Birth and death process.

In practical queue systems, the request arrivals result in resource allocation and eventually users get served and leave the system. This process belongs to a wider class of stochastic processes commonly referred to as the birth and death.
- incoming request $\rightarrow$ birth
- user leaving the system $\rightarrow$ death

Formal definition of a birth and death process.

Definition: Consider a stochastic process $(N_t)$ that is continuous in time domain but discrete in amplitude domain. Suppose that the process describes a physical system that may be in state $E_n$, $n=0,1,2,\ldots$ at time $t$ if and only if $N(t)=n$. Then the system is described by the birth and death process if there exist nonnegative birth rates $\lambda_n$, $n=0,1,2,\ldots$ and nonnegative death rates $\mu_n$, $n=1,2,\ldots$, such that following properties are true:

1) State changes are only between state $E_n$ to state $E_{n+1}$ or between $E_n$ to $E_{n-1}$ if $n \geq 1$, but from state $E_0$ to state $E_1$ only.

2) At time $t$, the system is in state $E_n$. The probability that between time $t$ and time $t+h$ a transition from state $E_n$ to state $E_{n+1}$ occurs is given by $\lambda_n h + o(h)$, and the probability of transition from $E_n$ to $E_{n-1}$ is $\mu_n h + o(h)$ (if $n \geq 1$).

3) The probability that in time interval from $t$ to $t+h$ there are more than one transition is given by $o(h)$.

The birth and death process is usually represented using a state diagram.
Example: State of the system: Number of resources occupied at the cellular system.

- Birth rate = number of resource requests per unit time
- Death rate = number of serviced requests per unit time

Of special interest to us is $P_{n+1} P_n(n+1) = n$. That is probability of system occupying state $E_n$. To determine this probability one considers a set of differential equations.

If $n > 1$, we only consider $P_{n-1}$, and at time $t$, the system occupies state $E_n$ has four components listed as follows.

1) The system was in state $E_{n-1}$ at time $t$ and no births or deaths have occurred.

$$P_{n-1}^{(1)}(t+h) = P_{n-1}(t) \cdot [1 - \mu n h + o(n)] \cdot [1 - \mu n h + o(n)]$$

$$= P_{n-1} [1 - \mu n h - \mu n h + o(n)] - \text{probability of transition } E_{n-1} \rightarrow E_n$$

2) The system was in state $E_{n-1}$ at time $t$ and a birth has occurred, but no deaths.

$$P_{n-1}^{(2)}(t+h) = P_{n-1}(t) \cdot [\mu n h + o(n)] \cdot [1 - \mu n h + o(n)] =$$

$$P_{n-1}(t) \cdot \mu n h + o(n) - \text{probability of transition } E_{n-1} \rightarrow E_n$$

3) The system was in state $E_n$ at time $t$ and a death has occurred, but no births.

$$P_{n}^{(3)}(t+h) = P_n(t) \cdot [1 - \mu n h + o(h)] \cdot [\mu n h + o(h)] =$$
= \rho_{n+1}(t) \cdot \mu_{n+1} t + \sigma(h)

4) Two random transitions occurred. By the definition of the birth and death process, the net probability is

\[
P_n^{(k)}(t) = \mathcal{O}(h)
\]

Summing \( P_0^{(k)} \), \( k = 1, 2 \), one obtains

\[
P_n(t + h) = \frac{4}{\lambda} P_0^{(k)} = \left[ 1 - \lambda \mu - \rho \nu \right] P_0(t) + \lambda n \mu P_{n-1}(t) + \mu n \nu P_{n+1}(t) + \mathcal{O}(h)
\]

or

\[
\frac{P_n(t + h) - P_n(t)}{h} = \left( \lambda n + \mu n \right) P_{n+1}(t) + \lambda n \mu P_{n-1}(t) + \mu n \nu P_{n+1}(t) + \mathcal{O}(h)
\]

Letting \( h \to 0 \) one obtains difference-differential equation that describes birth and death process:

\[
\frac{dP_n(t)}{dt} = \left( \lambda n + \mu n \right) P_{n+1}(t) + \lambda n \mu P_{n-1}(t) + \mu n \nu P_{n+1}(t) , \quad n \geq 1
\]

For \( n = 0 \),

\[
\frac{dP_0(t)}{dt} = -2 \frac{P_0(t)}{\mu} + \lambda \mu P_1(t) + \mu \nu P_2(t)
\]

If the initial condition is \( F_i \) that is \( P_i(0) = 1, P_i(0) = 0, i > 1 \), one can solve set of equations \( (x) \) for any given time and any given set of birth and death rates. The solution can always be determined numerically (and almost never analytically).

Of a special interest are systems that are in a steady state. This means that all birth and death rates are constant and that probabilities have converged to their steady state values. In such a case, \( \frac{dP_n(t)}{dt} = 0 \), and one obtains
\[ O = \sum_{n=1}^{\infty} P_{n-1} + \mu_0 \, P_{n+1} - (\lambda_0 + \mu_0) \, P_n, \quad n \geq 1 \]

\[ O = \lambda_1 \, P_1 - \lambda_0 \, P_0, \quad \text{for } n = 0 \]

Or,

\[ n = 0 \quad P_1 = \frac{\lambda_0}{\mu_1} \cdot P_0 \]

\[ n = 1 \quad O = \lambda_0 \, P_0 + \mu_2 \, P_2 - (\lambda_1 + \mu_1) \cdot P_1 \]

\[ \frac{P_2}{\mu_2} = \frac{1}{\mu_2} \left[ \lambda_0 \, P_0 - (\lambda_1 + \mu_1) \cdot P_1 \right] = \frac{1}{\mu_2} \left( \mu_1 \cdot P_1 - \lambda_1 \, P_1 \right) \cdot P_1 \]

\[ P_2 = \frac{\lambda_1}{\mu_2} \cdot P_1 \]

\[ n = 3 \quad P_3 = \frac{\lambda_2}{\mu_3} \cdot P_3, \quad \text{and so on.} \]

In general, \[ P_{n+1} = \frac{\lambda_n \cdot P_0}{\mu_{n+1}}, \quad n = 0, 1, \ldots \]

Therefore, one way write

\[ P_1 = \frac{\lambda_0}{\mu_1} \cdot P_0 = C_1 \cdot P_0 \]

\[ P_2 = \frac{\lambda_1}{\mu_2} \cdot P_1 = \frac{\lambda_1 \cdot \lambda_0}{\mu_2 \cdot \mu_1} \cdot P_0 = C_2 \cdot P_0 \]

\[ \vdots \]

\[ P_n = \frac{\lambda_{n-1} \cdot P_{n-1}}{\mu_n} = \frac{\lambda_{n-1} \cdot \lambda_{n-2} \cdots \lambda_0}{\mu_n \cdot \mu_{n-1} \cdots \mu_1} \cdot P_0 = C_n \cdot P_0 \]

Given that sum of all probabilities must be equal to 1, one obtains

\[ \sum_{n=0}^{\infty} P_n = P_0 \left( 1 + C_1 + C_2 + \cdots + C_n \cdots \right) = P_0 \leq 1 \]
Theorem:

$P_0 = 1/\lambda$, \text{ closed form solution for state probabilities}

$P_n = C_n/\lambda$

Where $C_n = \frac{\lambda^n \cdot \lambda^{-1} \cdots \lambda^{-n}}{\mu_n \cdot \mu_{n-1} \cdots \mu_1}$, $S = 1 + C_1 + C_2 + \cdots + C_n$

Example: Consider a queuing system with one server. Assume that the arrivals are described with a Poisson process having average rate $\lambda = 1$ arrival/min. The average service time is $WS = 0.5$ min. Estimate probability of having exactly 3 users in the queue assuming that the queue capacity is infinite.

![Queue diagram]

Death rate $\mu = \frac{1}{WS} = \frac{1}{0.5\text{ min}} = 2 \text{ /min}$

$C_n = \frac{\lambda^n \cdot \lambda^{-1} \cdots \lambda^{-n}}{\mu_n \cdot \mu_{n-1} \cdots \mu_1} = \left(\frac{\lambda}{\mu}\right)^n = \left(\frac{1}{2}\right)^n$

$S = 1 + C_1 + C_2 + \cdots + C_n = 1 + \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \cdots = \frac{1}{1 - \frac{1}{2}} = 2$

Therefore:

$P_0 = 1/\lambda = 1/2$

$P_n = C_n \cdot P_0 = \left(\frac{1}{2}\right)^n \cdot \frac{1}{2}$

For example, the probability of having 3 users in the queue is given by
\[ P_3 = \left(\frac{1}{2}\right)^3 \times \frac{1}{2} = \frac{1}{2^4} = 0.0625 \]

Kendall's notation

Kendall's notation is frequently used to describe a queuing system. The notation has the following form:

A/B/C/K/W/2

- A: distribution of interarrival times
- B: distribution of service times
- C: number of servers in the system
- K: maximum number of users in the system
- W: size of user population
- 2: queuing discipline

* For exponentially distributed service times, the used symbol is M
* For exponentially distributed interarrival times, the used symbol is M

Example. The queuing system described in previous example can be represented using Kendall's notation as:

\[ M/M/1/\infty/\infty/FIFO \]

If the service is \( \infty \), this may be left out of the notation.
If the population is \( \infty \), this may be left out of the notation.
If the discipline is FIFO, this may be left out of the notation.

The shorthand notation for the queuing system in the previous example is given by:

\[ M/M/1 \]
**Erlang B Formula** - lost call cleared.

- Erlang B Formula is used to calculate blocking probability for circuit-switched service in cellular networks.
- The formula is applied on a per-cell basis assuming that each cell may be modeled as a queuing system of a type.

\[ \text{H/M/c/C} \]

H - users are requesting service in a manner described using Poisson process.
H - call holding time is exponentially distributed variable.
\( e \) - number of channels.
\( c \) - maximum number of users in the system. \( c \) large is no queuing in the system.

The calls that are placed and that do not find available channel are rejected and need to be placed again.

State diagram for \( \text{H/M/c/C} \) system is presented as follows.

![State Diagram](image)

Therefore,

\[ C_n = \frac{\lambda_0 \lambda_1 \lambda_2 \cdots \lambda_0}{\mu_0 \mu_1 \mu_2 \cdots \mu_0} = \frac{\lambda^n}{n! \mu} = \frac{\alpha^n}{n!} \quad \text{where} \quad \alpha = \frac{\lambda}{\mu} \]

and

\[ \phi = 1 + \phi_1 + \phi_2 + \cdots + \phi_n = \sum_{k=0}^{n} \frac{\alpha^k}{k!} \]

Probability of finding exactly \( n \) occupied resources is given by.
\[ q_n = \frac{C_0}{S} = \frac{a^n/n!}{\sum_{k=0}^{\infty} \frac{a^k}{k!}} \]

An incoming call is blocked if all of the resources are occupied. That is, the probability of blocked call is given by:

\[ P(\text{blocked call}) = B [a,c] = \frac{a^c/c!}{\sum_{k=0}^{\infty} \frac{a^k}{k!}} \Rightarrow \text{Erlang B formula} \]