Populationsgenetik - von Mathematik zu Biologie und zurück

Maite Wilke Berenguer
Humboldt-Universität zu Berlin

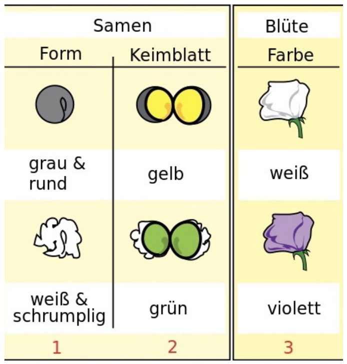
Berliner Mathematische Gesellschaft
Vortrag im Rahmen der Jahresvollversammlung
08.02.24

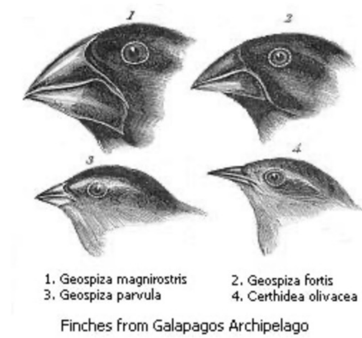
DER PLAN

- Ursprung der Populationsgenetik
- Wozu Zufall und was meinen wir damit?
- Das Modell von Wright & Fisher (nicht historisch präsentiert)
  - ↳ vorwärts in der Zeit: Allelfrequenzen
  - ↳ rückwärts in der Zeit: Genealogien
- Limiten für große Populationen:
  Wright-Fisher diffusion und Kingman-Koaleszent
  (Dualität)
- Mathematik ⇄ Biologie

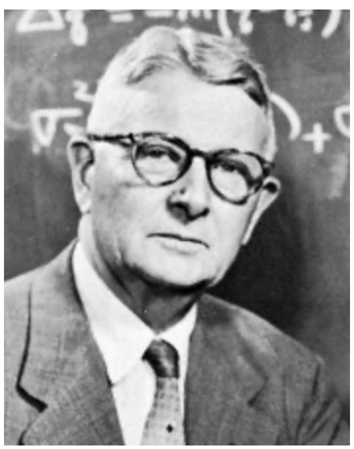
The Origin of Population Genetics

Mendel v. Darwin





Modern Synthesis



Wright

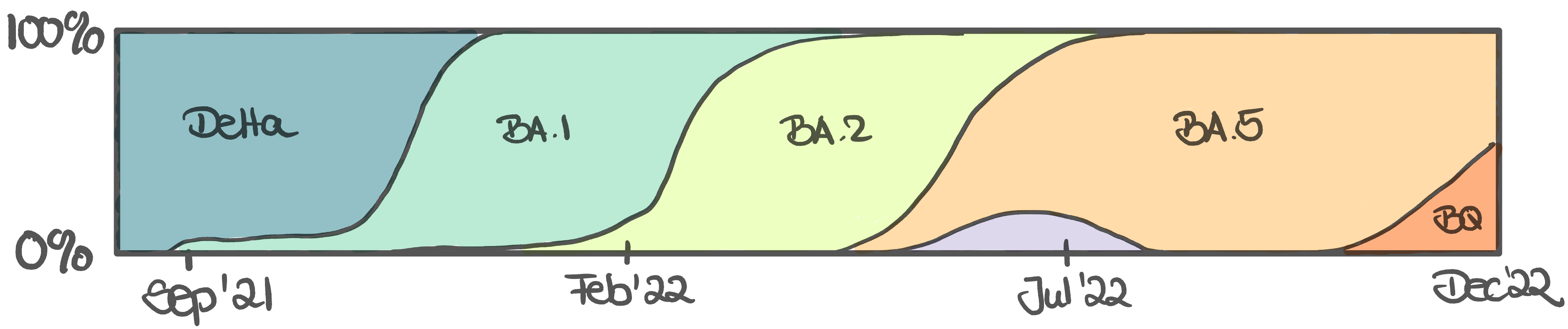
Fischer

Haldane

WHY IS RANDOMNES NECESSARY (TO MODEL EVOLUTION)?

* Randomness is inherent in reproduction! *

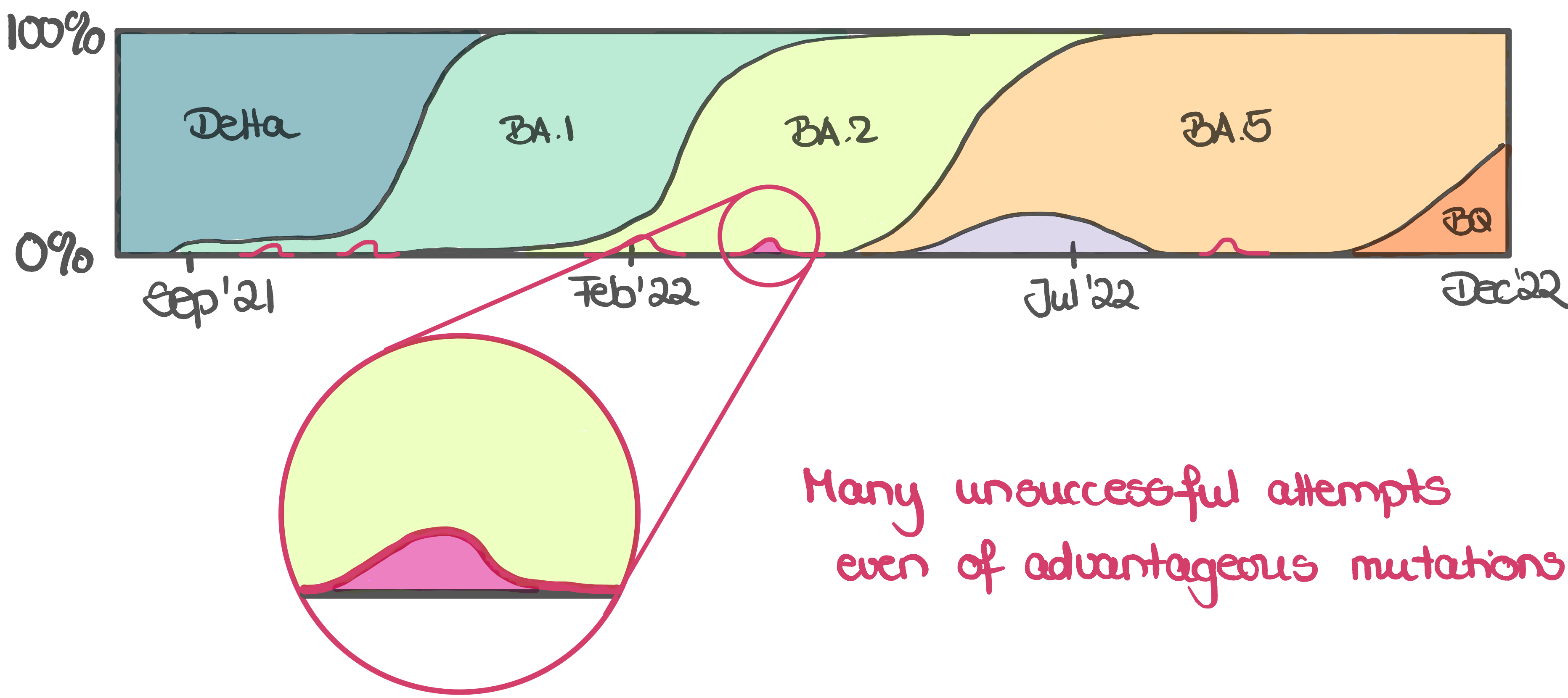
Sketch of Covid-19 evolution:



WHY IS RANDOMNES NECESSARY (TO MODEL EVOLUTION)?

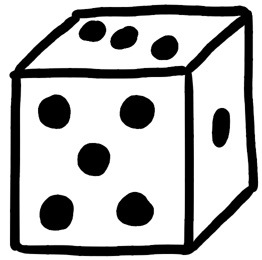
* Randomness is inherent in reproduction! *

Sketch of Covid-19 evolution:



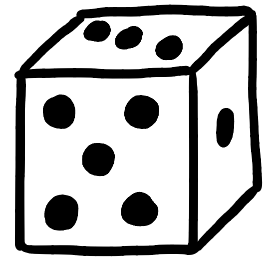
Many unsuccessful attempts even of advantageous mutations!

WHAT IS RANDOMNESS?



WHAT IS RANDOMNESS?

Maybe there is a function f such that

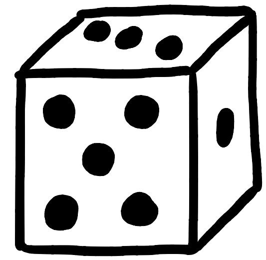


$$f(\omega) = \text{„eyes of the die"}$$

all the information about the state & history of the universe

WHAT IS RANDOMNESS?

Maybe there is a function f such that



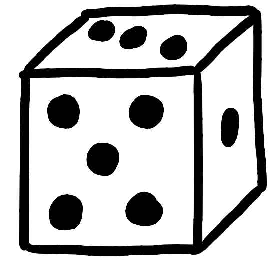
$f(\omega)$ = „eyes of the die"

all the information about the state & history of the universe

- $f: \mathbb{R} \to \mathbb{R}$
  $x \mapsto x^2 = f(x)$
- $f: \{\text{audience}\} \to \{\text{words}\}$
  $x \mapsto \text{name}(x) = f(x)$
- $f: \Omega \to E$

WHAT IS RANDOMNESS?

Maybe there is a function f such that



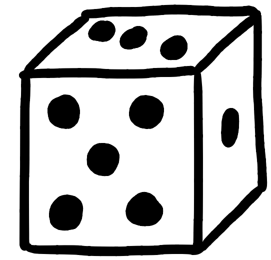
$f(\omega)$ = „eyes of the die"

all the information about the state & history of the universe

- $f: \mathbb{R} \to \mathbb{R}$
  $x \mapsto x^2 = f(x)$
- $f: \{\text{audience}\} \to \{\text{words}\}$
  $x \mapsto \text{name}(x) = f(x)$
- $f: \Omega \to E$

What can I say about $f(\omega)$ if I don't know „ω" or „$\mapsto$" but I do know „how often $f(\omega) = 5$"?

WHAT IS RANDOMNESS?



Maybe there is a function f such that

$$f(\omega) = \text{„eyes of the die"}$$

all the information about the state & history of the universe

- $f: \mathbb{R} \to \mathbb{R}$
  $x \mapsto x^2 = f(x)$
- $f: \{\text{audience}\} \to \{\text{words}\}$
  $x \mapsto \text{name}(x) = f(x)$
- $f: \Omega \to E$

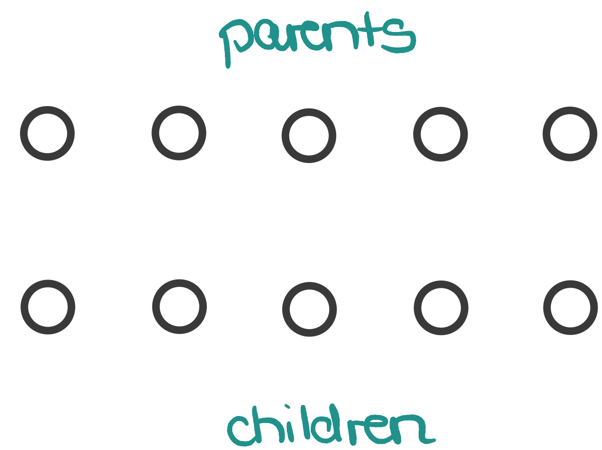
What can I say about $f(\omega)$
if I don't know „ω" or „$\mapsto$"
but I do know „how often $f(\omega) = 5$"?

$\rightsquigarrow$ $\boxed{\mathbb{P}(f = 5)}$ mostly: $\cancel{f}$ $X, Y, Z, \Pi \ldots$

The Wright-Fischer Random Graph

* neutral *

Set-up: haploid population,
discrete generations,
fixed population size N



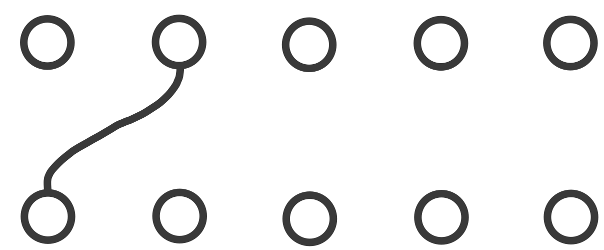
Reproduction:

The Wright-Fischer Random Graph

Set-up: haploid population,
discrete generations,
fixed population size N

Reproduction: multinomial sampling

- Each offspring individual chooses a parent (1,. . , N) uniformly at random, independently (with replacement)
-

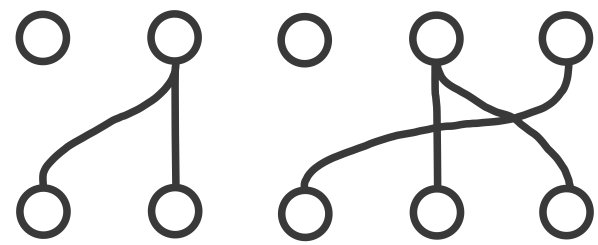


THE WRIGHT-FISCHER RANDOM GRAPH

Set-up: haploid population,
discrete generations,
fixed population size N

Reproduction: multinomial sampling

- Each offspring individual chooses a parent $(1, .., N)$ uniformly at random, independently (with replacement)
-

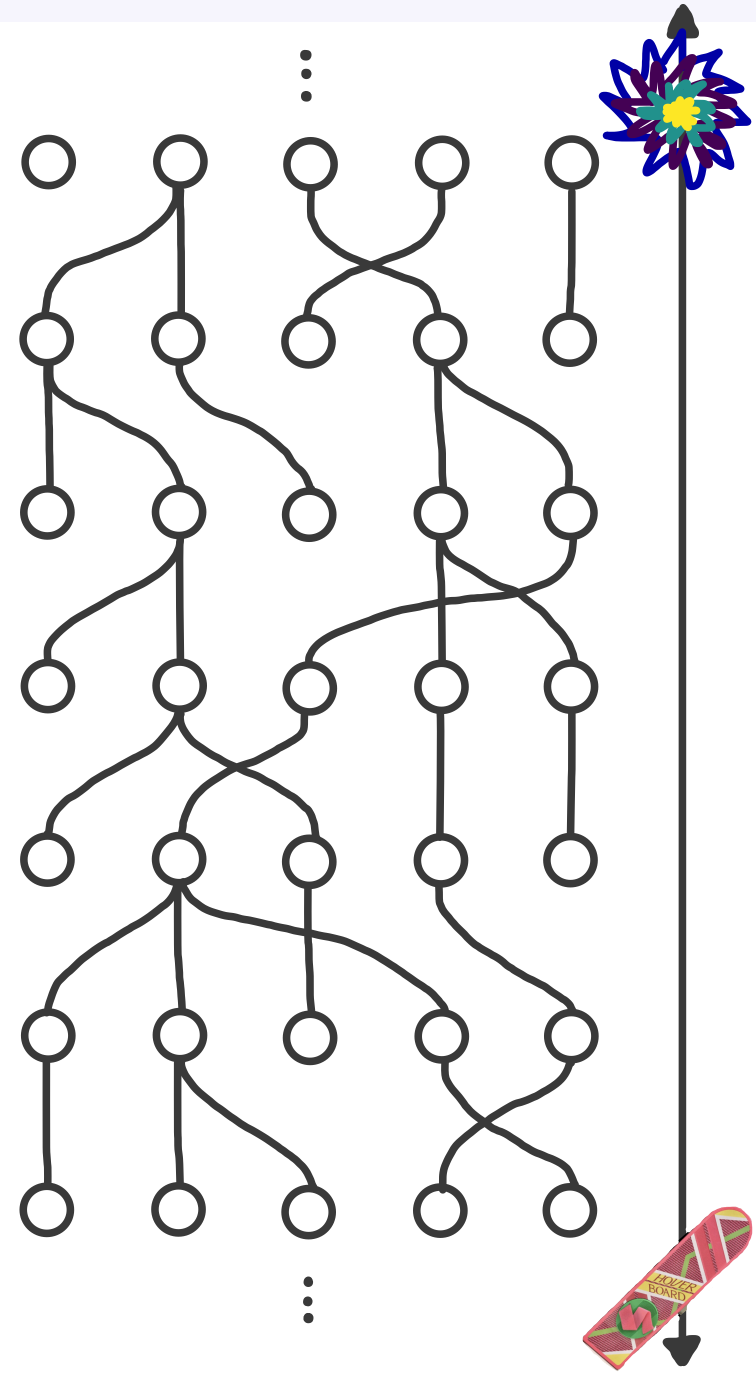


The Wright-Fischer Random Graph

Set-up: haploid population,
discrete generations,
fixed population size N

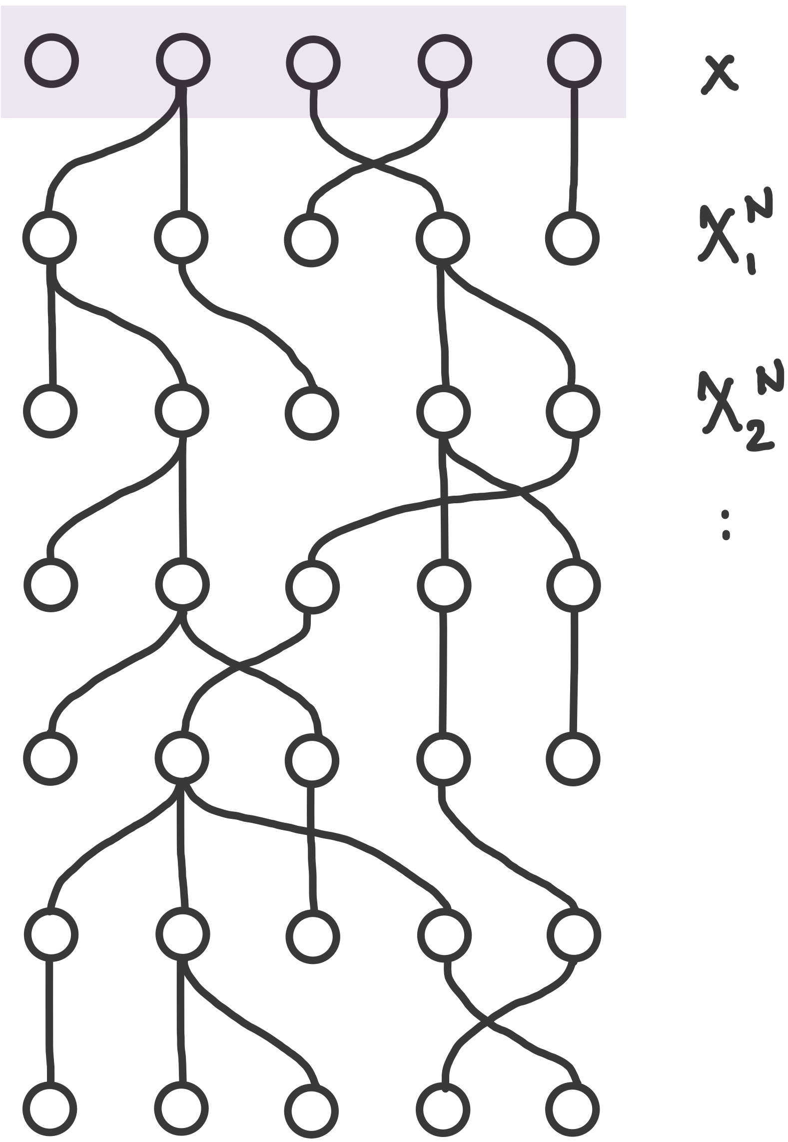
Reproduction: multinomial sampling

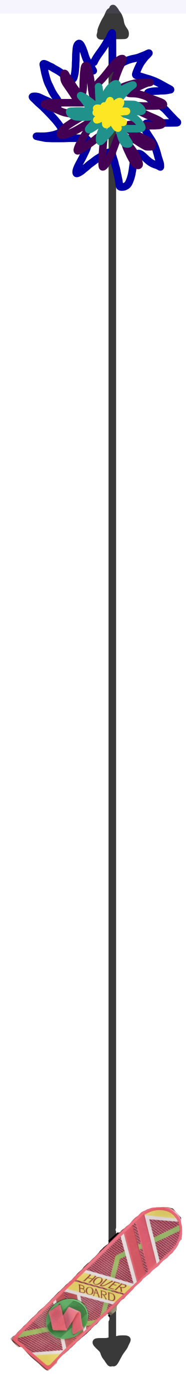
- Each offspring individual chooses a parent (1, . . , N) uniformly at random, independently (with replacement)
- Repeat independently



FORWARD: 2 types ○ ●

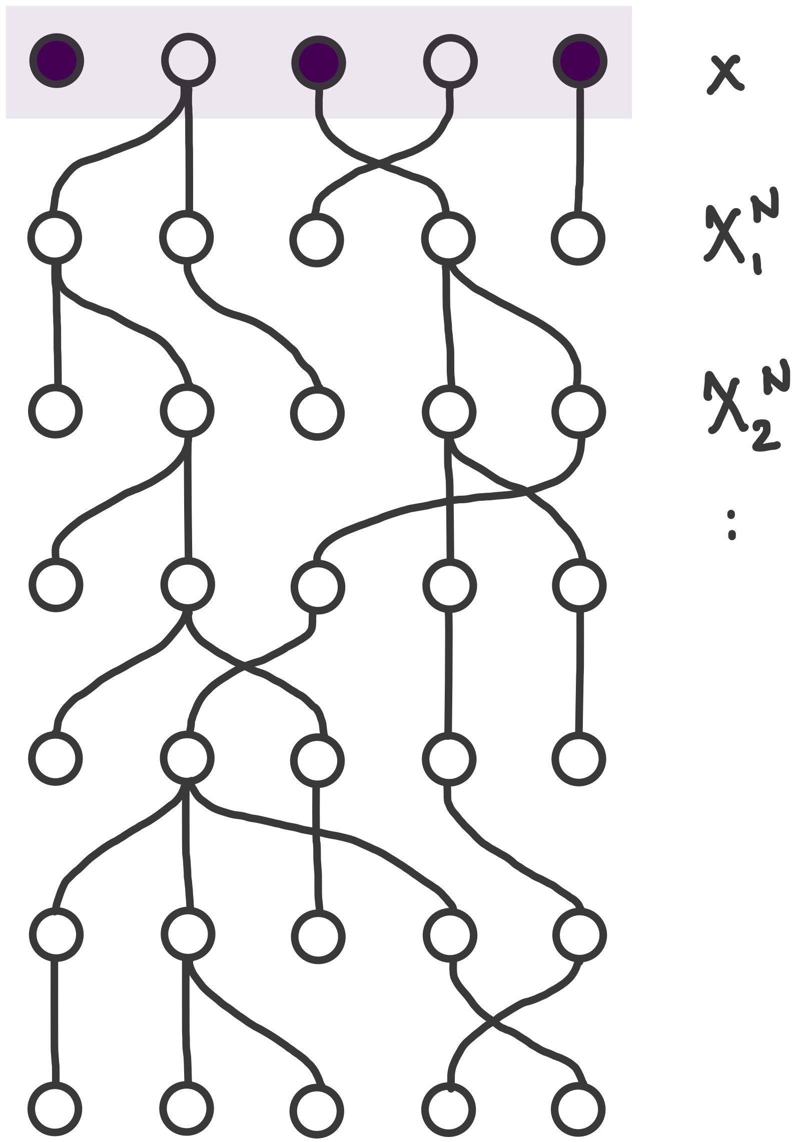
Frequency of ●

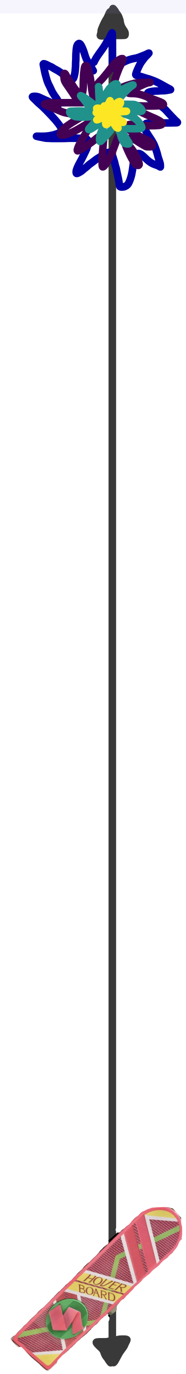




FORWARD: 2 types ○ ●

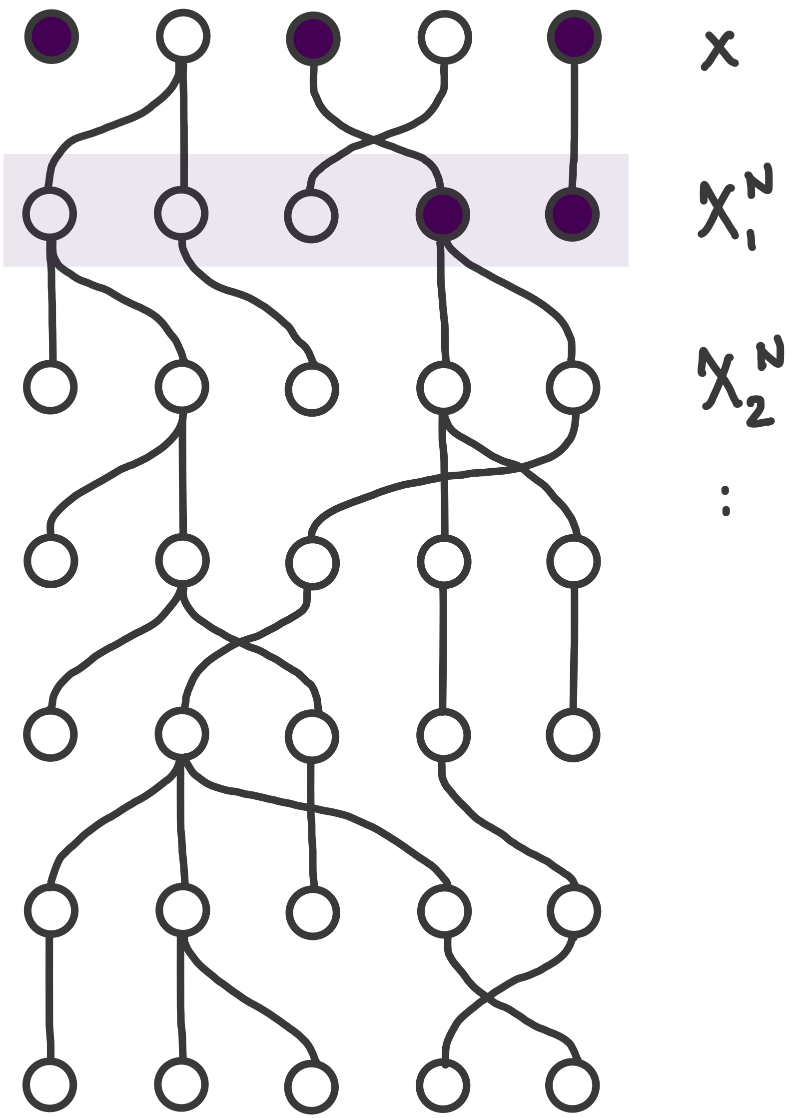
Frequency of ●

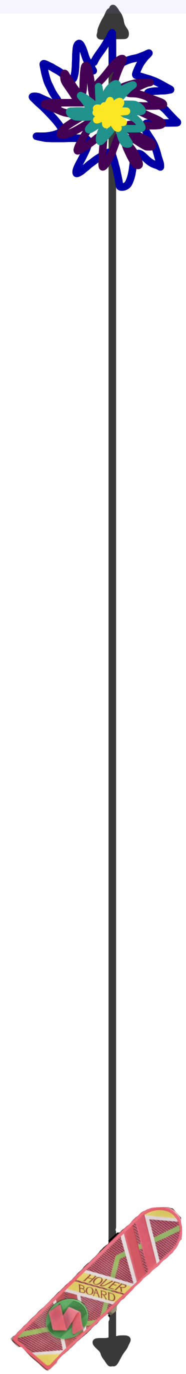




FORWARD: 2 types ○ ●

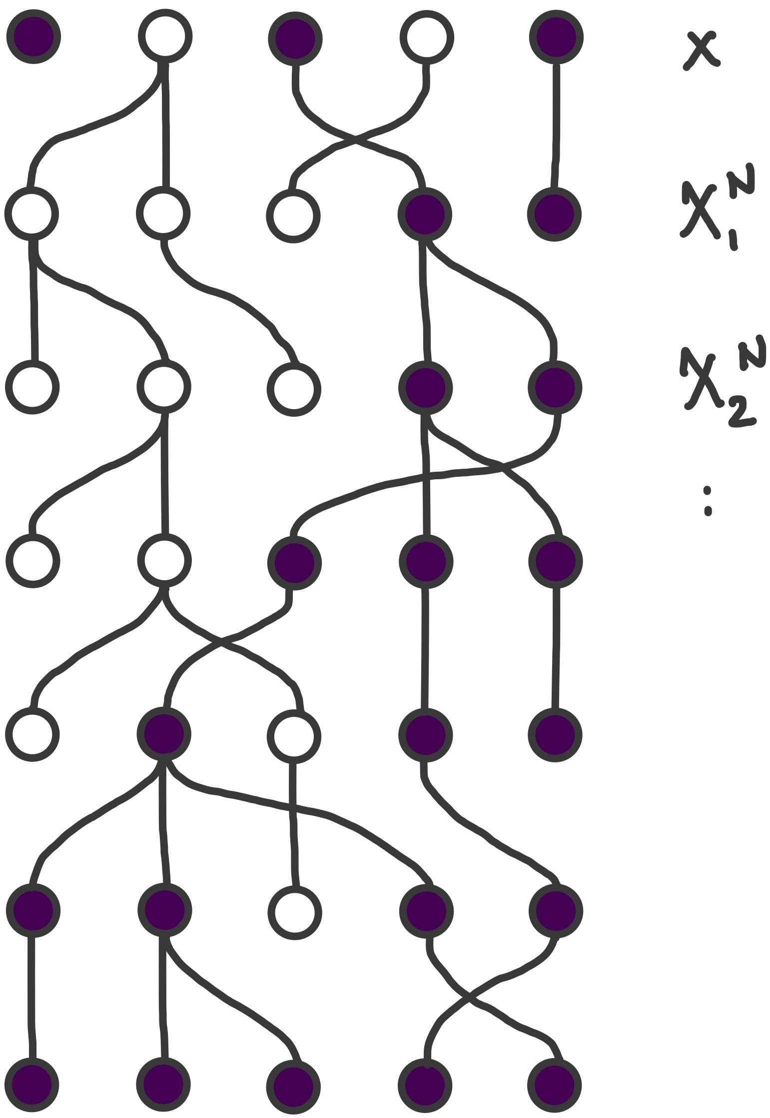
Frequency of ●

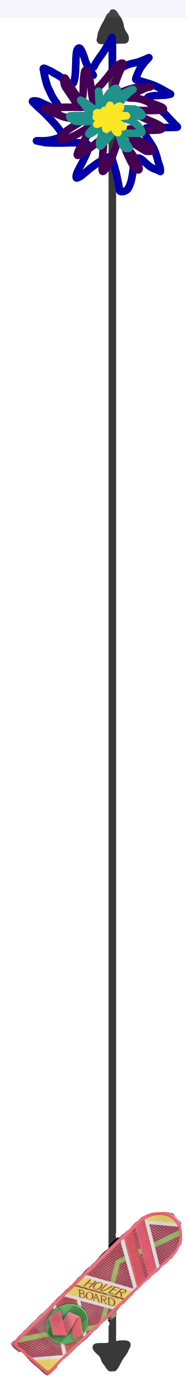




FORWARD: 2 types ○ ●

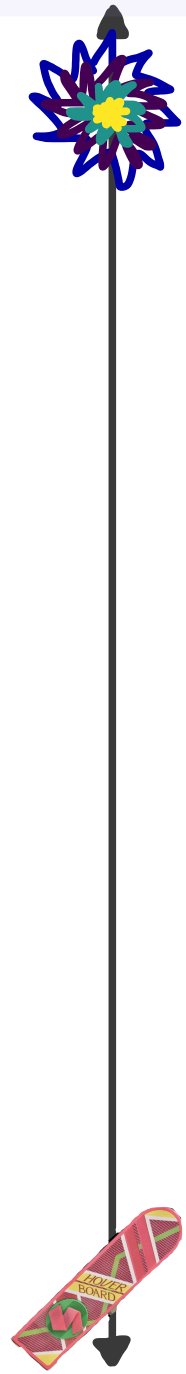
Frequency of ●



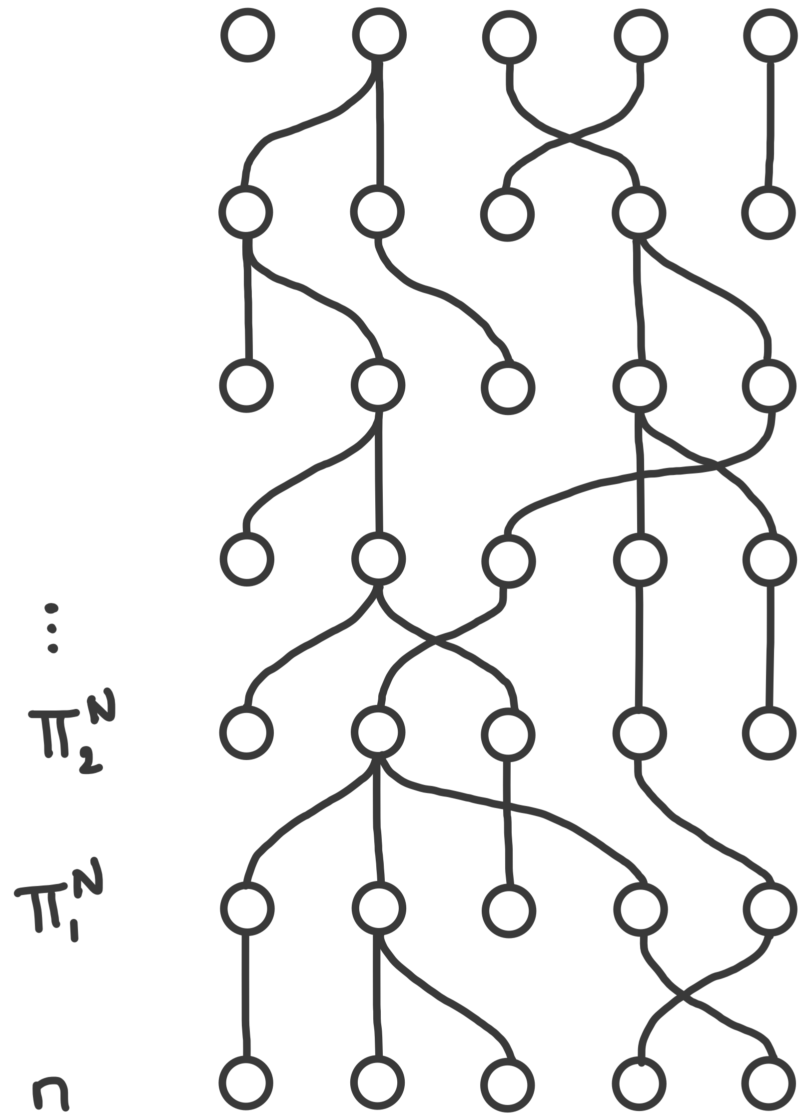


FORWARD: 2 types ○ ● | BACKWARD

Frequency of ●

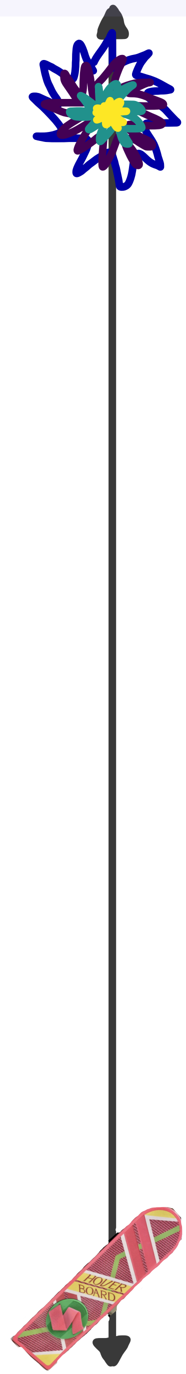


Genealogy of a sample

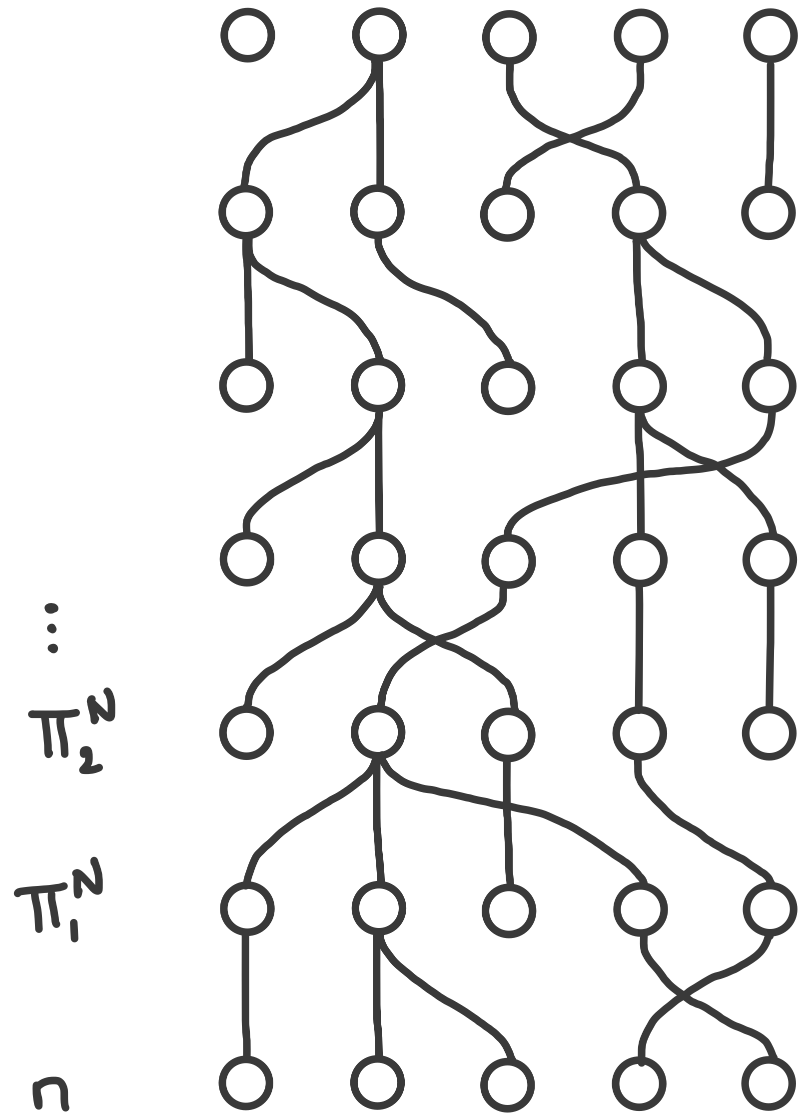


FORWARD: 2 types ○ ● | BACKWARD

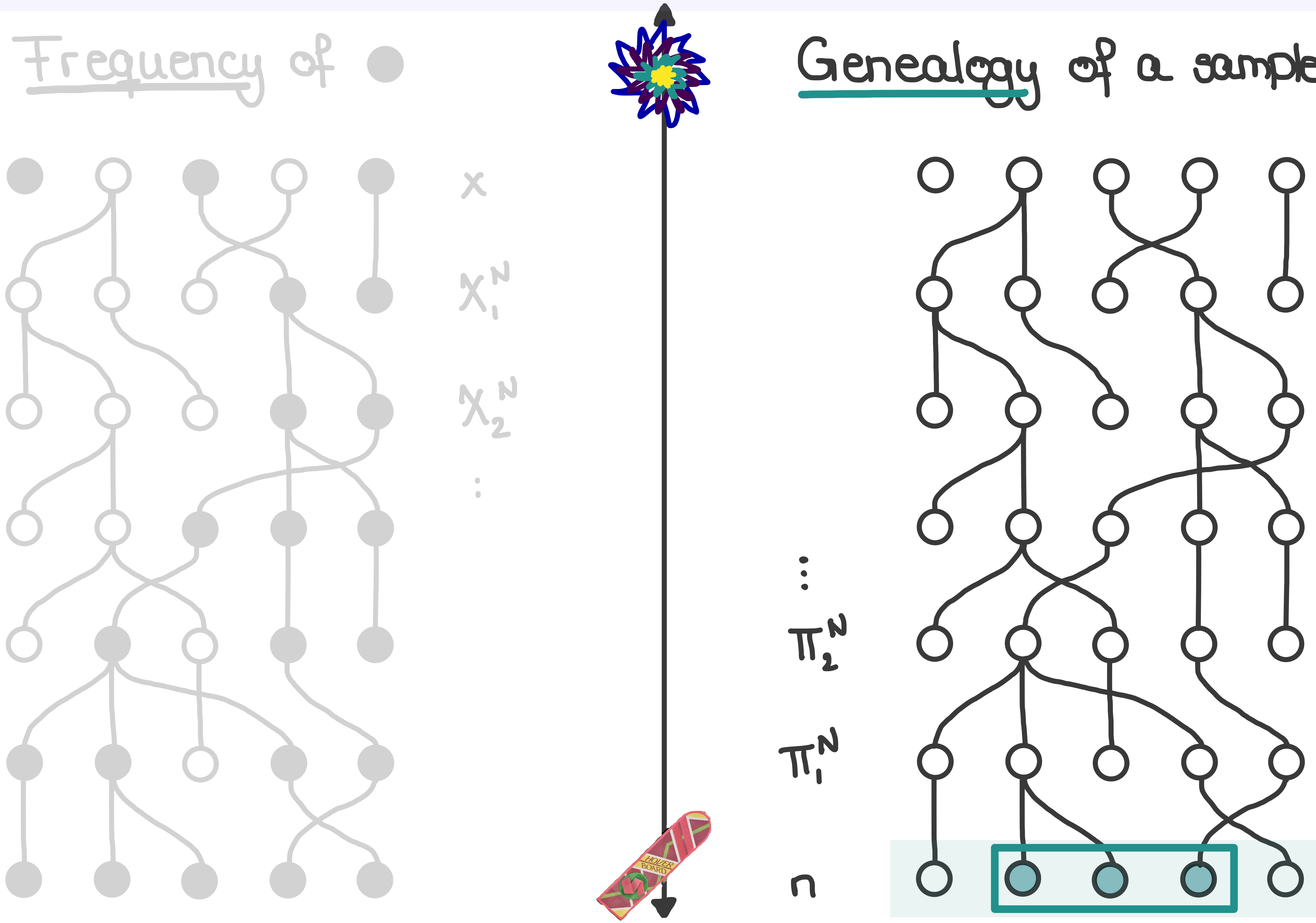
Frequency of ●



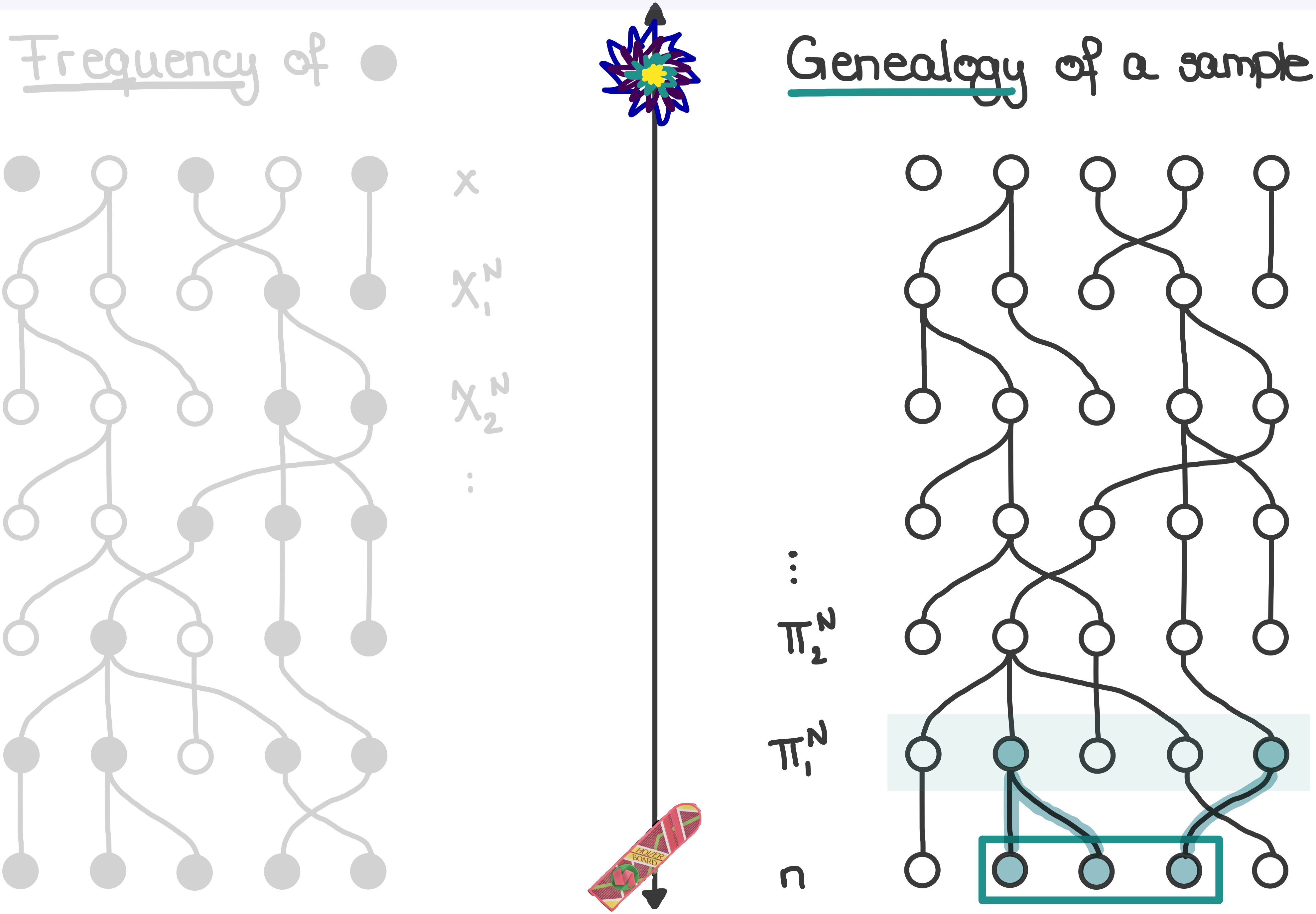
Genealogy of a sample



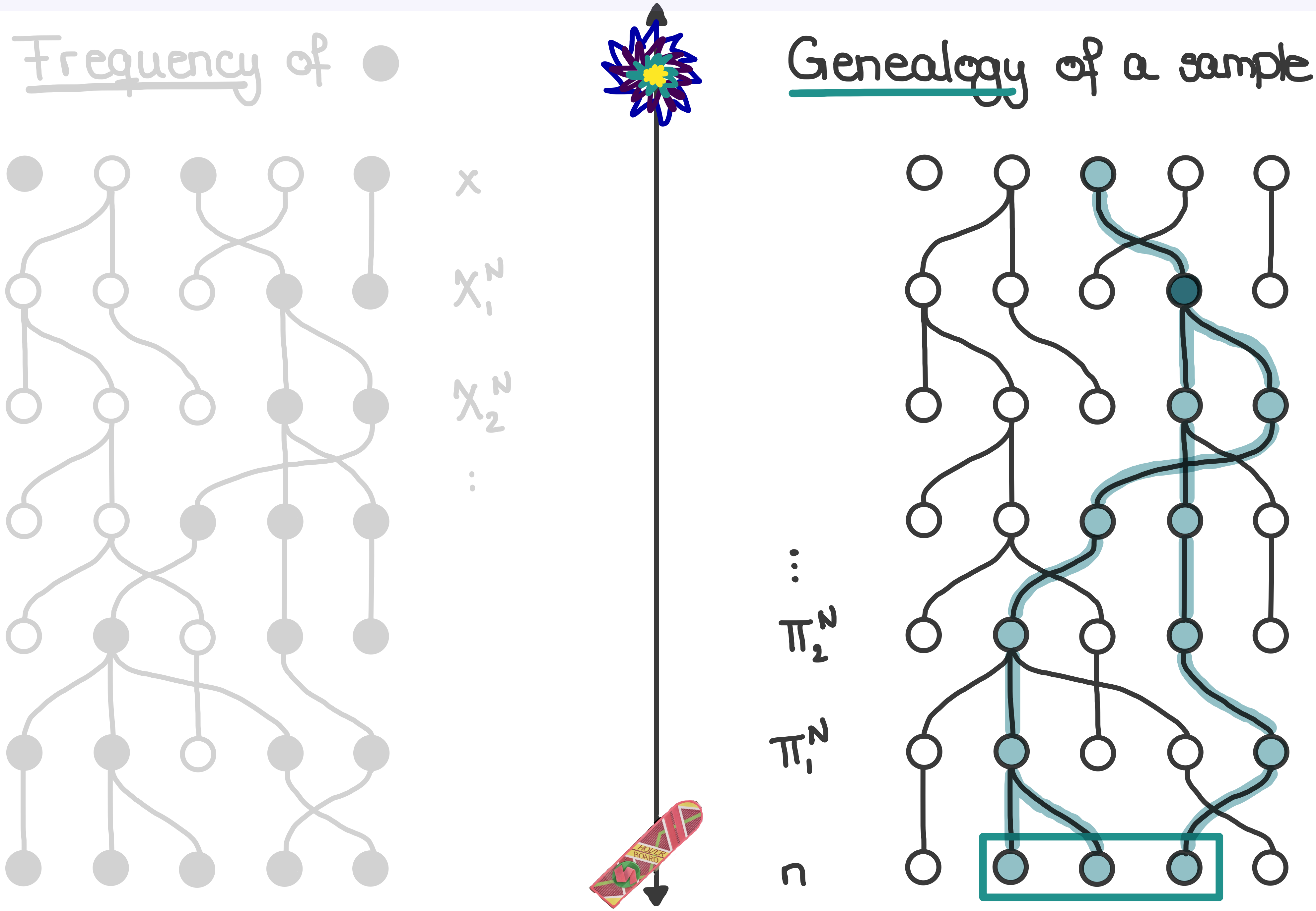
FORWARD: 2 types ○ ● | BACKWARD



FORWARD: 2 types ○ ● | BACKWARD

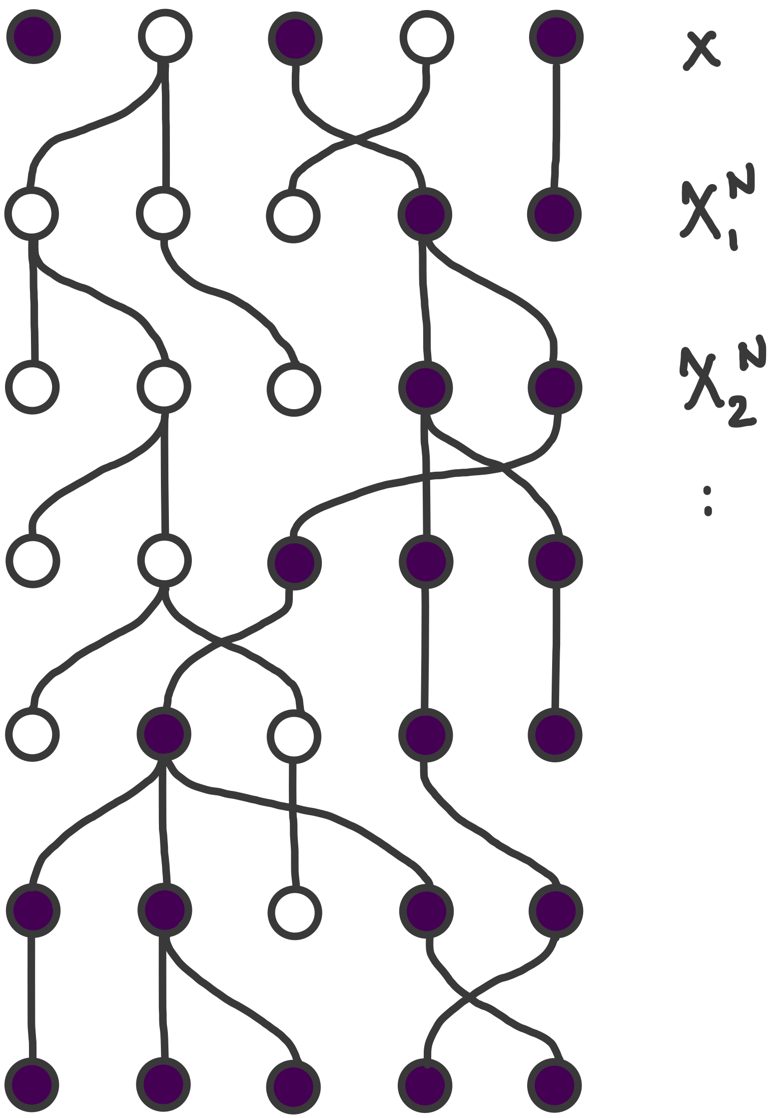


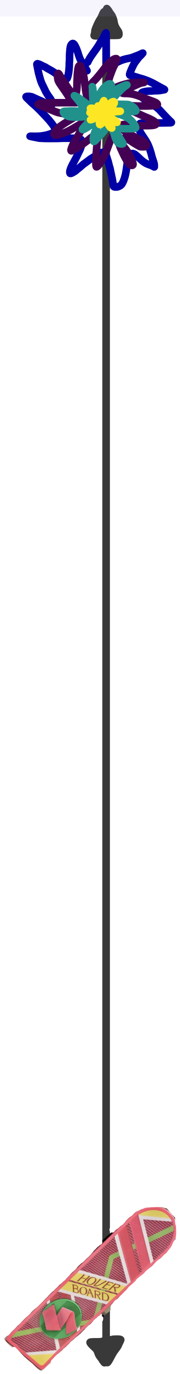
FORWARD: 2 types ○ ● | BACKWARD



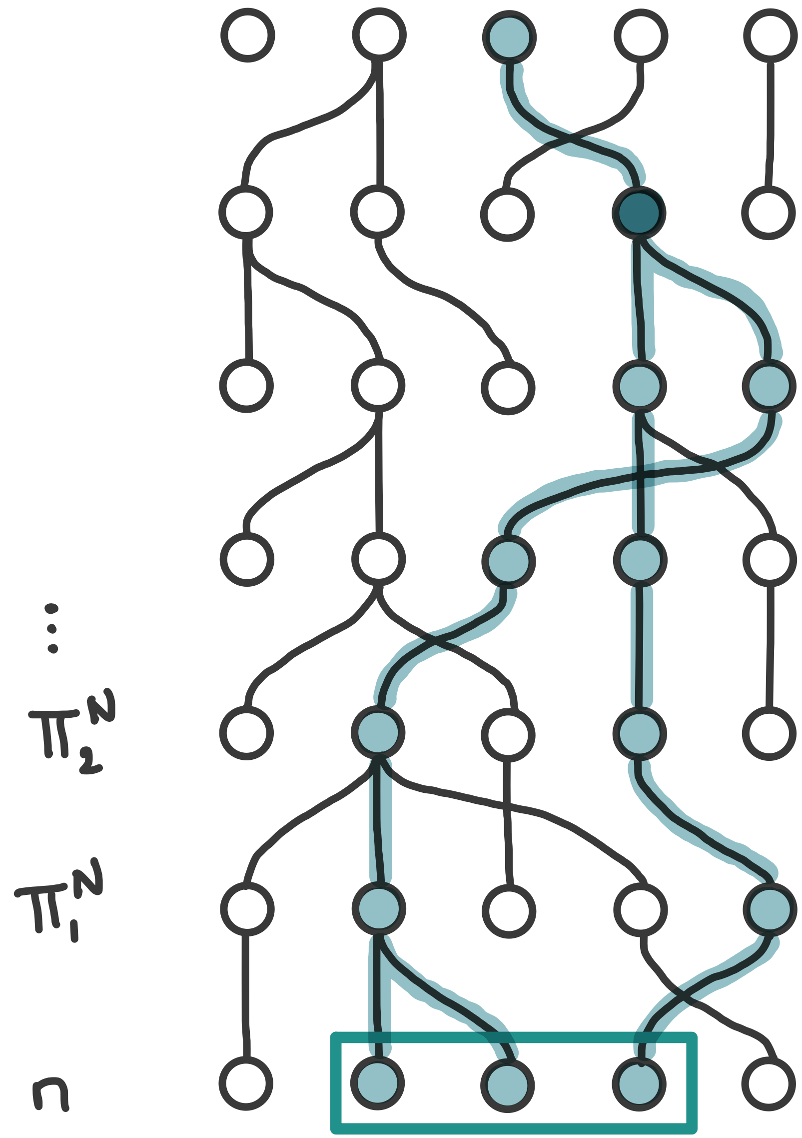
FORWARD: 2 types ○ ● | BACKWARD

Frequency of ●





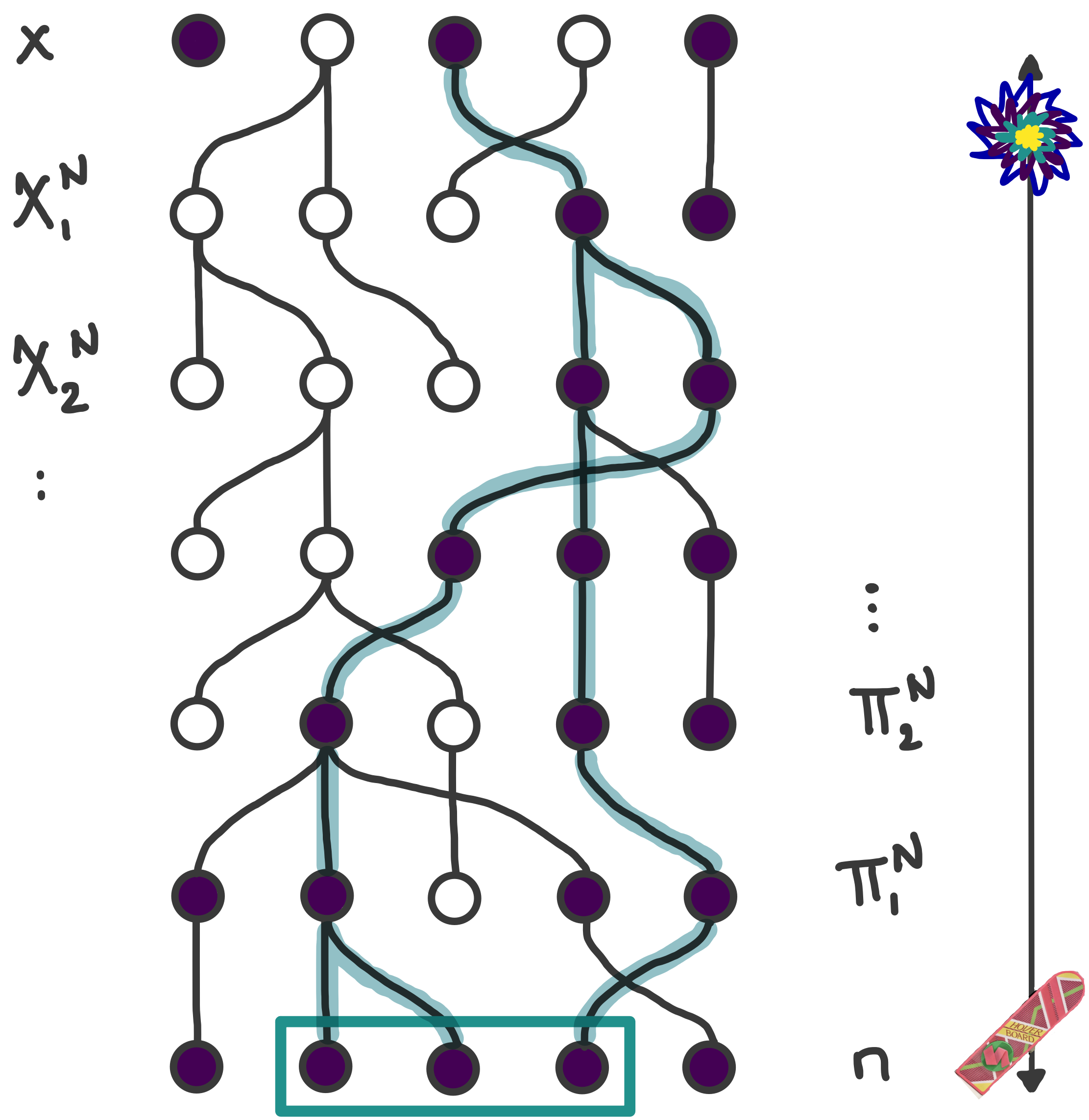
Genealogy of a sample



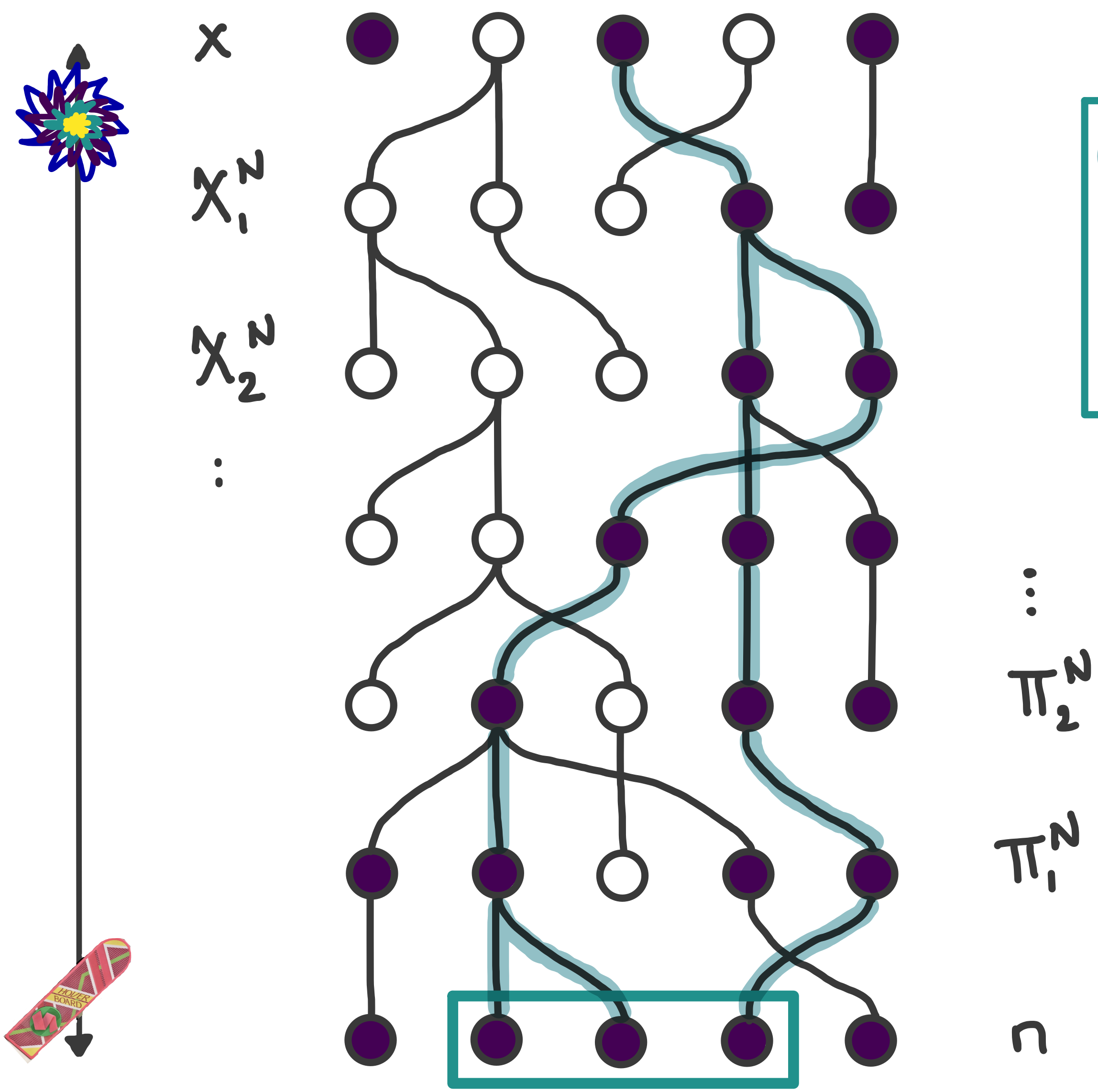
FORWARD: 2 types ○ ● | BACKWARD

Frequency of ●

Genealogy of a sample



DUALITY



LEMMA:

$\forall k \in \mathbb{N} \;\; \forall x \in [0,1] \;\; \forall n \in \mathbb{N}_0$

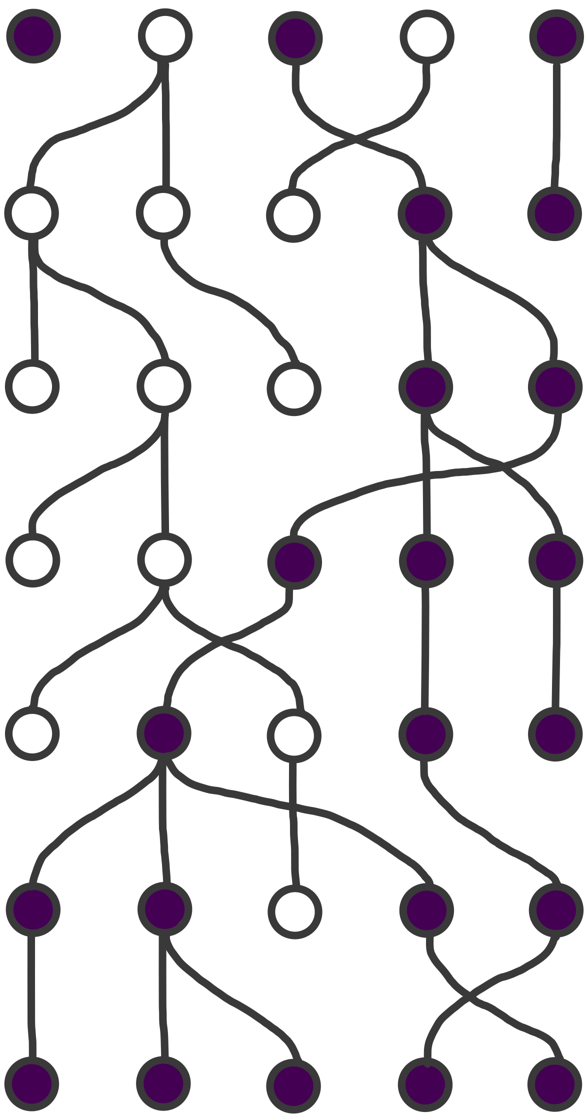
$$\mathbb{E}_x\left[X_k^n\right] = \mathbb{E}^n\left[x^{|\pi_k|}\right]$$

MOMENT DUALITY

$h(x,n) = x^n$

SCALING LIMITS: $N \to \infty$

Frequency of ●



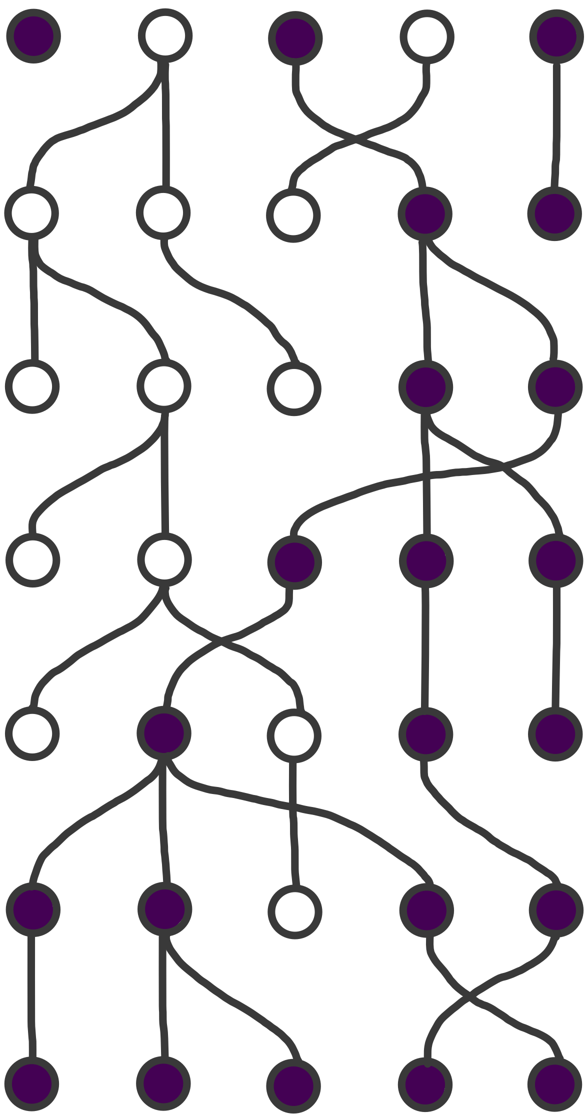
x

X_1^N

$$X_1^N = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\{(1,i) \text{ is purple}\}}$$

SCALING LIMITS: $N \to \infty$

Frequency of ●



x

X_1^N

$$X_1^N = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\{(1,i) \text{ is purple}\}}$$

stochastic process $(X_k^N)_{k \in \mathbb{N}}$

- MARKOV CHAIN

  Given the present, the past and the future are independent.

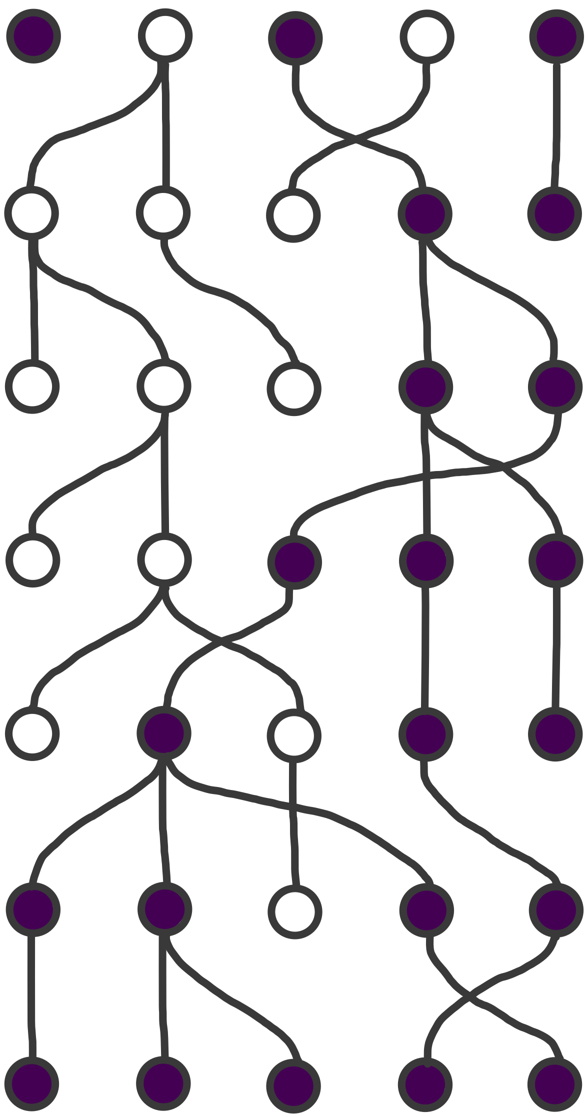
- MARTINGALE

  Fair game

SCALING LIMITS: $N \to \infty$

Frequency of ●

$$X_1^N = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{(1,i) \text{ is purple}\}}$$



x

X_1^N

stochastic process $(X_k^N)_{k \in \mathbb{N}}$

- MARKOV CHAIN

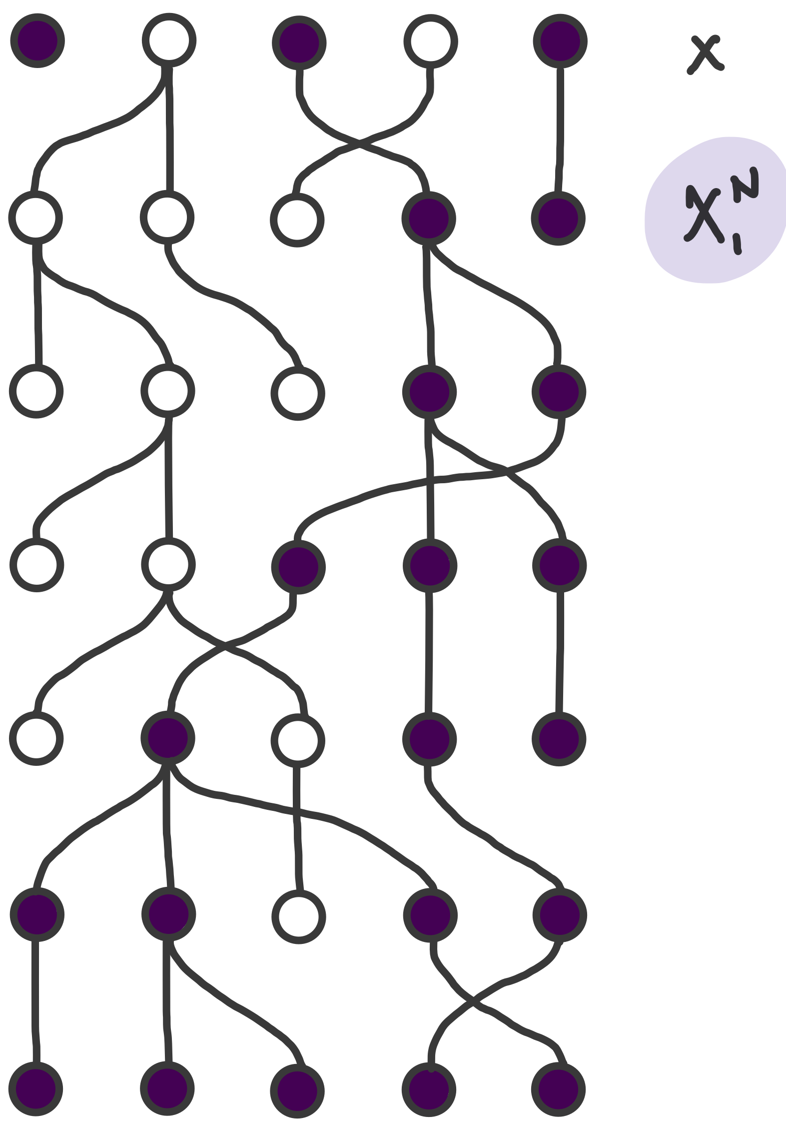
  Given the present, the past and the future are independent.

- MARTINGALE

  Fair game

SCALING LIMITS: $N \to \infty$

Frequency of ●



$$X_1^N = \frac{1}{N}\sum_{i=1}^{N} \underbrace{\mathbb{1}_{\{(1,i) \text{ is purple}\}}}_{\text{iid Bernoulli}(x)}$$

$\hat{=}$ unfair coins!

stochastic process $(X_k^N)_{k\in\mathbb{N}}$

- MARKOV CHAIN

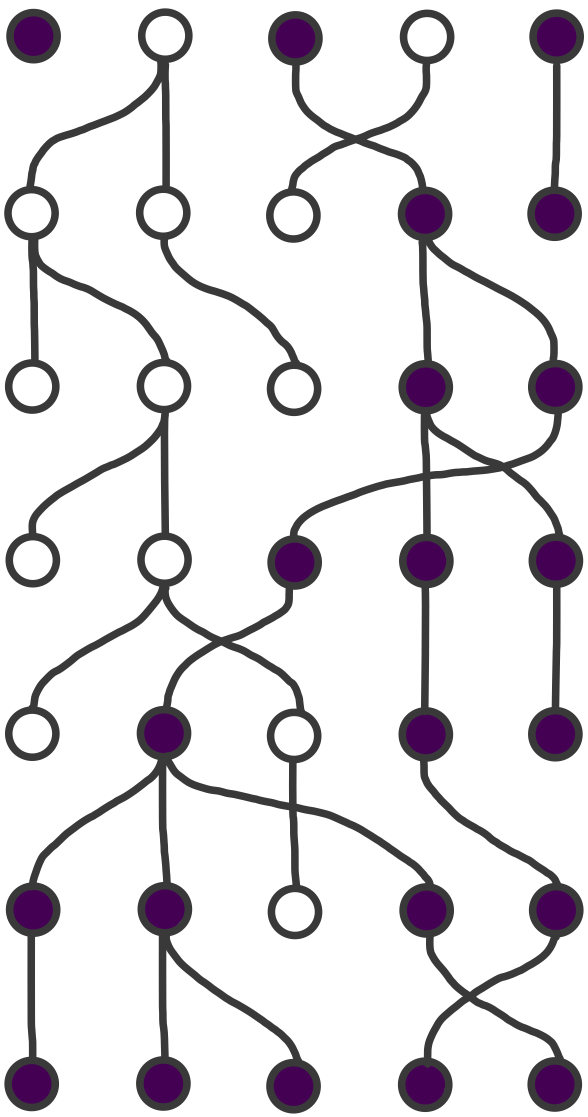
  Given the present, the past and the future are independent.

- MARTINGALE

  Fair game

SCALING LIMITS: $N \to \infty$

Frequency of ●



x

X_1^N

$$X_1^N = \frac{1}{N}\sum_{i=1}^{N} \underbrace{\mathbb{1}_{\{(1,i) \text{ is purple}\}}}_{\text{iid Bernoulli}(x)} \xrightarrow{\text{LLN}} x$$

$\hat{=}$ unfair coins!

stochastic process $(X_k^N)_{k\in\mathbb{N}}$

- MARKOV CHAIN

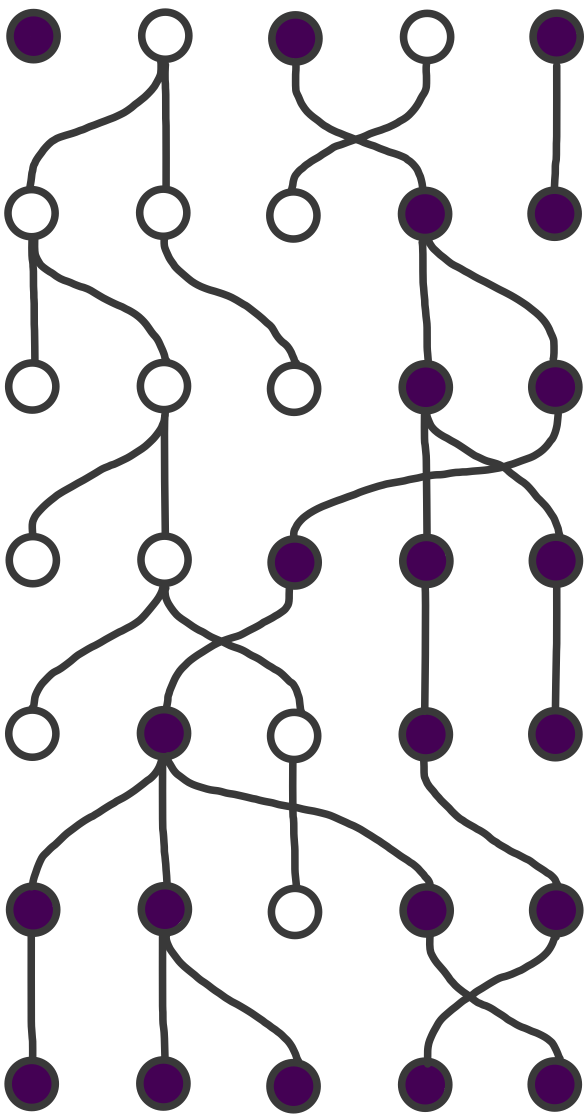
  Given the present, the past and the future are independent.

- MARTINGALE

  Fair game

SCALING LIMITS: $N \to \infty$

Frequency of ●



x

X_1^N

$$X_1^N = \frac{1}{N}\sum_{i=1}^{N} \underbrace{\mathbb{1}_{\{(1,i) \text{ is purple}\}}}_{\text{iid Bernoulli}(x)} \xrightarrow{\text{LLN}} x$$

$\hat{=}$ unfair coins!

$\Rightarrow$ RESCALE TIME, TOO, TO „COLLECT" RANDOMNESS!

stochastic process $(X_k^N)_{k \in \mathbb{N}}$

- MARKOV CHAIN

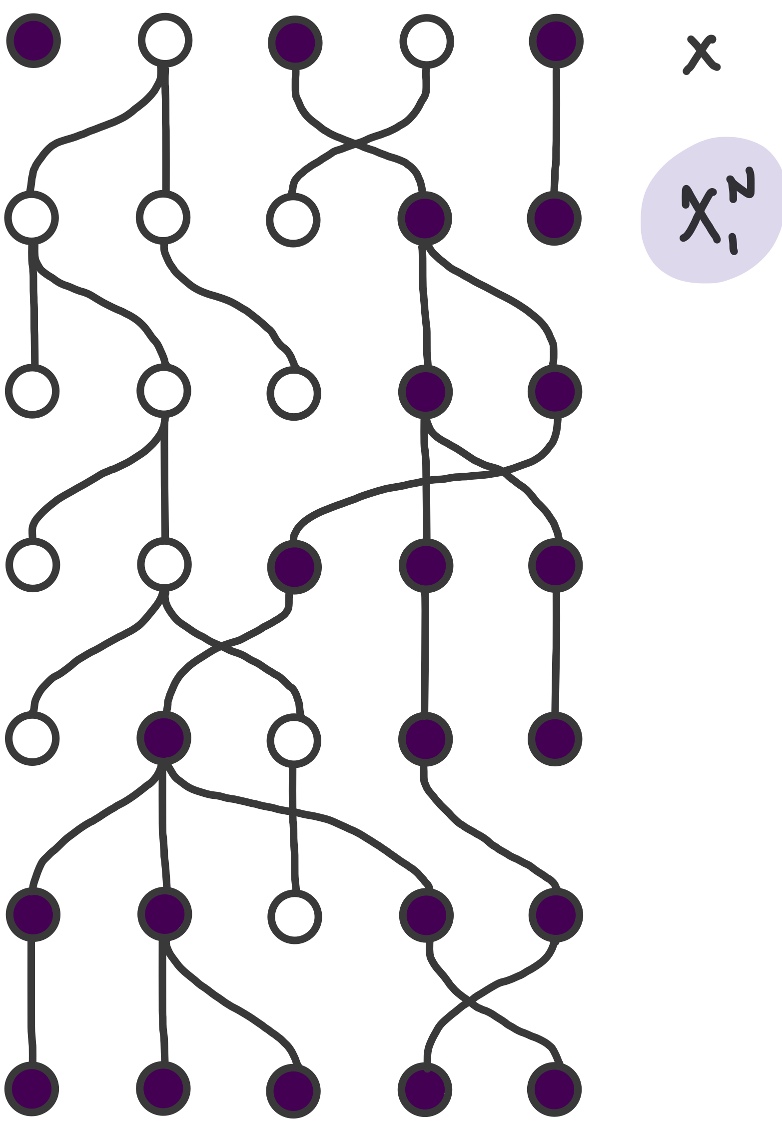
  Given the present, the past and the future are independent.

- MARTINGALE

  Fair game

SCALING LIMITS: $N \to \infty$

Frequency of ●



$$X_1^N = \frac{1}{N}\sum_{i=1}^{N} \underbrace{\mathbb{1}_{\{(1,i) \text{ is purple}\}}}_{\text{iid Bernoulli}(x)} \xrightarrow{\text{LLN}} x$$

$\hat{=}$ unfair coins!

THEOREM: WRIGHT '33, KIMURA '54

$$(X^N_{\lfloor Nt \rfloor})_{t\geq 0} \xrightarrow[N\to\infty]{w} (X_t)_{t\geq 0}$$

where $(X_t)_{t\geq 0}$ is the solution to

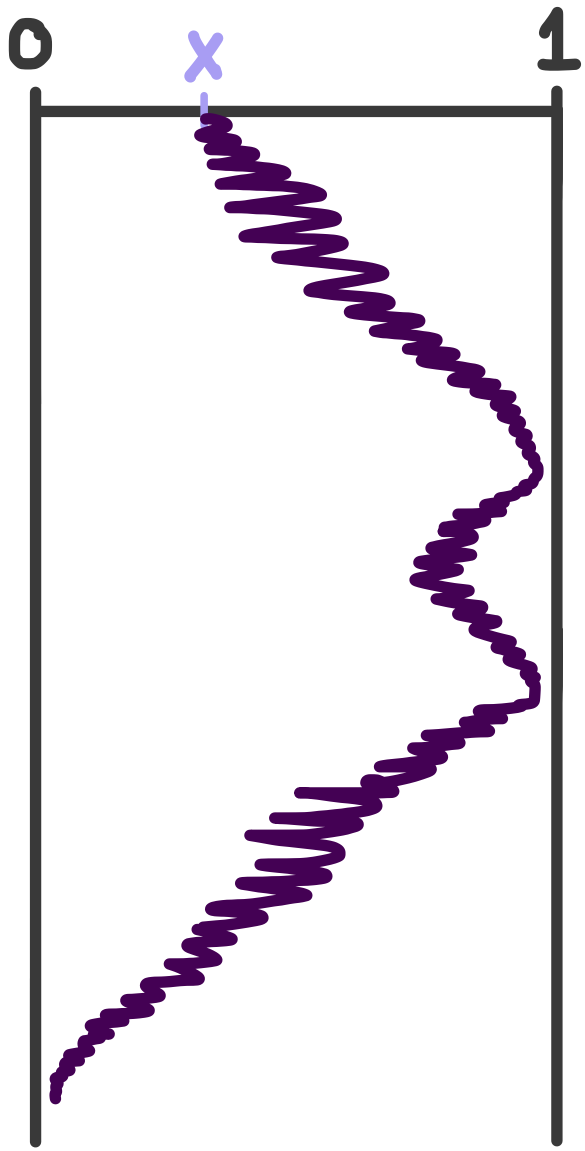
$$dX_t = \sqrt{X_t(1-X_t)}\, dB_t \, .$$

WRIGHT FISHER DIFFUSION

BROWNIAN MOTION

SCALING LIMITS: $N \to \infty$

Frequency of ●



$$X_1^N = \frac{1}{N} \sum_{i=1}^{N} \underbrace{\mathbb{1}_{\{(1,i) \text{ is purple}\}}}_{\text{iid Bernoulli}(x)} \xrightarrow{\text{LLN}} x$$

$\hat{=}$ unfair coins!

THEOREM: WRIGHT '33, KIMURA '54

$$(X^N_{\lfloor Nt \rfloor})_{t \geq 0} \xrightarrow[N \to \infty]{w} (X_t)_{t \geq 0}$$

where $(X_t)_{t \geq 0}$ is the solution to

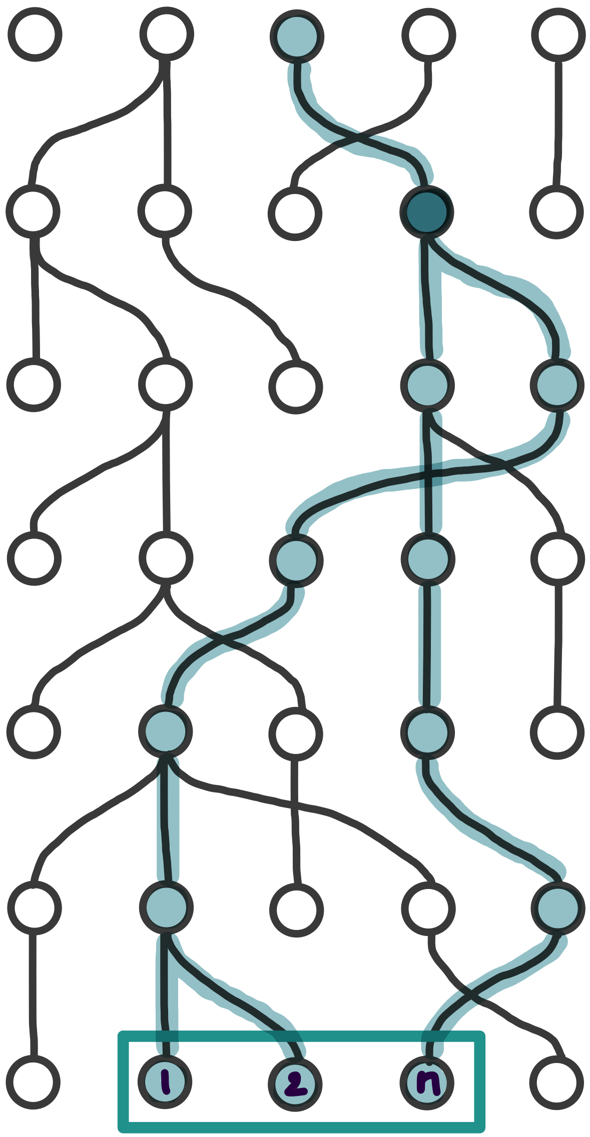
$$dX_t = \sqrt{X_t(1 - X_t)}\, dB_t \, .$$

WRIGHT FISHER DIFFUSION

BROWNIAN MOTION

SCALING LIMITS: $N \to \infty$

Genealogy of a sample

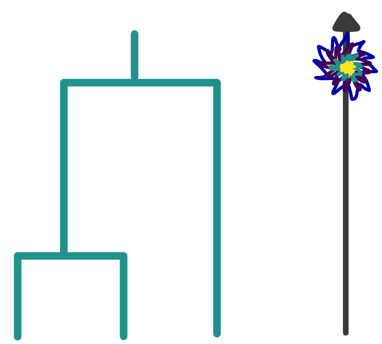


$\Pi_k^N \hat{=}$ ancestors of sample $\{1,..,n\}$ in generation $-k$

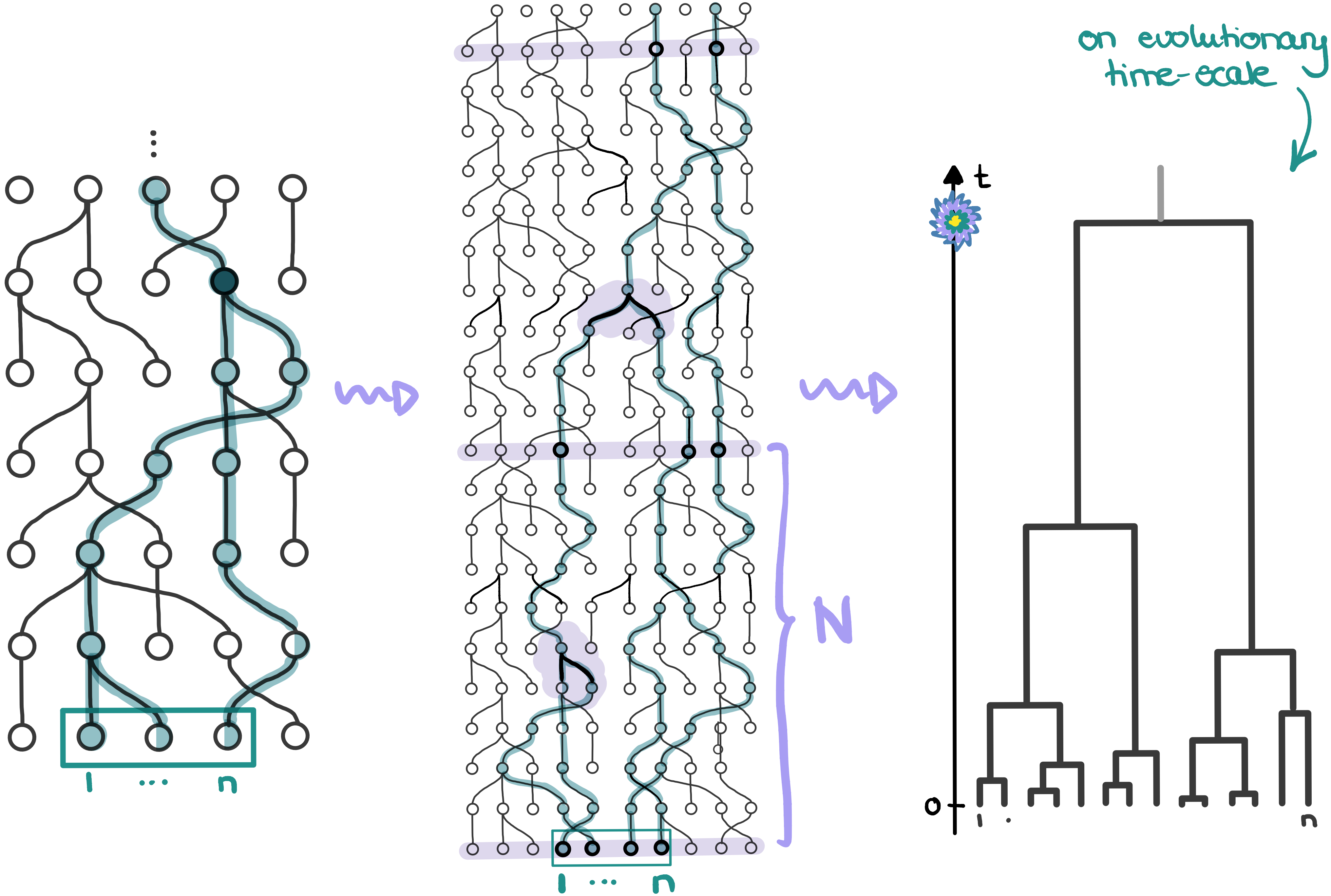
THEOREM. KINGMAN '82

$$\left(\Pi^N_{\lfloor Nt \rfloor}\right)_{t\geq 0} \xrightarrow[N\to\infty]{w} \left(\Pi_t\right)_{t\geq 0}$$

KINGMAN('S) COALESCENT



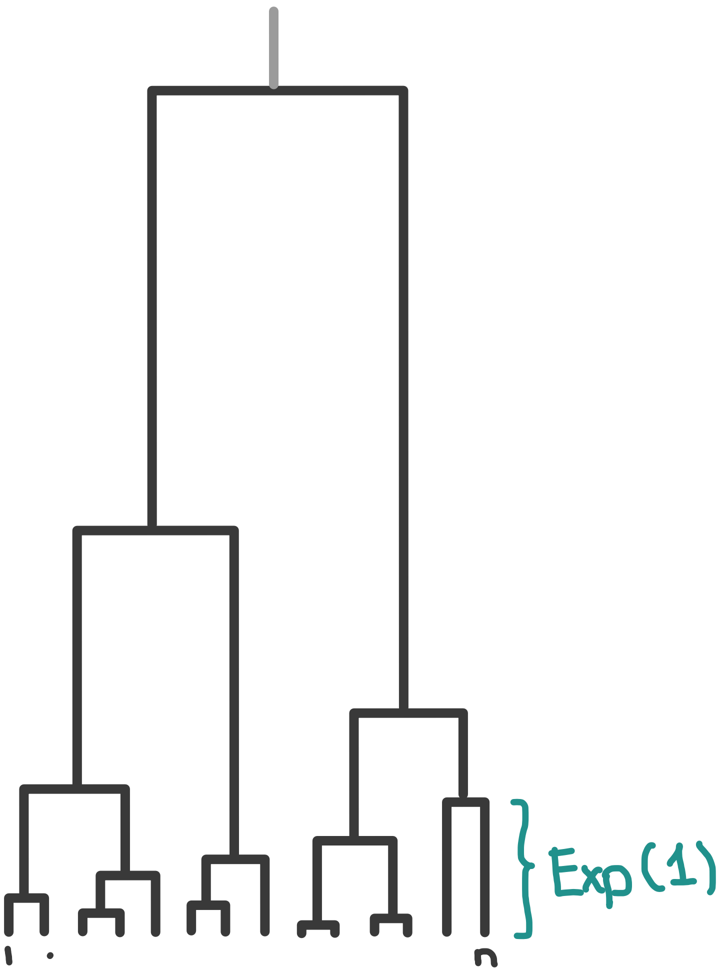
SCALING LIMITS: $N \to \infty$



Scaling Limits: $N \to \infty$

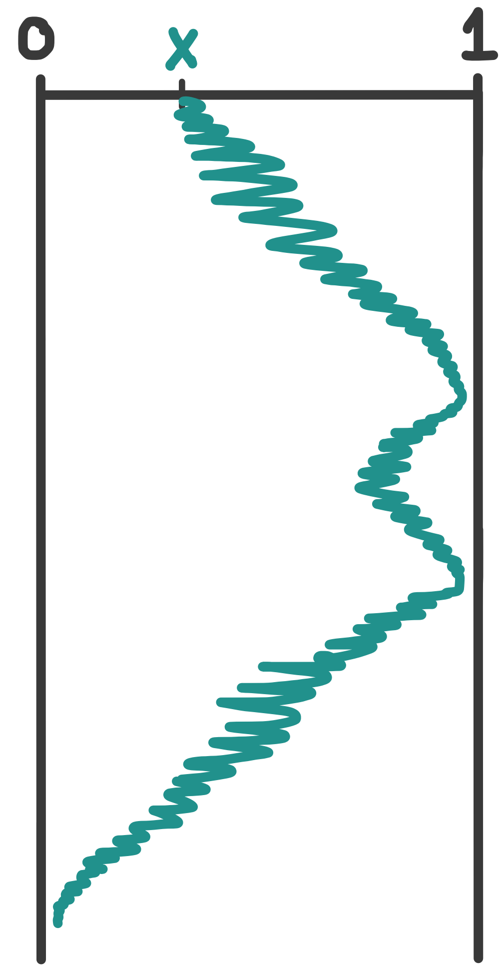
Kingman Coalescent

Genealogy of a sample



Wright-Fischer Diffusion

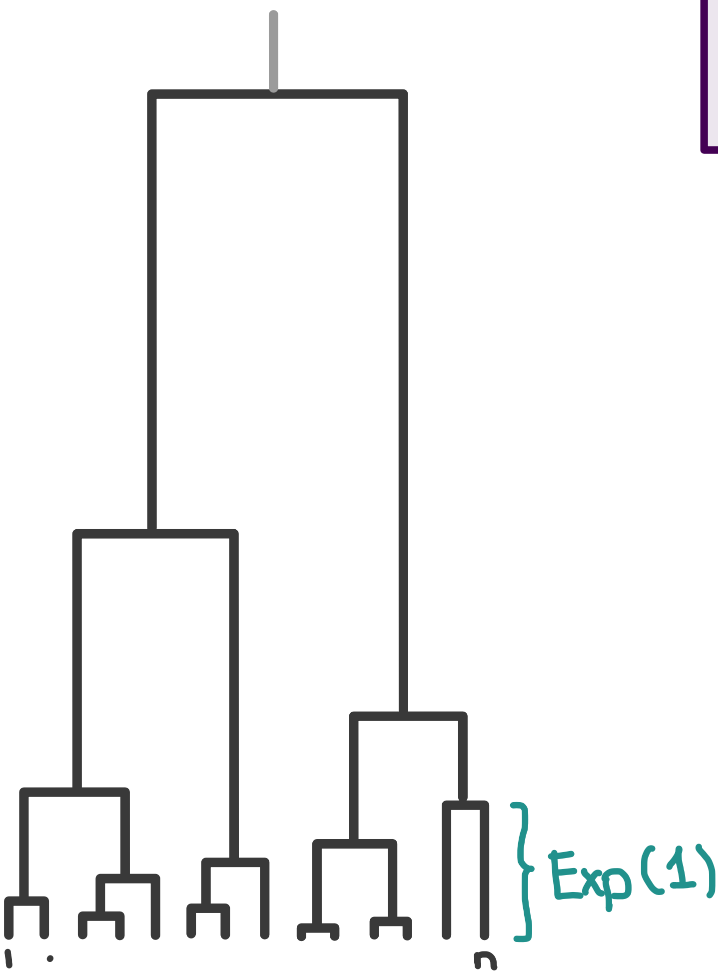
Frequency of ●



SCALING LIMITS: $N \to \infty$

KINGMAN COALESCENT

Genealogy of a sample

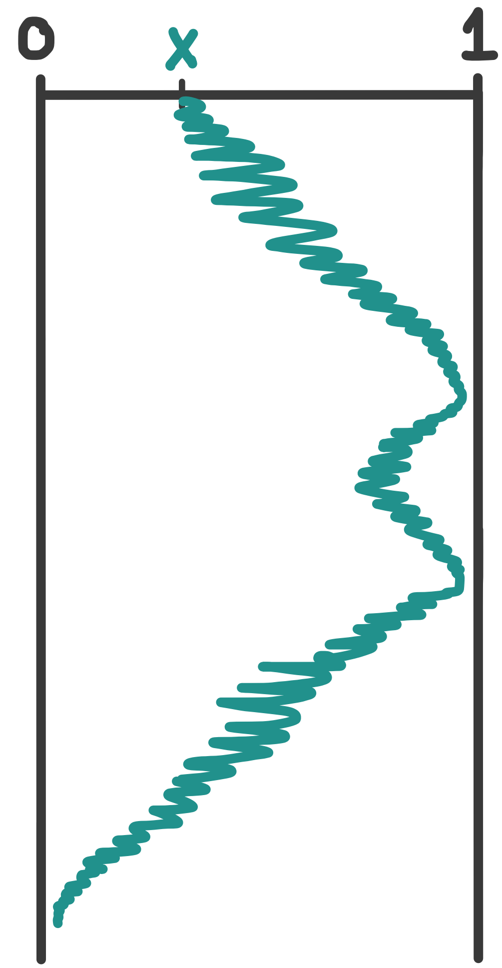


* ROBUST *

ARE UNIVERSAL SCALING LIMITS

WRIGHT-FISCHER DIFFUSION

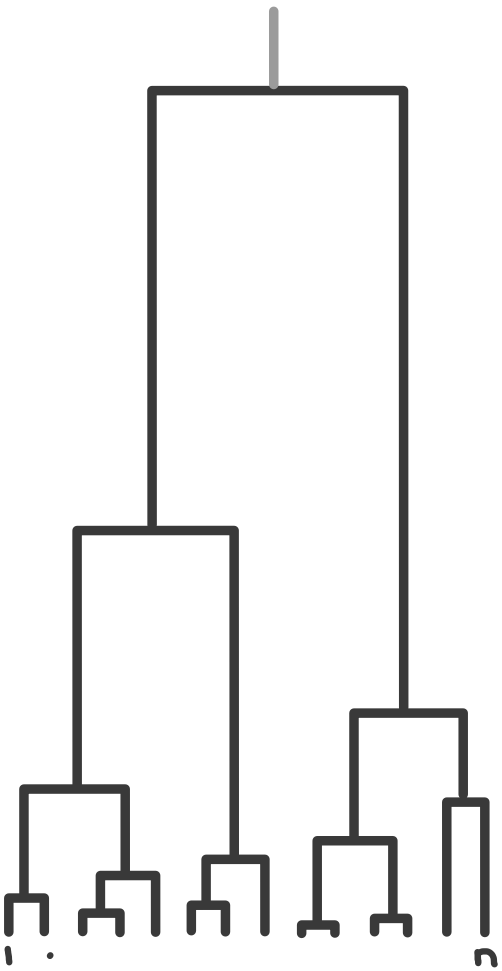
Frequency of ●



SCALING LIMITS: $N \to \infty$

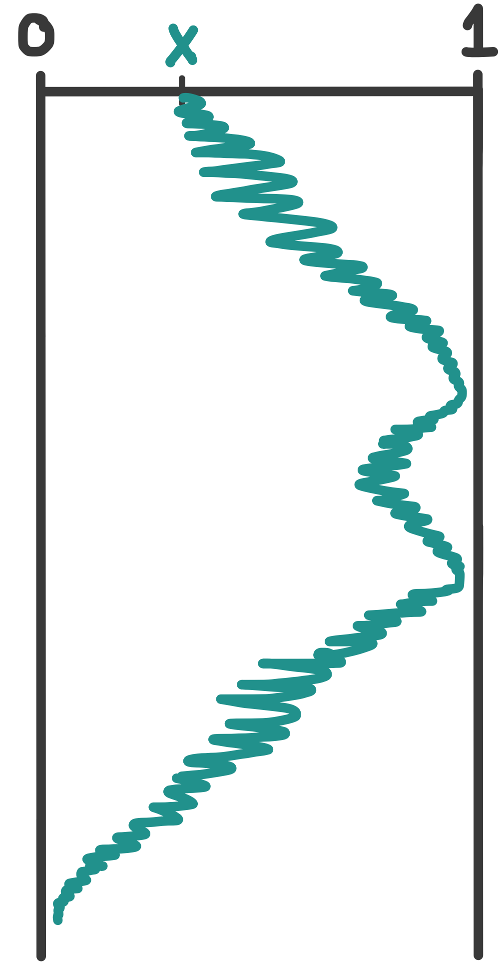
KINGMAN COALESCENT

Genealogy of a sample



WRIGHT-FISCHER DIFFUSION

Frequency of ●



* ROBUST *

ARE UNIVERSAL SCALING LIMITS

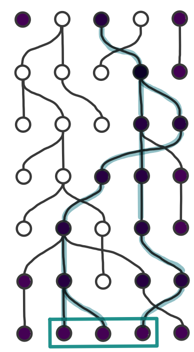
and moment duals:

LEMMA:

$\forall t \geq 0 \ \forall x \in [0,1] \ \forall n \in \mathbb{N}_0$

$$\mathbb{E}_x[X_t^n] = \mathbb{E}^n[x^{|\Pi_t|}]$$

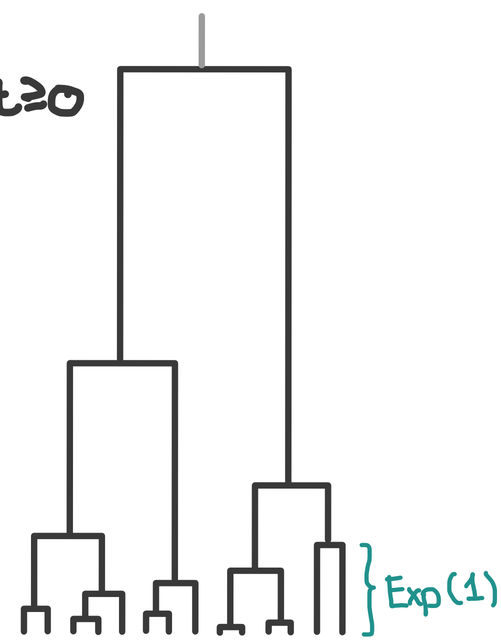
The complete picture



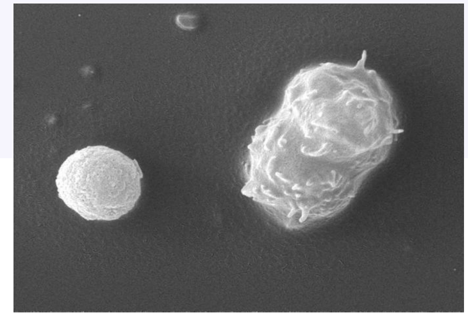
$$(X_k^N)_{k\in\mathbb{N}_0} \qquad\qquad (\Pi_k^N)_{k\in\mathbb{N}_0}$$

$$(X_t)_{t\ge 0} \xleftrightarrow[\text{DUALITY}]{\text{MOMENT}} (\Pi_t)_{t\ge 0}$$

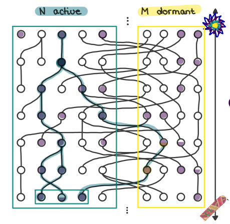
$$dX_t = \sqrt{X_t(1-X_t)}\, dB_t$$



DORMANCY



$(X_k^N, Y_k^N)_{k\in\mathbb{N}}$



$(\Pi_k^N)_{k\in\mathbb{N}}$

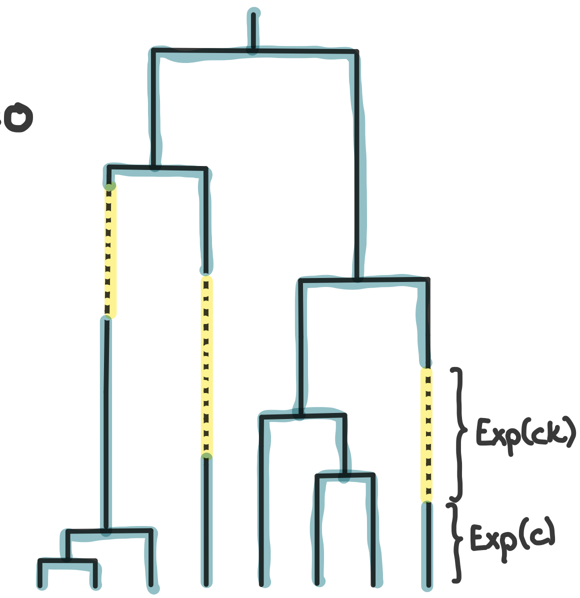
Frequency of ●

Genealogy of a sample

$(X_t, Y_t)_{t\geq 0}$ MOMENT DUALITY $(\Pi_t)_{t\geq 0}$

$$dX_t = c(Y_t - X_t)\,dt + \sqrt{X_t(1-X_t)}\,dB_t$$

$$dY_t = cK(X_t - Y_t)\,dt$$

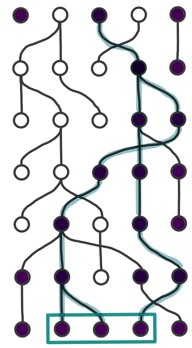


LEMMA: MOMENT DUALITY

$\forall t\geq 0 \ \forall x,y\in[0,1] \ \forall n,m\in\mathbb{N}_0$

$$\mathbb{E}_{x,y}[X_t^n Y_t^m] = \mathbb{E}^{n,m}[x^{N_t} y^{M_t}]$$

Rare Selection

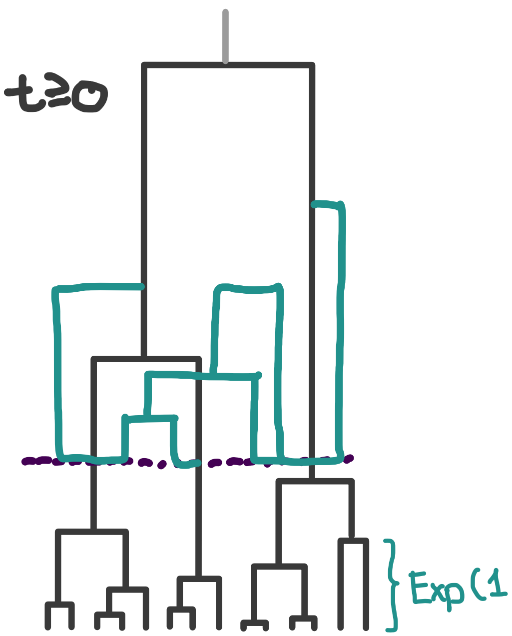


$$(X^N_k)_{k\in\mathbb{N}_0} \qquad \qquad (\Pi^N_k)_{k\in\mathbb{N}_0}$$

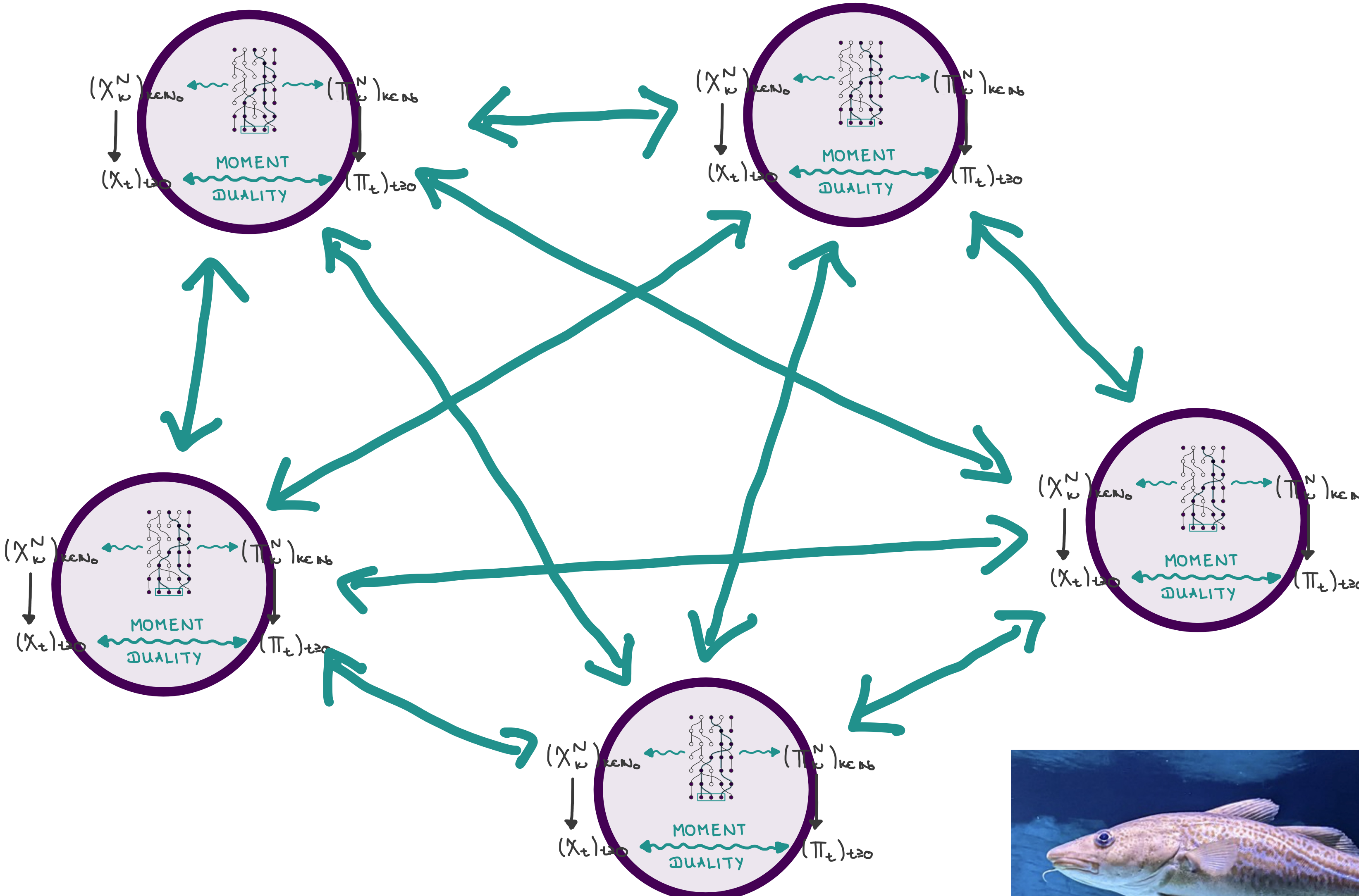
$$\Big\downarrow \qquad \qquad \qquad \Big\downarrow$$

$$(X_t)_{t\geq 0} \quad \xleftrightarrow[\text{DUALITY}]{\text{MOMENT}} \quad (\Pi_t)_{t\geq 0}$$

$$dX_t = \sqrt{X_t(1-X_t)}\,dB_t + \underbrace{\int \left(\mathbb{E}\left[X_{t-}^{K,y} \mid X_{t-}\right] - X_{t-}\right) \mathcal{N}_s(dt,dy)}_{\text{jumps!}}$$



Islands + Bottlenecks



TAKE HOME: POP GEN

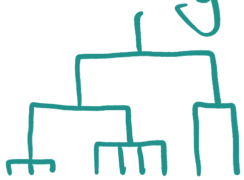
$dX_t = \sqrt{X_t(1-X_t)}\, dW_t$

non-Lipschitz coefficients

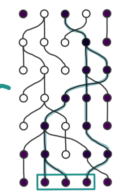
Feller: 1-d diffusion theory

classification of boundaries

Coalescent-theory:

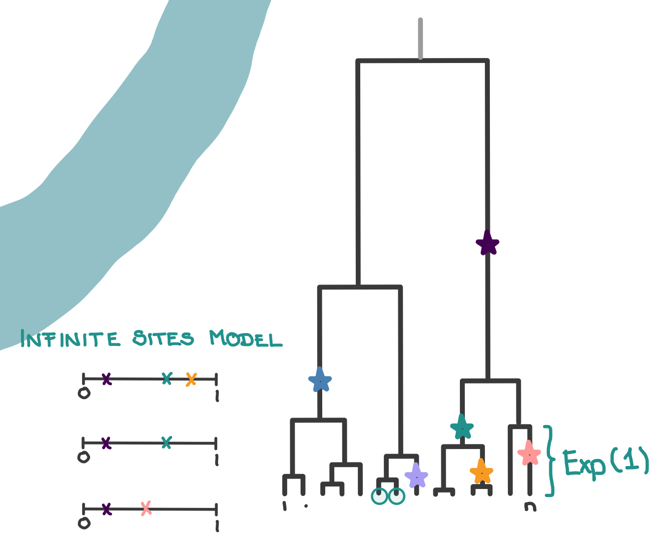


BIO

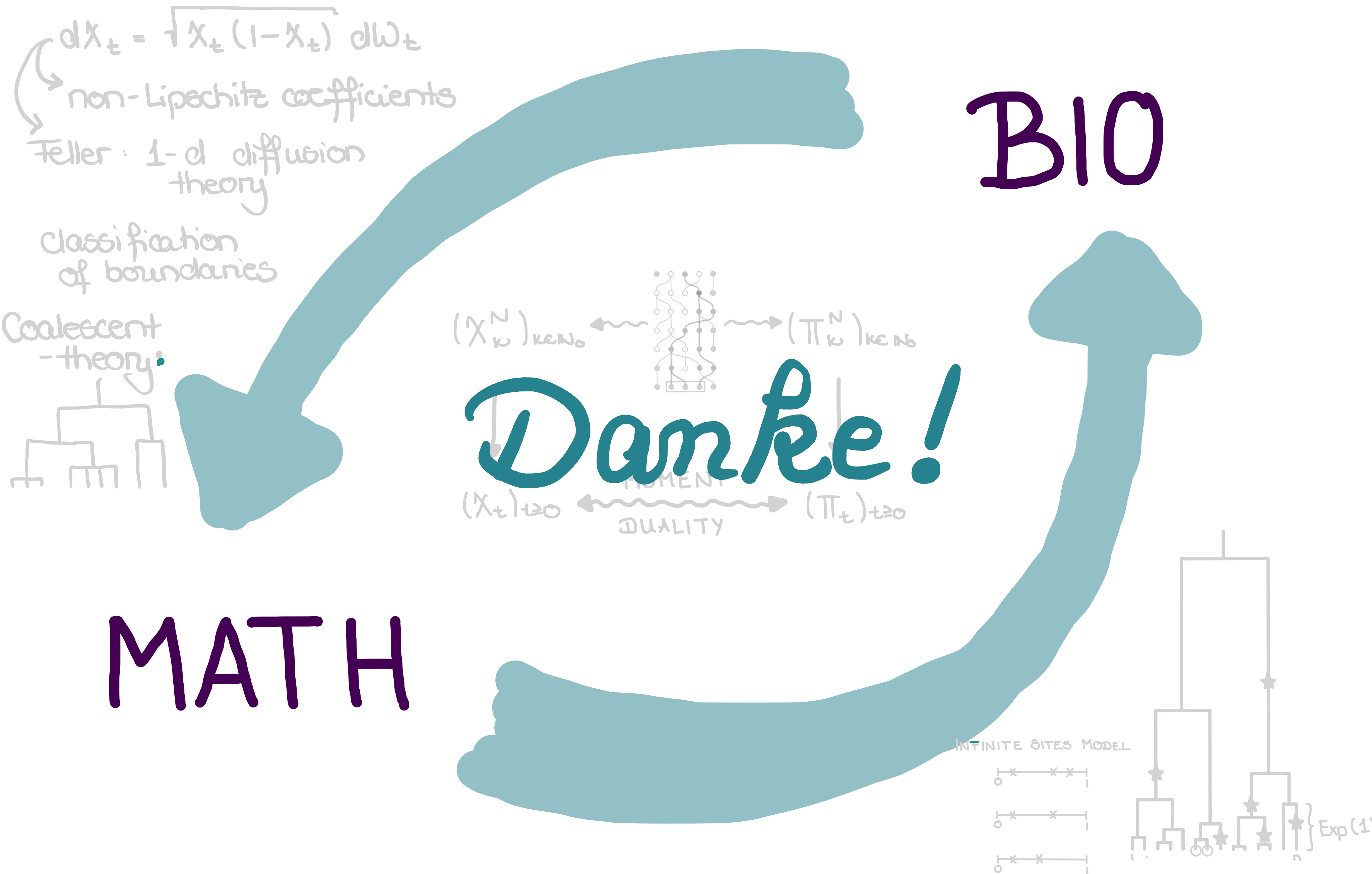
$(X^N_k)_{k\in\mathbb{N}_0}$  $(\Pi^N_k)_{k\in\mathbb{N}_0}$

$(X_t)_{t\geq 0}$ $\xleftrightarrow[\text{DUALITY}]{\text{MOMENT}}$ $(\Pi_t)_{t\geq 0}$

MATH



TAKE HOME : POP GEN

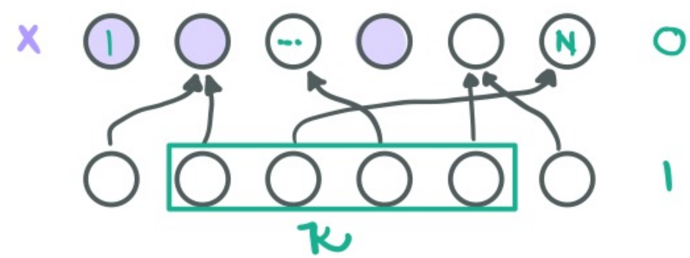


Momentendualität

<u>Momentendualität</u>: $\underline{z}\ \mathbb{E}_x[X_g^k] = \mathbb{E}^k[x^{K_g}]$ „$\forall g \in \mathbb{N}_0,\ \forall x \in \{0, \frac{1}{N}, \dots, 1\},\ \forall k \in \{1, \dots, N\}$"

Vorüberlegung:

Was ist die W'keit $p(x,k)$, dass alle Individuen einer Probe lila sind, wenn die Probe Größe k hat und der Anteil der Lilanen an der vorangegangenen Generation x ist?

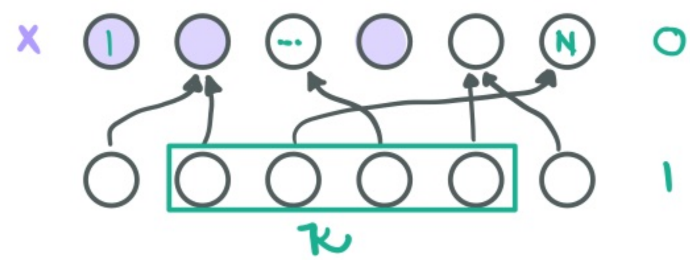


$p(x,k) =$

Momentendualität

Momentendualität: Z $\mathbb{E}_x[\cancel{X}_g^k] = \mathbb{E}^k[x^{K_g}]$ „$\forall g \in \mathbb{N}_0$, $\forall x \in \{0, \frac{1}{N}, \dots, 1\}$, $\forall k \in \{1, \dots, N\}$"

Vorüberlegung:



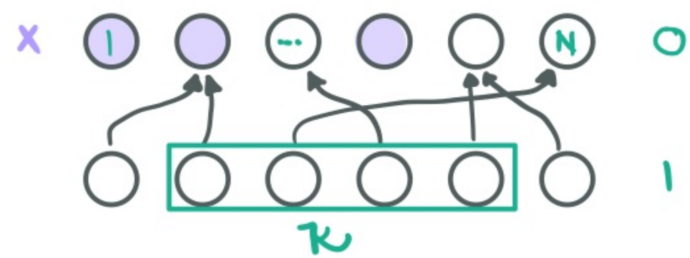
Was ist die W'keit $p(x, k)$, dass alle Individuen einer Probe lila sind, wenn die Probe Größe k hat und der Anteil der lilanen an der vorangegangenen Generation x ist?

$$p(x, k) = x^k$$

MOMENTENDUALITÄT

MOMENTENDUALITÄT: $Z\quad \mathbb{E}_x[X_g^k] = \mathbb{E}^k[x^{K_g}]$ „$\forall g \in \mathbb{N}_0,\ \forall x \in \{0, \frac{1}{N}, \dots, 1\},\ \forall k \in \{1, \dots, N\}$"

Vorüberlegung:



Was ist die W'keit $p(x,k)$, dass alle Individuen einer Probe lila sind, wenn die Probe Größe k hat und der Anteil der Lilanen an der vorangegangenen Generation x ist?

$$p(x,k) = x^k$$

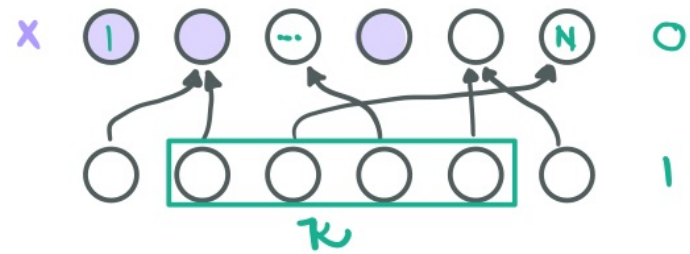
jedes Individuum hat W'keit x und sie sind unabhängig!

Momentendualität

<u>Momentendualität</u>: Ƶ $\mathbb{E}_x[X_g^k] = \mathbb{E}^k[x^{K_g}]$ „$\forall g \in \mathbb{N}_0, \forall x \in \{0, \frac{1}{N}, \dots, 1\}, \forall k \in \{1, \dots, N\}$"

Vorüberlegung:

Was ist die W'keit $p(x,k)$, dass alle Individuen einer Probe lila sind, wenn die Probe Größe k hat und der Anteil der Lilanen an der vorangegangenen Generation x ist?



$$p(x,k) = x^k$$

jedes Individuum hat W'keit x und sie sind unabhängig!

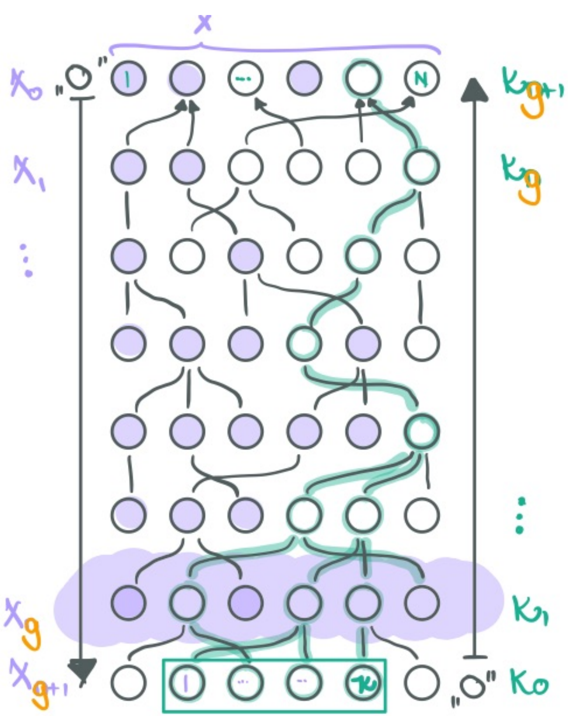
$$p(x,k) = \mathbb{P}(\text{„alle Individuen der Probe sind lila"} \mid X_0 = x, K_0 = k)$$

Momentendualität

Momentendualität: $\mathbb{Z}$ $\mathbb{E}_x[X_g^k] = \mathbb{E}^k[x^{K_g}]$ „$\forall g \in \mathbb{N}_0$, $\forall x \in \{0, \frac{1}{N}, \dots, 1\}$, $\forall k \in \{1, \dots, N\}$"

$p(x,k) = \mathbb{P}($„alle Individuen der Probe sind lila" $| X_0 = x, K_0 = k) = x^k$

Beweis: Gesucht ist die W'keit $p_g(x,k)$, dass alle Individuen einer Stichprobe lila sing, gegeben, dass die Stichprobe Größe k hat und der Anteil an lila $g+1$ Generationen zuvor gleich x ist.
Starte $(X_n)_{n\in\mathbb{N}_0}$ und $(K_n)_{n\in\mathbb{N}_0}$ wie im Bild unabhängig voneinander.

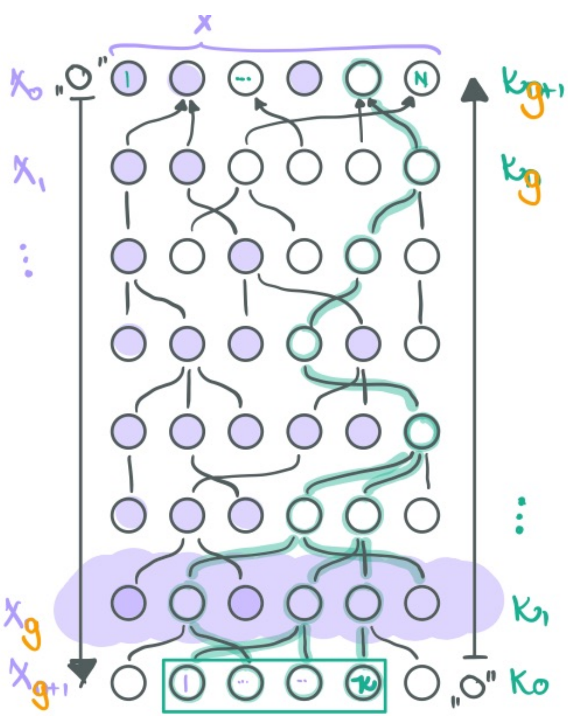


Momentendualität

Momentendualität: Z $\mathbb{E}_x[X_g^{k}] = \mathbb{E}^{k}[x^{K_g}]$ „$\forall g \in \mathbb{N}_0$, $\forall x \in \{0, \frac{1}{N}, \dots, 1\}$, $\forall k \in \{1, \dots, N\}$"

$p(x,k) = \mathbb{P}($„alle Individuen der Probe sind lila" $| X_0 = x, K_0 = k) = x^{k}$

Beweis: Gesucht ist die W'keit $p_g(x,k)$, dass alle Individuen einer Stichprobe lila sing, gegeben, dass die Stichprobe Größe k hat und der Anteil an lila $g+1$ Generationen zuvor gleich x ist.
Starte $(X_n)_{n\in\mathbb{N}_0}$ und $(K_n)_{n\in\mathbb{N}_0}$ wie im Bild unabhängig voneinander.



Antwort A: Benutze $(X_n)_{n\in\mathbb{N}_0}$.

Idee: Alle in Probe zur Zeit $g+1$ sind lila, wenn sie alle lila Eltern aus der Generation g gewählt haben.
Anteil an lila in Generation g: X_g.

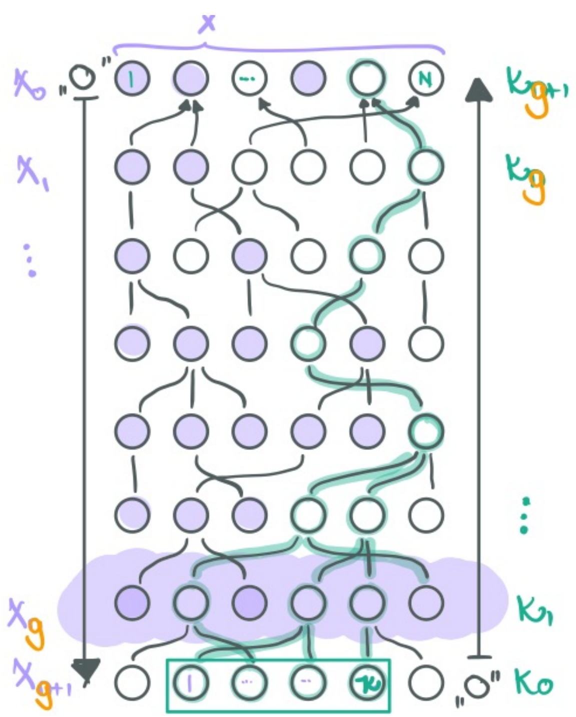
$$p_g(x,k) =$$

Momentendualität

Momentendualität: z.z. $\mathbb{E}_x[X_g^{\kappa}] = \mathbb{E}^{\kappa}[x^{K_g}]$ „$\forall g \in \mathbb{N}_0, \forall x \in \{0, \frac{1}{N}, \dots, 1\}, \forall \kappa \in \{1, \dots, N\}$"

$p(x,\kappa) = \mathbb{P}(\text{„alle Individuen der Probe sind lila"} \mid X_0 = x, K_0 = \kappa) = x^{\kappa}$

Beweis: Gesucht ist die W'keit $p_g(x,\kappa)$, dass alle Individuen einer Stichprobe lila sing, gegeben, dass die Stichprobe Größe κ hat und der Anteil an lila g+1 Generationen zuvor gleich x ist.
Starte $(X_n)_{n\in\mathbb{N}_0}$ und $(K_n)_{n\in\mathbb{N}_0}$ wie im Bild unabhängig voneinander.



Antwort A: Benutze $(X_n)_{n\in\mathbb{N}_0}$.

Idee: Alle in Probe zur Zeit g+1 sind lila, wenn sie alle lila Eltern aus der Generation g gewählt haben.
Anteil an lila in Generation g: X_g.

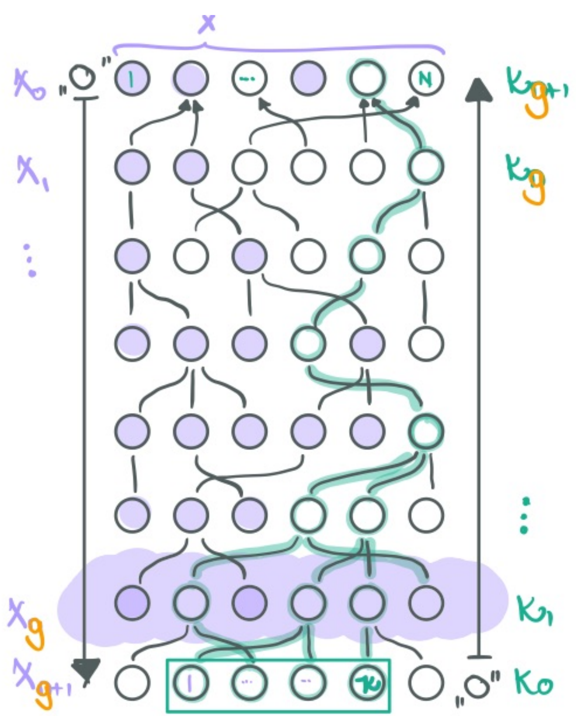
$$p_g(x,\kappa) = X_g^{\kappa}$$

Momentendualität

Momentendualität: Z $\mathbb{E}_x[X_g^k] = \mathbb{E}^k[x^{K_g}]$ „$\forall g \in \mathbb{N}_0$, $\forall x \in \{0, \frac{1}{N}, \ldots, 1\}$, $\forall k \in \{1, \ldots, N\}$"

$p(x,k) = \mathbb{P}($„alle Individuen der Probe sind lila" $| X_0 = x, K_0 = k) = x^k$

Beweis: Gesucht ist die W'keit $p_g(x,k)$, dass alle Individuen einer Stichprobe lila sing, gegeben, dass die Stichprobe Größe k hat und der Anteil an lila $g+1$ Generationen zuvor gleich x ist.
Starte $(X_n)_{n\in\mathbb{N}_0}$ und $(K_n)_{n\in\mathbb{N}_0}$ wie im Bild unabhängig voneinander.



Antwort A: Benutze $(X_n)_{n\in\mathbb{N}_0}$.

Idee: Alle in Probe zur Zeit $g+1$ sind lila, wenn sie alle lila Eltern aus der Generation g gewählt haben.
Anteil an lila in Generation g: X_g.

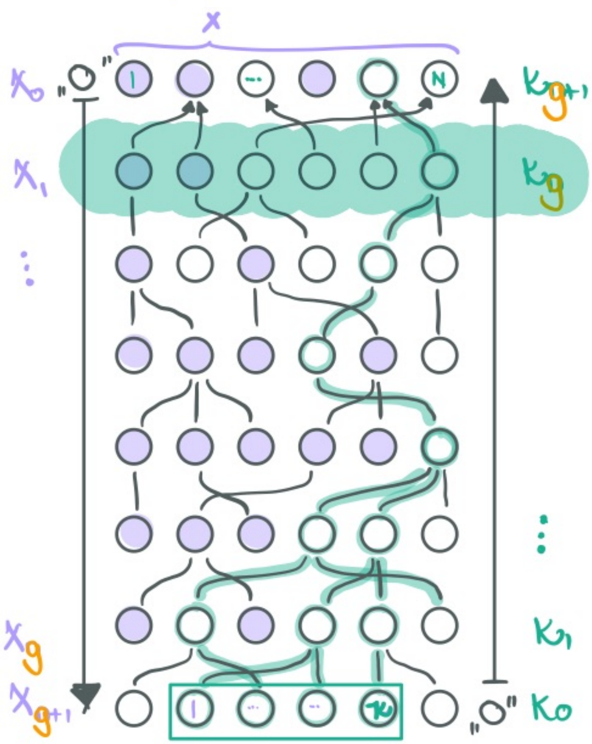
$$p_g(x,k) = \mathbb{E}[X_g^k \mid X_0 = x]$$

Momentendualität

Momentendualität: z.z. $\mathbb{E}_x[X_g^{k}] = \mathbb{E}^{k}[x^{K_g}]$ „$\forall g \in \mathbb{N}_0,\ \forall x \in \{0, \frac{1}{N}, \dots, 1\},\ \forall k \in \{1, \dots, N\}$"

$p(x,k) = \mathbb{P}($„alle Individuen der Probe sind lila" $\mid X_0 = x,\ K_0 = k) = x^{k}$

Beweis: Gesucht ist die W'keit $p_g(x,k)$, dass alle Individuen einer Stichprobe lila sing, gegeben, dass die Stichprobe Größe k hat und der Anteil an lila $g+1$ Generationen zuvor gleich x ist.
Starte $(X_n)_{n\in\mathbb{N}_0}$ und $(K_n)_{n\in\mathbb{N}_0}$ wie im Bild unabhängig voneinander.



Antwort B: Benutze $(K_n)_{n\in\mathbb{N}_0}$.

Idee: Alle in Probe zur Zeit $g+1$ sind lila, wenn sie alle lila Vorfahren in der Generation 1 gewählt haben.
Anzahl an Vorfahren vor g Generationen: K_g.

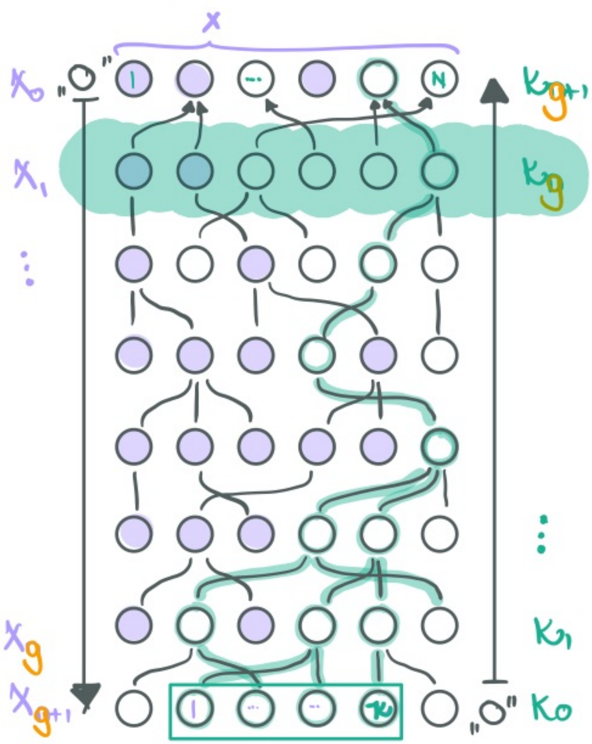
$$p_g(x,k) =$$

Momentendualität

Momentendualität: z $\mathbb{E}_x[X_g^k] = \mathbb{E}^k[x^{K_g}]$ „$\forall g \in \mathbb{N}_0$, $\forall x \in \{0, \frac{1}{N}, \dots, 1\}$, $\forall k \in \{1, \dots, N\}$"

$p(x,k) = \mathbb{P}(\text{„alle Individuen der Probe sind lila"} \mid X_0 = x, K_0 = k) = x^k$

Beweis: Gesucht ist die W'keit $p_g(x,k)$, dass alle Individuen einer Stichprobe lila sing, gegeben, dass die Stichprobe Größe k hat und der Anteil an lila $g+1$ Generationen zuvor gleich x ist.
Starte $(X_n)_{n \in \mathbb{N}_0}$ und $(K_n)_{n \in \mathbb{N}_0}$ wie im Bild unabhängig voneinander.



Antwort B: Benutze $(K_n)_{n \in \mathbb{N}_0}$.

Idee: Alle in Probe zur Zeit $g+1$ sind lila, wenn sie alle lila Vorfahren in der Generation 1 gewählt haben.

Anzahl an Vorfahren vor g Generationen: K_g.

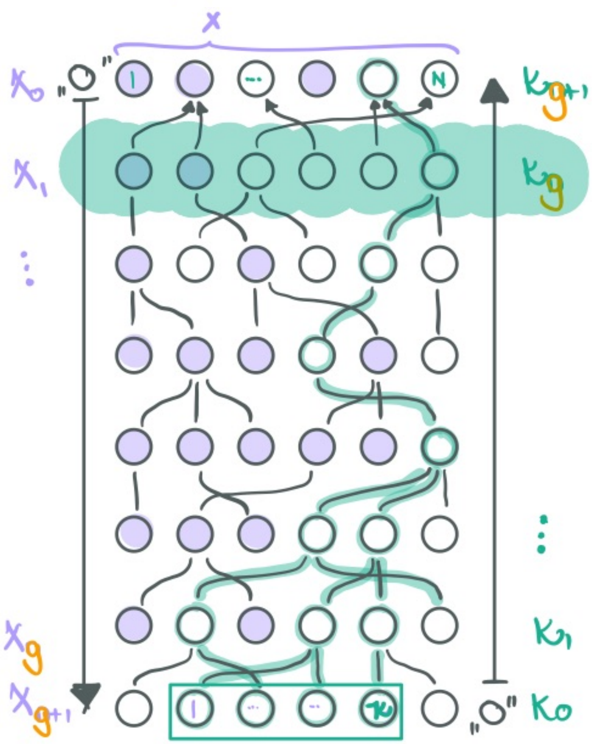
$$p_g(x,k) = x^{K_g}$$

Momentendualität

Momentendualität: z.z. $\mathbb{E}_x[X_g^k] = \mathbb{E}^k[x^{K_g}]$ „$\forall g \in \mathbb{N}_0$, $\forall x \in \{0, \frac{1}{N}, \dots, 1\}$, $\forall k \in \{1, \dots, N\}$"

$p(x,k) = \mathbb{P}(\text{„alle Individuen der Probe sind lila"} \mid X_0 = x, K_0 = k) = x^k$

Beweis: Gesucht ist die W'keit $p_g(x,k)$, dass alle Individuen einer Stichprobe lila sing, gegeben, dass die Stichprobe Größe k hat und der Anteil an lila $g+1$ Generationen zuvor gleich x ist.
Starte $(X_n)_{n \in \mathbb{N}_0}$ und $(K_n)_{n \in \mathbb{N}_0}$ wie im Bild unabhängig voneinander.



Antwort B: Benutze $(K_n)_{n \in \mathbb{N}_0}$.

Idee: Alle in Probe zur Zeit $g+1$ sind lila, wenn sie alle lila Vorfahren in der Generation 1 gewählt haben.

Anzahl an Vorfahren vor g Generationen: K_g.

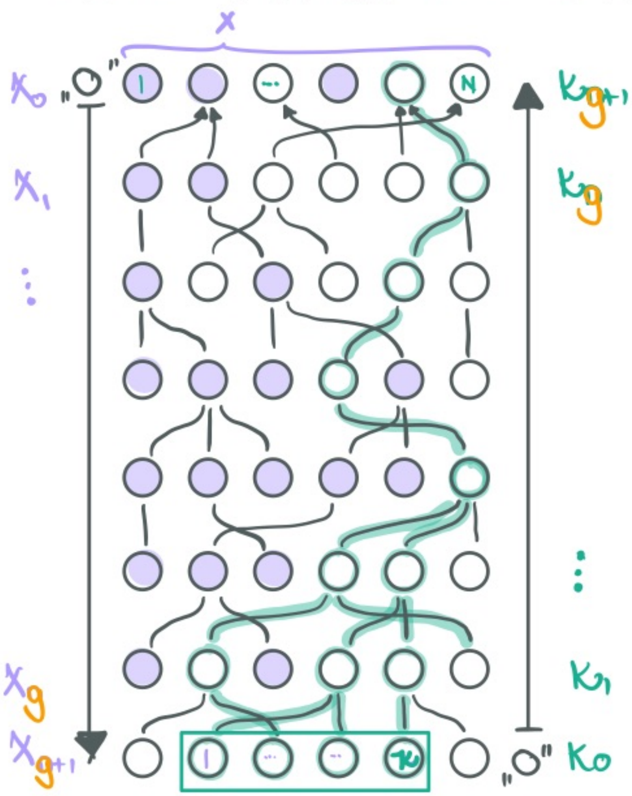
$$p_g(x,k) = \mathbb{E}[x^{K_g} \mid K_0 = k]$$

Momentendualität

<u>Momentendualität:</u> z.z. $\mathbb{E}_x[X_g^k] = \mathbb{E}^k[x^{K_g}]$ „$\forall g \in \mathbb{N}_0$, $\forall x \in \{0, \frac{1}{N}, \dots, 1\}$, $\forall k \in \{1, \dots, N\}$"

$p(x,k) = \mathbb{P}$(„alle Individuen der Probe sind lila" | $X_0 = x$, $K_0 = k$) $= x^k$

<u>Beweis</u>: Gesucht ist die W'keit $p_g(x,k)$, dass alle Individuen einer Stichprobe lila sing, gegeben, dass die Stichprobe Größe k hat und der Anteil an lila $g+1$ Generationen zuvor gleich x ist.
Starte $(X_n)_{n\in\mathbb{N}_0}$ und $(K_n)_{n\in\mathbb{N}_0}$ wie im Bild unabhängig voneinander.



<u>A)</u> Antwort mit $(X_n)_{n\in\mathbb{N}_0}$

$$p_g(x,k) = \mathbb{E}_x[X_g^k]$$

<u>B)</u> Antwort mit $(K_n)_{n\in\mathbb{N}_0}$

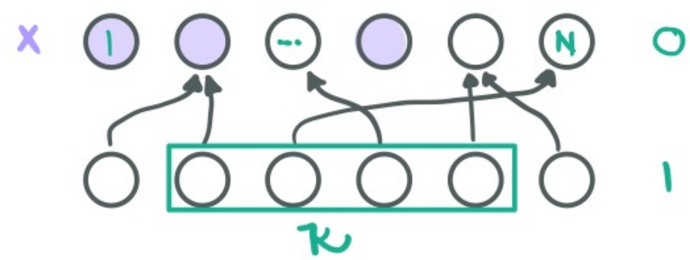
$$p_g(x,k) = \mathbb{E}^k[x^{K_g}]$$

Da beides die gleiche Größe bestimmt, gilt die Gleichheit!

(Technik: sampling Dualität)

MOMENTENDUALITÄT: Z $\mathbb{E}_x[Y_g^k] = \mathbb{E}^k[x^{K_g}]$ „$\forall g \in \mathbb{N}_0$, $\forall x \in \{0, \frac{1}{N}, \dots, 1\}$, $\forall k \in \{1, \dots, N\}$"

Vorüberlegung:



Was ist die W'keit $p(x, k)$, dass alle Individuen einer Probe lila sind, wenn die Probe Größe k hat und der Anteil der Lilanen an der vorangegangenen Generation x ist?

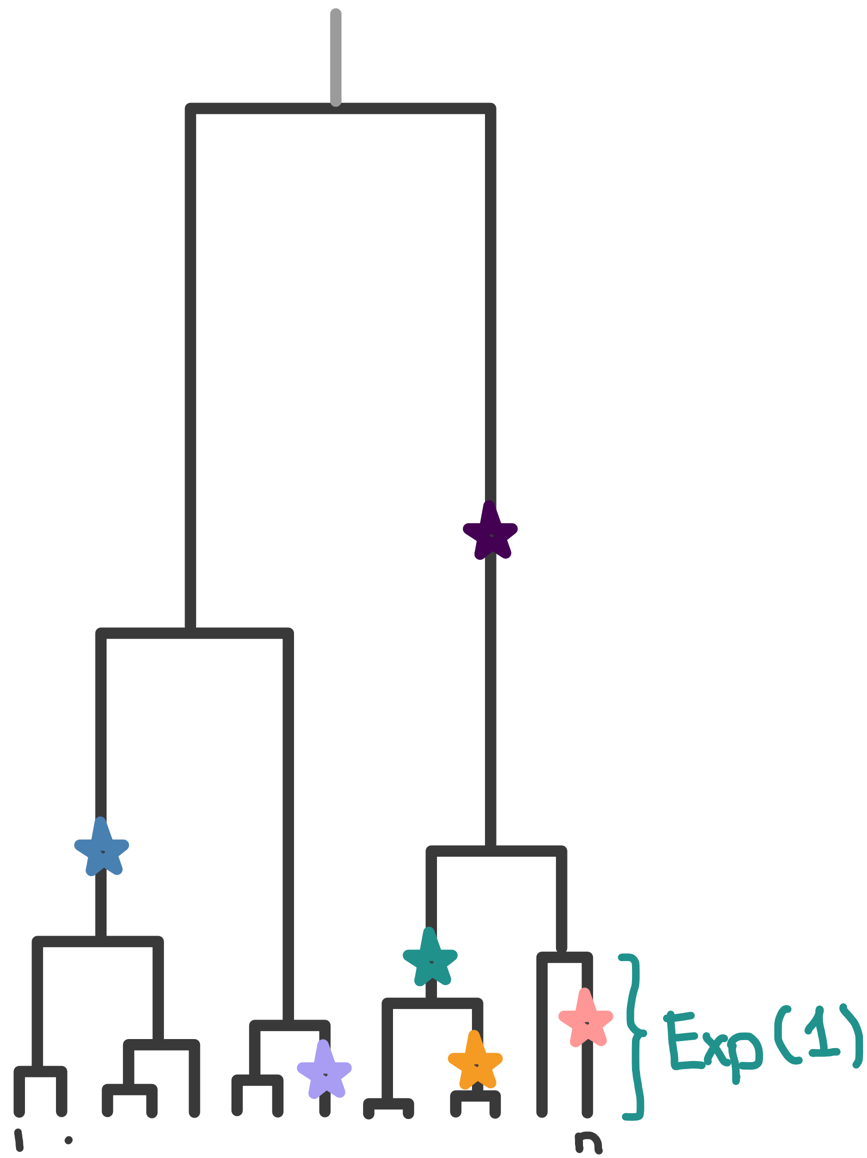
$$p(x, k) = x^k$$

jedes Individuum hat W'keit x und sie sind unabhängig!

$p(x, k) = \mathbb{P}($„alle Individuen der Probe sind lila" $| \; X_0 = x, \; K_0 = k)$

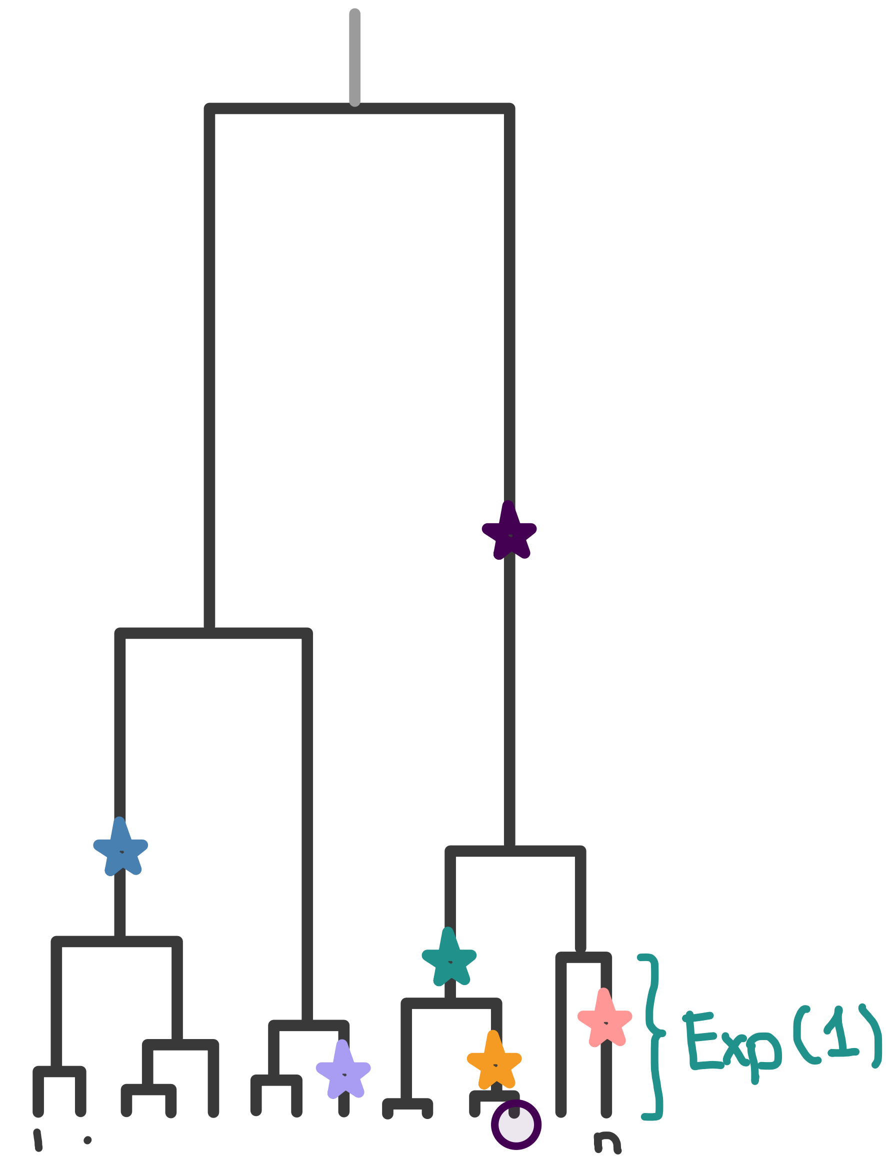
THE KINGMAN COALESCENT + MUTATION

Genealogy of a sample $\quad (\Pi_t)_{t\geq 0}$



The Kingman Coalescent + Mutation

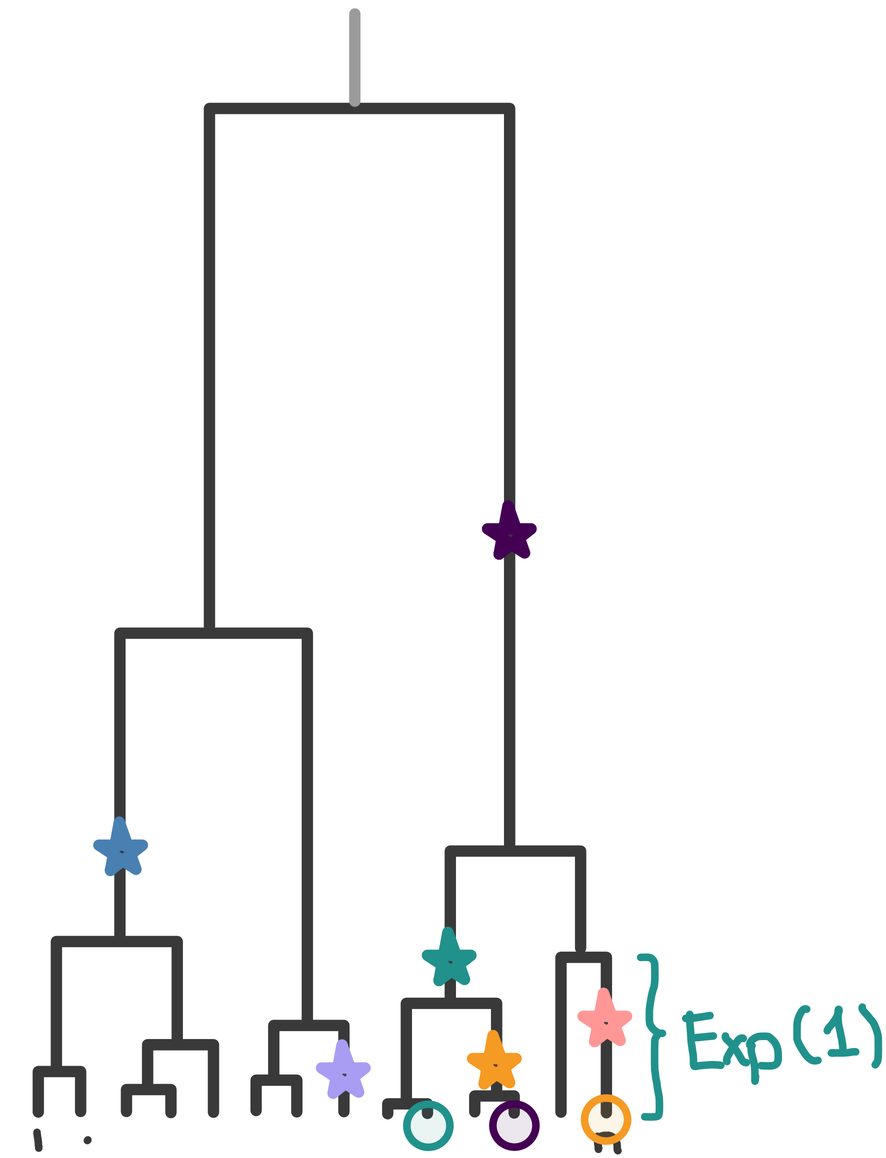
Genealogy of a sample $(\Pi_t)_{t\geq 0}$



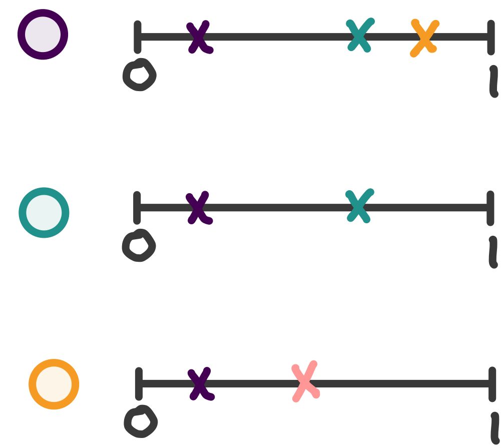
Infinite Sites Model

The Kingman Coalescent + Mutation

Genealogy of a sample $(\Pi_t)_{t\geq 0}$



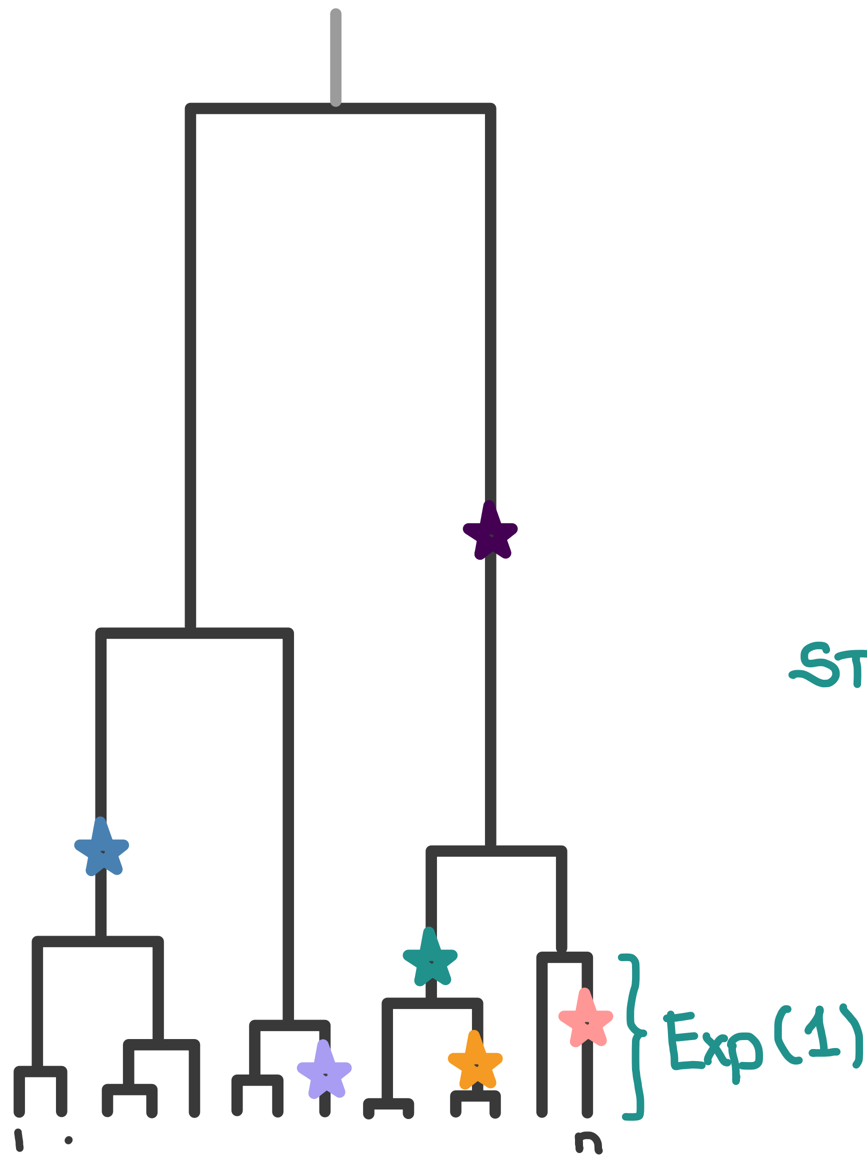
Infinite Sites Model



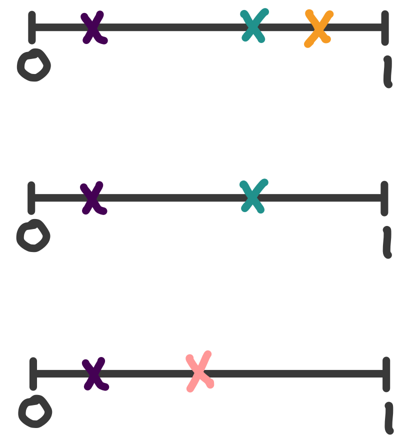
The Kingman Coalescent + Mutation

Genealogy of a sample

$(\Pi_t)_{t \geq 0}$



Infinite Sites Model

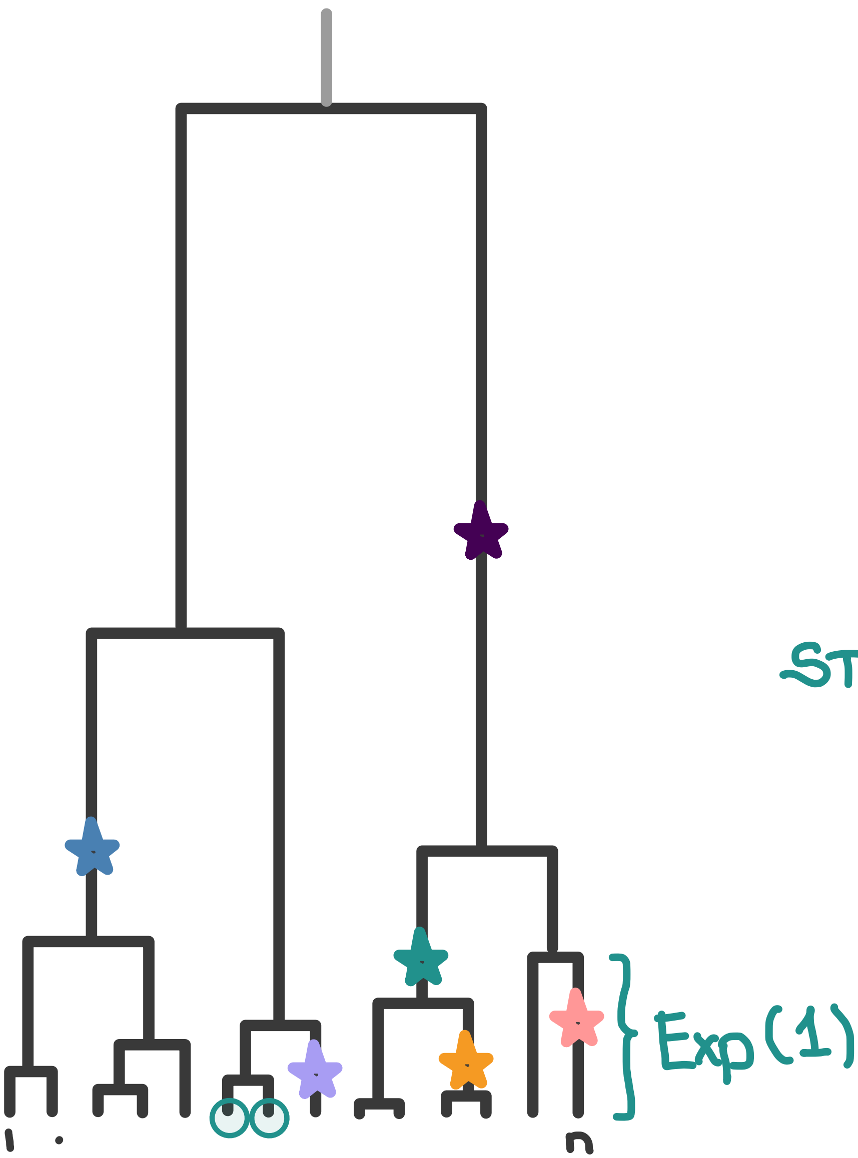


Statistic: Site Frequency Spectrum

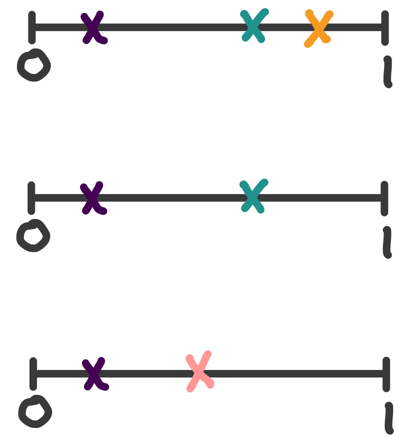
$(2, 7, 3, 2, 0, 0, 0, 0)$

The Kingman Coalescent + Mutation

Genealogy of a sample $\quad (\Pi_t)_{t\geq 0}$



Infinite Sites Model



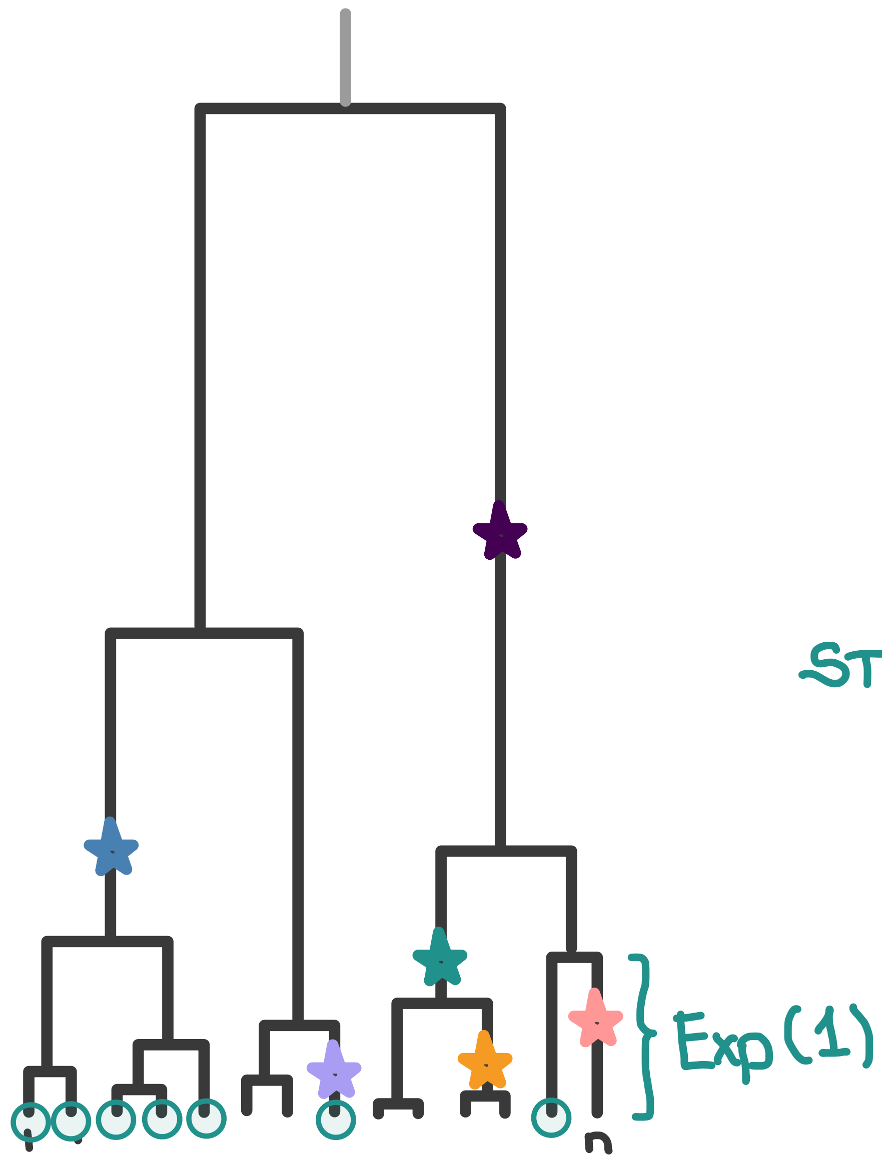
Statistic: Site Frequency Spectrum

(2, 7, 3, 2, 0, 0, 0, 0)

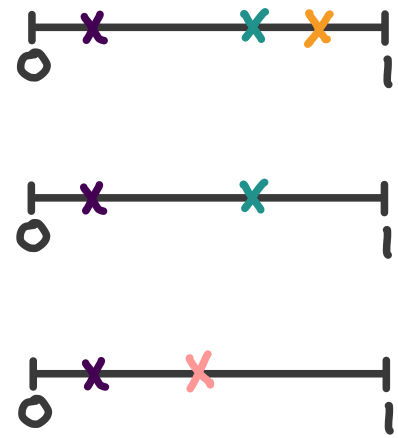
How many individuals have 0 mutations

THE KINGMAN COALESCENT + MUTATION

Genealogy of a sample $(\Pi_t)_{t\geq 0}$



INFINITE SITES MODEL



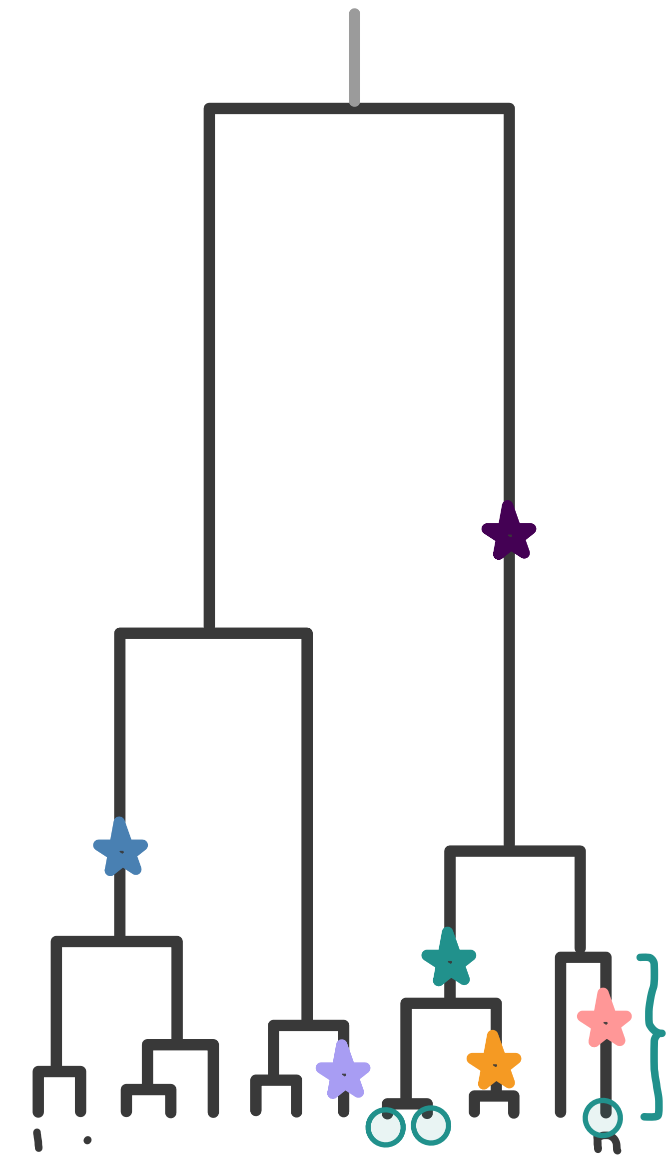
STATISTIC: SITE FREQUENCY SPECTRUM

$$(2, 7, 3, 2, 0, 0, 0, 0)$$

How many individuals have 1 mutations

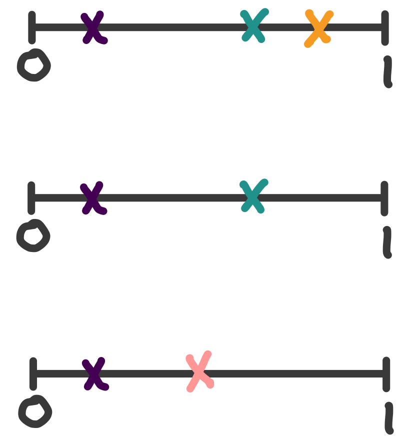
THE KINGMAN COALESCENT + MUTATION

Genealogy of a sample $\qquad (\Pi_t)_{t \geq 0}$



Exp(1)

INFINITE SITES MODEL



STATISTIC: SITE FREQUENCY SPECTRUM

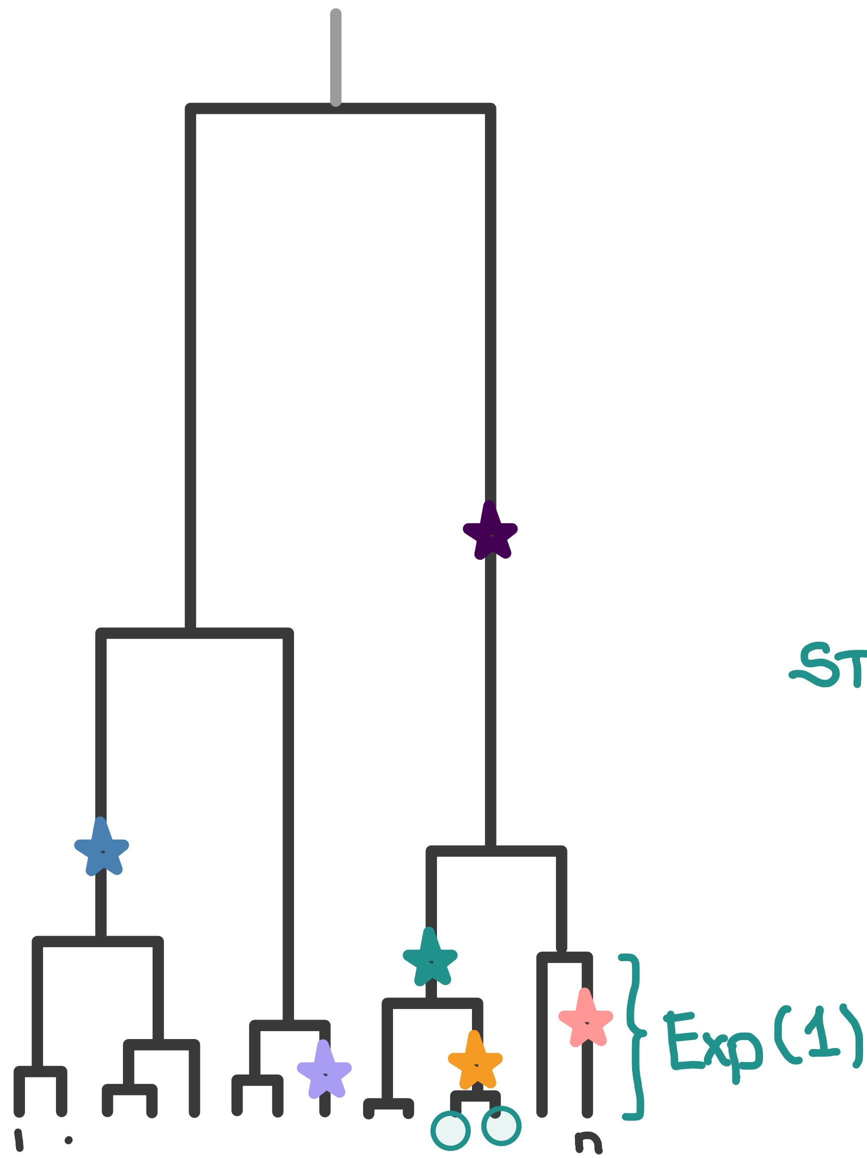
$$(2, 7, 3, 2, 0, 0, 0, 0)$$

How many individuals have 2 mutations

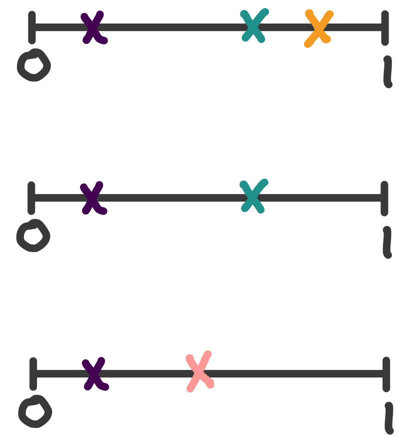
The Kingman Coalescent + Mutation

Genealogy of a sample

$(\Pi_t)_{t\geq 0}$



Infinite Sites Model



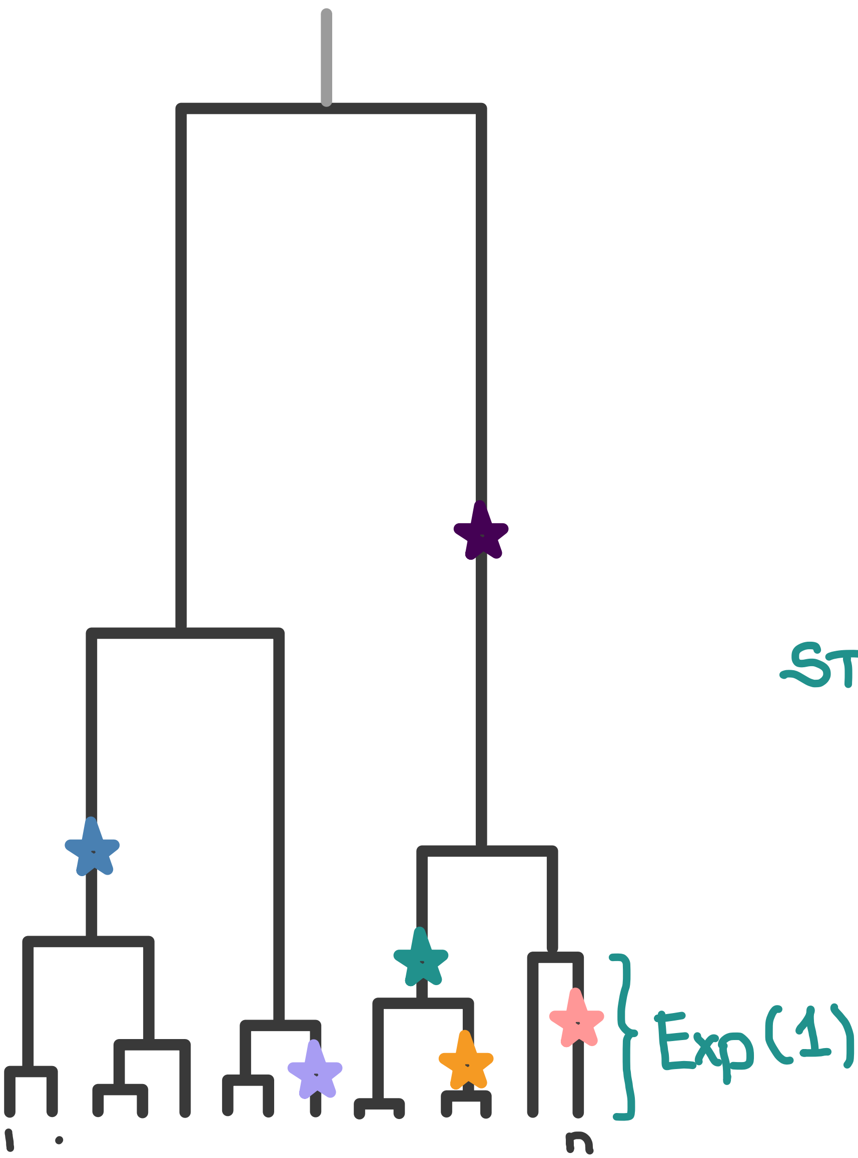
Statistic: Site Frequency Spectrum

$(2, 7, 3, 2, 0, 0, 0, 0)$

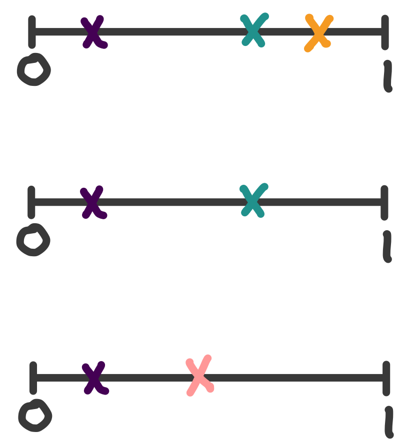
How many individuals have 3 mutations

The Kingman Coalescent + Mutation

Genealogy of a sample $(\Pi_t)_{t\geq 0}$



Infinite Sites Model



Statistic: Site Frequency Spectrum

$(2, 7, 3, 2, 0, 0, 0, 0)$

* SCALE RESULTS BACK TO FINITE POPULATIONS *