MTH 428/528

Lectures 21-26

Yevgeniy Kovchegov
Oregon State University
Topics:

- Continuous random variables
- The Coalescent
- Branching process
- Yule process
Stationary distribution for Moran process. Recall:

\[ \pi_k = \binom{n}{k} \frac{\mathcal{B}(n \cdot \frac{\beta}{1-\alpha-\beta} + k, n \cdot \frac{\alpha}{1-\alpha-\beta} + (n-k))}{\mathcal{B}(n \cdot \frac{\beta}{1-\alpha-\beta}, n \cdot \frac{\alpha}{1-\alpha-\beta} + n)} \pi_0 \]

and

\[ 1 = \frac{\mathcal{B}(n \cdot \frac{\beta}{1-\alpha-\beta}, n \cdot \frac{\alpha}{1-\alpha-\beta})}{\mathcal{B}(n \cdot \frac{\beta}{1-\alpha-\beta}, n \cdot \frac{\alpha}{1-\alpha-\beta} + n)} \pi_0. \]

Hence, we have exact solution:

\[ \pi_k = \binom{n}{k} \frac{\mathcal{B}(n \cdot \frac{\beta}{1-\alpha-\beta} + k, n \cdot \frac{\alpha}{1-\alpha-\beta} + (n-k))}{\mathcal{B}(n \cdot \frac{\beta}{1-\alpha-\beta}, n \cdot \frac{\alpha}{1-\alpha-\beta})}. \]
**Stationary distribution for Moran process.** Recall:

\[ B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)} = \int_0^1 x^{a-1}(1 - x)^{b-1} dx \]

and

\[ \pi_k = \binom{n}{k} \frac{B(n \cdot \frac{\beta}{1 - \alpha - \beta} + k, n \cdot \frac{\alpha}{1 - \alpha - \beta} + (n - k))}{B(n \cdot \frac{\beta}{1 - \alpha - \beta}, n \cdot \frac{\alpha}{1 - \alpha - \beta})}. \]

We use Stirling’s approximation

\[ \Gamma(z) = \sqrt{2\pi}e^{-z}z^{z-\frac{1}{2}} \left( 1 + O \left( \frac{1}{z} \right) \right) \quad \text{and} \quad m! = \sqrt{2\pi}e^{-m}m^{m+\frac{1}{2}} \left( 1 + O \left( \frac{1}{m} \right) \right) \]

and obtain for \( \rho := \frac{k}{n} \in [\epsilon, 1 - \epsilon], \quad \alpha = \frac{a}{n} \text{ and } \beta = \frac{b}{n}, \)

\[ \pi_k = \frac{1}{n} \cdot \frac{\rho^{a-1}(1 - \rho)^{b-1}}{B(a, b)} (1 + o(1)). \]
**Stationary distribution for Moran process.**
For \( \rho \equiv \frac{k}{n} \in [\epsilon, 1 - \epsilon] \), \( \alpha = \frac{a}{n} \) and \( \beta = \frac{b}{n} \),

\[
\pi_k = \frac{1}{n} \cdot \frac{\rho^{a-1}(1 - \rho)^{b-1}}{B(a, b)} (1 + o(1)).
\]
Continuous random variables.

**Definition.** We say that $X$ is a **continuous random variable** if there exists a nonnegative function $f(x)$ defined for all real $x$ such that for any $a \leq b$,

$$P(a \leq X \leq b) = \int_{a}^{b} f(x) \, dx$$

Such function $f(x)$ is the **probability density function** of $X$. 
Continuous random variables.

Definition. We say that $X$ is a **continuous random variable** if there exists a nonnegative function $f(x)$ defined for all real $x$ such that for any $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Properties:

- $\int_{-\infty}^{\infty} f(x)dx = P(-\infty < X < \infty) = 1$, where $\int_{-\infty}^{\infty} f(x)dx = \lim_{a \to \infty} \int_{-a}^{a} f(x)dx$

- $P(X = a) = \int_{-a}^{a} f(x)dx = 0$ for any real $a$.

- Hence

$$P(a < X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$$
Exponential random variables.

- Given $\lambda > 0$. Let $X$ be a continuous random variable with density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Then $X$ is said to be an exponential random variable with parameter $\lambda$. 
**Exponential random variables.**

- Given $\lambda > 0$. Let $X$ be a continuous random variable with density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Then $X$ is said to be an exponential random variable with parameter $\lambda$.

- Check:

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{0} 0 \cdot dx + \int_{0}^{\infty} \lambda e^{-\lambda x} dx = 0 + \left[ -e^{-\lambda x} \right]_{0}^{\infty} = 1$$

Thus $f(x)$ is indeed a probability density function.

- Note: Exponential random variable is a continuous analogue to the geometric random variable. In particular it also satisfies the following memorylessness property:

$$P(X > a + b \mid X > a) = P(X > b)$$

for any two positive $a$ and $b$. 
Uniform random variables.

- Consider an interval $[\alpha, \beta]$, where $\alpha < \beta$. Let $X$ be a continuous random variable with density function

$$
 f(x) = \begin{cases} 
 \frac{1}{\beta - \alpha} & \text{if } \alpha \leq x \leq \beta \\
 0 & \text{otherwise}
\end{cases}
$$

Then $X$ is said to be an uniform random variable over $[\alpha, \beta]$.

- Check:

$$
\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{x} 0 \cdot dx + \int_{x}^{\beta} \frac{dx}{\beta - \alpha} + \int_{\beta}^{\infty} 0 \cdot dx = 0 + \left[ \frac{x}{\beta - \alpha} \right]_{x}^{\beta} + 0
$$

$$
= \frac{\beta}{\beta - \alpha} - \frac{\alpha}{\beta - \alpha} = 1
$$

Thus $f(x)$ is indeed a probability density function.
Continuous random variables.

**Definition.** Let $X$ be a continuous random variable with density function $f(x)$. Then its **expectation** is

$$E[X] = \int_{-\infty}^{\infty} xf(x) \, dx$$

**Properties:**

- For any real valued function $g$, $g(X)$ will also be a random variable, and

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x) \, dx$$

- **Markov inequality.** If $X$ is a random variable that takes only nonnegative values, then for any $\alpha > 0$,

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha}$$

- **Chebyshev inequality.** If $X$ is a random variable with finite mean $\mu$ and variance, then for any $\kappa > 0$,

$$P(|X - \mu| \geq \kappa) \leq \frac{Var(X)}{\kappa^2}$$
Expectation and variance.

Now, let $X$ be a random variable with mean $E[X] = \mu$.

- **Definition.** The variance of $X$ is
  \[ \text{Var}(X) = E[(X - \mu)^2] \]

- **Definition.** The standard deviation of $X$ is
  \[ \text{SD}(X) = \sqrt{\text{Var}(X)} = \sqrt{E[(X - \mu)^2]} \]

Finally,

- **Theorem.** The variance of $X$ is
  \[ \text{Var}(X) = E[X^2] - \mu^2 \]
Normal random variables.

$X$ is a **Normal (Gaussian) random variable** with parameters $\mu$ and $\sigma^2$ if its density function is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty$$
A normal random variable $X$ is said to be **standard normal** if its parameters $\mu = 0$ and $\sigma^2 = 1$, and therefore

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad -\infty < x < \infty$$
Normal random variables.

\( \mathcal{N}(\mu, \sigma^2) \) denotes normal distribution with parameters \( \mu \) and \( \sigma^2 \). Here I plotted two normal densities, one standard normal \( \mathcal{N}(0, 1) \) and one \( \mathcal{N}(5, 4) \).
Exponential random variables.

Example. Given $\lambda > 0$. Let $X$ be a continuous random variable with density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Then $X$ is said to be an exponential random variable with parameter $\lambda$.

We want to find its expectation.

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x) \, dx = \int_{0}^{\infty} \lambda x \cdot e^{-\lambda x} \, dx$$

• Recall the integration by parts formula,

$$\int uv' \, dx = uv - \int u'v \, dx$$

Here we let $u(x) = x$ and $v(x) = -e^{-\lambda x}$. Then $u'(x) = 1$ and $v'(x) = \lambda e^{-\lambda x}$, and

$$\int_{0}^{\infty} \lambda x \cdot e^{-\lambda x} \, dx = \int_{0}^{\infty} u(x)v'(x) \, dx$$
Exponential random variables.

Example (continued).

- Recall the integration by parts formula,
  \[ \int uv'\,dx = uv - \int u'v\,dx \]

Here we let \( u(x) = x \) and \( v(x) = -e^{-\lambda x} \). Then \( u'(x) = 1 \) and \( v'(x) = \lambda e^{-\lambda x} \), and

\[ \int_0^\infty \lambda x \cdot e^{-\lambda x}\,dx = \int_0^\infty u(x)v'(x)\,dx \]

\[ \begin{align*}
E[X] &= \int_0^\infty \lambda x \cdot e^{-\lambda x}\,dx \\
&= \left[ -xe^{-\lambda x} \right]_0^\infty - \int_0^\infty (-e^{-\lambda x})\,dx \\
&= \int_0^\infty e^{-\lambda x}\,dx \\
&= \frac{1}{\lambda} \cdot \int_0^\infty \lambda e^{-\lambda x}\,dx \\
&= \frac{1}{\lambda} \\
\end{align*} \]

as we have shown that \( \int_{-\infty}^{\infty} f(x)\,dx = \int_0^\infty \lambda e^{-\lambda x}\,dx = 1 \).
The Coalescent.

Consider the Wright-Fisher model with \( n = 2N \) copies of the genetic locus with two alleles \( A \) and \( a \) that have the same fitness.

Consider a pair of distinct individuals \( A_0 \) and \( B_0 \) at time \( t \) out of the state space \( S = \{1, 2, \ldots, n\} \).

For each individual we track back ancestral history \( m \) generations back:

\[
A_m \rightarrow A_{m-1} \rightarrow \ldots \rightarrow A_1 \rightarrow A_0 \quad \text{gen. } t - m
\]

\[
B_m \rightarrow B_{m-1} \rightarrow \ldots \rightarrow B_1 \rightarrow B_0 \quad \text{gen. } t - m + 1
\]

The probability that the two individuals originated from the same ancestor exactly \( m \) generations ago is

\[
P(A_m = B_m, A_{m-1} \neq B_{m-1}, \ldots, A_1 \neq B_1) = \left(1 - \frac{1}{n}\right)^{m-1} \cdot \frac{1}{n}
\]
The Coalescent.

\[ A_m \quad \rightarrow \quad A_{m-1} \quad \rightarrow \quad \ldots \quad \rightarrow \quad A_1 \quad \rightarrow \quad A_0 \]

\[ B_m \quad \rightarrow \quad B_{m-1} \quad \rightarrow \quad \ldots \quad \rightarrow \quad B_1 \quad \rightarrow \quad B_0 \]

gen. \( t - m \quad \rightarrow \quad \ldots \quad \rightarrow \quad \text{gen.} \quad t - 1 \quad \rightarrow \quad \text{gen.} \quad t \]

The probability that the two individuals originated from different ancestors looking back \( m \) generations is

\[
P(A_m \neq B_m, A_{m-1} \neq B_{m-1}, \ldots, A_1 \neq B_1) = \left(1 - \frac{1}{n}\right)^m
\]

Let \( X_t \) denote the number of alleles \( A \) in the Wright-Fisher model. Define the heterozygosity at time \( t \) as the frequency of heterozygous species (‘Aa’) in a diploid population of size \( N \):

\[
H_t := \frac{X_t(n - X_t)}{\binom{n}{2}}.
\]
The Coalescent.

The probability that the two individuals originated from different ancestors looking back \( m \) generations is

\[
P(A_m \neq B_m, A_{m-1} \neq B_{m-1}, \ldots, A_1 \neq B_1) = \left(1 - \frac{1}{n}\right)^m
\]

Let \( X_t \) denote the number of alleles \( A \) in the Wright-Fisher model, in generation \( t \). Define the heterozygosity at time \( t \) as the frequency of heterozygous species (‘\( Aa \)’) in a diploid population of size \( N \):

\[
H_t := \frac{X_t(n - X_t)}{\binom{n}{2}}.
\]

Then

\[
E[H_t] = \left(1 - \frac{1}{n}\right)^t E[H_0].
\]
The Coalescent. Recall:

\[ P(A_m \neq B_m, A_{m-1} \neq B_{m-1}, \ldots, A_1 \neq B_1) = \left(1 - \frac{1}{n}\right)^m \]

Let

\[ H_t := \frac{X_t(n - X_t)}{{n \choose 2}}. \]

Then

\[ E[H_t] = \left(1 - \frac{1}{n}\right)^t E[H_0]. \]

Indeed, for a pair of individuals \( A_0 \) and \( B_0 \) at time \( t \), the only way it is a heterozygous pair (one is \( A \) and the other is \( a \)) is if they originated from different ancestors \( A_t \neq B_t \) looking back \( t \) generations, and one of the ancestors was \( A \) while the other was \( a \).
The Coalescent. Recall:

\[ H_t := \frac{X_t(n - X_t)}{\binom{n}{2}} \quad \text{and} \quad E[H_t] = \left(1 - \frac{1}{n}\right)^t E[H_0]. \]

Then \( P(T_f > t) = P(1 \leq X_t \leq n-1) = P\left(H_t \geq \frac{n-1}{\binom{n}{2}}\right) = P\left(H_t \geq \frac{2}{n}\right) \)

as \( \binom{n}{2} = \frac{n(n-1)}{2} \). Therefore, by Markov inequality,

\[ P(T_f > t) = P\left(H_t \geq \frac{2}{n}\right) \leq \frac{nE[H_t]}{2} = \frac{n \left(1 - \frac{1}{n}\right)^t E[H_0]}{2} \]

and \( P(T_f > t) \leq \frac{n^2}{4(n-1)} \left(1 - \frac{1}{n}\right)^t \leq \frac{n}{3} \left(1 - \frac{1}{n}\right)^t \)

as \( H_0 = \frac{X_0(n - X_0)}{\binom{n}{2}} \leq \frac{(\frac{n}{2})^2}{\binom{n}{2}} = \frac{n}{2(n-1)}. \)
The Coalescent. Recall:

\[ H_t \coloneqq \frac{X_t(n - X_t)}{\binom{n}{2}} \quad \text{and} \quad E[H_t] = \left(1 - \frac{1}{n}\right)^t E[H_0]. \]

and \( P(T_f > t) \leq \frac{n}{3} \left(1 - \frac{1}{n}\right)^t. \)

Now, since \( 1 - x \leq e^{-x} \), for \( t = n \cdot \left( \ln n + \ln \frac{1}{3\epsilon} \right) \),

\[ P(T_f > t) \leq \frac{n}{3} \left(1 - \frac{1}{n}\right)^t \leq \frac{n}{3} e^{-t/n} = \epsilon \]

for any \( \epsilon > 0 \) you choose.

This allows us to estimate the fixation time for Wright-Fisher model as

\[ T_f \leq n \cdot (\ln n + \text{Constant}). \]
The Coalescent.

Terminology: **lineage** is a sequence of species (organisms, populations, cells, or genes), that form a line of descent, connected by ancestor/descendent relationships.

Moran model:

\[
\begin{align*}
1 & \rightarrow 1 \rightarrow 1 \rightarrow 1 \\
2 & \rightarrow 2 \rightarrow 2 \rightarrow 2 \leftarrow 2 \\
3 & \rightarrow 5 \rightarrow 5 \rightarrow 5 \rightarrow 5 \\
4 & \rightarrow 4 \rightarrow 2 \rightarrow 2 \rightarrow 2 \\
5 & \rightarrow 5 \rightarrow 5 \rightarrow 6 \rightarrow 6 \leftarrow 6 \\
6 & \rightarrow 6 \rightarrow 6 \rightarrow 6 \rightarrow 6 \rightarrow 6 \\
\end{align*}
\]
The Coalescent. Recall:

\[ H_t := \frac{X_t(n - X_t)}{\binom{n}{2}} \quad \text{and} \quad E[H_t] = \left(1 - \frac{1}{n}\right)^t E[H_0]. \]

In the Wright-Fisher model let us scale time by the number of individuals \( n = 2N \) as follows:

Let \( \tau = \frac{t}{n} \).

For \( n \) large enough, we switch to continuous time \( \tau \). Here

\[ \left(1 - \frac{1}{n}\right)^t = \left(1 - \frac{1}{n}\right)^{n\tau} \approx e^{-\tau} \quad \text{as} \quad \lim_{n \to \infty} \left(1 - \frac{1}{n}\right)^n = e^{-1}. \]

Thus

\[ E[H_t] \approx e^{-\tau} E[H_0]. \]

Hence, a pair of lineages will coalesce after waiting for an exponentially distributed (scaled) time variable.
The Coalescent.

Let $\tau = \frac{t}{n}$. For $n$ large enough, we switch to continuous time $\tau$. Here

$$
\left(1 - \frac{1}{n}\right)^t = \left(1 - \frac{1}{n}\right)^{n\tau} \approx e^{-\tau} \quad \text{as} \quad \lim_{n \to \infty} \left(1 - \frac{1}{n}\right)^n = e^{-1}.
$$

Thus a pair of lineages will coalesce after waiting for an exponentially distributed (scaled) time variable.

The probability that the two individuals originated from the same ancestor exactly $t$ generations ago is

$$
\left(1 - \frac{1}{n}\right)^{t-1} \cdot \frac{1}{n} \approx \lambda e^{-\lambda t}, \quad \text{where} \quad \lambda = \frac{1}{n}.
$$

If at time $t$ there are $X_t = k$ lineages, the waiting time will be an exponential random variable with parameter

$$
\lambda = \frac{1}{n} \binom{k}{2}.
$$
Coalescent tree.
Coalescent tree.
Coalescent tree.
Coalescent tree.
Coalescent tree.
Coalescent tree.
Kingman's coalescent.

Kingman's coalescent process begins with $n$ particles (clusters) of mass one (singletons). A pair of clusters with masses $i$ and $j$ coalesces independently of the other pairs (at the same rate) to form a new cluster of mass $i + j$. The process continues until there is a single cluster of mass $n$.

If at time $t$ there are $X_t = k$ lineages, the waiting time will be an exponential random variable with parameter

$$
\lambda = \frac{1}{n \binom{k}{2}}.
$$

The expectation for the time of the $m$th merger, $T_m$:

$$
E[T_m] = \frac{2m}{n - m}.
$$
Kingman’s coalescent.

The expectation for the time of the $m$th merger, $T_m$:

$$E[T_m] = \frac{n}{\binom{n}{2}} + \frac{n}{\binom{n-1}{2}} + \cdots + \frac{n}{\binom{n-m+1}{2}}$$

$$= \frac{2n}{n(n-1)} + \frac{2n}{(n-1)(n-2)} + \cdots + \frac{2n}{(n-m+1)(n-m)}$$

$$= 2n \left( -\frac{1}{n} + \frac{1}{n-1} - \frac{1}{n-1} + \frac{1}{n-2} + \cdots - \frac{1}{n-m+1} + \frac{1}{n-m} \right)$$

$$= 2n \left( -\frac{1}{n} + \frac{1}{n-m} \right) = \frac{2m}{n-m}.$$  

In particular, all $n-1$ mergers occur at

$$E[T_{n-1}] = 2(n-1).$$
Kingman’s coalescent.

The expectation for the time of the $m$th merger, $T_m$:

$$E[T_m] = \frac{2m}{n - m}.$$ 

In particular, all $n - 1$ mergers occur at

$$E[T_{n-1}] = 2(n - 1).$$

Thus, the fixation time for the Wright-Fisher model is

$$E[T_f] \sim E[T_{n-1}] = 2(n - 1) \sim 2n.$$
Branching process.

Francis Galton (1873) in the Educational Times:

Problem 4001: A large nation, of whom we will only concern ourselves with the adult males, $N$ in number, and who each bear separate surnames, colonize a district. Their law of population is such that, in each generation, $a_0$ per cent of the adult males have no male children who reach adult life; $a_1$ have one such male child; $a_2$ have two; and so on up to $a_5$ who have five.

Find (1) what proportion of the surnames will have become extinct after $r$ generations; and (2) how many instances there will be of the same surname being held by $m$ persons.

Solution was given by Reverend H. W. Watson: answer was wrong (guaranteed extinction), but the approach was correct.
Branching process.

Problem of Extinction. Start in the 0th generation with 1 parent. In the first generation we shall have 0, 1, 2, ... offspring with respective probabilities $p_0, p_1, p_2, \ldots$.

If in the $t$th generation there are $Z_t = k$ offspring, then in the $(t + 1)$st generation there will be

$$Z_{t+1} = X_1 + X_2 + \ldots + X_k$$

offspring, where $X_1, X_2, \ldots, X_k$ are independent random variables, each with the same probability mass function $p_0, p_1, p_2, \ldots$.

This description enables us to construct a tree, and a tree measure, for any number of generations.
Branching process.

Let $d_m$ be the probability that the process dies out by the $m$th generation. Then

$$0 = d_0 \leq d_1 \leq d_2 \leq \ldots \leq 1.$$ 

Monotonicity implies $d_m \to d$, $(0 \leq d \leq 1)$. Here, $d$ is the probability of extinction.

$$d_m = p_0 + p_1 d_{m-1} + p_2 (d_{m-1})^2 + p_3 (d_{m-1})^3 + \ldots.$$ 

Let $h(z)$ be the generating function for the distribution $p_k$:

$$h(z) = p_0 + p_1 z + p_2 z^2 + p_3 z^3 + \ldots.$$ 

Then

$$d_m = h(d_{m-1})$$

Hence, as $d_m \to d$,

$$d = h(d).$$
Branching process.

Source: Grinstead and Laurie Snell (Chapter 10)

\[ d = h(d). \]
Branching process.

If $Z_t = k$, then

$$Z_{t+1} = X_1 + X_2 + \ldots + X_k$$

offspring,

where $X_i$ ($i = 1, 2, \ldots$) have probability mass function $p_0, p_1, p_2, \ldots$.

Recall:

$$h(z) = p_0 + p_1 z + p_2 z^2 + p_3 z^3 + \ldots.$$  

Then, the derivative

$$h'(z) = p_1 + 2p_2 z + 3p_3 z^2 + \ldots$$

and

$$h'(1) = p_1 + 2p_2 + 3p_3 + \ldots = E[X_i].$$

**Main result:** Suppose $p_1 \neq 1$. Then,

$$d = 1 \text{ (guaranteed extinction) if and only if } h'(1) = E[X_i] \leq 1.$$
Branching process.

$h(z)$ is a **convex** function for $z \geq 0$ as $h''(z) = 2p_2 + 6p_3z + \ldots \geq 0$.

$d = 1$ if and only if $h'(1) = E[X_i] \leq 1$. 
Branching process.

\[ d_m = h(d_{m-1}) \quad \text{and} \quad d = h(d). \]
Critical branching process.

Example. Consider a critical binary Galton-Watson (branching) process:

\[ p_0 = p_2 = \frac{1}{2} \]

It is critical: \( E[X_i] = p_1 + 2p_2 + 3p_3 + \ldots = 1 \).

Let \( N \) be the number vertices. Then,

\[ P(N < \infty) = 1 \quad \text{and} \quad E[N] = \infty \]

Example. Consider a Galton-Watson (branching) process with \( p_0 = \frac{1}{2}, \ p_1 = \frac{1}{4}, \ p_2 = \frac{1}{8}, \ldots, \ p_k = \frac{1}{2^{k+1}}, \ldots \).

It is critical: \( E[X_i] = p_1 + 2p_2 + 3p_3 + \ldots = 1 \).

Here too, for the number of vertices \( N \),

\[ P(N < \infty) = 1 \quad \text{and} \quad E[N] = \infty \]
Harris path.

(a) Tree $T$

(b) Harris path $H_T$
Critical branching process.

Example. Consider a critical binary Galton-Watson (branching) process:

\[ p_0 = p_2 = \frac{1}{2} \]

Corresponding random walk:
- probability \( \frac{1}{2} \) for up or down;
- jumps = independent exponential random variables.

Example. Consider a Galton-Watson (branching) process with \( p_0 = \frac{1}{2}, \ p_1 = \frac{1}{4}, \ p_2 = \frac{1}{8}, \ \ldots, \ p_k = \frac{1}{2^{k+1}}, \ \ldots \)

Corresponding random walk:
- probability \( \frac{1}{2} \) for up or down;
- jump sizes = 1.
Yule process.

Consider a branching process with $p_2 = 1$. That is, every individual leaves exactly two offsprings.

The time is continuous. The waiting times before splitting in two are independent exponential random variables with a given parameter $\lambda > 0$.

Let $Y(t)$ be the number of the individuals at time $t$. We begin with $Y(0) = y_0$.

Deterministic analysis: $Y(t + \delta) = Y(t) + \lambda \delta Y(t) + O(\delta^2)$

Thus,

$$Y'(t) = \lambda Y(t) \quad \text{and} \quad Y(t) = y_0 e^{\lambda t}.$$
Yule process.

Consider a branching process with $p_2 = 1$. That is, every individual leaves exactly two offsprings.

The time is continuous. The waiting times before splitting in two are independent exponential random variables with a given parameter $\lambda > 0$.

Let $Y_t$ be the number of the individuals at time $t$. We begin with $Y_0 = y_0$.

**Probabilistic analysis:** let $p_k(t) = P(Y_t = k)$. Then,

$$p_k(t+\delta) = p_k(t)e^{-\lambda k\delta} + p_{k-1}(t)(\lambda(k-1)\delta + O(\delta^2)) + O(\delta^2)$$

which rewrites as

$$p_k(t+\delta) - p_k(t) = -\lambda k\delta p_k(t) + \lambda(k-1)\delta p_{k-1}(t) + O(\delta^2)$$
Yule process.

Let $Y_t$ be the number of the individuals at time $t$. We begin with $Y_0 = y_0$.

**Probabilistic analysis:** let $p_k(t) = P(Y_t = k)$. Then,

$$p_k(t + \delta) = p_k(t) e^{-\lambda k \delta} + p_{k-1}(t) (\lambda (k-1)\delta + O(\delta^2)) + O(\delta^2)$$

which rewrites as

$$p_k(t + \delta) - p_k(t) = -\lambda k \delta p_k(t) + \lambda (k-1) \delta p_{k-1}(t) + O(\delta^2)$$

Thus,

$$p'_k(t) = -\lambda k p_k(t) + \lambda (k-1) p_{k-1}(t) \quad \text{with} \quad p_{y_0}(0) = 1.$$  

Answer: for $k = y_0, y_0 + 1, y_0 + 2, \ldots$,

$$p_k(t) = \binom{k-1}{k-y_0} e^{-\lambda ty_0} (1 - e^{-\lambda t})^{k-y_0}$$

Thus, $Y_t - y_0$ is **negative binomial**, $\text{NB}(y_0, 1 - e^{-\lambda t})$.  


**Yule process.** Let $Y_t$ be the number of the individuals at time $t$. We begin with $Y_0 = y_0$.

Probabilistic analysis: let $p_k(t) = P(Y_t = k)$. Then,

$$p'_k(t) = -\lambda k p_k(t) + \lambda (k-1) p_{k-1}(t) \quad \text{with} \quad p_y(0) = 1.$$  

Answer: for $k = y_0, y_0 + 1, y_0 + 2, \ldots$,

$$p_k(t) = \binom{k-1}{k-y_0} e^{-\lambda t y_0} \left(1 - e^{-\lambda t}\right)^{k-y_0}$$

Thus, $Y_t - y_0$ is **negative binomial**, $NB(y_0, 1 - e^{-\lambda t})$.

**Negative Binomial:** $X$ is $NB(r, p)$ if its p.m.f. is

$$P(X = m) = \binom{m+r-1}{m} (1-p)^r p^m \quad m = 0, 1, 2, \ldots$$

In terms of Bernoulli trials, it is the number of success before the $r^{th}$ failure.

Expectation: if $X$ is $NB(r, p)$, then $E[X] = \frac{pr}{1-p}$. 

Yule process.

Probabilistic analysis: let \( p_k(t) = P(Y_t = k) \).

\[
p_k(t) = \binom{k - 1}{k - y_0} e^{-\lambda t y_0} (1 - e^{-\lambda t})^{k - y_0}
\]

Thus, \( Y_t - y_0 \) is **negative binomial**, \( \text{NB}(y_0, 1 - e^{-\lambda t}) \).

**Negative Binomial:** \( X \) is \( \text{NB}(r, p) \) if its p.m.f. is

\[
P(X = m) = \binom{m + r - 1}{m} (1 - p)^r p^m \quad m = 0, 1, 2, \ldots
\]

In terms of Bernoulli trials, it is the number of success before the \( r^{th} \) failure.

Expectation: if \( X \) is \( \text{NB}(r, p) \), then \( E[X] = \frac{pr}{1-p} \).

Hence,

\[
E[Y_t] = y_0 + \frac{y_0(1 - e^{-\lambda t})}{e^{-\lambda t}} = y_0 e^{\lambda t}
\]

as in the deterministic analysis.
Yule process.

Tree structure: let $y_0 = 1$, then, conditioned on $Y_t = n$, the dendritic structure of the tree for Yule process matches that of the tree for the Kingman coalescent process that begins with $n$ singletons.