is a different concept than *genetic drift* in evolutionary biology.

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

Drift Analysis of Evolutionary Algorithms

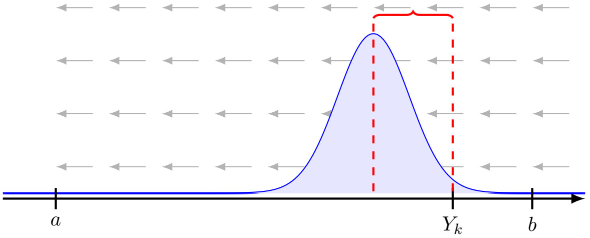
- Introduced to EC in 2001 by He and Yao [7, 8]
  - (1+1) EA on linear functions: $O(n \ln n)$ [7]
  - (1+1) EA on maximum matching by Giel and Wegener [5]
- Simplified drift in 2008 by Oliveto and Witt [18]
- Multiplicative drift by Doerr et al [2]
  - (1+1) EA on linear functions: $en\ln(n) + O(n)$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



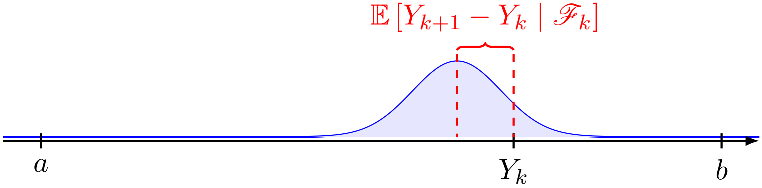
- X_k a stochastic process[2] in some general state space $\mathcal{X}$
- $Y_k := g(X_k)$ were $g : \mathcal{X} \to \mathbb{R}$ is a "distance function"
- Two stopping times τ_a and τ_b

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

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

[2] not necessarily Markovian.

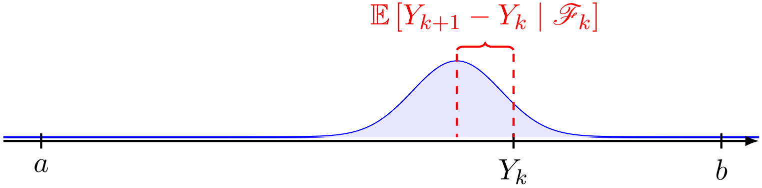
Overview of Tutorial



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]

[3]Some drift theorems need additional conditions.

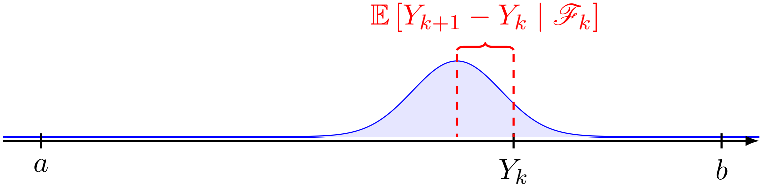
Overview of Tutorial



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]
$\mathbb{E}\left[Y_{k+1} \mid \mathscr{F}_k\right] \geq Y_k - \varepsilon_0$	$\leq \mathbb{E}\left[\tau_a\right]$	Additive drift (lower b.) [7, 9]

[3] Some drift theorems need additional conditions.

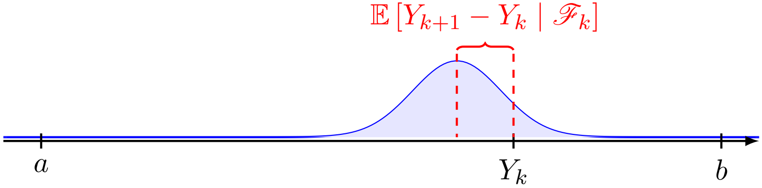
Overview of Tutorial



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]
$\mathbb{E}\left[Y_{k+1} \mid \mathscr{F}_k\right] \geq Y_k - \varepsilon_0$	$\leq \mathbb{E}\left[\tau_a\right]$	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]

[3]Some drift theorems need additional conditions.

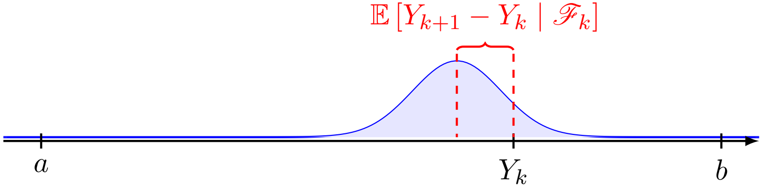
Overview of Tutorial



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]
$\mathbb{E}\left[Y_{k+1} \mid \mathscr{F}_k\right] \geq Y_k - \varepsilon_0$	$\leq \mathbb{E}\left[\tau_a\right]$	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]
	$\Pr\left(\tau_a > B_3\right) \leq$	[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]

[3] Some drift theorems need additional conditions.

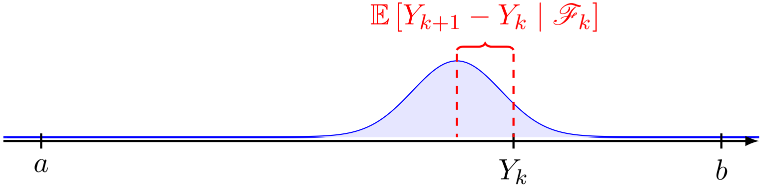
Overview of Tutorial



Drift Condition[3]	Statement	Note
$\mathbb{E}[Y_{k+1} \mid \mathscr{F}_k] \leq Y_k - \varepsilon_0$	$\mathbb{E}[\tau_a] \leq$	Additive drift [7, 10]
$\mathbb{E}[Y_{k+1} \mid \mathscr{F}_k] \geq Y_k - \varepsilon_0$	$\leq \mathbb{E}[\tau_a]$	Additive drift (lower b.) [7, 9]
$\mathbb{E}[Y_{k+1} \mid \mathscr{F}_k] \leq Y_k$	$\mathbb{E}[\tau_a] \leq$	Supermartingale [16]
$\mathbb{E}[Y_{k+1} \mid \mathscr{F}_k] \leq (1-\delta)Y_k$	$\mathbb{E}[\tau_a] \leq$	Multiplicative drift [2, 4]
	$\Pr(\tau_a > B_3) \leq$	[1]
$\mathbb{E}[Y_{k+1} \mid \mathscr{F}_k] \geq (1-\delta)Y_k$	$\leq \mathbb{E}[\tau_a]$	Multipl. drift (lower b.) [13]
$\mathbb{E}[Y_{k+1} \mid \mathscr{F}_k] \leq Y_k - h(Y_k)$	$\mathbb{E}[\tau_a] \leq$	Variable drift [11]

[3] Some drift theorems need additional conditions.

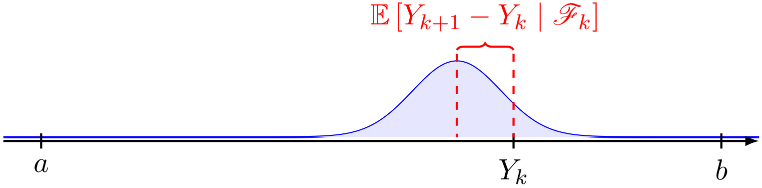
Overview of Tutorial



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]
	$\Pr\left(\tau_a > B_1\right) \leq$	[6]
	$\Pr\left(\tau_b < B_2\right) \leq$	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]
	$\Pr\left(\tau_a > B_3\right) \leq$	[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]

[3] Some drift theorems need additional conditions.

Overview of Tutorial

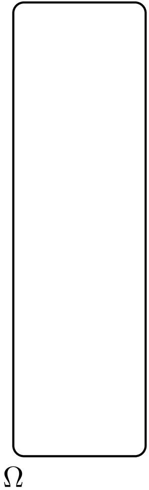


Drift Condition[3]	Statement	Note
$\mathbb{E}\left[Y_{k+1} \mid \mathscr{F}_k\right] \leq Y_k - \varepsilon_0$	$\mathbb{E}\left[\tau_a\right] \leq$	Additive drift [7, 10]
	$\Pr\left(\tau_a > B_1\right) \leq$	[6]
	$\Pr\left(\tau_b < B_2\right) \leq$	Simplified drift [6, 17]
$\mathbb{E}\left[Y_{k+1} \mid \mathscr{F}_k\right] \geq Y_k - \varepsilon_0$	$\leq \mathbb{E}\left[\tau_a\right]$	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]
	$\Pr\left(\tau_a > B_3\right) \leq$	[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

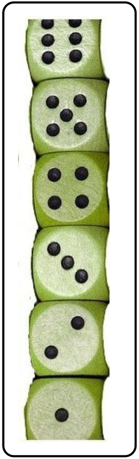
Basic Probability Theory



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

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

Basic Probability Theory

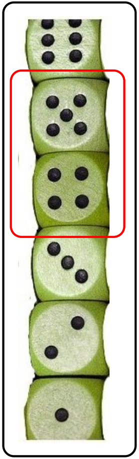


Ω

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

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

Basic Probability Theory



Ω

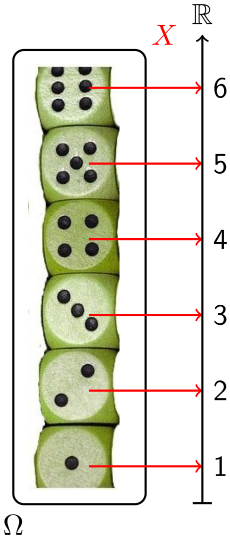
Probability Triple $(\Omega, \mathscr{F}, \Pr)$

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

Events

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

Basic Probability Theory



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

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

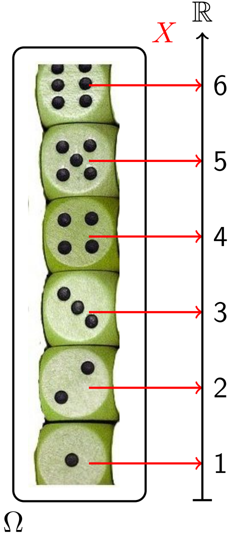
Events

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

Random Variable

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

Basic Probability Theory



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

- Ω : Sample space
- $\mathscr{F}$: σ-algebra (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 $X^{-1} : \mathcal{B} \to \mathscr{F}$
- $X = y \iff \{\omega \in \Omega \mid X(\omega) = y\}$

Expectation

- $\mathbb{E}[X] := \sum_y y \Pr(X = y)$.

Conditional Expectation

$$\Pr(X = x \mid \mathcal{E}) = \frac{\Pr(X = x \wedge \mathcal{E})}{\Pr(\mathcal{E})}$$

Conditional Expectation

$$\mathbb{E}\left[X \mid \mathcal{E}\right] := \sum_x x \Pr\left(X = x \mid \mathcal{E}\right) = \sum_x x \frac{\Pr\left(X = x \wedge \mathcal{E}\right)}{\Pr\left(\mathcal{E}\right)}$$

$\mathbb{E}\left[X \mid Z = z\right]$

Conditional Expectation

$$\mathbb{E}\left[X \mid \mathcal{E}\right] := \sum_x x \Pr\left(X = x \mid \mathcal{E}\right) = \sum_x x \frac{\Pr\left(X = x \wedge \mathcal{E}\right)}{\Pr\left(\mathcal{E}\right)}$$

$\mathbb{E}\left[X \mid Z = z\right]$

$\mathbb{E}\left[X \mid Z\right](\omega) = \mathbb{E}\left[X \mid Z = z\right]$, where $z = Z(\omega)$

Conditional Expectation

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

$\mathbb{E}[X \mid Z = z]$

$\mathbb{E}[X \mid Z](\omega) = \mathbb{E}[X \mid Z = z]$, where $z = Z(\omega)$

Definition

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

1. Y is $\mathscr{G}$-*measurable*, ie, $Y^{-1}(A) \in \mathscr{G}$ for all $A \in \mathcal{B}$
2. $\mathbb{E}[|Y|] < \infty$
3. $\mathbb{E}[Y I_F] = \mathbb{E}[X I_F]$ for all $F \in \mathscr{G}$

Three Properties

- $\mathbb{E}[a_1 X_1 + a_2 X_2 \mid \mathscr{G}] = a_1 \mathbb{E}[X_1 \mid \mathscr{G}] + a_2 \mathbb{E}[X_2 \mid \mathscr{G}]$
- If X is $\mathscr{G}$-measurable then $\mathbb{E}[X \mid \mathscr{G}] = X$
- If $\mathscr{H} \subset \mathscr{G}$, then $\mathbb{E}[\mathbb{E}[X \mid \mathscr{G}] \mid \mathscr{H}] = \mathbb{E}[X \mid \mathscr{H}]$

Stochastic Processes and Filtration

Definition

- A *stochastic process* is a sequence of rv $Y_1, Y_2, \ldots$
- A *filtration* is an increasing family of sub-σ-algebras 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

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

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 decide whether the event $\{\tau \leq k\}$ is true or not.

Example

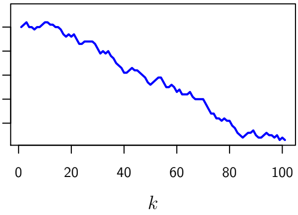
- The smallest k such that $Y_k < a$ in a stochastic process.
- The runtime of an evolutionary algorithm

Martingales

Definition (Supermartingale)

Any process Y st $\forall k$

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

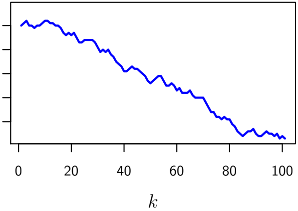


Martingales

Definition (Supermartingale)

Any process Y st $\forall k$

1. Y is adapted to $\mathscr{F}$
2. $\mathbb{E}[|Y_k|] < \infty$
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

$$Y_k := \Delta_1 + \cdots + \Delta_k$$

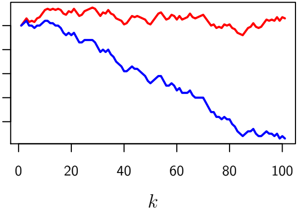
$$\begin{aligned}
\mathbb{E}[Y_{k+1} \mid \mathscr{F}_k] &= \Delta_1 + \cdots + \Delta_k + \mathbb{E}[\Delta_{k+1} \mid \mathscr{F}_k] \\
&\leq \Delta_1 + \cdots + \Delta_k - \varepsilon_0 \qquad < Y_k
\end{aligned}$$

Martingales

Definition (Supermartingale)

Any process Y st $\forall k$

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



Example

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

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

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

Martingales

Lemma

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

Martingales

Lemma

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

Proof.

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

Martingales

Lemma

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

Proof.

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

Example

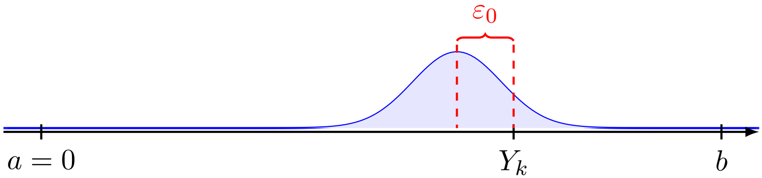
Where is the process Y in the previous example after k steps?

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

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

Part 2 - Additive Drift

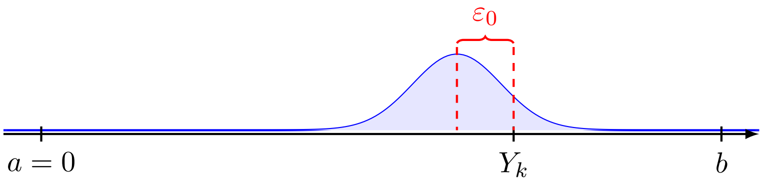
Additive Drift



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

Additive Drift



(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 ([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* $\mathbb{E}\left[\tau \mid \mathscr{F}_0\right] < \infty$*.*

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

Obtaining Supermartingales from Drift Conditions

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

- Y_k not necessarily a supermartingale, because (C1) assumes $Y_k > a$

Obtaining Supermartingales from Drift Conditions

Definition (Stopped Process)

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

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

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

- Y_k not necessarily a supermartingale, because (C1) assumes $Y_k > a$
- But the "stopped process" $Y_{k \wedge \tau_a}$ is a supermartingale, so

$$\forall k \quad Y_0 \geq \mathbb{E}\left[Y_{k \wedge \tau_a} \mid \mathscr{F}_0\right]$$
$$\forall k \quad Y_0 \geq \mathbb{E}\left[Y_{k \wedge \tau_a} + (k \wedge \tau_a)\varepsilon_0 \mid \mathscr{F}_0\right]$$

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* $\mathbb{E}\left[\tau \mid \mathscr{F}_0\right] < \infty$*.*

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

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* $\mathbb{E}\left[\tau \mid \mathscr{F}_0\right] < \infty$.

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

Proof.

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

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

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

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

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* $\mathbb{E}\left[\tau \mid \mathscr{F}_0\right] < \infty$*.*

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

Proof.

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

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

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

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

Examples: (1+1) EA

Algorithm 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

Law of Total Probability

$$\mathbb{E}\left[X\right] = \Pr\left(\mathcal{E}\right)\mathbb{E}\left[X \mid \mathcal{E}\right] + \Pr\left(\overline{\mathcal{E}}\right)\mathbb{E}\left[X \mid \overline{\mathcal{E}}\right]$$

Example 1: (1+1) EA on LeadingOnes

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

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

Left-most 0-bit

Example 1: (1+1) EA on LeadingOnes

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

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

Left-most 0-bit

- Let $Y_k := n - \textsc{Lo}(x^{(k)})$ be the "remaining" bits in step $k \geq 0$.

Example 1: (1+1) EA on LeadingOnes

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

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

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.

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

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

Example 1: (1+1) EA on LeadingOnes

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

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

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

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

- 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

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

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

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

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

$$\mathbb{E}\left[Y_{k+1} - Y_k \mid \mathcal{E}_i \wedge \mathscr{F}_k \right]$$

$$\leq \qquad w_i \left(x_i^{(k)} - 1 \right)$$

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

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

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

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 $c(n) \rightarrow (e-1)/2$ [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])

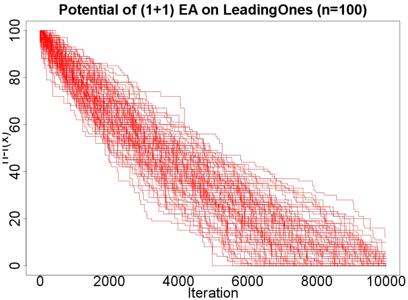
Assume Y is a homogeneous Markov chain, and τ the time to absorption. Then the function $g(x) := \mathbb{E}[\tau \mid Y_0 = x]$, satisfies

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

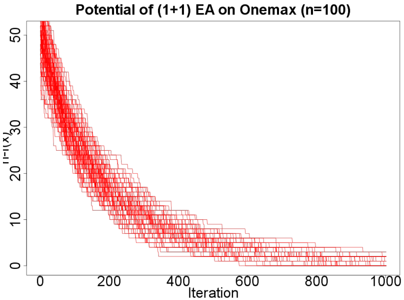
- Distance function g gives exact expected runtime!
- But g requires complete knowledge of the expected runtime!
- Still provides insight into what is a good distance function:
  - a good approximation (or guess) for the remaining runtime

Part 3 - Variable Drift

Drift may be Position-Dependant



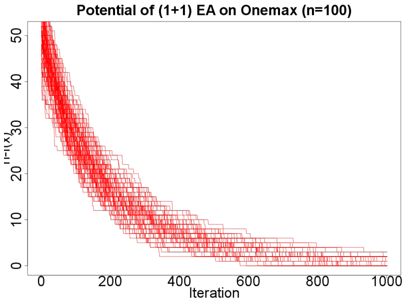
Constant Drift



Variable Drift

Drift may be Position-Dependant

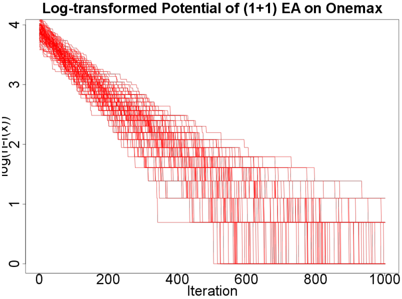
Idea: Find a function $g : \mathbb{R} \rightarrow \mathbb{R}$ st. the transformed stochastic process $g(X_1), g(X_2), g(X_3), \ldots$ has constant drift.



Variable Drift

Drift may be Position-Dependant

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



Constant Drift

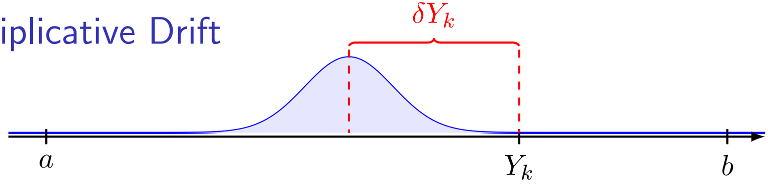
Jensen's Inequality

Theorem

If $g : \mathbb{R} \to \mathbb{R}$ *concave, then* $\mathbb{E}\left[g(X) \mid \mathscr{F}\right] \leq g(\mathbb{E}\left[X \mid \mathscr{F}\right])$*.*

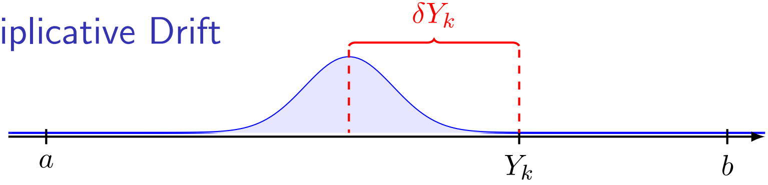
- If $g''(x) < 0$, then g is concave.

Multiplicative Drift



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

Multiplicative Drift



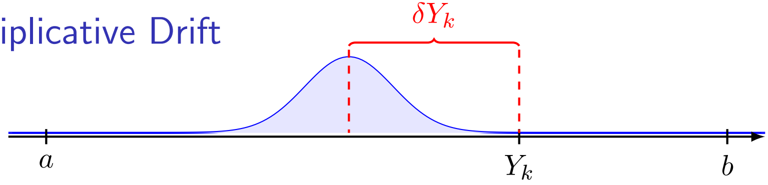
$$\text{(M)} \quad \forall k \quad \mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > a \wedge \mathscr{F}_k\right] \leq -\delta Y_k$$

Theorem ([2, 4])

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

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

Multiplicative Drift



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

Theorem ([2, 4])

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

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

Proof.

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

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

Example: Linear Functions Revisited

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

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

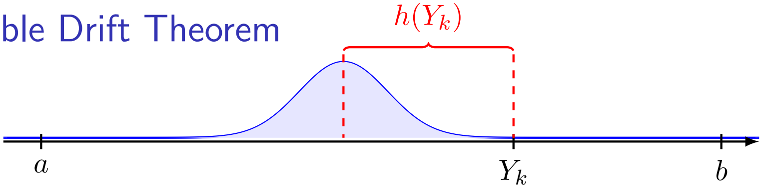
- We have already seen that

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

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

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

Variable Drift Theorem



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

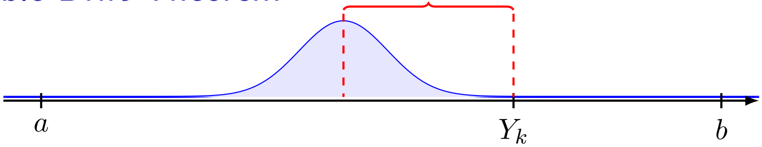
Theorem ([15, 11])

Given a sequence $(Y_k, \mathscr{F}_k)$ over an interval $[a,b] \subset \mathbb{R}, a > 0$. Define $\tau_a := \min\{k \geq 0 \mid Y_k = a\}$, 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*

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

Variable Drift Theorem



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

Theorem ([15, 11])

Given a sequence $(Y_k, \mathscr{F}_k)$ *over an interval* $[a, b] \subset \mathbb{R}, a > 0$.
Define $\tau_a := \min\{k \geq 0 \mid Y_k = a\}$, *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*

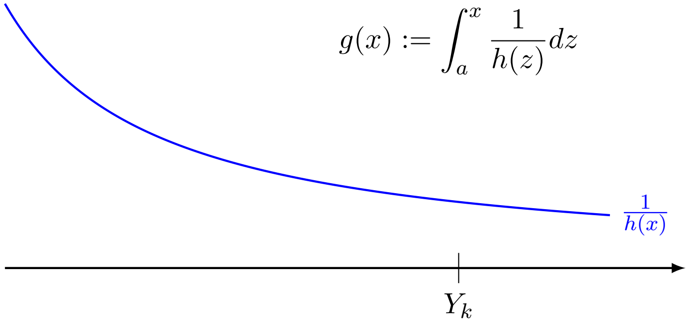
- $h(x) > 0$ *and* $h'(x) > 0$ *for all* $x \in [a, b]$, *and*
- *drift condition (V) holds, then*

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

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

Variable Drift Theorem: Proof

$$g(x) := \int_a^x \frac{1}{h(z)} dz$$

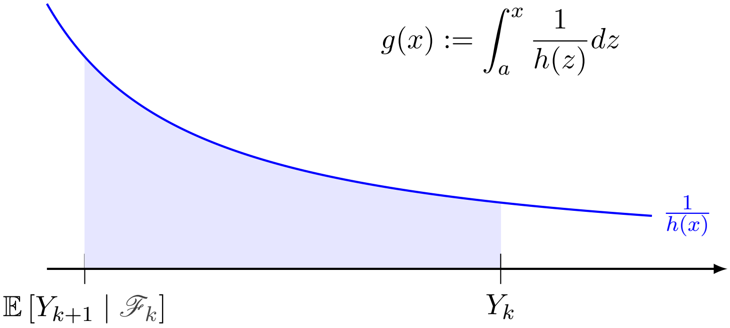


Proof.

$$\mathbb{E}\left[g(Y_k) - g(Y_{k+1}) \mid \mathscr{F}_k\right]$$

□

Variable Drift Theorem: Proof



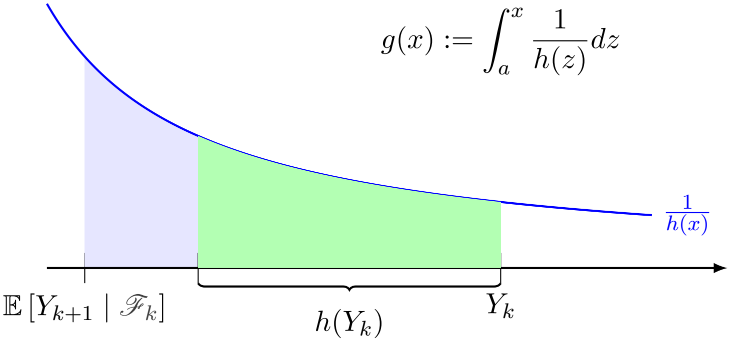
Proof.

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

$$\mathbb{E}[g(Y_k) - g(Y_{k+1}) \mid \mathscr{F}_k] \geq g(Y_k) - g(\mathbb{E}[Y_{k+1} \mid \mathscr{F}_k])$$

$\square$

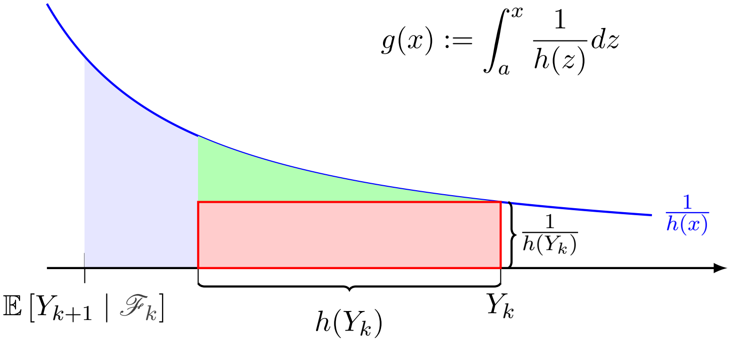
Variable Drift Theorem: Proof



Proof.

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

Variable Drift Theorem: Proof

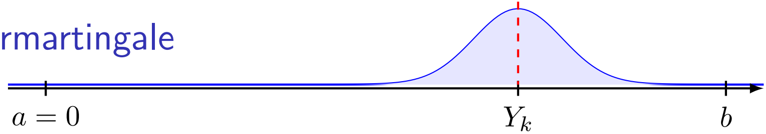


Proof.

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

Part 4 - Supermartingale

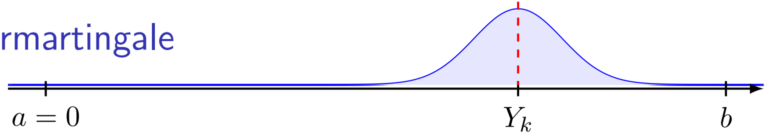
Supermartingale



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

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

Supermartingale



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

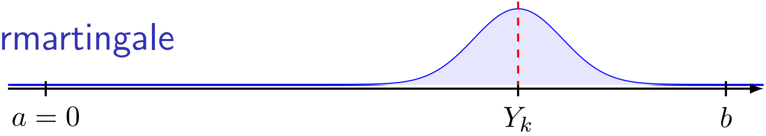
(S2) $\forall k \quad \mathbf{Var}\left[Y_{k+1} \mid Y_k > 0 \wedge \mathscr{F}_k\right] \geq \sigma^2$

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* $\mathbb{E}\left[\tau \mid \mathscr{F}_0\right] \leq \frac{Y_0(2b - Y_0)}{\sigma^2}$

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 $b - Y_k \leq \mathbb{E}\left[b - Y_{k+1} \mid \mathscr{F}_k\right]$.

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

Part 5 - Hajek's Theorem

Hajek's Theorem[4]

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

(*) $\Pr\left(\tau_b < B \mid Y_0 < a\right) \leq \frac{BD}{(1-\delta)\eta\varepsilon_0} \cdot e^{\eta(a-b)}$

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

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

Hajek's Theorem[4]

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

(*) $\Pr\left(\tau_b < B \mid Y_0 < a\right) \leq \frac{BD}{(1-\delta)\eta\varepsilon_0} \cdot e^{\eta(a-b)}$

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

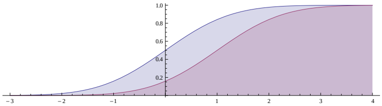
$$\Pr\left(\tau_b \leq e^{cn} \mid \mathscr{F}_0\right) \leq e^{-\Omega(n)}$$

[4] 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

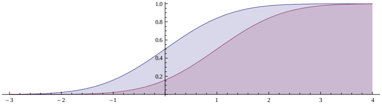
$Y \prec Z$ if $\Pr(Z \le c) \le \Pr(Y \le c)$ for all $c \in \mathbb{R}$



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 $c \in \mathbb{R}$



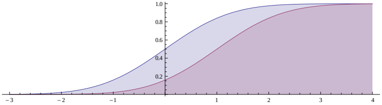
Example

1. If $Y \le Z$, then $Y \prec Z$.

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

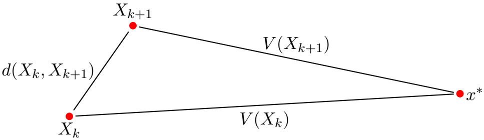
Definition

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



Example

1. If $Y \leq Z$, 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 $j \geq 0$,

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

Markov's inequality

- If $X \geq 0$, then $\Pr\left(X \geq k\right) \leq \mathbb{E}\left[X\right]/k$.

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 $M_X^{(n)}(0) = \mathbb{E}\left[X^n\right]$, 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

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

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 $M_X^{(n)}(0) = \mathbb{E}\left[X^n\right]$, 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 $\Pr\left(X_i = 1\right) = p$ and $\Pr\left(X_i = 0\right) = 1 - p$. Then

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

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	$X \sim N(\mu, \sigma^2)$	$\exp(t\mu + \frac{1}{2}\sigma^2 t^2)$

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

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

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** $x^{(k+1)} := x^{(k)}$

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 $\lambda = \ln(2)$

Simple Application (1+1) EA on Needle

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

$$\textsc{Needle}(x) := \prod_{i=1}^{n} x_i$$

$$Y_k := H(x^{(k)}, 0^n)$$
$$a := (3/4)n$$
$$b := n$$

Condition (C2) satisfied[5] 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

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

$$\Pr\left(\tau_b < e^{n/200} \mid \mathscr{F}_0\right) = e^{-\Omega(n)}$$

[5] See previous slide.

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

(*) $\Pr\left(\tau_b < B \mid Y_0 < a\right) \leq e^{\eta(a-b)} BD/(1-\rho)$

(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* $\rho < 1$

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

Proof.

Theorem

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

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)

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

Theorem

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

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,

$$\mathbb{E}\left[e^{\eta Y_{k+1}} \mid \mathscr{F}_0\right] = \mathbb{E}\left[\mathbb{E}\left[e^{\eta Y_{k+1}} \mid \mathscr{F}_k\right] \mid \mathscr{F}_0\right]$$

□

Theorem

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

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),

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

$\square$

Theorem

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

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

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

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

Proof.

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

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

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

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

□

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

Proof.

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

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

$\square$

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

(*) $\Pr\left(\tau_b < B \mid Y_0 < a\right) \leq e^{\eta(a-b)} BD/(1-\rho)$

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

(*) $\Pr\left(\tau_b < B \mid Y_0 < a\right) \leq e^{\eta(a-b)} BD/(1-\rho)$

Proof.

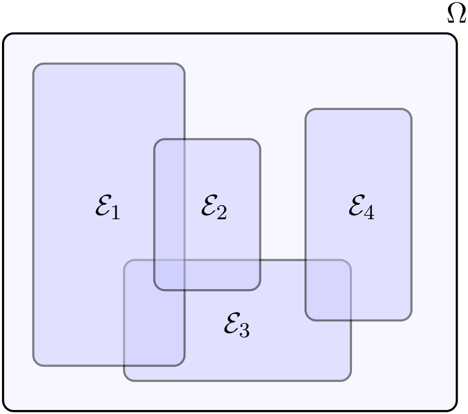
By the *union bound* and (2.8)

$$\Pr\left(\tau_b < B \mid Y_0 < a \wedge \mathscr{F}_0\right) \leq \sum_{k=1}^{B} \Pr\left(Y_k \geq b \mid Y_0 < a \wedge \mathscr{F}_0\right)$$

$$\leq \sum_{k=1}^{B} De^{\eta(a-b)}\left(\rho^k + \frac{1-\rho^k}{1-\rho}\right) \leq \frac{BDe^{\eta(a-b)}}{1-\rho}$$

□

Union Bound



$$\Pr\left(\mathcal{E}_1 \vee \mathcal{E}_2 \vee \cdots \vee \mathcal{E}_k\right) \leq \Pr\left(\mathcal{E}_1\right) + \Pr\left(\mathcal{E}_2\right) + \cdots + \Pr\left(\mathcal{E}_k\right)$$

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

Proof of (2.9)

Theorem

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

Assume that (D1) hold. Then

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

Proof.

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

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

Proof of (2.9)

Theorem

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

Assume that (D1) hold. Then

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

Proof.

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

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

By (2) and the law of total probability

$$e^{\eta Y_0} \geq \Pr\left(\tau_a > k \mid \mathscr{F}_0\right)\mathbb{E}\left[e^{\eta Y_{k\wedge\tau}}\rho^{-k\wedge\tau} \mid \tau_a > k \wedge \mathscr{F}_0\right]$$

$\square$

Proof of (2.9)

Theorem

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

Assume that (D1) hold. Then

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

Proof.

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

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

By (2) and the law of total probability

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

□

Proof of (2.9)

Theorem

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

Assume that (D1) hold. Then

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

Proof.

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

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

By (2) and the law of total probability

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

□

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

(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 $\lambda > 0$.

(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

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

$\square$

(C2) $\Longrightarrow$ (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) $\Longrightarrow$ (D1) and (D2)

(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 $\lambda > 0$.

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

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

Lemma

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

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

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

Corollary

Assume (C1), (C2) and $0 < \delta < 1$. *Then*

$$\eta := \min\{\lambda, \delta\varepsilon_0/c\}$$

$$\rho := 1 - (1 - \delta)\eta\varepsilon_0$$

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

(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 $\lambda > 0$.

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

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

Lemma

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

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

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

Corollary

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

$$\eta := \min\{\lambda, \delta\varepsilon_0/c\} \Longrightarrow \delta\varepsilon_0 \geq \eta c$$

$$\rho := 1 - (1-\delta)\eta\varepsilon_0 = 1 - \eta\varepsilon_0 + \eta\delta\varepsilon_0 \geq 1 - \eta\varepsilon + \eta^2 c$$

Reformulation of Hajek's Theorem

Theorem

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

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

(*) $\Pr\left(\tau_b < B \mid Y_0 < a\right) \leq \frac{BD}{(1-\rho)} \cdot e^{\eta(a-b)}$

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

Reformulation of Hajek's Theorem

Theorem

If there exist $\lambda, \varepsilon > 0$ *and* $1 < D < \infty$ *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) $\Pr\left(\tau_a > B \mid \mathscr{F}_0\right) \leq e^{\eta(Y_0 - a)}\rho^B$

(*) $\Pr\left(\tau_b < B \mid Y_0 < a\right) \leq \frac{BD}{(1-\rho)} \cdot e^{\eta(a-b)}$

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

1. Note that $\ln(\rho) \leq \rho - 1 = -(1-\delta)\eta\varepsilon_0$ so

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

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

Reformulation of Hajek's Theorem

Theorem

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

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

(*) $\Pr\left(\tau_b < B \mid Y_0 < a\right) \leq \frac{BD}{(1-\delta)\eta\varepsilon_0} \cdot e^{\eta(a-b)}$

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

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 $j \in \mathbb{N}_0$.

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

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[6]

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

- 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 in EC to point out that the much simpler to verify (C1), along with (S) is sufficient.

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

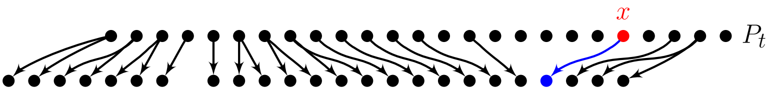
Part 6 - Population Drift

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

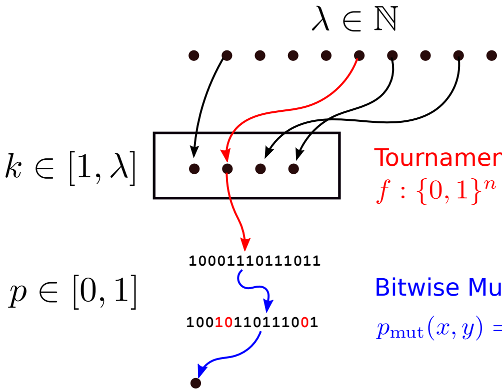


Require: ,

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

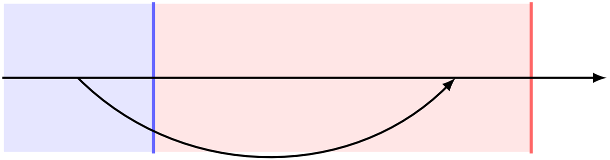


Tournament Selection wrt
$f : \{0,1\}^n \to \mathbb{R}$

Bitwise Mutation

$p_{\text{mut}}(x, y) = p^{H(x,y)}(1-p)^{n-H(x,y)}$

Population Drift



Central Parameters

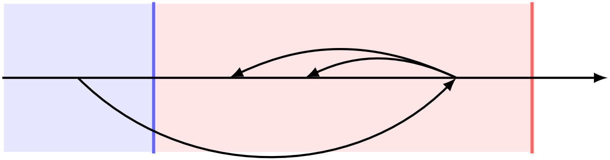
- Reproductive rate of selection mechanism p_{sel}

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

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

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

Population Drift



Central Parameters

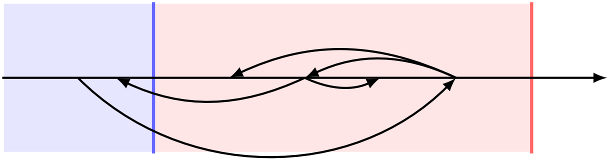
- Reproductive rate of selection mechanism p_{sel}

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

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

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

Population Drift



Central Parameters

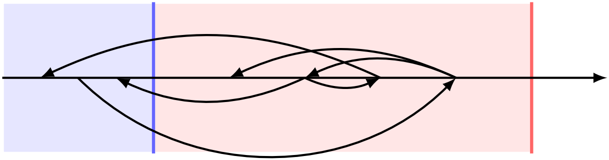
- Reproductive rate of selection mechanism p_{sel}

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

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

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

Population Drift



Central Parameters

- Reproductive rate of selection mechanism p_{sel}

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

- 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,[7] *then for some constants* $c, c' > 0$

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

[7] 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

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

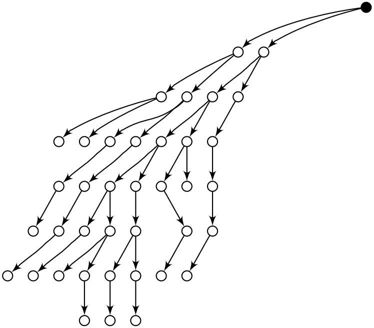
where

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

and

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

then the runtime is exponential.



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.