



#### 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/

Copyright is held by the author/owner(s). 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?



 $\blacktriangleright$  Prediction of the long term behaviour of a process X

• hitting time, stability, occupancy time etc. from properties of  $\Delta$ .





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

<sup>1</sup>NB! Drift is a different concept than *genetic drift* in evolutionary biology.

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



Doerr, Johannsen, and Winzen (GECCO 2010)

	the progress with respect to val. The aim is to prove that steps where the actual strin	ig Then $D_{\delta}^{*}(x) := \min\{1, d_{\delta}^{*}(x)\}$ . If $p > n/2$ , we have to be more cateful, because e	Since j <sub>1</sub> /w ≤ 1/2, Prob(d)C) ≤ (1/2) - d for some positive constant d, and we :
	changes lead to an expected gain which is bounded below by a positive constant	a, ery bit flipping from zero to one has only a contribution of 1. Hence, we defin	to if $\pi^2 < \pi/2$ , we again use the estimation $Prob(A C) \in 1/2$ . In this case the upp
	Therefore, it is necessary that the potential function contains the information that the	$\log  D_{3}^{*}(x) := \min\{1, val(x') - val(x)\}.$	on the expected number of bits flipping from 0 to 1 can be improved. The
	first half positions are those with the larger weights.	Is in obvious that, in both cases, $D_3^+(x)$ is an integer, $D_3^+(x) \in I$ , and $D_3^+(x) \in D_2(x)$	c) of pairs which are not allowed to flip is at least no <sup>2</sup> for some constant a>0
	The initial bit string a has the value sa(x) which is not negative. The value of	of implying that $Prob(D_3^*(x) \in r) \ge Prob(D_3(x) \in r)$ . We have to prove that $E(D_3^*(x)) \ge c$	" probability that exactly one of these pairs tries to flip from 1 to 0 is bounded I
	the global optimum equals (3/2)0. A step of the (1 + 1) EA working on the lines	ar for some positive constant d'.	some constant $x' > 0$ . Hence, the expected contribution of the bits flipping fro
	function f is called successful if it produces from x a matated string x' which replace	Case 1. $p \le n/2$ : We know that the mutation from x to x' is a successful one as	if to $D_S^*(x)$ can be lower bounded by $-(2f_2 + f_1)/n + n^n \ge -3/2 + n^n$ for some
	s (i.e., $x' \neq x$ and $f(x') \ge f(x)$ ). Let s be a bit string with $val(x) = i < (3/2)a$ . W	Te that S is the set of all indices J where $x_i = 0$ and $x'_i = 1$ . Hence, $x_i = 0$ and $T \notin S$ impli-	is $\pi^{\prime\prime} > 0$ . Altogether, we have proved that $E(D_{2}^{\prime}(x)) \ge d^{\prime\prime}$ for some positive con-
	are interested in the expected number of successful steps until we obtain for the fir	set $\mathbf{x}_{i}^{r} = 0$ ,	in Case 1.
	time a bit string a' where vak(a')>val(x). Later we prove that we can upper boar	I Let $j_1$ be the number of positions $i \le n/2$ where $n_i = 1$ and the matation whe	Case 2: p>n/2: Let s =  3 . First, we consider the subcase s > 4. In order
	the expected number independently from x by a constant c*. Afterwards, it is easy t	to only s, flips from 1 to 0 is successful, and let j1 be the corresponding number -	H a lower bound on $E(D_2^{*}(x))$ , we work with the following assumptions which i
	bound the expected number of successful and unsuccessful steps. This can be dear i	in positions $i > n/2$ . The expected number of flipping bits aroung these $j_1 + j_2$ bits (and	of to smaller values of E(D)(s)). We do not make use of the fact that the s
	the same way as in the pooef of Lemma 11 by proving a lower bound on the probabilit	ty no condition) is $(f_2 + f_1)/n$ which leads to an expected contribution of $-(2f_2 + f_1)/n$	to x to x' has to be successful and we assume that $x_i = 1$ for all $i \notin S$ . Then
	of successful steps if the value of the bit string is bounded above by 1<(3/2)s. The	in $d_{2}^{*}(x)$ . Given the set S of bits flipping from 0 to 1, if we assume in addition that the	w $n/2$ positions $i \le n/2$ where $n = 1$ and less then $n/2$ positions $i > n/2$ where $n$
THE REPORT OF TH	the number of zeros in the bit string is at least $r_1 = [(3/4)n - (1/2)i]$ and there exist	at mutation is successful, this may only decrease the number of hits flipping from 1 to	0. overestimate the number of bits flipping from 1 to 0 by assuming that we
Theorem 12. The especied enounty time of the (1+1) Ed on the class of theorem	at least vy different one bit matations which increase f. Hence, the success probability	by Therefore, we may estimate the expected corribution of these bits by $-(2j_2 + j_1)$	* positions $t > a/2$ where $n = 1$ . The random number Z' of bits in the first half
Jacobias with mon-tere weights or minimus.	is at least $r_i(1/n)(1-1/n)^{n-1} \ge e^{-1}r_i/n$ . Altogether, the expected number of steps t	to We have shown that $\mathcal{E}(d_1^*(x)) \ge 2 - (2j_2 + j_1)/\epsilon$ . If no bit flips from 1 to 0, $d_1^*(x) =$	2 from 1 to 0 is binomially distributed with respect to the parameters of 2 and
Beard. The lower bound follows down I server 10. We show the server bound for or	optimize f is bounded by	but D'(x) = 1. We have to consider the conditional probability that no bit flips from	I same holds by our possimistic assumptions for the random number Z' of bi
Frint, the state thank interes she takens to, we prove the opper tokens for an		to 0. It is sufficient to consider the $h + h$ selected positions. Let C be the event th	if around half, it is sufficient to mean a lower bound on E(D2(x1)) under the av-
afteliary knew functions /. We have already discussed that without loss of generality	Without 1 1 and a family 1 - Overhead	the mutation is successful and if he the errort that no hit flins from 1 to 0. Since	d that the reaches of hits flipping from 1 to 0 are sizes by Z' and Z' erest
$t = 0$ and $w_1 \oplus w_2 \oplus \cdots \oplus w_n > 0$ , we below the main ideas of the proof of Lemma 11	Z ·	implies (" not have Probe #2") - Probe (1) Probe("). (Decisionly, Probe () - (1 - 1) a) ["	The bin smill distribution with accounting of and his one has connected
but the general statution is much more complicated.		We are interested in some breach on Back (17) and therefore in breach breach a	The territory of the parameter is and it is an apprendice
We measure the progress of the (1+1) EA on the function / by the arbitral liness	The only claim we have to prove is the existence of a constant upper bound on th	Packer 1 Be definition	· rondoù anterenten was paraneer z = 1/2. Let zj and zj re moepeneen
Fanctoon	expectation of the number of successful steps to increase the value. Let a be a bit strin	I THEFT I BY BUILTING	valuents while another in runnes will partitely 2 - 1/2. We will be
11 A	(not the global optimum) and a' the random bit string produced by a successfal ste	P ( 1) 11 1 1) 1) 1 1 1)	$Z_1$ insical of $Z_2$ and $Z_1$ , respectively. Later we discuss the motatic caused
$\operatorname{val}(x) = 2\sum_{i} x_i + \sum_{i} x_i$	based on x. Let S be the random set of indices i where $x_i = 0$ and $x'_i = 1$ . Since the ste	$p = Prob(C) \ge (1) + (n + \mu) - (1)$	replacement.
111	was successful, $S \neq \emptyset$ . Let $D_{\beta}(x) = val(x') - val(x)$ be the random variable describin	4	We know that
This attificial fitness function plays the role of a potential function # in the analysis	for fixed S the gain with respect to the value function. Our main step is to defin	or implying that	1 /112 1 /112
of data structures and algorithms. We investigate the (1+1) EA on f but we measure	a random variable D2(x) which takes integer values, D2(x) < 1, and has the propert	N THE REAL PROPERTY AND A DESCRIPTION OF A DESCRIPTIONO OF A DESCRIPTION OF A DESCRIPTION OF A DESCRIPANTI OF A DESCRIPTION OF A DESCRIPTION OF A DESCRIPTION O	$Prob(Z_1 - z_1, Z_1 - z_2) = = [-] e^{-1/2} (-] e^{-1/2}$
	that $Prob(D(x) \leq r) \geq Prob(D_n(x) \leq r)$ for all r. We prove that there is some positive	Probalici < <	221 (2) 211 (2)
	constant d" (which does neither depend on x nor on S), such that E(D2(x)) > d" hold	$1 - 1/4 + (f_2 + f_1)/4 = 1 + (f_2 + f_1)/4$	to the following an entropy for Wiley's order the susception that the same
	In other to do so, we consider finitely many cours and chance at the end the smaller	w We conclude that	in the deriving we control a to average and the method
	of the considered constants		hipping from 1 to 0 are given by $Z_2$ and $Z_1$ , respectively. This new random
	The definition of Office) is controlicated 1 at an control of D_deb = val(y') - val(y') Th	$E(D_{j}^{*}(x)) \ge E(d_{j}^{*}(x)) - Preb(d C) \ge 2 - \frac{4d_{2} + d_{1}}{2} - \frac{1}{2}$	is called $D_{2}^{-}(x)$ , the part $(z_{2}, z_{1})$ of finishing variables faces values from
	The entropy of population of the product of the entropy of the product of the pro	4 1 + (1 + 1/2)	Hence,
	concentration of postation 7 to 17211 equation	This lower bound for $E(D_{i}^{2}(x))$ is a decreasing function of $j_{1}$ and $j_{2}$ . Since $j_{2} \leq (n/2)$ -	A CONTRACT OF AND
	• • • • • • • • • • • • • • • • • • •	by assumption, we can consider the lower bound for $i_1 = n/2$ leading to	$E(D_5^{**}(x)) = \sum E(D_5^{**}(x)) Z_1 = z_2, Z_1 = z_1)\operatorname{Prob}(Z_2 = z_2, Z_1 = z_1)$
	• $-2$ , if $x_i = 1$ , $x_i = 0$ , and $1 \le n/2$ .		1919
	• +1, if $x_1 = 0$ , $x_1 = 1$ , and $x > 0/2$ ,	manual A	We partition No x No into disjoint sets:
It is easy to see that	For the subcases 3.6 (1.2,3), we are more caterial by excitating cases which nee	• $A_1 = \{(1,z_1)   z_1 \ge 1\},$	((0.m))
<ul> <li>D<sup>**</sup><sub>3</sub>(s) = 1, if (z<sub>2</sub>, z<sub>1</sub>) ∈ A<sub>1</sub>.</li> </ul>	essarily lead to unsuccessful steps. A step cannot be successful if $z_2 > s$ or $z_2 = s$ and	$d \bullet A_4 = \{(2, z_1)   z_1 \ge 0\},\$	f = f(0, x) f(x, y)
<ul> <li>D<sup>**</sup><sub>2</sub>(x) ≥ 4 - z<sub>1</sub>, if (z<sub>1</sub>, z<sub>1</sub>) ∈ A<sub>1</sub>,</li> </ul>	$z_1 > 0$ . Hence, we consider only steps where $z_2 \in x - 1$ or $z_2 = x$ and $z_1 = 0$ . Let $x_i$ be	$M = A_5 = \{(3,0)\},\$	and - fusicities and
• $D_{4}^{**}(x) \ge 2 - z_1$ , if $(z_2, z_1) \in A_3$ .	the probability that $z_2 \le s - 1$ or $z_2 = s$ and $z_1 = 0$ . We follow the same approach a	$a \cdot A_5 = [(1, 0), (0, 1), (0, 2)].$	• $A_1 = \{(1, 2_1), (2_1), (3_2), \dots, (3_n), \dots$
• $D_{4}^{++}(x) \ge 4 - 2z_2 - z_1$ , if $(z_2, z_1) \in A_4$ .	in the subcase $s \ge 4$ , the probability of the event $(Z_2 = z_2, Z_1 = z_1)$ under the condition	st is easy to see that	• $A_4 = \{(2_2, 2_1, 3), 2_2, 3, 2, 2_1, 3, 0\},\$
<ul> <li>D(*(x) &gt; 0, if (x, y) ) &lt; d_0.</li> </ul>	$Z_2 \le s - 1$ or $Z_2 = s$ and $Z_1 = 0$ is equal to the product of $x_1^{-1}$ and the unconditions	al • $D_{x}^{**}(x) = 1$ , if $(z_2, z_1) \in A_1$ .	• $a_3 = m_3 \times m_0/(a_1 \cap a_2 \cap a_3 \cap a_4)$
Let con., 1 6 / 6.5, be the contribution of all terms (z (z)) 6.4, to the sum (+). Then	probability of the event $(Z_2 = z_2, Z_1 = z_1)$ .	• $D_1^{**}(x) = 3 - z_1$ , if $(z_2, z_1) \in A_2$ .	
• con, 2 c <sup>-1</sup> .	For $s = 1$ , we partition the set of pairs $(z_2, z_1)$ where $z_2 \in 0$ or $z_2 = 1$ and $z_1 = 0$ int	$D = D_{1}^{**}(x) = 1 - z_{1}$ , if $(z_{2}, z_{1}) \in A_{1}$ .	
<ul> <li>consist Y = 14 = contractivity interface = 1 = 1.</li> </ul>	• $A_1 = \{(0,0)\},\$	• $D_{1}^{(n)}(z) = -1 - z_{1}$ if $(z_{2}, z_{1}) \in A_{2}$ .	
$\pi = \cos 2 \sum_{i=1}^{n} (2 - \cos (1 - \sin (1 - 1))^{-1}) = -1 e^{-1}$	• $A_1 = \{(1,0)\},\$	• $D_{i}^{*}(x) = -3$ , if $(x_{i}, x_{i}) \in A_{i}$ .	
The second secon	• $d_{1} = l(0, z_{1})   z_{2} \ge 1$	· Differing if its solution	
· contra Transition and - stationarially a stationary	It is easy to see that	This lands to	
i conjunt.	• $D_{1}^{*}(x) = 1$ if $(x, y_{1}) \in A_{1}$	- can	Hanne
THEORY,	a difficient a difficuencia de	start St. of antiputation in the last	Index,
$E(D_{1}^{*}(x)) \ge \frac{1349}{1000} e^{-1}$	· Dran	• cong = () = contraction of e = n = () =	matters of the scatters of the scatters and the scatters of the
1920	Then	• $\cos_1 = \sum_{j=0,1} (1 - z_j)(1/z_j)(\frac{1}{2})$ , $d = \frac{1}{2} - \frac{1}{2} - \frac{1}{2} - \frac{1}{2} + \frac{1}{2} - \frac{1}{2} - \frac{1}{2} + \frac{1}{2} - \frac{1}{2} + \frac{1}{2} - \frac{1}{2} + \frac{1}{2} - \frac{1}{2} + \frac{1}{2} - \frac{1}{2} - \frac{1}{2} + \frac{1}{2} + \frac{1}{2} - \frac{1}{2} + \frac{1}{2} - \frac{1}{2} + 1$	$E(1, 0,0) = 1 + Hosto (1) \neq 11 - E(1, 0,0) - \sum_{i=1}^{n} Hosto (1) = \# He$
However, we are interested in an estimate of $E(D_{j}^{*}(x))$ . We estimate the mistake by	y tous	• $\cos a_1 = -\sum_{m \ge 0} (1 + z_1)(1/z_1 - 2!)(\frac{1}{2}! - 2!)(\frac{1}{2}! - \frac{1}{2}!)(\frac{1}{2}! - \frac{1}{2}!)(\frac{1}{2}!)($	- The second
replacing the random variables $Z'_1$ and $Z'_2$ which are binomially distributed with respect	* com, - c 's, .	• com,	$= 1 + \operatorname{Prob}(D^*(x) \neq 1) \cdot K(T^*(x)) - \sum \operatorname{Prob}(D^*(x) = d)de^*$
to a/2 and 1/s with the random variables Z <sub>1</sub> and Z <sub>2</sub> , respectively, with a Poisson	• cong =	· congitt 0.	
distribution with respect to 1/2. In the subcase s > 4 as in the later subcases we call	$y = \cos_1 \ln \sum_{n \ge 1} (1 - z_1) (1/z_1) (\frac{1}{2})^n e^{-1} x_1^{-1} \ln - \frac{1}{2} e^{-1} a_1^{-1}$	as contributions for the subcase $x = 3$ . Obviously, $E(D_2^{+}(x))$ is bounded below by	$+c^* \operatorname{Prob}(D^*(x) = 1)$
consider terms	Hence, $E(D_3^*(x))$ is also in this subcase bounded below by a positive constant, if $x$ is	<sup>16</sup> positive constant also in this last subcase.	
Barbar R' - who and a Barbar R' - who first source and A	large enough.	Altegether, \$7(1) has all desired properties and in particular, (202(x)) & a feature	$= 1 + E(T'(x)) - (T(x)) = (T(x)) + E(D'(x))k^{2}$
The same a strange, and the same a state	For $s = 2$ we partition the set of pairs $(z_2, z_1)$ where $z_2 < 1$ or $z_2 = 2$ and $z_1 = 1$ int	$I \cap C \cap C \cap X \cap X$	
$\sum \operatorname{Prob}(Z'_i = j)$ and $\sum j\operatorname{Prob}(Z'_i = j)$ for some $r \leq 5$	• $A_1 = \{(0,0)\},$	Lat 1 by the relation whether description the title maint of the where the	
124	• $A_2 = \{(1,0), (2,0)\},\$	unless of the hit string anothered he concerned stress of the (1 + 1) EA on the lin	
and finite combinations of these terms, It is well-known that for constant i, and param	• $d_1 = \{(0, z_1), (1, z_1)   z_1 \ge 1\}.$	and functions of stanting of a in larger them solved. We chain that \$777(a)) of all 110	This implics
steps a and $\phi(n)$ for the binomial distribution where $n \phi(n) \rightarrow \lambda$ as $n \rightarrow \infty$ the binomia	d it is not difficult to see that we have	We seems this close he considering the changes of ad. We investigate the desider	
doinbution converges against the Poisson distribution. In our case the product of th	$r = D_{1}^{(2)}(r) = 1$ , if $(r_{1}, r_{2}) \in A_{1}$ .	the prover related to the (1-1) EA on ( where the initial string is F = 1.	$E(T^*(x))$ Prob $(D^*(x) = 1) \le 1 - E(D^*(x))x^* + x^*$ Prob $(D^*(x) = 1)$
parameters of the binomial distribution equals $\lambda = 1/2$ . Hence, for each $\epsilon > 0$ and	$v = DT'(x) = 2 - 2z_0$ if $(z_0, z_1) \in A_0$	T. Y. F. As the experiment of orbital strings constand he economical string. The	
large encash	<ul> <li>D<sup>22</sup>(x) = 2 = 2x = x if (x, x) = 4x</li> </ul>	spiration of the second second straigs created by successful steps. The	$\leq e^{-Prob}(D^*(x) = 1).$
	Then	we seek the the first point of time t where $T_t := val(X_t) - val(X_t)$ takes a positive	and the second se
$ \operatorname{Prot}(Z_j = x) - \operatorname{Prob}(Z_j = x)  \le x$	and the second sec	same. The sequence Ye, Y1, Y2, is a stochastic process on Z. The distribution of	This last meganity follows, since E(D <sup>*</sup> (x)) ≥ d <sup>*</sup> = 1/c <sup>*</sup> . Since Prob(D <sup>*</sup> (x) = 1
and	a second s	$T_{i+1} = T_i$ for some given value of $X_i = x^i$ (and, therefore, also $F_i$ ) is described by	7 obtain the desired result E(T'(x))≤x". This proves the theorem.
LANDARY - CO. CREATE - CO. C.	· · · · · · · · · · · · · · · · · · ·	the random variable $D(x')$ which equals $D_2(x')$ if S is the set of positions flippin	8
$ r $ Prob $(Z_i = r) - r$ Prob $(Z_i = r) \in \mathcal{L}$	<ul> <li>An example in the example in</li> </ul>	from 0 to 1. We have shown that we obtain an upper bound on $T^*(x)$ if we re-	
for all r < 4. Since	I TANK A I TANKIN A	place $D(x')$ with $D^*(x')$ (and $D_2(x')$ with $D_2^*(x')$ ). Since $D^*(x) \leq 1$ , the first pos-	
Vindent in Vindent in 1	$cee_1 \ge \left[\sum_{i=1}^{n} (2-z_1) \frac{1}{z_1} \left[\frac{1}{x_1}\right] e^{-1} s_2^{-1}\right] + \left[\sum_{i=1}^{n} -z_1 \frac{1}{z_1 - u} \left(\frac{1}{x_1}\right) e^{-1} s_2^{-1}\right]$	sive value reached by this new process is the point 1. We cannot apply Wald	x
$\sum_{i=1}^{n} \operatorname{recei} x_i = j_1 = \sum_{i=1}^{n} \operatorname{rece} x_i = j_2 = 1$	1432 MONES 2/ 1631 1218 122 2/	identity, since D'(s') depends on s'. Therefore, we need a slight generalization of	ď.
	4 1 AN	Wald's identity. We prove the claim by induction on value). The base of induction	

## 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]

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

### 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<sub>k</sub> is a stochastic process<sup>2</sup> in some general state space 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

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

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

<sup>2</sup>We do not require  $X_k$  to be a Markov process.

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

### Overview of Tutorial

### Part 1 - Basic Probability Theory

#### <sup>3</sup>Some drift theorems need additional conditions.

### Basic Probability Theory



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

- $\Omega$  : Sample space
- $\mathscr{F}$  :  $\sigma$ -algebra on  $\Omega$  (family of events)
- Pr: 𝔅 → ℝ probability function (satisfying probability axioms)

#### Events

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

#### **Random Variable**

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

### Stochastic Processes and Filtration

#### Definition

- ► A *stochastic process* is a sequence of random variables *Y*<sub>1</sub>, *Y*<sub>2</sub>,... on the same probability space.
- A *filtration* is an increasing family of sub  $\sigma$ -algebras<sup>4</sup> 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  $k: \forall A \in \mathcal{B}\{\omega \in \Omega \mid Y_k(\omega) \in A\} \in \mathscr{F}_k$
- $\implies \text{Informally, } \mathscr{F}_k \text{ represents the information that has been revealed about the process during the first $k$ steps, including the value of $Y_k$.}$

<sup>4</sup>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  $k\geq 0$ 

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

► The information obtained until step k is sufficient to decided whether the event {τ ≤ 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

$$\mathbb{E}\left[X\right] := \sum_{x} x \Pr\left(X = x\right)$$

#### Independence

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

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

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  $a, b \in \mathbb{R}$ 

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

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

### Markov's Inequality

#### Theorem (Markov's Inequality)

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



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

### 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

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

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

## Jensen's Inequality

### Conditional Probability and Expectation

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

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

Conditional Expectation

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

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

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

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

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

#### 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

$$\mathbb{E}\left[\varphi(X)\right] \le \varphi(\mathbb{E}\left[X\right])$$

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

### Martingales



#### Example

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

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

$$\mathbb{E}\left[Y_{k+1}Z_{k+1} \mid \mathscr{F}_k\right] = \Delta_1 + \dots + \Delta_k + \mathbb{E}\left[\Delta_{k+1} \mid \mathscr{F}_k\right] + (k+1)\varepsilon_0$$
$$\leq \Delta_1 + \dots + \Delta_k - \varepsilon_0 + (k+1)\varepsilon_0 < Y_k = Z_k$$

## 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.

$$\mathbb{E}\left[Y_{k} \mid \mathscr{F}_{0}\right] = \mathbb{E}\left[\mathbb{E}\left[Y_{k} \mid \mathscr{F}_{k}\right] \mid \mathscr{F}_{0}\right]$$
$$\leq \mathbb{E}\left[Y_{k-1} \mid \mathscr{F}_{0}\right] \underbrace{\leq \cdots \leq \mathbb{E}\left[Y_{0} \mid \mathscr{F}_{0}\right]}_{\text{by induction on } k} = Y_{0}$$

#### Example

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

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

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





 $\begin{array}{ll} (\mathsf{C1+}) \ \forall k \quad \mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > 0 \land \mathscr{F}_k\right] \leq -\varepsilon_0 \\ (\mathsf{C1-}) \ \forall k \quad \mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > 0 \land \mathscr{F}_k\right] \geq -\varepsilon_0 \end{array}$ 

### Theorem ([7, 9, 10])

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

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

Part 2 - Additive Drift

### Obtaining Supermartingales from Drift Conditions

Definition (Stopped Process)

Let Y be a stochastic process and  $\tau$  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  $\tau_a := \min\{k \ge 0 \mid Y_k \le a\}$ , and assume (C1)  $\mathbb{E}[Y_{k+1} - Y_k \mid Y_k > a \land \mathscr{F}_k] \le -\varepsilon_0$ 

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

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

### Proof of Additive Drift Theorem

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

#### Theorem

Given a sequence  $(Y_k, \mathscr{F}_k)$  over an interval  $[0, b] \subset \mathbb{R}$ Define  $\tau := \min\{k \ge 0 \mid Y_k = 0\}$ , 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  $\mathbb{E}[\tau \mid \mathscr{F}_0] \ge Y_0/\varepsilon_0$ .

#### Proof.

By (C1+),  $Z_k := Y_{k\wedge au} + arepsilon_0(k\wedge au)$  is a super-martingale, so

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

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

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

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

$$|X_k| \le Y.$$

Then it holds

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

### Proof of Additive Drift Theorem

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

#### Theorem

Given a sequence  $(Y_k, \mathscr{F}_k)$  over an interval  $[0, b] \subset \mathbb{R}$ Define  $\tau := \min\{k \ge 0 \mid Y_k = 0\}$ , 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  $\mathbb{E}[\tau \mid \mathscr{F}_0] \ge Y_0/\varepsilon_0$ .

#### 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  $\tau$  has finite expectation, the dominated convergence theorem applies and

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

### Example: (1+1) EA on LEADINGONES

#### 1 (1+1) EA

1: Sample  $x^{(0)}$  uniformly at random from  $\{0,1\}^n$ . 2: for  $k = 0, 1, 2, \dots$  do 3: Set  $y := x^{(k)}$ , and flip each bit of y with probability 1/n. 4:  $x^{(k+1)} := \begin{cases} y & \text{if } f(y) \ge f(x^{(k)}) \\ x^{(k)} & \text{otherwise.} \end{cases}$ 

5: end for

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

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

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

• 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

$$\mathbb{E}\left[Y_{k+1} - Y_k \mid \mathscr{F}_k\right] \le \sum_{i=1}^n \Pr\left(\mathcal{E}_i \mid \mathscr{F}_k\right) \mathbb{E}\left[Y_{k+1} - Y_k \mid \mathcal{E}_i \land \mathscr{F}_k\right]$$
$$\le \left(\frac{1}{n}\right) \left(1 - \frac{1}{n}\right)^{n-1} \sum_{i=1}^n w_i \left(x_i^{(k)} - 1\right) \le -\frac{Y_k}{en} \le -\frac{w_{\min}}{en}$$

• By the additive drift theorem,  $\mathbb{E}\left[\tau \mid \mathscr{F}_{0}\right] \leq en^{2}(w_{\max}/w_{\min}).$ 

### Example 1: (1+1) EA on LEADINGONES

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



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

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

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

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

 $\implies$  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  $\tau$  the time to absorption. Then the function  $g(x) := \mathbb{E} [\tau | Y_0 = x]$ , satisfies

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

- ▶ 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











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

#### Proof.

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

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



### 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  $\mathbb{E}[\tau_a \mid \mathscr{F}_0] < \infty$ . 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

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

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

### 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

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

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

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

Variable Drift Theorem: Proof



#### Proof.

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

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

1248







# Hajek's Theorem<sup>5</sup>

#### Theorem

If there exist  $\lambda, \varepsilon_0 > 0$  and  $D < \infty$  such that for all  $k \ge 0$ (C1)  $\mathbb{E}[Y_{k+1} - Y_k | Y_k > a \land \mathscr{F}_k] \le -\varepsilon_0$ (C2)  $(|Y_{k+1} - Y_k| | \mathscr{F}_k) \prec Z$  and  $\mathbb{E}[e^{\lambda Z}] = D$ then for any  $\delta \in (0, 1)$ (2.9)  $\Pr(\tau_a > B | \mathscr{F}_0) \le e^{\eta(Y_0 - a - B(1 - \delta)\varepsilon_0)}$ (\*)  $\Pr(\tau_b < B | \mathscr{F}_0) \le \frac{BD}{(1 - \delta)\eta\varepsilon_0} \cdot e^{\eta(a - b)}$ for some  $\eta \ge \min\{\lambda, \delta\varepsilon_0\lambda^2/D\} > 0$ .

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

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

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

#### 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  $(\Omega, d)$  be a metric space, and  $V(x) := d(x, x^*)$ . Then  $|V(X_{k+1}) - V(X_k)| \prec d(X_{k+1}, X_k)$



# 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<sup>(n)</sup><sub>X</sub>(0) = E [X<sup>n</sup>], hence M<sub>X</sub> 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

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

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

#### Assume that

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

Then for any  $j \ge 0$ ,

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

Moment Generating Functions

Distribution		mgf
Bernoulli	$\Pr\left(X=1\right)=p$	$1 - p + pe^t$
Binomial	$X \sim \operatorname{Bin}(n, p)$	$(1-p+pe^t)^n$
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  $X \sim Bin(n, p)$  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  $f(x') \ge f(x^{(k)})$ , 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  $Z := H(x^{(k)}, x')$
- $Z \sim \operatorname{Bin}(n,p)$  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  $D \ge 1$  such that (D1)  $\mathbb{E} \left[ e^{\eta Y_{k+1}} \mid Y_k > a \land \mathscr{F}_k \right] \le \rho e^{\eta Y_k}$ (D2)  $\mathbb{E} \left[ e^{\eta Y_{k+1}} \mid Y_k \le a \land \mathscr{F}_k \right] \le D e^{\eta a}$ Then

 $\begin{array}{l} (2.6) \ \mathbb{E}\left[e^{\eta Y_{k+1}} \mid \mathscr{F}_{0}\right] \leq \rho^{k} e^{\eta Y_{0}} + D e^{\eta a} (1-\rho^{k})/(1-\rho). \\ (2.8) \ \Pr\left(Y_{k} \geq b \mid \mathscr{F}_{0}\right) \leq \rho^{k} e^{\eta(Y_{0}-b)} + D e^{\eta(a-b)} (1-\rho^{k})/(1-\rho). \\ (*) \ \Pr\left(\tau_{b} < B\right) \leq e^{\eta(a-b)} B D/(1-\rho) \\ (2.9) \ \Pr\left(\tau_{a} > k \mid \mathscr{F}_{0}\right) \leq e^{\eta(Y_{0}-a)} \rho^{k} \end{array}$ 

#### Lemma

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

(C1)  $\mathbb{E}[Y_{k+1} - Y_k | Y_k > a \land \mathscr{F}_k] \leq -\varepsilon_0$ (C2)  $(|Y_{k+1} - Y_k| | \mathscr{F}_k) \prec Z$  and  $\mathbb{E}[e^{\lambda Z}] = D < \infty$  for a  $\lambda > 0$ . then (D1) and (D2) hold for some  $\eta$  and  $\rho < 1$ 

### Simple Application (1+1) EA on NEEDLE

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

NEEDLE
$$(x) := \prod_{i=1}^{n} x_i$$
  
 $b := n$   
 $Y_k := H(x^{(k)}, 0^n)$   
 $a := (3/4)n$   
 $b := n$ 

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

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

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

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

<sup>6</sup>See previous slide.

#### Theorem

(D1)  $\mathbb{E}\left[e^{\eta(Y_{k+1}-Y_k)} \mid Y_k > a \land \mathscr{F}_k\right] \leq \rho$ (D2)  $\mathbb{E}\left[e^{\eta(Y_{k+1}-a)} \mid Y_k \leq a \land \mathscr{F}_k\right] \leq D$ 

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.

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

$$\mathbb{E}\left[e^{\eta Y_{k+1}} \mid \mathscr{F}_{0}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{\eta Y_{k+1}} \mid \mathscr{F}_{k}\right] \mid \mathscr{F}_{0}\right] \\ \leq \rho \mathbb{E}\left[e^{\eta Y_{k}} \mid \mathscr{F}_{0}\right] + De^{\eta a} \\ \leq \rho^{k}e^{\eta Y_{0}} + (1+\rho+\rho^{2}+\dots+\rho^{k-1})De^{\eta a}.$$

# Proof of (2.8)

#### Theorem

(D1) 
$$\mathbb{E} \left[ e^{\eta(Y_{k+1}-Y_k)} \mid Y_k > a \land \mathscr{F}_k \right] \leq \rho$$
  
(D2)  $\mathbb{E} \left[ e^{\eta(Y_{k+1}-a)} \mid Y_k \leq a \land \mathscr{F}_k \right] \leq D$   
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)

$$\Pr\left(Y_{k+1} \ge b \mid \mathscr{F}_{0}\right) = \Pr\left(e^{\eta Y_{k+1}} \ge e^{\eta b} \mid \mathscr{F}_{0}\right)$$
$$\leq \mathbb{E}\left[e^{\eta Y_{k+1}} \mid \mathscr{F}_{0}\right] e^{-\eta b}$$

# Proof of (2.9)

#### Theorem

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

Assume that (D1) hold. Then (2.9)  $\Pr(\tau_a > k \mid \mathscr{F}_0) \le e^{\eta(Y_0 - a)} \rho^k$ 

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

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

By (2) and the law of total probability

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

# Proof of (\*)

### Theorem

(D1) 
$$\mathbb{E}\left[e^{\eta(Y_{k+1}-Y_k)} \mid Y_k > a \land \mathscr{F}_k\right] \leq \rho$$
  
(D2)  $\mathbb{E}\left[e^{\eta(Y_{k+1}-a)} \mid Y_k \leq a \land \mathscr{F}_k\right] \leq D$   
Assume that (D1) and (D2) hold for  $D \geq 1$ . Then  
(2.8)  $\Pr\left(Y_k \geq b \mid \mathscr{F}_0\right) \leq \rho^k e^{\eta(Y_0-b)} + De^{\eta(a-b)}(1-\rho^k)/(1-\rho).$   
(\*)  $\Pr\left(\tau_b < B\right) \leq e^{\eta(a-b)}BD/(1-\rho)$ 

### Proof.

By the union bound and (2.8)  $\Pr(\tau_b < B \mid Y_0 < a \land \mathscr{F}_0) \le \sum_{k=1}^{B} \Pr(Y_k \ge b \mid Y_0 < a \land \mathscr{F}_0)$   $\le \sum_{k=1}^{B} De^{\eta(a-b)} \left(\rho^k + \frac{1-\rho^k}{1-\rho}\right) \le \frac{BDe^{\eta(a-b)}}{1-\rho}$ 

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

(C1)  $\mathbb{E}[Y_{k+1} - Y_k | Y_k > a \land \mathscr{F}_k] \leq -\varepsilon_0$ (C2)  $(|Y_{k+1} - Y_k| | \mathscr{F}_k) \prec Z$  and  $\mathbb{E}[e^{\lambda Z}] = D < \infty$  for a  $\lambda > 0$ . (D1)  $\mathbb{E}[e^{\eta(Y_{k+1} - Y_k)} | Y_k > a \land \mathscr{F}_k] \leq \rho$ 

Lemma

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

#### Proof.

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

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

1252

 $(C2) \Longrightarrow (D2)$ 

(C2) 
$$(|Y_{k+1} - Y_k| | \mathscr{F}_k) \prec Z$$
 and  $\mathbb{E}\left[e^{\lambda Z}\right] = D < \infty$  for a  $\lambda > 0$ .  
(D2)  $\mathbb{E}\left[e^{\eta(Y_{k+1}-a)} | Y_k \leq a \land \mathscr{F}_k\right] \leq D$ 

#### Theorem

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

#### Proof.

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

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

Furthermore, by (C2)

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

### Reformulation of Hajek's Theorem

#### Theorem

If there exist  $\lambda, \varepsilon > 0$  and  $1 < D < \infty$  such that for all  $k \ge 0$ (C1)  $\mathbb{E}[Y_{k+1} - Y_k \mid Y_k > a \land \mathscr{F}_k] \le -\varepsilon_0$ (C2)  $(|Y_{k+1} - Y_k| \mid \mathscr{F}_k) \prec Z$  and  $\mathbb{E}[e^{\lambda Z}] = D$ then for any  $\delta \in (0, 1)$ (2.9)  $\Pr(\tau_a > B \mid \mathscr{F}_0) \le e^{\eta(Y_0 - a)}\rho^B$ (\*)  $\Pr(\tau_b < B \mid \mathscr{F}_0) \le \frac{BD}{(1 - \rho)} \cdot e^{\eta(a - b)}$ where  $\eta := \min\{\lambda, \delta\varepsilon_0/c\}$  and  $\rho := 1 - (1 - \delta)\eta\varepsilon_0$ 

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

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

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

(C1) and (C2)  $\Longrightarrow$  (D1) and (D2) (C1)  $\mathbb{E}[Y_{k+1} - Y_k | Y_k > a \land \mathscr{F}_k] \leq -\varepsilon_0$ (C2)  $(|Y_{k+1} - Y_k| | \mathscr{F}_k) \prec Z$  and  $\mathbb{E}[e^{\lambda Z}] = D < \infty$  for a  $\lambda > 0$ . (D1)  $\mathbb{E}[e^{\eta(Y_{k+1} - Y_k)} | Y_k > a \land \mathscr{F}_k] \leq \rho$ (D2)  $\mathbb{E}[e^{\eta(Y_{k+1} - a)} | Y_k \leq a \land \mathscr{F}_k] \leq D$ 

#### Lemma

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

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

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

#### Corollary

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

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

### Reformulation of Hajek's Theorem

#### Theorem

If there exist  $\lambda, \varepsilon > 0$  and  $1 < D < \infty$  such that for all  $k \ge 0$ (C1)  $\mathbb{E}[Y_{k+1} - Y_k \mid Y_k > a \land \mathscr{F}_k] \le -\varepsilon_0$ (C2)  $(|Y_{k+1} - Y_k| \mid \mathscr{F}_k) \prec Z$  and  $\mathbb{E}[e^{\lambda Z}] = D$ then for any  $\delta \in (0, 1)$ (2.9)  $\Pr(\tau_a > B \mid \mathscr{F}_0) \le e^{\eta(Y_0 - a - B(1 - \delta)\varepsilon_0)}$ (\*)  $\Pr(\tau_b < B \mid \mathscr{F}_0) \le \frac{BD}{(1 - \delta)\eta\varepsilon_0} \cdot e^{\eta(a - b)}$ for some  $\eta \ge \min\{\lambda, \delta\varepsilon_0\lambda^2/D\}$ .

# Simplified Drift Theorem [17]

We have already seen that

(C2)  $(|Y_{k+1} - Y_k| | \mathscr{F}_k) \prec Z \text{ and } \mathbb{E}\left[e^{\lambda Z}\right] = D$ implies  $\Pr\left(|Y_{k+1} - Y_k| \ge j\right) \le De^{-\lambda j}$  for all  $j \in \mathbb{N}_0$ .

The simplified drift theorem replaces (C2) with

(S)  $\Pr(Y_{k+1} - Y_k \ge j \mid Y_k < b) \le r(n)(1+\delta)^{-j}$  for all  $j \in \mathbb{N}_0$ . and with some additional assumptions, provides a bound of type<sup>7</sup>

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

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

 $^{7}$ See [17] for the exact statement.

### 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
- ⇒ This part of the tutorial focuses on a drift theorem for populations which alleviates the state aggregation problem.

Part 6 - Population Drift

### Population-based Evolutionary Algorithms



#### Require: ,

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

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



# Population Drift



### **Central Parameters**

▶ Reproductive rate of selection mechanism  $p_{sel}$ 

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

 $\blacktriangleright$  Random walk process corresponding to variation operator  $p_{\rm mut}$ 

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

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

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

- $p_{sel}$  has reproductive rate less than  $\alpha_0$
- ▶ the random walk process corresponding to  $p_{mut}$  satisfies (C1P) and some other conditions hold,<sup>8</sup> then for some constants c, c' > 0

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

# Population Drift: Decoupling Selection & Variation

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

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

where

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

and

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

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

### Acknowledgements

- 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, ...

#### Thanks to

- David Hodge
- Carsten Witt, and
- Daniel Johannsen

for insightful discussions.

# References I

- 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.

In *GECCO '10: Proceedings of the 12th annual conference on Genetic and evolutionary computation*, pages 1449–1456, New York, NY, USA, 2010. ACM.

- Stefan Droste, Thomas Jansen, and Ingo Wegener.
   On the analysis of the (1+1) Evolutionary Algorithm. Theoretical Computer Science, 276:51–81, 2002.
- Simon Fischer, Lars Olbrich, and Berthold Vöcking. Approximating wardrop equilibria with finitely many agents. Distributed Computing, 21(2):129–139, 2008.
- 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.
- 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.
- 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 V

# 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.

#### [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.