Lecture 12

Birth and Death Processes

The subsequent material has been partially taken from V. Andasari, course notes Stochastic Modeling, at Boston University and from Wikipedia.

Transition Probabilities, Kolmogorov's equations

The continuous-time birth and death Markov chain $\{X(t) : t \in [0, \infty)\}$ may have either a finite $\{0, 1, 2, \cdots, N]$ or infinite $\{0, 1, 2, \cdots\}$ state space. Assume that its transition probabilities $P_y(t)$ are stationary, i.e.

$$P_{ij}(t) = P|X(t+s) = j|X(s) = i\}, \quad \text{for all } \Delta t \geq 0.$$

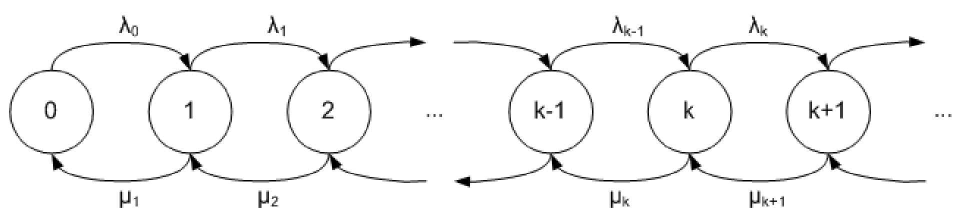
In addition, assume the infinitesimal transition probabilities for this process are

$$\begin{aligned} P_{i,i+j}(\Delta t) &= P\{X(t+\Delta t) - X(t) = j \mid X(t) = i\} \\ &= \begin{cases} \lambda_i \Delta t + o(\Delta t), & j = 1 \\ \mu_i \Delta t + o(\Delta t), & j = -1 \\ 1 - (\lambda_i + \mu_i)\,\Delta t + o(\Delta t), & j = 0 \\ o(\Delta t), & j \neq -1, 0, 1 \end{cases} \end{aligned}$$

for Δt sufficiently small, $\mu_0 = 0, \lambda_0 > 0$, and $\lambda_i > 0, \mu_i > 0$ for $i = 1, 2, \cdots$. It is often the case that $\lambda_0 = 0$, except when there is immigration. The initial conditions are

$$P_{ij}(0) = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

In a small time interval Δt, at most one change in state can occur, either a birth, $i \to i+1$ or a death, $i \to i-1$.



Source Wikipedia

In the same way as for the Poisson process, we define $P_i(t) = P\{X(t) = i\}$ and assume $X(0) = 0$. If $\Delta t > 0, i \geq 1$, by invoking the law of total probability and the Markov property, we obtain

$$\begin{aligned}
P_{ij}(t+\Delta t) &= \sum_{k=0}^{\infty} P_{ik}(t) P_{kj}(\Delta t) \\
&= \sum_{k=0}^{\infty} P_{ik}(t) \cdot P\{X(t+\Delta t) = j \mid X(t) = k\} \\
&= \sum_{k=0}^{\infty} P_{ik}(t) \cdot P\{X(t+\Delta t) - X(t) = j - k \mid X(t) = k\} \\
&= \sum_{k=0}^{k=j} P_{ik}(t) \cdot P\{X(t+\Delta t) - X(t) = j - k \mid X(t) = k\}.
\end{aligned}$$

Now, for $k = j$, for the right hand side we have

$$\begin{aligned}
&= P_{ij}(t) \cdot P\{X(t+\Delta t) - X(t) = j - j \mid X(t) = j\} \\
&= P_{ij}(t) \cdot P\{X(t+\Delta t) - X(t) = 0 \mid X(t) = j\} \\
&= P_{ij}(t) \cdot [1 - (\lambda_j + \mu_j)\,\Delta t + o(\Delta t)]
\end{aligned}$$

for $k = j - 1$,

$$\begin{aligned}
&= P_{i,j-1}(t) \cdot P\{X(t+\Delta t) - X(t) = j - (j-1) \mid X(t) = j-1\} \\
&= P_{i,j-1}(t) \cdot P\{X(t+\Delta t) - X(t) = 1 \mid X(t) = j-1\} \\
&= P_{i,j-1}(t) \cdot [\lambda_{j-1}\Delta t + o(\Delta t)]
\end{aligned}$$

for $k = j + 1$

$$\begin{aligned}
&= P_{i,j+1}(t) \cdot P\{X(t+\Delta t) - X(t) = j - (j+1) \mid X(t) = j+1\} \\
&= P_{i,j+1}(t) \cdot P\{X(t+\Delta t) - X(t) = -1 \mid X(t) = j+1\} \\
&= P_{i,j+1}(t) \cdot [\mu_{j+1}\Delta t + o(\Delta t)]
\end{aligned}$$

whereas for $k \neq j, j-1, j+1$, that is for $k \leq j - 2$

$$\begin{aligned}
&= P_{ik}(t) \cdot P\{X(t+\Delta t) - X(t) \geq 2 \mid X(t) = k\} \\
&= P_{ik}(t) \cdot o(\Delta t)
\end{aligned}$$

and $k \geq j + 2$

$$\begin{aligned}
&= P_{ik}(t) \cdot P\{X(t+\Delta t) - X(t) \leq -2 \mid X(t) = k\} \\
&= P_{ik}(t) \cdot o(\Delta t)\,.
\end{aligned}$$

Collecting all together,

$$\begin{aligned}
P_{ij}(t+\Delta t) =& P_{ij}(t) \cdot [1 - (\lambda_j + \mu_j)\,\Delta t + o(\Delta t)] + P_{i,j-1}(t) \cdot [\lambda_{j-1}\Delta t + o(\Delta t)] + \\
& P_{i,j+1}(t) \cdot [\mu_{j+1}\Delta t + o(\Delta t)] + P_{ik}(t) \cdot o(\Delta t) \\
=& P_{i,j-1}(t)\lambda_{j-1}\Delta t + P_{i,j+1}(t)\mu_{j+1}\Delta t + P_{ij}(t)\,[1 - (\lambda_j + \mu_j)\,\Delta t] + o(\Delta t),
\end{aligned}$$

which holds for all i and j in the state space with the exception of $j = 0$ and $j = N$ (if the population size is finite).

If $j = 0$, then (and due to $\mu_0 = 0$)

$$P_{i0}(t+\Delta t) = P_{i1}(t)\mu_1\Delta t + P_{i0}(t)\,[1 - \lambda_0\Delta t] + o(\Delta t)$$

If $j = N$, which is the maximum population size, then

$$P_{iN}(t+\Delta t) = P_{i,N-1}(t)\lambda_{N-1}\Delta t + P_{iN}(t)\left[1-\mu_N)\,\Delta t\right] + o(\Delta t),$$

where $\lambda_N = 0$ and $P_{kN} = 0$ for $k > N$. Subtracting $P_{ij}(t), P_{i0}(t)$, and $P_{iN}(t)$ from the preceding three equations, respectively, dividing by Δt, and taking the limit as $\Delta t \to 0$, yields the *forward Kolmogorov differential equations* for the general birth and death process,

$$\begin{aligned}
\frac{dP_{ij}(t)}{dt} &= \lambda_{j-1}P_{i,j-1}(t) - (\lambda_j+\mu_j)\,P_{ij}(t) + \mu_{j+1}P_{i,j+1}(t) \\
\frac{dP_{i0}(t)}{dt} &= -\lambda_0 P_{i0}(t) + \mu_1 P_{i1}(t), \quad \text{for } j = 0 \\
\frac{dP_{iN}(t)}{dt} &= \lambda_{N-1}P_{i,N-1}(t) - \mu_N P_{iN}(t), \quad \text{for } j = N\,.
\end{aligned}$$

The forward Kolmogorov differential equations can be written in matrix notation,

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{P}(t)\mathbf{Q}$$

or

$$\begin{bmatrix} dP_{i0}/dt \\ dP_{i1}/dt \\ dP_{i2}/dt \\ dP_{i3}/dt \\ \vdots \end{bmatrix} = [P_{i0}\; P_{i1}\; P_{i2}\; P_{i3}\; \cdots]\cdot \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & \cdots \\ 0 & 0 & \mu_3 & -(\lambda_3+\mu_3) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

where the generator matrix $\mathbf{Q}$ for the infinite state space is

$$\mathbf{Q} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & \cdots \\ 0 & 0 & \mu_3 & -(\lambda_3+\mu_3) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix},$$

and for the finite state space is

$$\mathbf{Q} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \mu_{N-1} & -(\lambda_{N-1}+\mu_{N-1}) & \lambda_{N-1} \\ 0 & 0 & 0 & \mu_N & -\mu_N \end{bmatrix}.$$

Similarly, to obtain the backward Kolmogorov differential equations, we start from

$$\begin{aligned}
P_{ij}(\Delta t + t) &= \sum_{k=0}^{\infty} P_{ik}(\Delta t)P_{kj}(t) \\
&= \underbrace{P_{i,i-1}(\Delta t)P_{i-1,j}(t)}_{k=i-1} + \underbrace{P_{ii}(\Delta t)P_{ij}(t)}_{k=i} + \underbrace{P_{i,i+1}(\Delta t)P_{i+1,j}(t)}_{k=i+1} + \sum_{k} P_{ik}(\Delta t)P_{kj}(t),
\end{aligned}$$

where the last summation is over all $k \neq i-1, i+1, i$. For $k = i$

$$P_{ii}(\Delta t) P_{ij}(t) = P_{ij}(t) \cdot [1 - (\lambda_i + \mu_i)\, \Delta t + o(\Delta t)]$$

for $k = i - 1$

$$P_{i,i-1}(\Delta t) P_{i-1,j}(t) = P_{i-1,j}(t) \cdot [\mu_i \Delta t + o(\Delta t)]$$

for $k = i + 1$

$$P_{i,i+1}(\Delta t) P_{i+1,j}(t) = P_{i+1,j}(t) \cdot [\lambda_i \Delta t + o(\Delta t)]$$

and for $k \neq j, j-1, j+1$, we have

$$\sum_k P_{ik}(\Delta t) P_{kj}(t) = P_{kj}(t) \cdot o(\Delta t)\,.$$

Piecing everything together and rearranging the resulting equation , it turns out that

$$P_{ij}(t + \Delta t) - P_{ij}(t) = \mu_i \Delta t P_{i-1,j}(t) - (\lambda_i + \mu_i)\, \Delta t P_{ij}(t) + \lambda_i \Delta t P_{i+1,j}(t) + o(\Delta t)$$

Dividing by Δt and taking the limit as $\Delta t \to 0$, we obtain the *backward Kolmogorov differential equations*

$$\begin{aligned}
\frac{dP_{ij}(t)}{dt} &= \mu_i P_{i-1,j}(t) - (\lambda_i + \mu_i)\, P_{ij}(t) + \lambda_i P_{i+1,j}(t) \\
\frac{dP_{0j}(t)}{dt} &= -\lambda_0 P_{0j}(t) + \lambda_0 P_{1j}(t), \qquad \text{for } i = 0 \\
\frac{dP_{Nj}(t)}{dt} &= \mu_N P_{N-1,j}(t) - \mu_N P_{Nj}(t), \qquad \text{for } i = N
\end{aligned}$$

where $\mu_0 = 0$ and $\lambda_N = 0$, and $j \geq 0, i \geq 0$ The backward Kolmogorov differential equations can be written in matrix notation,

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{QP}(t)$$

or

$$\begin{bmatrix} dP_{0j}/dt \\ dP_{1j}/dt \\ dP_{2j}/dt \\ \vdots \\ dP_{Nj}/dt \end{bmatrix} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & \mu_N & -\mu_N \end{bmatrix} \begin{bmatrix} P_{0j} \\ P_{1j} \\ P_{2j} \\ \vdots \\ P_{Nj} \end{bmatrix}.$$

The Limiting Behavior of Birth and Death Processes

For a general birth and death process that has no absorbing states, the limits

$$\lim_{t \to \infty} P_{ij}(t) = \pi_j \geq 0$$

exist and are independent of the initial state i. The limits form the limiting distribution (or limiting probability) of the process, which is at the same time the stationary distribution.

To determine if a limiting distribution exists and what its values are, we rewrite the forward Kolmogorov differential equations,

$$\begin{aligned}\frac{dP_{i0}(t)}{dt} &= -\lambda_0 P_{i0}(t) + \mu_1 P_{i1}(t)\\ \frac{dP_{ij}(t)}{dt} &= \lambda_{j-1}P_{i,j-1}(t) - (\lambda_j + \mu_j)\, P_{ij}(t) + \mu_{j+1}P_{i,j+1}(t), \quad j \geq 1,\end{aligned}$$

with initial conditions $P_{ij}(0) = \delta_{ij.}$ If we take the limit as $t \to \infty$ to these equations, the limit of the right hand sides exists. The limit on the left hand side, the derivatives $P'_{ij}(t)$, also exists. Since the probabilities are converging to a constant, the limit of these derivatives must be zero. These equations can be solved recursively. For the first equation, or when $j = 0$

$$\lim_{t\to\infty}\frac{dP_{i0}(t)}{dt} = -\lambda_0 \lim_{t\to\infty} P_{i0}(t) + \mu_1 \lim_{t\to\infty} P_{i1}(t)$$

$$0 = -\lambda_0\pi_0 + \mu_1\pi_1$$

Or

$$\pi_1 = \frac{\lambda_0}{\mu_1}\pi_0$$

From the second equation,

$$\lim_{t\to\infty}\frac{dP_{ij}(t)}{dt} = \lambda_{j-1}\lim_{t\to\infty} P_{i,j-1}(t) - (\lambda_j + \mu_j)\lim_{t\to\infty} P_{ij}(t) + \mu_{j+1}\lim_{t\to\infty} P_{i,j+1}(t)$$

$$0 = \lambda_{j-1}\pi_{j-1} - (\lambda_j + \mu_j)\,\pi_j + \mu_{j+1}\pi_{j+1}\,,$$

when $j = 1$

$$\pi_2 = \frac{\lambda_1}{\mu_2}\pi_1 = \frac{\lambda_0\lambda_1}{\mu_1\mu_2}\pi_0\,,$$

when $j = 2$,

$$\pi_3 = \frac{\lambda_2}{\mu_3}\pi_2 = \frac{\lambda_0\lambda_1\lambda_2}{\mu_1\mu_2\mu_3}\pi_0$$

and so on. Thus,

$$\pi_{i+1} = \frac{\lambda_i}{\mu_{i+1}}\pi_i\,.$$

Now we aim to show by induction that the stationary probability distribution equals

$$\pi_i = \frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i}\pi_0\,, \quad i = 1, 2, 3, \cdots.$$

We have to verify this relation for $i+1$ instead of i. Then

$$\begin{aligned}\mu_{i+1}\pi_{i+1} &= (\lambda_i + \mu_i)\,\pi_i - \lambda_{i-1}\pi_{i-1}\\ &= \left(\frac{\lambda_0\lambda_1\cdots\lambda_{i-1}\,(\lambda_i+\mu_i)}{\mu_1\mu_2\cdots\mu_i} - \frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_{i-1}}\right)\pi_0\\ &= \frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_{i-1}}\left(\frac{\lambda_i+\mu_i}{\mu_i} - 1\right)\pi_0\end{aligned}$$

or,

$$\pi_{i+1} = \frac{\lambda_0\lambda_1\cdots\lambda_i}{\mu_1\mu_2\cdots\mu_{i+1}}\pi_0\,.$$

Thus, if the state space is infinite, $\{0,1,2,\cdots\}$, a unique positive stationary probability distribution π for a general birth and death process exists, with

$$\mu_i > 0 \quad \text{and} \quad \lambda_{i-1} > 0 \quad \text{for } i = 1,2,3,\cdots$$

Since $\sum_{i=0}^{\infty}\pi_i = 1$, which we expand

$$\sum_{i=0}^{\infty}\pi_i = 1$$

$$\pi_0 + \sum_{i=1}^{\infty}\pi_i = 1$$

$$\pi_0\left(1 + \sum_{i=1}^{\infty}\frac{\pi_i}{\pi_0}\right) = 1,$$

from which we solve for π_0,

$$\pi_0 \;=\; \frac{1}{1 + \sum_{i=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i}}\,. \tag{1}$$

Let us summarize our work.

Theorem Suppose $\mu_i > 0$ and $\lambda_{i-1} > 0$ for all $i \in \mathbb{N}$ and

$$\sum_{i=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i} < \infty\,.$$

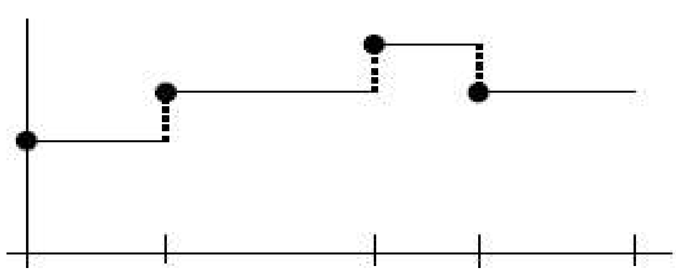
Then there exists a unique limiting distribution

$$\pi_i = \frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i}\pi_0\,, \quad i \in \mathbb{N},$$

where π_0 is given by (1).

Construction by Trajectories: Waiting times and Jumps

Now suppose we are given a birth and death process $\{X(t) : t \in [0,\infty)\}$ with state space $S = \mathbb{Z}_+$. Given $X(0) = i \in S$, introduce the *waiting time* $T_i := \inf\{t > 0 : X(t) \neq i\}$.



Theorem It holds that $T_0 \sim Exp(\lambda_0)$ and $T_i \sim Exp(\lambda_i + \mu_i)$ $i \in \mathbb{N}$.

Proof. Let us carry out the proof for $i \in \mathbb{N}$, the proof for $i = 0$ is similar. Introduce

$$G_i(t) := P(T_i > t | X(0) = i)\,, \quad t > 0.$$

By the definition of conditional probability and the fact that $\{T_i > t + h, T_i > t, X(0) = i\} = \{T_i > t + h, X(0) = i\}$ we have for $h > 0$

$$\begin{aligned} P(T_i > t + h | X(0) = i) &= \frac{P(T_i > t, X(0) = i)}{P(X(0) = 0)} \cdot \frac{P(T_i > t + h, T_i > t, X(0) = i)}{P(T_i > t, X(0) = i)} \\ &= P(T_i > t | X(0) = i) \cdot P(T_i > t + h | T_i > t, X(0) = i)\,. \end{aligned}$$

Thus, by the Markov principal, and time homogeneity,

$$\begin{aligned} P(T_i > t + h | X(0) = i) &= P(T_i > t | X(0) = i) \cdot P(T_i > t + h | X(t) = i) \\ &= P(T_i > t | X(0) = i) \cdot P(T_i > h | X(0) = i)\,. \end{aligned}$$

According to the infinitesimal transition probabilities, this says

$$G_i(t + h) = G_i(t) \cdot G_i(h) = G_i(t) \cdot (1 - (\lambda_i + \mu_i)\, h) + o(h)\,.$$

where we consider now h as an infinitesimal time increment. Rearranging this and dividing by h we obtain

$$\frac{G_i(t + h) - G_i(t)}{h} = -(\lambda_i + \mu_i) G_i(t) + \frac{o(h)}{h}\,.$$

Letting $h \to 0$ this implies

$$\frac{d}{dt} G_i(t) = -(\lambda_i + \mu_i) G_i(t)\,, \quad G_i(0) = 1\,,$$

which has the unique solution

$$P(T_i > t | X(0) = i) = G_i(t) = e^{-(\lambda_i + \mu_i)\, t}\,, \quad t \geq 0\,.$$

However the right-hand side is $1-$ cumulative distribution function of $Exp(\lambda_i + \mu_i)$. □

Remark For the $Exp(\lambda_i + \mu_i)$-distribution of the waiting times between the $(n-1)$st and nth jump we recall the *memoryless property* of the exponential distribution. Let $X \sim Exp(\lambda)$. We have

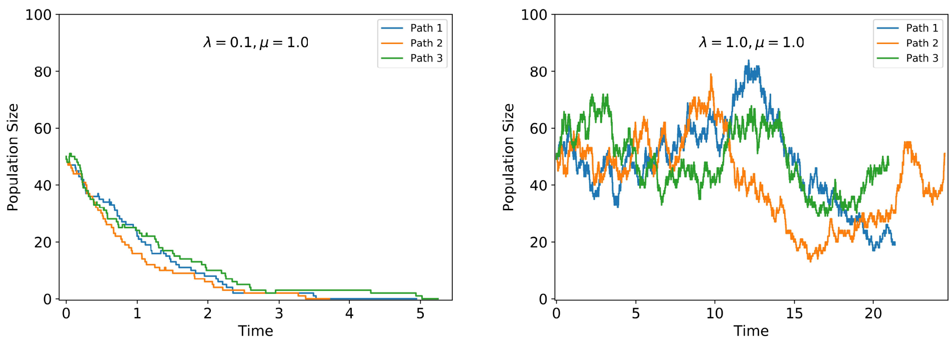
$$\begin{aligned} P(X > t + x | X > t) &= \frac{P(X > t + x, X > t)}{P(X > t)} \\ &= \frac{P(X > t + x)}{P(X > t)} = \frac{1 - F_X(t + x)}{1 - F_X(t)} = \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} \\ &= e^{-\lambda x} = P(X > x)\,. \end{aligned}$$

Theorem Let $X(t) = i \in \mathbb{N}$ for some $t \geq 0$. Then the the probability to jump to $i - 1$ (resp. $i + 1$) next is $\frac{\mu_i}{\lambda_i + \mu_i}$ (resp. $\frac{\lambda_i}{\lambda_i + \mu_i}$).

Example The so-called *simple birth and death process* $\{X(t) : t \in [0, \infty)\}$ has transition probabilities

$$P_{i,i+j}(\Delta t) = P\{X(t + \Delta t) - X(t) = j \mid X(t) = i\}$$
$$= \begin{cases} 1 - i(\mu + \lambda)\Delta t + o(\Delta t) & j = 0 \\ i\lambda\Delta t + o(\Delta t) & j = 1 \\ i\mu\Delta t + o(\Delta t) & j = -1 \\ o(\Delta t) & j \neq -1, 0, 1 \end{cases}$$

for all $i \in \mathbb{Z}_+$. Simple birth and death processes are also known as birth and death processes with absorbing states. For these processes, the zero state is an absorbing state, where when the population size becomes zero, it remains zero thereafter.

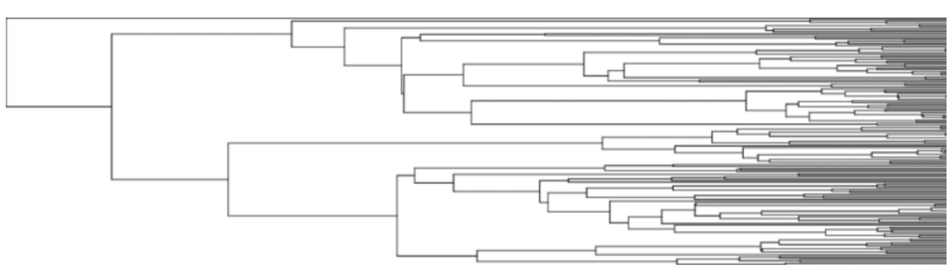


V. Andasari, course notes Stochastic Modeling, BU

Plots of three sample paths for the simple birth and death process when $\mu = 1.0$, $X(0) = 50$, and $\lambda = 0.1$ (left) and $\lambda = 1.0$ (right). In the case $\lambda << \mu$ (left figure), the population size becomes zero and it remains in zero forever.

Example The *Yule Process* or *simple birth process* is defined by its birth rates $\lambda_i := i\,\lambda$, $i \in \mathbb{N}$ for some $\lambda > 0$. There is no death, i. e. $\mu_i := 0$, $i \in \mathbb{N}$. The process stars in state 1, i. e. $X(0) = 1$.

In other words, for a simple birth process the birth rates are proportional to the size of the current population. This is, for example, a reasonable assumption on the growth of a virus population in an infected individual until the production of anti-bodies sets in. As time proceeds according to the horizontal axis we get a branching image as below.



Federico Polito, research gate, modified

The number of births at time t of a simple birth process of population size n is given by

$$p_{n,n+m}(t) = \binom{n}{m} (\lambda t)^m (1-\lambda t)^{n-m} + o(h)\,, \quad t \geq 0.$$

In exact form, the number of births is the negative binomial distribution with parameters n and $e^{-\lambda t}$, recall that on Wikipedia.

For the special case $n = 1$, this is the geometric distribution with success rate $e^{-\lambda t}$. The expectation of the process grows exponentially. In particular, if $X_0 = 1$ then $E[X(t)] = e^{\lambda t}$, $t \geq 0$.

Jörg-Uwe Löbus, `jorg-uwe.lobus@liu.se`