Open paths on the hypercube

Éric Brunet

Laboratoire de Physique Statistique, ENS, UPMC, Paris

Eurandom 2014

In collaboration with Julien Berestycki and Zhan Shi (LPMA UPMC)

Population with asexual reproduction

- A genome with L loci (= location of genes)

Population with asexual reproduction

- A genome with L loci (= location of genes)

- There are two viable types (alleles) for each gene: the **wild type** (0) and the **mutated type** (1)

Population with asexual reproduction

- A genome with L loci (= location of genes)

- There are two viable types (alleles) for each gene: the **wild type** (0) and the **mutated type** (1)

Genome of a wild individual

0	0	0	0	0	0	0	0	0	0	0	0

0	0	0	0	0	0	0	0	0	1	0	0

With one mutation

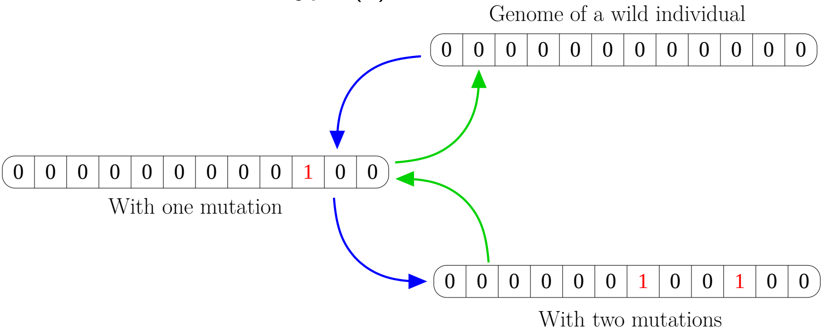
0	0	0	0	0	0	1	0	0	1	0	0

With two mutations

Population with asexual reproduction

- A genome with L loci (= location of genes)

- There are two viable types (alleles) for each gene: the **wild type** (0) and the **mutated type** (1)

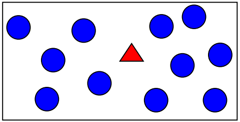


- During reproduction, when a mutation occurs, **only one gene is affected**.

$0 \longrightarrow 1$: forward mutation $\quad$ $1 \longrightarrow 0$: backward mutation

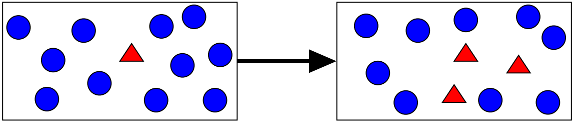
Low mutation rate, population not too large

When a mutation occurs,



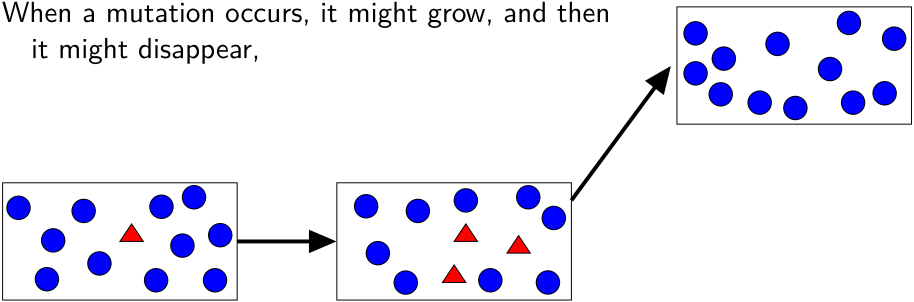
Low mutation rate, population not too large

When a mutation occurs, it might grow, and then



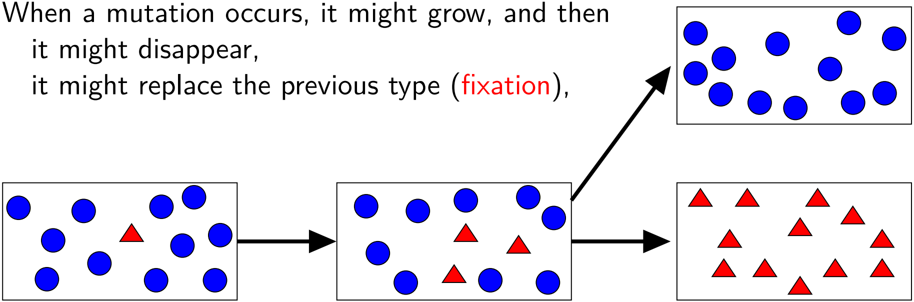
Low mutation rate, population not too large

When a mutation occurs, it might grow, and then it might disappear,



Low mutation rate, population not too large

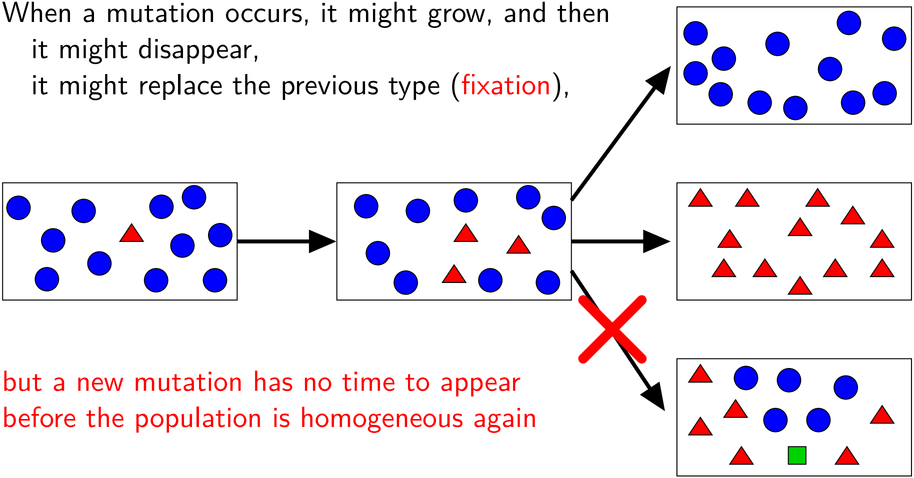
When a mutation occurs, it might grow, and then
- it might disappear,
- it might replace the previous type (fixation),



Low mutation rate, population not too large

When a mutation occurs, it might grow, and then

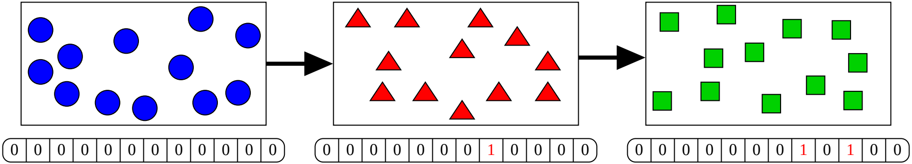
- it might disappear,
- it might replace the previous type (fixation),



but a new mutation has no time to appear before the population is homogeneous again

Evolutionary paths and Hypercube

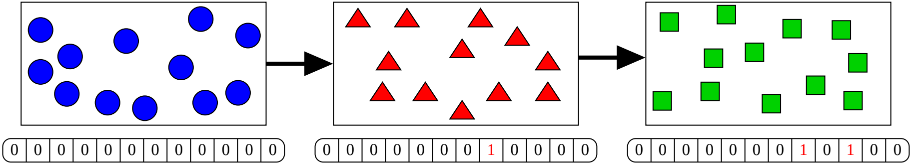
Big simplification:



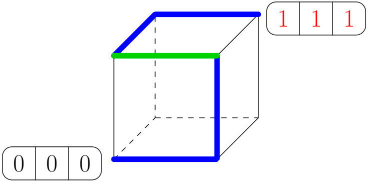
Gillespie 1983, Kauffman Levin 1987

Evolutionary paths and Hypercube

Big simplification:



Evolutionary path = walk on the hypercube



($0 \longrightarrow 1$: forward mutation $1 \longrightarrow 0$: backward mutation)

Gillespie 1983, Kauffman Levin 1987

Fitness and selection

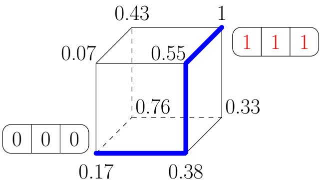
Evolutionary path = walk on the hypercube

- To each of the 2^L genomes one associates a fitness value
- Assume strong selection

Fitness and selection

Evolutionary path = walk on the hypercube

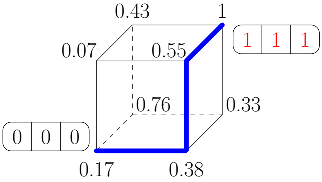
- To each of the 2^L genomes one associates a fitness value
- Assume strong selection
- A transition (= a mutation fixates) may occur only if the fitness value increases



Open or **accessible** evolutionary path =
walk on the hypercube such that fitness values increase along the walk

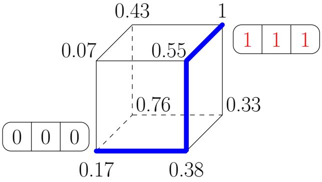
Choosing the fitness values

- Flat landscape: fitness value proportional to number of mutations. All forward paths are accessible.
- Rough landscape: no clear relationship between fitness value and number of mutations. Lots of local extrema, valleys and dead ends.



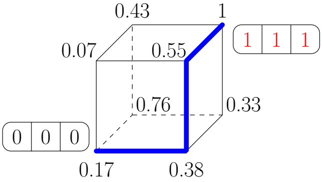
Choosing the fitness values

- Flat landscape: fitness value proportional to number of mutations. All forward paths are accessible.
- Rough landscape: no clear relationship between fitness value and number of mutations. Lots of local extrema, valleys and dead ends.



Choosing the fitness values

- Flat landscape: fitness value proportional to number of mutations. All forward paths are accessible.
- Rough landscape: no clear relationship between fitness value and number of mutations. Lots of local extrema, valleys and dead ends.



Roughest landscape of all

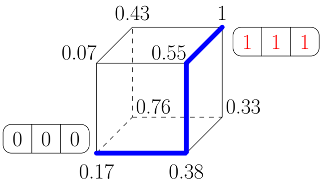
the House of Cards model

Fitness values are independent random numbers

Kingman 1978

Choosing the fitness values

- Flat landscape: fitness value proportional to number of mutations. All forward paths are accessible.
- Rough landscape: no clear relationship between fitness value and number of mutations. Lots of local extrema, valleys and dead ends.



Roughest landscape of all

the House of Cards model

Fitness values are independent random numbers

Kingman 1978

The question: can the population reach the fittest possible state?

Summary of the model

For an asexual population which is not too large, has low mutation rate and where selection is high, if one assumes that fitness values are distributed according to the House of Cards model, is there an *accessible* or open evolutionary path which leads to the fittest possible genome?

Summary of the model

For an asexual population which is not too large, has low mutation rate and where selection is high, if one assumes that fitness values are distributed according to the House of Cards model, is there an *accessible* or open evolutionary path which leads to the fittest possible genome?

- Consider a L-hypercube.
- Each site is assigned an independent random value, its fitness.
- A path is said to be open if the fitness values increase along it.
- One starts from site $(0, 0, 0, \dots, 0)$.

Is there an open path to the fittest site ? And how many are there ?

Summary of the model

For an asexual population which is not too large, has low mutation rate and where selection is high, if one assumes that fitness values are distributed according to the House of Cards model, is there an *accessible* or open evolutionary path which leads to the fittest possible genome?

- Consider a L-hypercube.
- Each site is assigned an independent random value, its fitness.
- A path is said to be open if the fitness values increase along it.
- One starts from site $(0, 0, 0, \dots, 0)$.

Is there an open path to the fittest site ? And how many are there ?

Remarks:

- The answer does not depend on the (continuous) distribution
- The fittest site is uniformly chosen among the 2^L sites

Summary of the model

For an asexual population which is not too large, has low mutation rate and where selection is high, if one assumes that fitness values are distributed according to the House of Cards model, is there an *accessible* or open evolutionary path which leads to the fittest possible genome?

- Consider a L-hypercube.
- Each site is assigned an independent random value, its fitness.
- A path is said to be open if the fitness values increase along it.
- One starts from site $(0, 0, 0, \dots, 0)$.

Is there an open path to the fittest site ? And how many are there ?

Remarks:

- The answer does not depend on the (continuous) distribution
- The fittest site is uniformly chosen among the 2^L sites

Summary of the model

For an asexual population which is not too large, has low mutation rate and where selection is high, if one assumes that fitness values are distributed according to the House of Cards model, is there an *accessible* or open evolutionary path which leads to the fittest possible genome?

- Consider a L-hypercube.
- Each site is assigned an independent random value, its fitness.
- Choose location of the fittest site; give it a fitness value 1
- A path is said to be open if the fitness values increase along it.
- One starts from site $(0, 0, 0, \dots, 0)$.

Is there an open path to the fittest site ? And how many are there ?

Remarks:

- The answer does not depend on the (continuous) distribution
- The fittest site is uniformly chosen among the 2^L sites

Summary of the model

For an asexual population which is not too large, has low mutation rate and where selection is high, if one assumes that fitness values are distributed according to the House of Cards model, is there an *accessible* or open evolutionary path which leads to the fittest possible genome?

- Consider a L-hypercube.
- Each site is assigned an independent random value, its fitness.
- Choose location of the fittest site; give it a fitness value 1
- The other sites get independent uniform fitness values between 0 and 1
- A path is said to be open if the fitness values increase along it.
- One starts from site $(0, 0, 0, \ldots, 0)$.

Is there an open path to the fittest site ? And how many are there ?

Remarks:

- The answer does not depend on the (continuous) distribution
- The fittest site is uniformly chosen among the 2^L sites

Results

- When one allows only forward mutations
- When one allows both forward and backward mutations

Only forward mutations

- No backward mutation, only $0 \to 1$ and never $1 \to 0$, path length is L.
- Starting from $(0, 0, 0, \ldots, 0)$, assume fittest site is $(1, 1, 1, \ldots, 1)$

Nowak Krug 2013, Hegarty Martinsson 2012

Only forward mutations

- No backward mutation, only $0 \to 1$ and never $1 \to 0$, path length is L.
- Starting from $(0, 0, 0, \ldots, 0)$, assume fittest site is $(1, 1, 1, \ldots, 1)$
- Total number of paths is $L!$

Nowak Krug 2013, Hegarty Martinsson 2012

Only forward mutations

- No backward mutation, only $0 \to 1$ and never $1 \to 0$, path length is L.
- Starting from $(0, 0, 0, \ldots, 0)$, assume fittest site is $(1, 1, 1, \ldots, 1)$
- Total number of paths is $L!$
- Probability a given path is open is $1/L!$

Nowak Krug 2013, Hegarty Martinsson 2012

Only forward mutations

- No backward mutation, only $0 \to 1$ and never $1 \to 0$, path length is L.
- Starting from $(0, 0, 0, \dots, 0)$, assume fittest site is $(1, 1, 1, \dots, 1)$
- Total number of paths is $L!$
- Probability a given path is open is $1/L!$

$$\boxed{\mathbb{E}(\text{nb of open paths}) = 1}$$

But...

Nowak Krug 2013, Hegarty Martinsson 2012

Only forward mutations

- No backward mutation, only $0 \to 1$ and never $1 \to 0$, path length is L.
- Starting from $(0, 0, 0, \ldots, 0)$, assume fittest site is $(1, 1, 1, \ldots, 1)$
- Total number of paths is $L!$
- Probability a given path is open is $1/L!$

$$\boxed{\mathbb{E}(\text{nb of open paths}) = 1}$$

But... Conditionally on the event that starting position has given fitness x

Nowak Krug 2013, Hegarty Martinsson 2012

Only forward mutations

- No backward mutation, only $0 \to 1$ and never $1 \to 0$, path length is L.
- Starting from $(0, 0, 0, \dots, 0)$, assume fittest site is $(1, 1, 1, \dots, 1)$
- Total number of paths is $L!$
- Probability a given path is open is $1/L!$

$$\boxed{\mathbb{E}(\text{nb of open paths}) = 1}$$

But... Conditionally on the event that starting position has given fitness x

- Probability a given path is open is $(1 - x)^{L-1}/(L - 1)!$

Nowak Krug 2013, Hegarty Martinsson 2012

Only forward mutations

- No backward mutation, only $0 \to 1$ and never $1 \to 0$, path length is L.
- Starting from $(0, 0, 0, \ldots, 0)$, assume fittest site is $(1, 1, 1, \ldots, 1)$
- Total number of paths is $L!$
- Probability a given path is open is $1/L!$

$$\boxed{\mathbb{E}(\text{nb of open paths}) = 1}$$

But... Conditionally on the event that starting position has given fitness x

- Probability a given path is open is $(1 - x)^{L-1}/(L - 1)!$

$$\boxed{\mathbb{E}^x(\text{nb of open paths}) = L(1 - x)^{L-1}}$$

Nowak Krug 2013, Hegarty Martinsson 2012

Only forward mutations

- No backward mutation, only $0 \to 1$ and never $1 \to 0$, path length is L.
- Starting from $(0,0,0,\dots,0)$, assume fittest site is $(1,1,1,\dots,1)$
- Total number of paths is $L!$
- Probability a given path is open is $1/L!$

$$\boxed{\mathbb{E}(\text{nb of open paths}) = 1}$$

But... Conditionally on the event that starting position has given fitness x

- Probability a given path is open is $(1-x)^{L-1}/(L-1)!$

$$\boxed{\mathbb{E}^x(\text{nb of open paths}) = L(1-x)^{L-1}} \quad \begin{cases} \propto L & \text{If } x \lesssim \frac{1}{L} \\ \propto 1 & \text{If } x \approx \frac{\ln L}{L} \\ \ll 1 & \text{If } x \gg \frac{\ln L}{L} \end{cases}$$

Nowak Krug 2013, Hegarty Martinsson 2012

Only forward mutations

- No backward mutation, only $0 \to 1$ and never $1 \to 0$, path length is L.
- Starting from $(0,0,0,\ldots,0)$, assume fittest site is $(1,1,1,\ldots,1)$
- Total number of paths is $L!$
- Probability a given path is open is $1/L!$

$$\boxed{\mathbb{E}(\text{nb of open paths}) = 1}$$

But... Conditionally on the event that starting position has given fitness x

- Probability a given path is open is $(1-x)^{L-1}/(L-1)!$

$$\boxed{\mathbb{E}^x(\text{nb of open paths}) = L(1-x)^{L-1}} \quad \begin{cases} \propto L & \text{If } x \lessapprox \frac{1}{L} \\ \propto 1 & \text{If } x \approx \frac{\ln L}{L} \\ \ll 1 & \text{If } x \gg \frac{\ln L}{L} \end{cases}$$

$$\mathbb{P}(\text{nb of open paths} \neq 0) \leq \frac{\ln L + \text{Cste}}{L}$$

Nowak Krug 2013, Hegarty Martinsson 2012

Only forward mutations

Assume fittest site is $(1, 1, 1, \ldots, 1)$.

Theorem (Hegarty-Martinsson 2012)

As $L \to \infty$,

$$\mathbb{P}(\text{nb of open paths} \neq 0) \sim \frac{\ln L}{L},$$

with a sharp transition for existence of paths around starting fitness $\dfrac{\ln L}{L}$

Only forward mutations

Assume fittest site is $(1, 1, 1, \dots, 1)$.

Theorem (Hegarty-Martinsson 2012)

As $L \to \infty$,

$$\mathbb{P}(\text{nb of open paths} \neq 0) \sim \frac{\ln L}{L},$$

with a sharp transition for existence of paths around starting fitness $\dfrac{\ln L}{L}$

If $a(L) \to \infty$ (but, typically, $a(L) \ll \ln L$),

$$\mathbb{P}^{\frac{\ln L - a(L)}{L}}(\text{nb of open paths} \neq 0) \to 1 \qquad \left(\begin{array}{c}\text{If starting position has a fitness below} \\ (\ln L)/L\text{, there are some open paths.}\end{array}\right)$$

$$\mathbb{P}^{\frac{\ln L + a(L)}{L}}(\text{nb of open paths} \neq 0) \to 0 \qquad \left(\begin{array}{c}\text{If starting position has a fitness above} \\ (\ln L)/L\text{, there are no open paths.}\end{array}\right)$$

Only forward mutations — summary

Assume fittest site is $(1, 1, 1, \ldots, 1)$.

$\mathbb{E}(\text{nb of open paths}) = 1$	(a lie: typical nb of open paths $\neq 1$)
$\mathbb{E}^x(\text{nb of open paths}) = L(1-x)^{L-1}$	(truth: correct order of magnitude)
$\mathbb{P}(\text{nb of open paths} \neq 0) \sim \dfrac{\ln L}{L}$	(value of x for which $\mathbb{E}^x(\ldots) \approx 1$)

Only forward mutations — summary

Assume fittest site is $(1, 1, 1, \ldots, 1)$.

$\mathbb{E}(\text{nb of open paths}) = 1$	(a lie: typical nb of open paths $\neq 1$)
$\mathbb{E}^x(\text{nb of open paths}) = L(1-x)^{L-1}$	(truth: correct order of magnitude)
$\mathbb{P}(\text{nb of open paths} \neq 0) \sim \dfrac{\ln L}{L}$	(value of x for which $\mathbb{E}^x(\ldots) \approx 1$)

Theorem (Berestycki-Brunet-Shi 2013)

If $x = \frac{X}{L}$, as $L \to \infty$,

$$\frac{\text{nb of open paths}}{L} \xrightarrow{\text{in law}} e^{-X} \times \mathcal{E} \times \mathcal{E}'$$

where $\mathcal{E}$ and $\mathcal{E}'$ are two independent exponential numbers.

Results

- When one allows only forward mutations
- When one allows both forward and backward mutations

Results

- When one allows only forward mutations
- When one allows both forward and backward mutations

For large L, when the location of the fittest site is random

- There are no open paths is starting fitness is larger than $0.27818\ldots$.
- There are open paths otherwise. (Only a conjecture!)

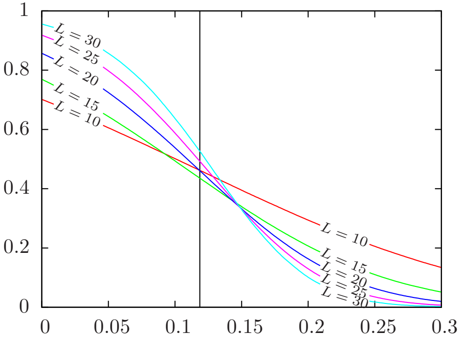
Results

- When one allows only forward mutations
- When one allows both forward and backward mutations

For large L, when the location of the fittest site is random

- There are no open paths is starting fitness is larger than $0.27818\ldots$.
- There are open paths otherwise. (Only a conjecture!)

When fittest site is $(1, 1, 1, \ldots, 1)$, critical value is $0.11863\ldots$



Paths with forward and backward mutations

We allow paths to do $0 \to 1$ or $1 \to 0$. Assume fittest site is $(1, 1, 1, \dots, 1)$.

0 backstep	length L
1 backstep	length $L + 2$
2 backsteps	length $L + 4$
p backsteps	length $L + 2p$

Paths with forward and backward mutations

We allow paths to do $0 \to 1$ or $1 \to 0$. Assume fittest site is $(1, 1, 1, \ldots, 1)$.

		nb of self-avoiding paths
0 backstep	length L	$a_{L,0} = L!$
1 backstep	length $L+2$	
2 backsteps	length $L+4$	
p backsteps	length $L+2p$	

Paths with forward and backward mutations

We allow paths to do $0 \to 1$ or $1 \to 0$. Assume fittest site is $(1, 1, 1, \ldots, 1)$.

		nb of self-avoiding paths
0 backstep	length L	$a_{L,0} = L!$
1 backstep	length $L+2$	$a_{L,1} = L! \times \frac{L(L-1)(L-2)}{6}$
2 backsteps	length $L+4$	
p backsteps	length $L+2p$	

Paths with forward and backward mutations

We allow paths to do $0 \to 1$ or $1 \to 0$. Assume fittest site is $(1, 1, 1, \ldots, 1)$.

nb of self-avoiding paths

0 backstep	length L	$a_{L,0} = L!$
1 backstep	length $L + 2$	$a_{L,1} = L! \times \frac{L(L-1)(L-2)}{6}$
2 backsteps	length $L + 4$	$a_{L,2} = L! \times \frac{(L-1)(L-2)(5L^4+3L^3+34L^2-264L+180)}{360}$
p backsteps	length $L + 2p$	

Paths with forward and backward mutations

We allow paths to do $0 \to 1$ or $1 \to 0$. Assume fittest site is $(1, 1, 1, \ldots, 1)$.

		nb of self-avoiding paths
0 backstep	length L	$a_{L,0} = L!$
1 backstep	length $L+2$	$a_{L,1} = L! \times \frac{L(L-1)(L-2)}{6}$
2 backsteps	length $L+4$	$a_{L,2} = L! \times \frac{(L-1)(L-2)(5L^4+3L^3+34L^2-264L+180)}{360}$
p backsteps	length $L+2p$	$a_{L,p} \sim L! \times \frac{L^{3p}}{6^p p!}$ (p fixed, L large)

Paths with forward and backward mutations

We allow paths to do $0 \to 1$ or $1 \to 0$. Assume fittest site is $(1, 1, 1, \dots, 1)$.

		nb of self-avoiding paths
0 backstep	length L	$a_{L,0} = L!$
1 backstep	length $L+2$	$a_{L,1} = L! \times \frac{L(L-1)(L-2)}{6}$
2 backsteps	length $L+4$	$a_{L,2} = L! \times \frac{(L-1)(L-2)(5L^4+3L^3+34L^2-264L+180)}{360}$
p backsteps	length $L+2p$	$a_{L,p} \sim L! \times \frac{L^{3p}}{6^p p!}$ (p fixed, L large)

$a_L = a_{L,0} + a_{L,1} + a_{L,2} + \cdots =$ total nb of self-avoiding paths.

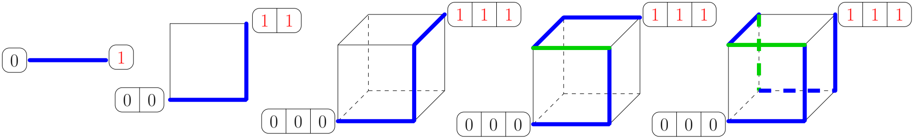
Paths with forward and backward mutations

We allow paths to do $0 \to 1$ or $1 \to 0$. Assume fittest site is $(1, 1, 1, \dots, 1)$.

nb of self-avoiding paths

0 backstep	length L	$a_{L,0} = L!$
1 backstep	length $L+2$	$a_{L,1} = L! \times \frac{L(L-1)(L-2)}{6}$
2 backsteps	length $L+4$	$a_{L,2} = L! \times \frac{(L-1)(L-2)(5L^4+3L^3+34L^2-264L+180)}{360}$
p backsteps	length $L+2p$	$a_{L,p} \sim L! \times \frac{L^{3p}}{6^p p!}$ (p fixed, L large)

$a_L = a_{L,0} + a_{L,1} + a_{L,2} + \cdots =$ total nb of self-avoiding paths.



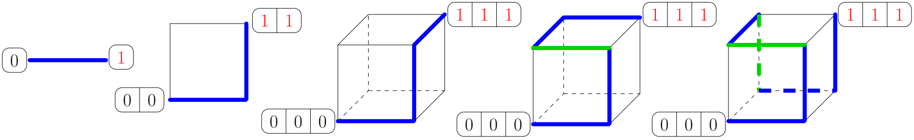
$a_1 = 1, \quad a_2 = 2, \quad a_3 = 18$

Paths with forward and backward mutations

We allow paths to do $0 \to 1$ or $1 \to 0$. Assume fittest site is $(1, 1, 1, \ldots, 1)$.

		nb of self-avoiding paths
0 backstep	length L	$a_{L,0} = L!$
1 backstep	length $L+2$	$a_{L,1} = L! \times \frac{L(L-1)(L-2)}{6}$
2 backsteps	length $L+4$	$a_{L,2} = L! \times \frac{(L-1)(L-2)(5L^4+3L^3+34L^2-264L+180)}{360}$
p backsteps	length $L+2p$	$a_{L,p} \sim L! \times \frac{L^{3p}}{6^p p!}$ $\quad$ (p fixed, L large)

$a_L = a_{L,0} + a_{L,1} + a_{L,2} + \cdots =$ total nb of self-avoiding paths.



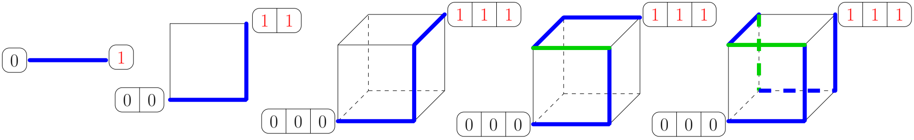
$$a_1 = 1, \quad a_2 = 2, \quad a_3 = 18, \quad a_4 = 6\,432, \quad a_5 = 18\,651\,552\,840$$

Paths with forward and backward mutations

We allow paths to do $0 \to 1$ or $1 \to 0$. Assume fittest site is $(1, 1, 1, \ldots, 1)$.

		nb of self-avoiding paths
0 backstep	length L	$a_{L,0} = L!$
1 backstep	length $L+2$	$a_{L,1} = L! \times \frac{L(L-1)(L-2)}{6}$
2 backsteps	length $L+4$	$a_{L,2} = L! \times \frac{(L-1)(L-2)(5L^4+3L^3+34L^2-264L+180)}{360}$
p backsteps	length $L+2p$	$a_{L,p} \sim L! \times \frac{L^{3p}}{6^p p!}$ (p fixed, L large)

$a_L = a_{L,0} + a_{L,1} + a_{L,2} + \cdots =$ total nb of self-avoiding paths.



$$a_1 = 1, \quad a_2 = 2, \quad a_3 = 18, \quad a_4 = 6\,432, \quad a_5 = 18\,651\,552\,840$$

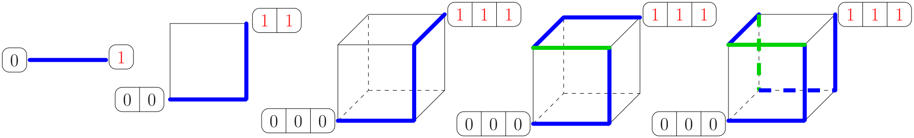
Asymptotically, $e^{c \times 2^L} \leq a_L \leq e^{c' \times (\ln L) 2^L}$

Paths with forward and backward mutations

We allow paths to do $0 \to 1$ or $1 \to 0$. Assume fittest site is $(1, 1, 1, \ldots, 1)$.

		nb of self-avoiding paths
0 backstep	length L	$a_{L,0} = L!$
1 backstep	length $L+2$	$a_{L,1} = L! \times \frac{L(L-1)(L-2)}{6}$
2 backsteps	length $L+4$	$a_{L,2} = L! \times \frac{(L-1)(L-2)(5L^4+3L^3+34L^2-264L+180)}{360}$
p backsteps	length $L+2p$	$a_{L,p} \sim L! \times \frac{L^{3p}}{6^p p!}$ (p fixed, L large)

$a_L = a_{L,0} + a_{L,1} + a_{L,2} + \cdots =$ total nb of self-avoiding paths.



$$a_1 = 1, \quad a_2 = 2, \quad a_3 = 18, \quad a_4 = 6\,432, \quad a_5 = 18\,651\,552\,840$$

Asymptotically, $e^{c \times 2^L} \leq a_L \leq e^{c' \times (\ln L) 2^L}$

How many are open ?

Paths with forward and backward mutations

Fittest site is $(1, 1, 1, \ldots, 1)$

$$\boxed{\mathbb{E}(\text{nb of open paths}) = \sum_p a_{L,p} \frac{1}{(L+2p)!}}$$

But. . .

Paths with forward and backward mutations

Fittest site is $(1, 1, 1, \ldots, 1)$

$$\mathbb{E}(\text{nb of open paths}) = \sum_p a_{L,p} \frac{1}{(L+2p)!}$$

But... Conditionally on the event that starting position has given fitness x

Paths with forward and backward mutations

Fittest site is $(1, 1, 1, \dots, 1)$

$$\mathbb{E}(\text{nb of open paths}) = \sum_p a_{L,p} \frac{1}{(L+2p)!}$$

But... Conditionally on the event that starting position has given fitness x

$$\mathbb{E}^x(\text{nb of open paths}) = \sum_p a_{L,p} \frac{(1-x)^{L+2p-1}}{(L+2p-1)!}$$

Paths with forward and backward mutations

Fittest site is $(1, 1, 1, \ldots, 1)$

$$\mathbb{E}(\text{nb of open paths}) = \sum_p a_{L,p} \frac{1}{(L+2p)!}$$

But... Conditionally on the event that starting position has given fitness x

$$\mathbb{E}^x(\text{nb of open paths}) = \sum_p a_{L,p} \frac{(1-x)^{L+2p-1}}{(L+2p-1)!}$$

Theorem (Berestycki-Brunet-Shi 2013)

$$\Big[\mathbb{E}^x(\text{nb of open paths})\Big]^{1/L} \xrightarrow[L\to\infty]{} \sinh(1-x).$$

Corollary: if $x > \underbrace{1 - \sinh^{-1}(1)}_{0.11863\ldots}$, $\quad \mathbb{P}^x(\text{nb of open paths} \neq 0) \to 0.$

Unproved speculation: $\mathbb{E}^x(\text{nb of open paths}) \sim \phi(x) L \big[\sinh(1-x)\big]^L.$

Generalization and conjecture

Fittest site is $(1,1,1,\dots,1)$: $\left[\mathbb{E}^x(\text{nb of open paths})\right]^{\frac{1}{L}} \to \sinh(1-x)$

$$\text{No open path if } x > \underbrace{1-\sinh^{-1}(1)}_{0.11863\ldots}$$

Generalization and conjecture

Fittest site is $(1,1,1,\dots,1)$: $\left[\mathbb{E}^x(\text{nb of open paths})\right]^{\frac{1}{L}} \to \sinh(1-x)$

$$\text{No open path if } x > \underbrace{1-\sinh^{-1}(1)}_{0.11863\ldots}$$

Fittest site at distance αL from $(0,0,0,\dots,0)$:

$$\left[\mathbb{E}^x(\text{nb of open paths})\right]^{\frac{1}{L}} \to \sinh(1-x)^{\alpha}\cosh(1-x)^{1-\alpha}$$

Generalization and conjecture

Fittest site is $(1,1,1,\dots,1)$: $\left[\mathbb{E}^x(\text{nb of open paths})\right]^{\frac{1}{L}} \to \sinh(1-x)$

$$\text{No open path if } x > \underbrace{1-\sinh^{-1}(1)}_{0.11863\ldots}$$

Fittest site at distance αL from $(0,0,0,\dots,0)$:

$$\left[\mathbb{E}^x(\text{nb of open paths})\right]^{\frac{1}{L}} \to \sinh(1-x)^{\alpha}\cosh(1-x)^{1-\alpha}$$

Fittest site is randomly chosen: $\left[\mathbb{E}^x(\text{nb of open paths})\right]^{\frac{1}{L}} \to \sqrt{\dfrac{\sinh(2-2x)}{2}}$

Generalization and conjecture

Fittest site is $(1,1,1,\dots,1)$: $\left[\mathbb{E}^x(\text{nb of open paths})\right]^{\frac{1}{L}} \to \sinh(1-x)$

$$\text{No open path if } x > \underbrace{1-\sinh^{-1}(1)}_{0.11863\ldots}$$

Fittest site at distance αL from $(0,0,0,\dots,0)$:

$$\left[\mathbb{E}^x(\text{nb of open paths})\right]^{\frac{1}{L}} \to \sinh(1-x)^{\alpha}\cosh(1-x)^{1-\alpha}$$

Fittest site is randomly chosen: $\left[\mathbb{E}^x(\text{nb of open paths})\right]^{\frac{1}{L}} \to \sqrt{\dfrac{\sinh(2-2x)}{2}}$

$$\text{No open path if } x > \underbrace{1-\frac{1}{2}\sinh^{-1}(2)}_{0.27818\ldots}$$

Generalization and conjecture

Fittest site is $(1,1,1,\dots,1)$: $\left[\mathbb{E}^x(\text{nb of open paths})\right]^{\frac{1}{L}} \to \sinh(1-x)$

$$\text{No open path if } x > \underbrace{1-\sinh^{-1}(1)}_{0.11863\ldots}$$

Fittest site at distance αL from $(0,0,0,\dots,0)$:

$$\left[\mathbb{E}^x(\text{nb of open paths})\right]^{\frac{1}{L}} \to \sinh(1-x)^{\alpha}\cosh(1-x)^{1-\alpha}$$

Fittest site is randomly chosen: $\left[\mathbb{E}^x(\text{nb of open paths})\right]^{\frac{1}{L}} \to \sqrt{\dfrac{\sinh(2-2x)}{2}}$

$$\text{No open path if } x > \underbrace{1-\frac{1}{2}\sinh^{-1}(2)}_{0.27818\ldots}$$

Conjecture

Expectations are telling the truth. $\mathbb{P}^x(\text{nb of open paths} \neq 0) \to 1$ if $x < x^*$ with x^* given above. Furthermore, $\mathbb{P}(\text{nb of open paths} \neq 0) \to x^*$

Outline of proof

Forward and backward mutations, fittest is $(1, 1, 1, \dots, 1)$.

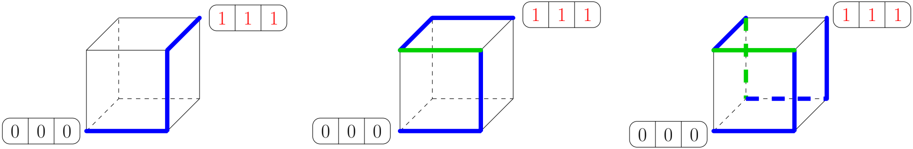
$$\left[\mathbb{E}^x(\text{nb of open paths})\right]^{\frac{1}{L}} = \left[\sum_p a_{L,p} \frac{(1-x)^{L+2p-1}}{(L+2p-1)!}\right]^{\frac{1}{L}} \to \sinh(1-x)$$

Outline of proof

Forward and backward mutations, fittest is $(1, 1, 1, \ldots, 1)$.

$$\left[\mathbb{E}^x(\text{nb of open paths})\right]^{\frac{1}{L}} = \left[\sum_p a_{L,p} \frac{(1-x)^{L+2p-1}}{(L+2p-1)!}\right]^{\frac{1}{L}} \to \sinh(1-x)$$

- Code paths as sequence of numbers



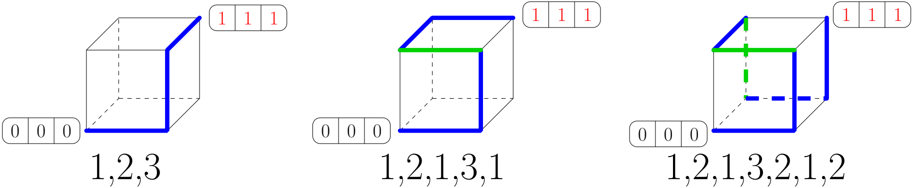
1,2,3   1,2,1,3,1   1,2,1,3,2,1,2

Outline of proof

Forward and backward mutations, fittest is $(1, 1, 1, \dots, 1)$.

$$\left[\mathbb{E}^x(\text{nb of open paths})\right]^{\frac{1}{L}} = \left[\sum_p a_{L,p} \frac{(1-x)^{L+2p-1}}{(L+2p-1)!}\right]^{\frac{1}{L}} \to \sinh(1-x)$$

- Code paths as sequence of numbers



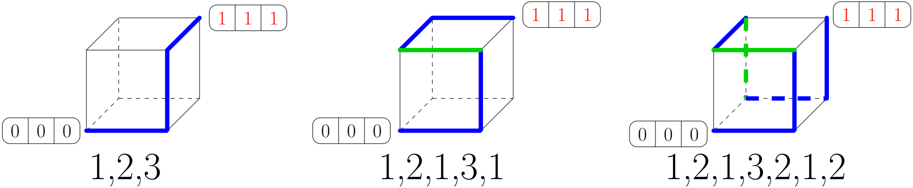
- A path in $a_{L,p}$ has a sequence of length $L + 2p$

Outline of proof

Forward and backward mutations, fittest is $(1, 1, 1, \ldots, 1)$.

$$\left[\mathbb{E}^x(\text{nb of open paths})\right]^{\frac{1}{L}} = \left[\sum_p a_{L,p} \frac{(1-x)^{L+2p-1}}{(L+2p-1)!}\right]^{\frac{1}{L}} \to \sinh(1-x)$$

- Code paths as sequence of numbers



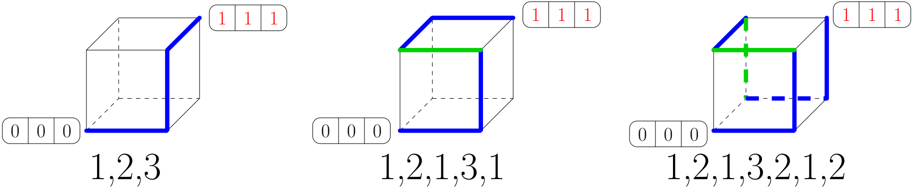
- A path in $a_{L,p}$ has a sequence of length $L + 2p$
- A path reaches $(1, 1, 1, \ldots, 1)$ if each number between 1 and L appears oddly many times in the sequence

Outline of proof

Forward and backward mutations, fittest is $(1, 1, 1, \dots, 1)$.

$$\left[\mathbb{E}^x(\text{nb of open paths})\right]^{\frac{1}{L}} = \left[\sum_p a_{L,p} \frac{(1-x)^{L+2p-1}}{(L+2p-1)!}\right]^{\frac{1}{L}} \to \sinh(1-x)$$

- Code paths as sequence of numbers



- A path in $a_{L,p}$ has a sequence of length $L + 2p$
- A path reaches $(1, 1, 1, \dots, 1)$ if each number between 1 and L appears oddly many times in the sequence
- A path is self-avoiding if in any non-empty substring, at least one number appears oddly many times

Outline of proof

Strategy: $m_{L,p} \leq a_{L,p} \leq M_{L,p}$

Outline of proof

Strategy: $m_{L,p} \leq a_{L,p} \leq M_{L,p}$

- $M_{L,p}$ counts all the paths of length $L + 2p$ to $(1, 1, 1, \ldots, 1)$, even self-intersecting ones

Outline of proof

Strategy: $m_{L,p} \leq a_{L,p} \leq M_{L,p}$

- $M_{L,p}$ counts all the paths of length $L + 2p$ to $(1, 1, 1, \dots, 1)$, even self-intersecting ones
- $m_{L,p}$ counts all the paths such that L does not appear on two consecutive positions and such that if all the L are removed, what remains is in $m_{L-1,p'}$

Outline of proof

Strategy: $m_{L,p} \leq a_{L,p} \leq M_{L,p}$

- $M_{L,p}$ counts all the paths of length $L+2p$ to $(1,1,1,\dots,1)$, even self-intersecting ones
- $m_{L,p}$ counts all the paths such that L does not appear on two consecutive positions and such that if all the L are removed, what remains is in $m_{L-1,p'}$

Example

$m_{3,0}$:	123 132 213 231 312 321
$a_{3,0}$:	123 132 213 231 312 321
$M_{3,0}$:	123 132 213 231 312 321

Outline of proof

Strategy: $m_{L,p} \leq a_{L,p} \leq M_{L,p}$

- $M_{L,p}$ counts all the paths of length $L + 2p$ to $(1, 1, 1, \ldots, 1)$, even self-intersecting ones
- $m_{L,p}$ counts all the paths such that L does not appear on two consecutive positions and such that if all the L are removed, what remains is in $m_{L-1,p'}$

Example

$m_{3,0}$:	123 132 213 231 312 321
$a_{3,0}$:	123 132 213 231 312 321
$M_{3,0}$:	123 132 213 231 312 321

$m_{3,1}$:	31323 32313
$a_{3,1}$:	12131 13121 21232 23212 31323 32313
$M_{3,1}$:	12131 13121 21232 23212 31323 32313 11123 12113 ...

Outline of proof

Strategy: $m_{L,p} \leq a_{L,p} \leq M_{L,p}$

- $M_{L,p}$ counts all the paths of length $L + 2p$ to $(1, 1, 1, \ldots, 1)$, even self-intersecting ones
- $m_{L,p}$ counts all the paths such that L does not appear on two consecutive positions and such that if all the L are removed, what remains is in $m_{L-1,p'}$

Example

$m_{3,0}$:	123 132 213 231 312 321
$a_{3,0}$:	123 132 213 231 312 321
$M_{3,0}$:	123 132 213 231 312 321

$m_{3,1}$:	31323 32313
$a_{3,1}$:	12131 13121 21232 23212 31323 32313
$M_{3,1}$:	12131 13121 21232 23212 31323 32313 11123 12113 ...

$m_{3,2}$:	
$a_{3,2}$:	1213212 1312313 2123121 2321323 3132131 3231232
$M_{3,2}$:	1213212 1312313 2123121 2321323 3132131 3231232 1211333 ...

$(M_{3,1} = 60,\ M_{3,2} = 4920\ldots)$

Outline of proof

Strategy: $m_{L,p} \leq a_{L,p} \leq M_{L,p}$

The paths in $m_{L,p}$ (resp. $M_{L,p}$) have the property that if all occurrence of L is removed, what remains is in $m_{L-1,p'}$ (resp. $M_{L-1,p'}$).

Outline of proof

Strategy: $m_{L,p} \leq a_{L,p} \leq M_{L,p}$

The paths in $m_{L,p}$ (resp. $M_{L,p}$) have the property that if all occurrence of L is removed, what remains is in $m_{L-1,p'}$ (resp. $M_{L-1,p'}$).

$$M_{L+1,p} = \sum_{q=0}^{p} \binom{L+1+2p}{2q+1} M_{L,p-q},$$

($M_{L+1,p}$: length $L+1+2p$. Number $L+1$ appears $2q+1$ times (odd). Fill in the remaining with a path in $M_{L,p-q}$ of length $L+2p-2q$.)

Outline of proof

Strategy: $m_{L,p} \leq a_{L,p} \leq M_{L,p}$

The paths in $m_{L,p}$ (resp. $M_{L,p}$) have the property that if all occurrence of L is removed, what remains is in $m_{L-1,p'}$ (resp. $M_{L-1,p'}$).

$$M_{L+1,p} = \sum_{q=0}^{p} \binom{L+1+2p}{2q+1} M_{L,p-q}, \quad m_{L+1,p} = \sum_{q=0}^{p} \begin{bmatrix} L+1+2p \\ 2q+1 \end{bmatrix} m_{L,p-q}$$

($M_{L+1,p}$: length $L+1+2p$. Number $L+1$ appears $2q+1$ times (odd). Fill in the remaining with a path in $M_{L,p-q}$ of length $L+2p-2q$.)

$$\begin{bmatrix} N \\ P \end{bmatrix} := \begin{pmatrix} \text{nb of ways of choosing } P \text{ items out of } N \\ \text{without taking two consecutive items} \end{pmatrix}$$

Outline of proof

Strategy: $m_{L,p} \leq a_{L,p} \leq M_{L,p}$

The paths in $m_{L,p}$ (resp. $M_{L,p}$) have the property that if all occurrence of L is removed, what remains is in $m_{L-1,p'}$ (resp. $M_{L-1,p'}$).

$$M_{L+1,p} = \sum_{q=0}^{p} \binom{L+1+2p}{2q+1} M_{L,p-q}, \quad m_{L+1,p} = \sum_{q=0}^{p} \begin{bmatrix} L+1+2p \\ 2q+1 \end{bmatrix} m_{L,p-q}$$

($M_{L+1,p}$: length $L+1+2p$. Number $L+1$ appears $2q+1$ times (odd). Fill in the remaining with a path in $M_{L,p-q}$ of length $L+2p-2q$.)

$$\begin{bmatrix} N \\ P \end{bmatrix} := \begin{pmatrix} \text{nb of ways of choosing } P \text{ items out of } N \\ \text{without taking two consecutive items} \end{pmatrix} = \binom{N-P+1}{P}$$

Outline of proof

Strategy: $m_{L,p} \le a_{L,p} \le M_{L,p}$

The paths in $m_{L,p}$ (resp. $M_{L,p}$) have the property that if all occurrence of L is removed, what remains is in $m_{L-1,p'}$ (resp. $M_{L-1,p'}$).

$$M_{L+1,p} = \sum_{q=0}^{p} \binom{L+1+2p}{2q+1} M_{L,p-q}, \quad m_{L+1,p} = \sum_{q=0}^{p} \begin{bmatrix} L+1+2p \\ 2q+1 \end{bmatrix} m_{L,p-q}$$

($M_{L+1,p}$: length $L+1+2p$. Number $L+1$ appears $2q+1$ times (odd). Fill in the remaining with a path in $M_{L,p-q}$ of length $L+2p-2q$.)

$$\begin{bmatrix} N \\ P \end{bmatrix} := \begin{pmatrix} \text{nb of ways of choosing } P \text{ items out of } N \\ \text{without taking two consecutive items} \end{pmatrix} = \binom{N-P+1}{P}$$

$$G_L(X) := \sum_p M_{L,p} \frac{X^{L+2p}}{(L+2p)!}$$

Outline of proof

Strategy: $m_{L,p} \le a_{L,p} \le M_{L,p}$

The paths in $m_{L,p}$ (resp. $M_{L,p}$) have the property that if all occurrence of L is removed, what remains is in $m_{L-1,p'}$ (resp. $M_{L-1,p'}$).

$$M_{L+1,p} = \sum_{q=0}^{p} \binom{L+1+2p}{2q+1} M_{L,p-q}, \quad m_{L+1,p} = \sum_{q=0}^{p} \begin{bmatrix} L+1+2p \\ 2q+1 \end{bmatrix} m_{L,p-q}$$

($M_{L+1,p}$: length $L+1+2p$. Number $L+1$ appears $2q+1$ times (odd). Fill in the remaining with a path in $M_{L,p-q}$ of length $L+2p-2q$.)

$$\begin{bmatrix} N \\ P \end{bmatrix} := \left(\begin{matrix} \text{nb of ways of choosing } P \text{ items out of } N \\ \text{without taking two consecutive items} \end{matrix} \right) = \binom{N-P+1}{P}$$

$$G_L(X) := \sum_p M_{L,p} \frac{X^{L+2p}}{(L+2p)!} = \big[\sinh X\big]^L$$

Outline of proof

Strategy: $m_{L,p} \leq a_{L,p} \leq M_{L,p}$

The paths in $m_{L,p}$ (resp. $M_{L,p}$) have the property that if all occurrence of L is removed, what remains is in $m_{L-1,p'}$ (resp. $M_{L-1,p'}$).

$$M_{L+1,p} = \sum_{q=0}^{p} \binom{L+1+2p}{2q+1} M_{L,p-q}, \quad m_{L+1,p} = \sum_{q=0}^{p} \begin{bmatrix} L+1+2p \\ 2q+1 \end{bmatrix} m_{L,p-q}$$

($M_{L+1,p}$: length $L+1+2p$. Number $L+1$ appears $2q+1$ times (odd). Fill in the remaining with a path in $M_{L,p-q}$ of length $L+2p-2q$.)

$$\begin{bmatrix} N \\ P \end{bmatrix} := \begin{pmatrix} \text{nb of ways of choosing } P \text{ items out of } N \\ \text{without taking two consecutive items} \end{pmatrix} = \binom{N-P+1}{P}$$

$$G_L(X) := \sum_p M_{L,p} \frac{X^{L+2p}}{(L+2p)!} = \big[\sinh X\big]^L$$

$$\mathbb{E}^x(\text{nb of open paths}) = \sum_p a_{L,p} \frac{(1-x)^{L+2p-1}}{(L+2p-1)!}$$

Outline of proof

Strategy: $m_{L,p} \le a_{L,p} \le M_{L,p}$

The paths in $m_{L,p}$ (resp. $M_{L,p}$) have the property that if all occurrence of L is removed, what remains is in $m_{L-1,p'}$ (resp. $M_{L-1,p'}$).

$$M_{L+1,p} = \sum_{q=0}^{p} \binom{L+1+2p}{2q+1} M_{L,p-q}, \quad m_{L+1,p} = \sum_{q=0}^{p} \begin{bmatrix} L+1+2p \\ 2q+1 \end{bmatrix} m_{L,p-q}$$

($M_{L+1,p}$: length $L+1+2p$. Number $L+1$ appears $2q+1$ times (odd). Fill in the remaining with a path in $M_{L,p-q}$ of length $L+2p-2q$.)

$$\begin{bmatrix} N \\ P \end{bmatrix} := \left(\begin{matrix} \text{nb of ways of choosing } P \text{ items out of } N \\ \text{without taking two consecutive items} \end{matrix} \right) = \binom{N-P+1}{P}$$

$$G_L(X) := \sum_p M_{L,p} \frac{X^{L+2p}}{(L+2p)!} = \big[\sinh X\big]^L$$

$$\mathbb{E}^x(\text{nb of open paths}) = \sum_p a_{L,p} \frac{(1-x)^{L+2p-1}}{(L+2p-1)!} \le G_L'(1-x)$$

When only forward steps are allowed

- Forward steps only are allowed
- Fittest site is $(1, 1, 1, \dots, 1)$
- Starting site $(0, 0, 0, \dots, 0)$ has fitness $x = X/L$
- $L \to \infty$

$$\boxed{\frac{1}{L}\left(\text{nb of open paths if starting fitness is } x = \frac{X}{L}\right) \to e^{-X} \times \mathcal{E} \times \mathcal{E}'}$$

with $\mathcal{E}$ and $\mathcal{E}'$ two independent exponential variables

One already knows that

- $\mathbb{E}^{\frac{X}{L}}(\text{nb of open paths}) = L(1-x)^{L-1} \sim Le^{-X}$

When only forward steps are allowed

- Forward steps only are allowed
- Fittest site is $(1, 1, 1, \dots, 1)$
- Starting site $(0, 0, 0, \dots, 0)$ has fitness $x = X/L$
- $L \to \infty$

$$\boxed{\frac{1}{L}\left(\text{nb of open paths if starting fitness is } x = \frac{X}{L}\right) \to e^{-X} \times \mathcal{E} \times \mathcal{E}'}$$

with $\mathcal{E}$ and $\mathcal{E}'$ two independent exponential variables

One already knows that

- $\mathbb{E}^{\frac{X}{L}}(\text{nb of open paths}) = L(1-x)^{L-1} \sim Le^{-X}$
- There are indeed typically $\propto L$ open paths

Hypercube vs Tree

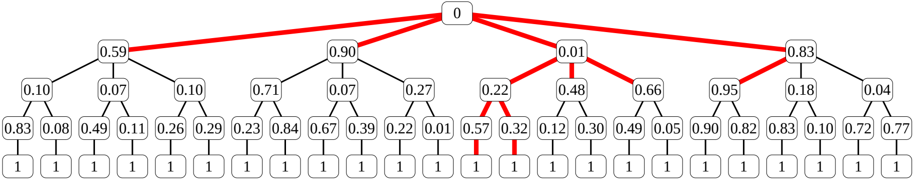
Hypercube is hard; try a tree!

1$^{\text{st}}$ step: L choices; 2$^{\text{nd}}$ step: $L-1$ choices; 3$^{\text{rd}}$ step: $L-2$ choices; ...

Hypercube vs Tree

Hypercube is hard; try a tree!

1st step: L choices; 2nd step: $L-1$ choices; 3rd step: $L-2$ choices; ...

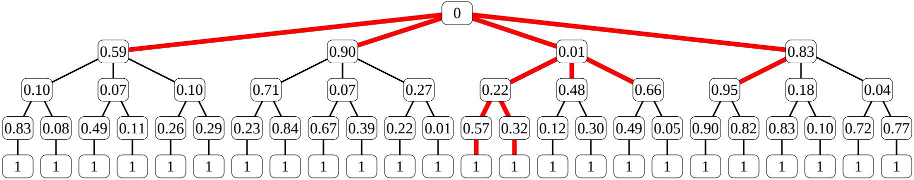


$\mathbb{E}^{\frac{X}{L}}\big[\text{nb of open paths}\big] = L(1-x)^{L-1} \sim Le^{-X}$ same for tree or hypercube!

Hypercube vs Tree

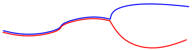
Hypercube is hard; try a tree!

1[st] step: L choices; 2[nd] step: $L-1$ choices; 3[rd] step: $L-2$ choices; ...



$$\mathbb{E}^{\frac{X}{L}}\left[\text{nb of open paths}\right] = L(1-x)^{L-1} \sim Le^{-X}$$ same for tree or hypercube!

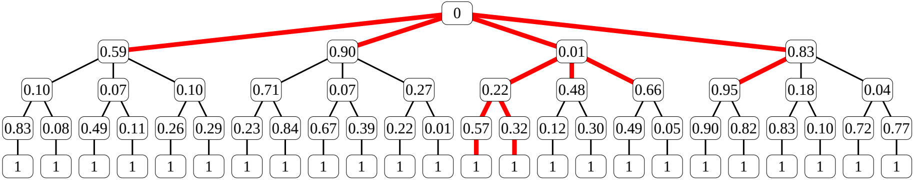
$$\mathbb{E}^{\frac{X}{L}}\left[(\text{nb of open paths})^2\right] \sim \begin{cases} 2L^2e^{-2X} & \text{(tree)} \\ & \end{cases}$$



Hypercube vs Tree

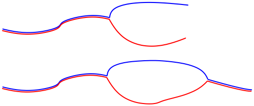
Hypercube is hard; try a tree!

1^{st} step: L choices; 2^{nd} step: $L-1$ choices; 3^{rd} step: $L-2$ choices; ...

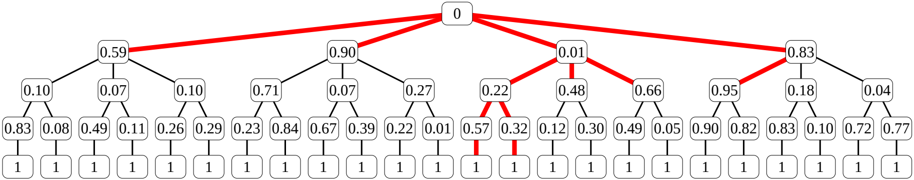


$$\mathbb{E}^{\frac{X}{L}}\big[\text{nb of open paths}\big] = L(1-x)^{L-1} \sim Le^{-X}$$ same for tree or hypercube!

$$\mathbb{E}^{\frac{X}{L}}\big[(\text{nb of open paths})^2\big] \sim \begin{cases} 2L^2e^{-2X} & \text{(tree)} \\ 4L^2e^{-2X} & \text{(hypercube)} \end{cases}$$

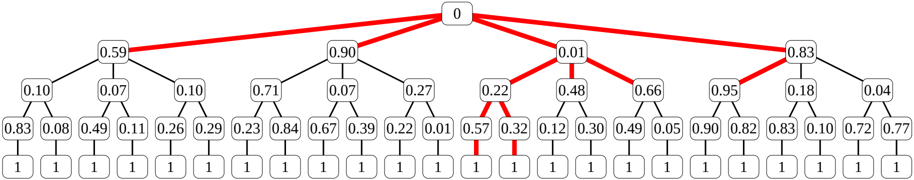


On the tree



$$(\text{Nb of open paths}) = \sum_{|\sigma|=1} (\text{nb of open paths going through } \sigma)$$

On the tree

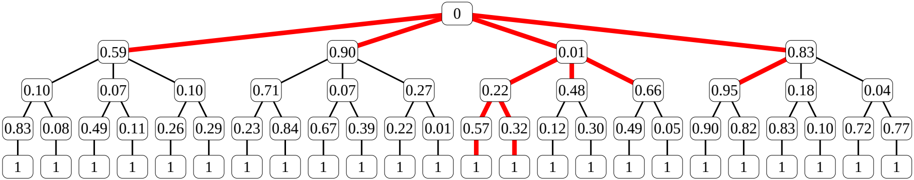


$$(\text{Nb of open paths}) = \sum_{|\sigma|=1} (\text{nb of open paths going through } \sigma)$$

Sum of uncorrelated terms (because it is a tree), generating function

$$G(\lambda, x, L) := \mathbb{E}^x\left(e^{-\lambda(\text{nb of open paths})}\right)$$

On the tree



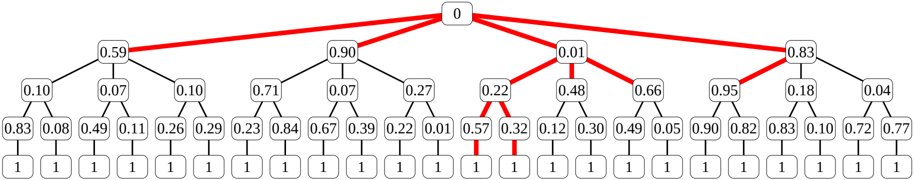
$$(\text{Nb of open paths}) = \sum_{|\sigma|=1} (\text{nb of open paths going through } \sigma)$$

Sum of uncorrelated terms (because it is a tree), generating function

$$G(\lambda, x, L) := \mathbb{E}^x\left(e^{-\lambda(\text{nb of open paths})}\right)$$

$$G(\lambda, x, L) = \Big[\qquad\qquad\qquad\qquad \Big]^L$$

On the tree



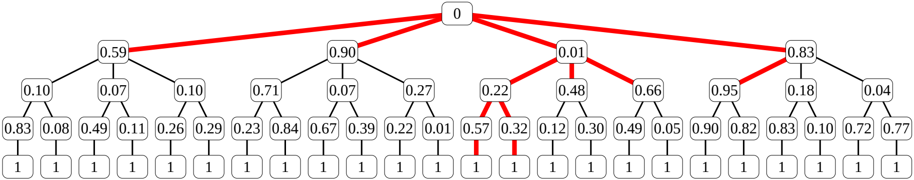
$$(\text{Nb of open paths}) = \sum_{|\sigma|=1} (\text{nb of open paths going through } \sigma)$$

Sum of uncorrelated terms (because it is a tree), generating function

$$G(\lambda, x, L) := \mathbb{E}^x\left(e^{-\lambda(\text{nb of open paths})}\right)$$

$$G(\lambda, x, L) = \Big[x + \qquad\qquad\qquad\qquad \Big]^L$$

On the tree



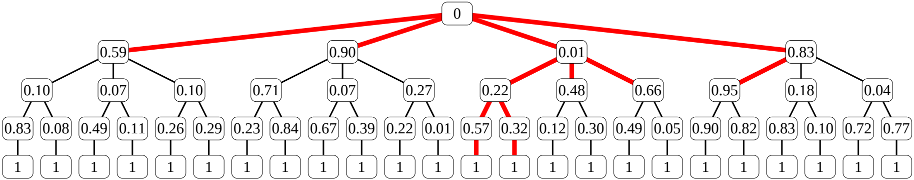
$$(\text{Nb of open paths}) = \sum_{|\sigma|=1} (\text{nb of open paths going through } \sigma)$$

Sum of uncorrelated terms (because it is a tree), generating function

$$G(\lambda, x, L) := \mathbb{E}^x\left(e^{-\lambda(\text{nb of open paths})}\right)$$

$$G(\lambda, x, L) = \left[x + \int_x^1 \mathrm{d}y\, G(\lambda, y, L-1)\right]^L$$

On the tree



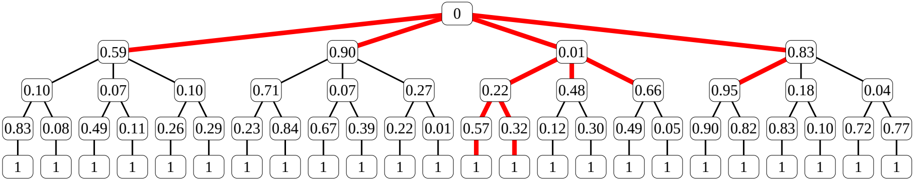
$$(\text{Nb of open paths}) = \sum_{|\sigma|=1} (\text{nb of open paths going through } \sigma)$$

Sum of uncorrelated terms (because it is a tree), generating function

$$G(\lambda, x, L) := \mathbb{E}^x\left(e^{-\lambda(\text{nb of open paths})}\right)$$

$$G(\lambda, x, L) = \left[x + \int_x^1 \mathrm{d}y\, G(\lambda, y, L-1)\right]^L, \qquad G(\lambda, x, 1) = e^{-\lambda}$$

On the tree



$$(\text{Nb of open paths}) = \sum_{|\sigma|=1} (\text{nb of open paths going through } \sigma)$$

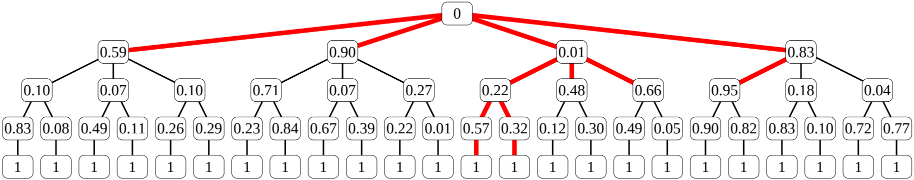
Sum of uncorrelated terms (because it is a tree), generating function

$$G(\lambda, x, L) := \mathbb{E}^x\left(e^{-\lambda(\text{nb of open paths})}\right)$$

$$G(\lambda, x, L) = \left[x + \int_x^1 \mathrm{d}y\, G(\lambda, y, L-1)\right]^L, \qquad G(\lambda, x, 1) = e^{-\lambda}$$

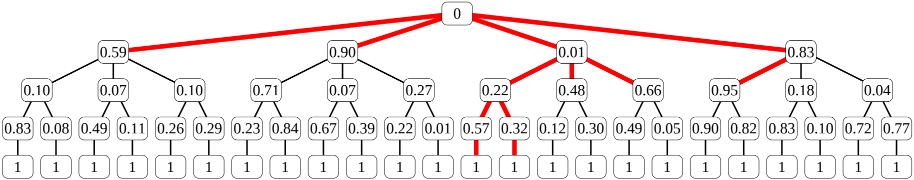
$$\lim_{L\to\infty} G\left(\frac{\mu}{L}, \frac{X}{L}, L\right) = ?$$

On the tree, second try



Idea: the first steps determine everything

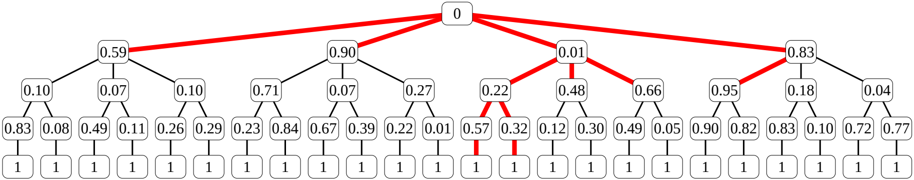
On the tree, second try



Idea: the first steps determine everything

$\Theta = (\text{nb of open paths})$, $\qquad \Theta_k = \mathbb{E}(\Theta|\mathcal{F}_k)$, $\qquad \mathcal{F}_k = (\text{info up to level } k)$

On the tree, second try



Idea: the first steps determine everything

$\Theta = $ (nb of open paths), $\quad \Theta_k = \mathbb{E}(\Theta | \mathcal{F}_k)$, $\quad \mathcal{F}_k = $ (info up to level k)

$$\Theta_k = \sum_{|\sigma|=k} \mathbb{1}_{\{\sigma \text{ open}\}} \underbrace{(L-k)(1-x_\sigma)^{L-k-1}}_{\substack{\text{expected nb of open} \\ \text{paths through } \sigma}}$$

$$\Theta_1 = 3(1-0.59)^2 + 3(1-0.90)^2 + 3(1-0.01)^2 + 3(1-0.83)^2 = 3.5613$$
$$\Theta_2 = 2(1-0.22)^1 + 2(1-0.48)^1 + 2(1-0.66)^1 + 2(1-0.95)^1 = 3.38$$

On the tree, second try

$\Theta = (\text{nb of open paths}), \qquad \Theta_k = \mathbb{E}(\Theta|\mathcal{F}_k), \qquad \mathcal{F}_k = (\text{info up to level } k)$

Intuitively, $\Theta_k \approx \Theta$ if $\quad \mathrm{Var}(\Theta|\mathcal{F}_k)$ is small

On the tree, second try

$\Theta = (\text{nb of open paths}), \qquad \Theta_k = \mathbb{E}(\Theta|\mathcal{F}_k), \qquad \mathcal{F}_k = (\text{info up to level } k)$

Intuitively, $\Theta_k \approx \Theta$ if $\mathbb{E}\big[\,\text{Var}(\Theta|\mathcal{F}_k)\big]$ is small

On the tree, second try

$\Theta = (\text{nb of open paths}), \qquad \Theta_k = \mathbb{E}(\Theta|\mathcal{F}_k), \qquad \mathcal{F}_k = (\text{info up to level } k)$

Intuitively, $\Theta_k \approx \Theta$ if $\mathbb{E}\big[\,\mathrm{Var}(\Theta|\mathcal{F}_k)\big]$ is small

$$\lim_{L\to\infty} \mathbb{P}^{\frac{X}{L}}\Big[\frac{\Theta}{L} < z\Big] = \lim_{k\to\infty}\lim_{L\to\infty} \mathbb{P}^{\frac{X}{L}}\Big[\frac{\Theta_k}{L} < z\Big] \text{ if}$$

On the tree, second try

$\Theta = (\text{nb of open paths}), \qquad \Theta_k = \mathbb{E}(\Theta|\mathcal{F}_k), \qquad \mathcal{F}_k = (\text{info up to level } k)$

Intuitively, $\Theta_k \approx \Theta$ if $\mathbb{E}\big[\,\mathrm{Var}(\Theta|\mathcal{F}_k)\big]$ is small

$$\lim_{L\to\infty} \mathbb{P}^{\frac{X}{L}}\Big[\frac{\Theta}{L} < z\Big] = \lim_{k\to\infty}\lim_{L\to\infty} \mathbb{P}^{\frac{X}{L}}\Big[\frac{\Theta_k}{L} < z\Big] \text{ if } \lim_{k\to\infty}\lim_{L\to\infty} \mathbb{E}^{\frac{X}{L}}\Big[\mathrm{Var}\Big[\frac{\Theta}{L}\Big|\mathcal{F}_k\Big]\Big] = 0$$

On the tree, second try

$\Theta = $ (nb of open paths), $\qquad \Theta_k = \mathbb{E}(\Theta|\mathcal{F}_k)$, $\qquad \mathcal{F}_k = $ (info up to level k)

Intuitively, $\Theta_k \approx \Theta$ if $\mathbb{E}\big[\,\mathrm{Var}(\Theta|\mathcal{F}_k)\big]$ is small

$$\lim_{L\to\infty} \mathbb{P}^{\frac{X}{L}}\left[\frac{\Theta}{L} < z\right] = \lim_{k\to\infty}\lim_{L\to\infty} \mathbb{P}^{\frac{X}{L}}\left[\frac{\Theta_k}{L} < z\right] \text{ if } \lim_{k\to\infty}\lim_{L\to\infty} \mathbb{E}^{\frac{X}{L}}\left[\mathrm{Var}\left[\frac{\Theta}{L}\middle|\mathcal{F}_k\right]\right] = 0$$

But (sum over pairs of paths):

$$\lim_{L\to\infty} \mathbb{E}^{\frac{X}{L}}\left[\mathrm{Var}\left[\frac{\Theta}{L}\middle|\mathcal{F}_k\right]\right] = \frac{e^{-2X}}{2^k}$$

In the $L \to \infty$, $k \to \infty$ limit, Θ/L and Θ_k/L have the same distribution

On the tree, second try

$\Theta = (\text{nb of open paths}), \qquad \Theta_k = \mathbb{E}(\Theta|\mathcal{F}_k), \qquad \mathcal{F}_k = (\text{info up to level } k)$

We want to write a generating function.

On the tree, second try

$\Theta = (\text{nb of open paths}), \qquad \Theta_k = \mathbb{E}(\Theta | \mathcal{F}_k), \qquad \mathcal{F}_k = (\text{info up to level } k)$

We want to write a generating function.

For Θ, we used

$$\Theta = \sum_{|\sigma|=1} (\text{nb of open paths through } \sigma)$$

On the tree, second try

$\Theta = (\text{nb of open paths}), \qquad \Theta_k = \mathbb{E}(\Theta|\mathcal{F}_k), \qquad \mathcal{F}_k = (\text{info up to level } k)$

We want to write a generating function.
For Θ, we used

$$\Theta = \sum_{|\sigma|=1} (\text{nb of open paths through } \sigma)$$

Now, for Θ_k, we use

$$\Theta_k = \sum_{|\sigma|=1} \mathbb{1}_{\{x_\sigma > x\}}(\text{"}\Theta_{k-1}\text{" of the } L-1 \text{ tree rooted on } \sigma)$$

On the tree, second try

$\Theta =$ (nb of open paths), $\qquad \Theta_k = \mathbb{E}(\Theta | \mathcal{F}_k)$, $\qquad \mathcal{F}_k =$ (info up to level k)

We want to write a generating function.

For Θ, we used

$$\Theta = \sum_{|\sigma|=1} (\text{nb of open paths through } \sigma)$$

Now, for Θ_k, we use

$$\Theta_k = \sum_{|\sigma|=1} \mathbb{1}_{\{x_\sigma > x\}} (\text{“}\Theta_{k-1}\text{” of the } L-1 \text{ tree rooted on } \sigma)$$

New generating function:

$$G_k(\lambda, x, L) := \mathbb{E}^x\left(e^{-\lambda \Theta_k}\right)$$

On the tree, second try

$\Theta =$ (nb of open paths), $\qquad \Theta_k = \mathbb{E}(\Theta|\mathcal{F}_k)$, $\qquad \mathcal{F}_k =$ (info up to level k)

We want to write a generating function.

For Θ, we used

$$\Theta = \sum_{|\sigma|=1} (\text{nb of open paths through } \sigma)$$

Now, for Θ_k, we use

$$\Theta_k = \sum_{|\sigma|=1} \mathbb{1}_{\{x_\sigma > x\}} (\text{"}\Theta_{k-1}\text{" of the } L-1 \text{ tree rooted on } \sigma)$$

New generating function:

$$G_k(\lambda, x, L) := \mathbb{E}^x\left(e^{-\lambda\Theta_k}\right) = \left[x + \int_x^1 \mathrm{d}y \, G_{k-1}(\lambda, y, L-1)\right]^L$$

On the tree, second try

$\Theta = $ (nb of open paths), $\qquad \Theta_k = \mathbb{E}(\Theta|\mathcal{F}_k)$, $\qquad \mathcal{F}_k = $ (info up to level k)

We want to write a generating function.
For Θ, we used

$$\Theta = \sum_{|\sigma|=1} (\text{nb of open paths through } \sigma)$$

Now, for Θ_k, we use

$$\Theta_k = \sum_{|\sigma|=1} \mathbb{1}_{\{x_\sigma > x\}} (\text{“}\Theta_{k-1}\text{” of the } L-1 \text{ tree rooted on } \sigma)$$

New generating function:

$$G_k(\lambda, x, L) := \mathbb{E}^x\left(e^{-\lambda\Theta_k}\right) = \left[x + \int_x^1 \mathrm{d}y\, G_{k-1}(\lambda, y, L-1)\right]^L$$
$$= \left[1 - \int_x^1 \mathrm{d}y \left(1 - G_{k-1}(\lambda, y, L-1)\right)\right]^L$$

On the tree, second try

$\Theta =$ (nb of open paths), $\quad \Theta_k = \mathbb{E}(\Theta|\mathcal{F}_k)$, $\quad \mathcal{F}_k =$ (info up to level k)

We want to write a generating function.

For Θ, we used

$$\Theta = \sum_{|\sigma|=1} (\text{nb of open paths through } \sigma)$$

Now, for Θ_k, we use

$$\Theta_k = \sum_{|\sigma|=1} \mathbb{1}_{\{x_\sigma > x\}}(\text{“}\Theta_{k-1}\text{” of the } L-1 \text{ tree rooted on } \sigma)$$

New generating function:

$$G_k(\lambda, x, L) := \mathbb{E}^x\left(e^{-\lambda\Theta_k}\right) = \left[x + \int_x^1 \mathrm{d}y\, G_{k-1}(\lambda, y, L-1)\right]^L$$
$$= \left[1 - \int_x^1 \mathrm{d}y\left(1 - G_{k-1}(\lambda, y, L-1)\right)\right]^L$$

$$G_0(\lambda, x, L) = e^{-\lambda L(1-x)^{L-1}}$$

On the tree, second try

$$G_k(\lambda, x, L) := \mathbb{E}^x\left(e^{-\lambda\Theta_k}\right) = \left[1 - \int_x^1 \mathrm{d}y\left(1 - G_{k-1}(\lambda, y, L-1)\right)\right]^L$$

$$G_0(\lambda, x, L) = e^{-\lambda L(1-x)^{L-1}}$$

On the tree, second try

$$G_k(\lambda, x, L) := \mathbb{E}^x\big(e^{-\lambda\Theta_k}\big) = \Big[1 - \int_x^1 \mathrm{d}y\,\Big(1 - G_{k-1}(\lambda, y, L-1)\Big)\Big]^L$$

$$G_0(\lambda, x, L) = e^{-\lambda L(1-x)^{L-1}}$$

$$G_k\Big(\frac{\mu}{L}, \frac{X}{L}, L\Big) = \mathbb{E}^{\frac{X}{L}}\big(e^{-\mu\frac{\Theta_k}{L}}\big) = \Big[1 - \frac{1}{L}\int_X^L \mathrm{d}Y\,\Big(1 - G_{k-1}\Big(\frac{\mu}{L}, \frac{Y}{L}, L-1\Big)\Big)\Big]^L$$

$$G_0\Big(\frac{\mu}{L}, \frac{X}{L}, L\Big) = e^{-\mu\left(1-\frac{X}{L}\right)^{L-1}}$$

On the tree, second try

$$G_k(\lambda, x, L) := \mathbb{E}^x\big(e^{-\lambda\Theta_k}\big) = \Big[1 - \int_x^1 \mathrm{d}y\, \Big(1 - G_{k-1}(\lambda, y, L-1)\Big)\Big]^L$$

$$G_0(\lambda, x, L) = e^{-\lambda L(1-x)^{L-1}}$$

$$G_k\Big(\frac{\mu}{L}, \frac{X}{L}, L\Big) = \mathbb{E}^{\frac{X}{L}}\big(e^{-\mu\frac{\Theta_k}{L}}\big) = \Big[1 - \frac{1}{L}\int_X^L \mathrm{d}Y\, \Big(1 - G_{k-1}\Big(\frac{\mu}{L}, \frac{Y}{L}, L-1\Big)\Big)\Big]^L$$

$$G_0\Big(\frac{\mu}{L}, \frac{X}{L}, L\Big) = e^{-\mu\left(1-\frac{X}{L}\right)^{L-1}}$$

One can then prove that $F_k(\mu, X) = \lim_{L\to\infty} G_k\big(\frac{\mu}{L}, \frac{X}{L}, L\big)$ exists and

$$F_k(\mu, X) = \exp\Big[-\int_X^\infty \mathrm{d}Y\, \big(1 - F_{k-1}(\mu, Y)\big)\Big], \quad F_0(\mu, X) = \exp\big(-\mu e^{-X}\big)$$

F_k is the generating function of $\lim_{L\to\infty} \frac{\Theta_k}{L}$ when starting fron $\frac{X}{L}$.

$$\lim_{L\to\infty} \mathbb{E}^{\frac{X}{L}}\big(e^{-\mu\frac{\Theta_k}{L}}\big) = \qquad F_k(\mu, X)$$

On the tree, second try

$$G_k(\lambda,x,L) := \mathbb{E}^x\left(e^{-\lambda\Theta_k}\right) = \left[1-\int_x^1 \mathrm{d}y\left(1-G_{k-1}(\lambda,y,L-1)\right)\right]^L$$

$$G_0(\lambda,x,L) = e^{-\lambda L(1-x)^{L-1}}$$

$$G_k\Big(\frac{\mu}{L},\frac{X}{L},L\Big) = \mathbb{E}^{\frac{X}{L}}\left(e^{-\mu\frac{\Theta_k}{L}}\right) = \left[1-\frac{1}{L}\int_X^L \mathrm{d}Y\left(1-G_{k-1}\Big(\frac{\mu}{L},\frac{Y}{L},L-1\Big)\right)\right]^L$$

$$G_0\Big(\frac{\mu}{L},\frac{X}{L},L\Big) = e^{-\mu\left(1-\frac{X}{L}\right)^{L-1}}$$

One can then prove that $F_k(\mu,X) = \lim_{L\to\infty} G_k\big(\frac{\mu}{L},\frac{X}{L},L\big)$ exists and

$$F_k(\mu,X) = \exp\left[-\int_X^\infty \mathrm{d}Y\left(1-F_{k-1}(\mu,Y)\right)\right],\quad F_0(\mu,X) = \exp\left(-\mu e^{-X}\right)$$

F_k is the generating function of $\lim_{L\to\infty}\frac{\Theta_k}{L}$ when starting fron $\frac{X}{L}$. Take $k\to\infty$:

$$\lim_{k\to\infty}\lim_{L\to\infty}\mathbb{E}^{\frac{X}{L}}\left(e^{-\mu\frac{\Theta_k}{L}}\right) = \lim_{k\to\infty} F_k(\mu,X)$$

On the tree, second try

$$G_k(\lambda, x, L) := \mathbb{E}^x\big(e^{-\lambda\Theta_k}\big) = \left[1 - \int_x^1 \mathrm{d}y\,\Big(1 - G_{k-1}(\lambda, y, L-1)\Big)\right]^L$$

$$G_0(\lambda, x, L) = e^{-\lambda L(1-x)^{L-1}}$$

$$G_k\Big(\frac{\mu}{L}, \frac{X}{L}, L\Big) = \mathbb{E}^{\frac{X}{L}}\big(e^{-\mu\frac{\Theta_k}{L}}\big) = \left[1 - \frac{1}{L}\int_X^L \mathrm{d}Y\,\Big(1 - G_{k-1}\Big(\frac{\mu}{L}, \frac{Y}{L}, L-1\Big)\Big)\right]^L$$

$$G_0\Big(\frac{\mu}{L}, \frac{X}{L}, L\Big) = e^{-\mu\left(1-\frac{X}{L}\right)^{L-1}}$$

One can then prove that $F_k(\mu, X) = \lim\limits_{L\to\infty} G_k(\frac{\mu}{L}, \frac{X}{L}, L)$ exists and

$$F_k(\mu, X) = \exp\left[-\int_X^\infty \mathrm{d}Y\,\big(1 - F_{k-1}(\mu, Y)\big)\right], \quad F_0(\mu, X) = \exp\big(-\mu e^{-X}\big)$$

F_k is the generating function of $\lim\limits_{L\to\infty}\frac{\Theta_k}{L}$ when starting fron $\frac{X}{L}$. Take $k\to\infty$:

$$\lim_{L\to\infty}\mathbb{E}^{\frac{X}{L}}\big(e^{-\mu\frac{\Theta}{L}}\big) = \lim_{k\to\infty}\lim_{L\to\infty}\mathbb{E}^{\frac{X}{L}}\big(e^{-\mu\frac{\Theta_k}{L}}\big) = \lim_{k\to\infty} F_k(\mu, X) = \frac{1}{1+\mu e^{-X}}$$

On the tree, second try

$$G_k(\lambda, x, L) := \mathbb{E}^x\big(e^{-\lambda\Theta_k}\big) = \Big[1 - \int_x^1 \mathrm{d}y \,\Big(1 - G_{k-1}(\lambda, y, L-1)\Big)\Big]^L$$

$$G_0(\lambda, x, L) = e^{-\lambda L(1-x)^{L-1}}$$

$$G_k\Big(\frac{\mu}{L}, \frac{X}{L}, L\Big) = \mathbb{E}^{\frac{X}{L}}\big(e^{-\mu\frac{\Theta_k}{L}}\big) = \Big[1 - \frac{1}{L}\int_X^L \mathrm{d}Y \,\Big(1 - G_{k-1}\Big(\frac{\mu}{L}, \frac{Y}{L}, L-1\Big)\Big)\Big]^L$$

$$G_0\Big(\frac{\mu}{L}, \frac{X}{L}, L\Big) = e^{-\mu\left(1-\frac{X}{L}\right)^{L-1}}$$

One can then prove that $F_k(\mu, X) = \lim_{L\to\infty} G_k(\frac{\mu}{L}, \frac{X}{L}, L)$ exists and

$$F_k(\mu, X) = \exp\Big[-\int_X^\infty \mathrm{d}Y\,\big(1 - F_{k-1}(\mu, Y)\big)\Big], \quad F_0(\mu, X) = \exp\big(-\mu e^{-X}\big)$$

F_k is the generating function of $\lim_{L\to\infty}\frac{\Theta_k}{L}$ when starting fron $\frac{X}{L}$. Take $k\to\infty$:

$$\lim_{L\to\infty}\mathbb{E}^{\frac{X}{L}}\big(e^{-\mu\frac{\Theta}{L}}\big) = \lim_{k\to\infty}\lim_{L\to\infty}\mathbb{E}^{\frac{X}{L}}\big(e^{-\mu\frac{\Theta_k}{L}}\big) = \lim_{k\to\infty} F_k(\mu, X) = \frac{1}{1+\mu e^{-X}}$$

On the tree, starting from $x = \frac{X}{L}$, $\qquad \dfrac{\Theta}{L} \xrightarrow[L\to\infty]{\text{in law}} e^{-X}\times\mathcal{E}$

Back to the hypercube

Same trick:

$$\Theta_k = \mathbb{E}(\Theta|\mathcal{F}_k), \qquad \mathcal{F}_k = (\text{info in the } k \text{ first and } k \text{ last levels})$$

Back to the hypercube

Same trick:

$$\Theta_k = \mathbb{E}(\Theta|\mathcal{F}_k), \qquad \mathcal{F}_k = (\text{info in the } k \text{ first and } k \text{ last levels})$$

Again

$$\lim_{L\to\infty} \mathbb{P}^{\frac{X}{L}}\left[\frac{\Theta}{L} < z\right] = \lim_{k\to\infty}\lim_{L\to\infty} \mathbb{P}^{\frac{X}{L}}\left[\frac{\Theta_k}{L} < z\right] \text{ if } \lim_{k\to\infty}\lim_{L\to\infty} \mathbb{E}^{\frac{X}{L}}\left[\text{Var}\left[\frac{\Theta}{L}\middle|\mathcal{F}_k\right]\right] = 0$$

The expectation of the conditional variance can be computed and it works.

Back to the hypercube

Same trick:

$$\Theta_k = \mathbb{E}(\Theta|\mathcal{F}_k), \qquad \mathcal{F}_k = (\text{info in the } k \text{ first and } k \text{ last levels})$$

Again

$$\lim_{L\to\infty} \mathbb{P}^{\frac{X}{L}}\left[\frac{\Theta}{L} < z\right] = \lim_{k\to\infty}\lim_{L\to\infty} \mathbb{P}^{\frac{X}{L}}\left[\frac{\Theta_k}{L} < z\right] \text{ if } \lim_{k\to\infty}\lim_{L\to\infty} \mathbb{E}^{\frac{X}{L}}\left[\text{Var}\left[\frac{\Theta}{L}\middle|\mathcal{F}_k\right]\right] = 0$$

The expectation of the conditional variance can be computed and it works.

$$\Theta_k = \sum_{|\sigma|=k} \sum_{|\tau|=L-k} n_\sigma m_\tau \mathbb{1}_{\{\tau \text{ reachable from } \sigma\}} \mathbb{1}_{\{x_\sigma < x_\tau\}} (L-2k)(x_\tau - x_\sigma)^{L-2k-1}$$

n_σ =nb of open paths from $(0,\dots,0)$ to σ; m_τ =nb from τ to $(1,\dots,1)$

Back to the hypercube

Same trick:

$$\Theta_k = \mathbb{E}(\Theta|\mathcal{F}_k), \qquad \mathcal{F}_k = (\text{info in the } k \text{ first and } k \text{ last levels})$$

Again

$$\lim_{L\to\infty} \mathbb{P}^{\frac{X}{L}}\left[\frac{\Theta}{L} < z\right] = \lim_{k\to\infty}\lim_{L\to\infty} \mathbb{P}^{\frac{X}{L}}\left[\frac{\Theta_k}{L} < z\right] \text{ if } \lim_{k\to\infty}\lim_{L\to\infty} \mathbb{E}^{\frac{X}{L}}\left[\operatorname{Var}\left[\frac{\Theta}{L}\middle|\mathcal{F}_k\right]\right] = 0$$

The expectation of the conditional variance can be computed and it works.

$$\Theta_k = \sum_{|\sigma|=k}\ \sum_{|\tau|=L-k} n_\sigma m_\tau \mathbb{1}_{\{\tau \text{ reachable from } \sigma\}} \mathbb{1}_{\{x_\sigma < x_\tau\}} (L-2k)(x_\tau - x_\sigma)^{L-2k-1}$$

n_σ =nb of open paths from $(0,\dots,0)$ to σ; m_τ =nb from τ to $(1,\dots,1)$

$$\tilde{\Theta}_k := \sum_{|\sigma|=k}\ \sum_{|\tau|=L-k} n_\sigma m_\tau\ L(x_\tau - x_\sigma x_\tau)^{L-2k-1}$$

Back to the hypercube

Same trick:

$$\Theta_k = \mathbb{E}(\Theta|\mathcal{F}_k), \qquad \mathcal{F}_k = (\text{info in the } k \text{ first and } k \text{ last levels})$$

Again

$$\lim_{L\to\infty} \mathbb{P}^{\frac{X}{L}}\left[\frac{\Theta}{L} < z\right] = \lim_{k\to\infty}\lim_{L\to\infty} \mathbb{P}^{\frac{X}{L}}\left[\frac{\Theta_k}{L} < z\right] \text{ if } \lim_{k\to\infty}\lim_{L\to\infty} \mathbb{E}^{\frac{X}{L}}\left[\operatorname{Var}\left[\frac{\Theta}{L}\middle|\mathcal{F}_k\right]\right] = 0$$

The expectation of the conditional variance can be computed and it works.

$$\Theta_k = \sum_{|\sigma|=k}\sum_{|\tau|=L-k} n_\sigma m_\tau \mathbb{1}_{\{\tau \text{ reachable from } \sigma\}} \mathbb{1}_{\{x_\sigma < x_\tau\}} (L-2k)(x_\tau - x_\sigma)^{L-2k-1}$$

n_σ =nb of open paths from $(0,\dots,0)$ to σ; m_τ =nb from τ to $(1,\dots,1)$

$$\tilde{\Theta}_k := \sum_{|\sigma|=k}\sum_{|\tau|=L-k} n_\sigma m_\tau \; L(x_\tau - x_\sigma x_\tau)^{L-2k-1}$$

$\tilde{\Theta}_k > \Theta_k$, but not that much: $\lim_{L\to\infty} \mathbb{E}^{\frac{X}{L}}\left[\frac{\Theta_k}{L}\right] = \lim_{L\to\infty} \mathbb{E}^{\frac{X}{L}}\left[\frac{\tilde{\Theta}_k}{L}\right]$

Back to the hypercube

Same trick:

$$\Theta_k = \mathbb{E}(\Theta|\mathcal{F}_k), \qquad \mathcal{F}_k = (\text{info in the } k \text{ first and } k \text{ last levels})$$

Again

$$\lim_{L\to\infty} \mathbb{P}^{\frac{X}{L}}\left[\frac{\Theta}{L} < z\right] = \lim_{k\to\infty}\lim_{L\to\infty} \mathbb{P}^{\frac{X}{L}}\left[\frac{\Theta_k}{L} < z\right] \text{ if } \lim_{k\to\infty}\lim_{L\to\infty} \mathbb{E}^{\frac{X}{L}}\left[\operatorname{Var}\left[\frac{\Theta}{L}\middle|\mathcal{F}_k\right]\right] = 0$$

The expectation of the conditional variance can be computed and it works.

$$\Theta_k = \sum_{|\sigma|=k}\ \sum_{|\tau|=L-k} n_\sigma m_\tau \mathbb{1}_{\{\tau \text{ reachable from } \sigma\}} \mathbb{1}_{\{x_\sigma < x_\tau\}} (L-2k)(x_\tau - x_\sigma)^{L-2k-1}$$

n_σ =nb of open paths from $(0,\ldots,0)$ to σ; m_τ =nb from τ to $(1,\ldots,1)$

$$\tilde{\Theta}_k := \sum_{|\sigma|=k}\ \sum_{|\tau|=L-k} n_\sigma m_\tau\ L(x_\tau - x_\sigma x_\tau)^{L-2k-1}$$

$\tilde{\Theta}_k > \Theta_k$, but not that much: $\lim_{L\to\infty} \mathbb{E}^{\frac{X}{L}}\left[\frac{\Theta_k}{L}\right] = \lim_{L\to\infty} \mathbb{E}^{\frac{X}{L}}\left[\frac{\tilde{\Theta}_k}{L}\right]$

$\tilde{\Theta}_k/L$ and Θ_k/L have the same distribution for large L.

Back to the hypercube

$$\tilde{\Theta}_k := \sum_{|\sigma|=k} \sum_{|\tau|=L-k} n_\sigma m_\tau L(x_\tau - x_\sigma x_\tau)^{L-2k-1}$$

Back to the hypercube

$$\tilde{\Theta}_k := \sum_{|\sigma|=k} \sum_{|\tau|=L-k} n_\sigma m_\tau L(x_\tau - x_\sigma x_\tau)^{L-2k-1}$$

$$\frac{\tilde{\Theta}_k}{L} = \left(\sum_{|\sigma|=k} n_\sigma (1 - x_\sigma)^{L-2k-1} \right) \left(\sum_{|\tau|=L-k} m_\tau (x_\tau)^{L-2k-1} \right)$$

Back to the hypercube

$$\tilde{\Theta}_k := \sum_{|\sigma|=k} \sum_{|\tau|=L-k} n_\sigma m_\tau L(x_\tau - x_\sigma x_\tau)^{L-2k-1}$$

$$\frac{\tilde{\Theta}_k}{L} = \left(\sum_{|\sigma|=k} n_\sigma (1 - x_\sigma)^{L-2k-1} \right) \left(\sum_{|\tau|=L-k} m_\tau (x_\tau)^{L-2k-1} \right)$$

First factor: beginning of the hypercube. Second factor: end of the hypercube. Terms are independent and symmetrical if $X = 0$.

Back to the hypercube

$$\tilde{\Theta}_k := \sum_{|\sigma|=k} \sum_{|\tau|=L-k} n_\sigma m_\tau L(x_\tau - x_\sigma x_\tau)^{L-2k-1}$$

$$\frac{\tilde{\Theta}_k}{L} = \left(\sum_{|\sigma|=k} n_\sigma (1 - x_\sigma)^{L-2k-1} \right) \left(\sum_{|\tau|=L-k} m_\tau (x_\tau)^{L-2k-1} \right)$$

First factor: beginning of the hypercube. Second factor: end of the hypercube. Terms are independent and symmetrical if $X = 0$.
Last step: prove that

$$\phi_k := \sum_{|\sigma|=k} n_\sigma (1 - x_\sigma)^{L-2k-1} \xrightarrow[L\to\infty \text{ then } k\to\infty]{\text{in law}} e^{-X}\mathcal{E}$$

Back to the hypercube

$$\tilde{\Theta}_k := \sum_{|\sigma|=k} \sum_{|\tau|=L-k} n_\sigma m_\tau L(x_\tau - x_\sigma x_\tau)^{L-2k-1}$$

$$\frac{\tilde{\Theta}_k}{L} = \left(\sum_{|\sigma|=k} n_\sigma (1 - x_\sigma)^{L-2k-1} \right) \left(\sum_{|\tau|=L-k} m_\tau (x_\tau)^{L-2k-1} \right)$$

First factor: beginning of the hypercube. Second factor: end of the hypercube. Terms are independent and symmetrical if $X = 0$.
Last step: prove that

$$\phi_k := \sum_{|\sigma|=k} n_\sigma (1 - x_\sigma)^{L-2k-1} \xrightarrow[L\to\infty \text{ then } k\to\infty]{\text{in law}} e^{-X}\mathcal{E}$$

Intuition: with k fixed and $L \to \infty$, loops become negligible, and the beginning of the hypercube looks like the beginning of the tree. So ϕ_k and Θ_k^{tree}/L have the same large L distribution.

Thank you

ArXiV 1304.0246   ArXiV 1401.6894