Appendix

A. Continuous time Markov chains

For the sake of completeness, we provide an informal discussion on continuous time Markov chains (CTMC). The reader is referred to Resnick (2002) for an in-depth explanation.

A CTMC $\{X(t)\}_{t>0}$ defined on a finite state space S is a stochastic process with the following two properties. Given that the process is in state $i \in S$, the holding time that the process stays in that state is exponentially distributed with parameter $\lambda_i > 0$. Furthermore, when the process leaves state i it transitions to state $j \in S$, $j \neq i$, with probability p_{ij} ($\sum_{j \neq i} p_{ij} = 1 \ \forall i \in S$). We call a state $0 \in S$ absorbing if, once the process enters state 0, it never exits the state and we say that $\lambda_0 = 0$.

A CTMC can be completely characterized by its starting probability vector $\boldsymbol{\tau}$ (with $\boldsymbol{\tau}_i := P(X(0) = i)$) and its infinitesimal generator matrix $\boldsymbol{U}$, defined as

$$\boldsymbol{U}_{ij} = \begin{cases} -\lambda_i & \text{if } i = j \\ \lambda_i p_{ij} & \text{otherwise.} \end{cases}$$

The entry $\boldsymbol{U}_{ij}, i \neq j$, can be interpreted as the rate at which the process transitions from i to j. By construction we have that, $\forall i \in S$, $\sum_{j \in S} \boldsymbol{U}_{ij} = 0$ (or in matrix notation, $\boldsymbol{U} \cdot \mathbf{1} = \mathbf{0}$). The generator matrix is used in different computations involving the CTMC, and in particular we have that $P\left(X(x) = j | X(0) = i\right) = \exp\left(\boldsymbol{U}x\right)_{ij} = \sum_{k=0}^{\infty} \frac{(\boldsymbol{U}x)^k}{k!}$.

For example, consider a CTMC defined on $S = \{1, 2, 0\}$ where the process starts in state 1, then after an exponentially distributed amount of time with rate λ it goes to state 2 and finally, after another exponentially distributed amount of time with rate λ, is absorbed into state 0. In such a process we have $p_{12} = p_{20} = 1$, $\boldsymbol{\tau}_1 = 1$ and the infinitesimal generator is given by:

$$\boldsymbol{U} = \begin{pmatrix} -\lambda & \lambda & 0 \\ 0 & -\lambda & \lambda \\ 0 & 0 & 0 \end{pmatrix}$$

Note that the time γ until absorption in state 0 is Erlang distributed with 2 phases and rate λ. In this case the matrix exponential is:

$$\exp\left(\boldsymbol{U}x\right) = \begin{pmatrix} e^{\lambda x} & \lambda x e^{\lambda x} & 1 - e^{\lambda x} - \lambda x e^{\lambda x} \\ 0 & e^{\lambda x} & 1 - e^{\lambda x} \\ 0 & 0 & 1 \end{pmatrix}$$

The probability that the absorption has occurred by time x is:

$$P\left(\gamma < x\right) = P\left(X(x) = 0 | X(0) = 1\right) = \exp\left(\boldsymbol{U}x\right)_{10} = 1 - \exp\left(\boldsymbol{U}x\right)_{11} - \exp\left(\boldsymbol{U}x\right)_{12} = 1 - e^{\lambda x} - \lambda x e^{\lambda x}$$

corresponding to the cumulative distribution function of the Erlang distribution, as expected.

B. PH properties

The properties of PH distributions exploited in this paper can be stated as (Latouche and Ramaswami 1999):

Property 1. If X is $PH(\boldsymbol{\tau}, \boldsymbol{Z})$ then its moment generating function is

$$E\left[X^\alpha\right] = (-1)^\alpha \alpha! \boldsymbol{\tau} \boldsymbol{Z}^{-\alpha} \mathbf{1}, \tag{1}$$

from which we can deduce its expected value

$$E\left[X\right] = -\boldsymbol{\tau} \boldsymbol{Z}^{-1} \mathbf{1}, \tag{2}$$

and variance

$$Var\left(X\right) = 2 \boldsymbol{\tau} \boldsymbol{Z}^{-2} \mathbf{1} - \left(\boldsymbol{\tau} \boldsymbol{Z}^{-1} \mathbf{1}\right)^2. \tag{3}$$

Its cumulative distribution function (CDF) is

$$F(x) = P(X < x) = 1 - \boldsymbol{\tau} \exp\left(\boldsymbol{Z}x\right) \mathbf{1}, \tag{4}$$

where $\exp\left(\boldsymbol{Z}x\right)$ denotes the matrix exponential, defined as:

$$\exp\left(\boldsymbol{Z}x\right) = \sum_{k=0}^{\infty} \frac{\left(\boldsymbol{Z}x\right)^k}{k!} \tag{5}$$

Several methods exist in the literature for computing the matrix exponential. Moler and Loan (2003) present an overview of the most common methods used and highlight those that provide better results.

PROPERTY 2. If X is $PH(\boldsymbol{\tau_1}, \boldsymbol{Z_1})$ with n phases and Y is $PH(\boldsymbol{\tau_2}, \boldsymbol{Z_2})$ with m phases then the sum $X+Y$ is $PH(\boldsymbol{\tau}, \boldsymbol{Z})$ with $n+m$ phases, such that

$$\boldsymbol{\tau} = [\boldsymbol{\tau_1}, \tau_{1_0} \boldsymbol{\tau_2}], \tag{6}$$

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{Z_1} & \boldsymbol{z_1} \cdot \boldsymbol{\tau_2} \\ \boldsymbol{0} & \boldsymbol{Z_2} \end{pmatrix}, \tag{7}$$

with

$$(\boldsymbol{z} \cdot \boldsymbol{\tau_2})_{ij} = z_i \tau_{2_j} \tag{8}$$
$$\boldsymbol{z} = -\boldsymbol{Z} \cdot \boldsymbol{1} \tag{9}$$

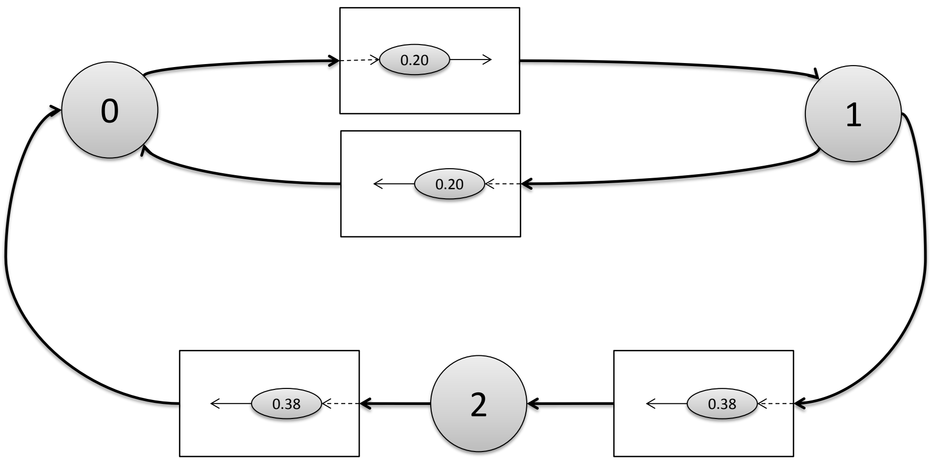
Computing the convolution of PH distribution is thus a trivial task. Furthermore, note that if $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$ are upper triangular matrices, $\boldsymbol{Z}$ is also upper triangular. Therefore, to compute the matrix exponential of such matrices, we will use Taylor's approximation with scaling and squaring, exploiting this structural property to speed up computations.

C. Monotonicity and Extensibility

Splitting procedures, to avoid evaluating a large number of routes, exploit the fact that constraints in $\mathcal{C}$ are monotonic (i.e., adding additional customers to an infeasible route will always yield an infeasible route). For instance, in a deterministic VRP, adding a customer to a route that violates the capacity (distance) constraint will result in a route that violates the capacity (distance) constraint by a larger amount. Consequently, extensions of infeasible routes need not be evaluated.

In a stochastic VRP, however, some constraints might be non-monotonic. Consider the example shown in Figure 1. The travel times $\tilde{t}_{0,1}$ and $\tilde{t}_{1,0}$ are exponentially distributed with rate 0.20, while $\tilde{t}_{1,2}$ and $\tilde{t}_{2,0}$ are exponentially distributed with rate 0.38, and there are no service times. We require that the probability of completing a route before $T = 16.5$ is greater than $\beta = 0.85$. This chance constraint is non-monotonic.

Figure 1 **Extensible route**



For route $r_1 = (0,1,0)$ we have that $E\left[\tilde{T}_{r_1}\right] = 10$ and $P\left(\tilde{T}_{r_1} \leq 16.5\right) = 0.84$, while for route $r_2 = (0,1,2,0)$ we have that $E\left[\tilde{T}_{r_2}\right] = 10.26$ and $P\left(\tilde{T}_{r_2} \leq 16.5\right) = 0.86$. Route r_1 violates the chance constraint, while r_2 does not, despite having a larger expected value. Consequently, extensions of infeasible routes cannot be discarded without evaluation, since such routes might be feasible.

We call a route $r = \left(0, v_{(1)}, \ldots, v_{(i)}, \ldots, v_{(n_r)}, 0\right)$ extensible when an extension of r may be feasible, and non-extensible when we know that all extensions of r are infeasible. In the case of this chance constraint, route r is extensible if and only if the probability of visiting the last customer before $T = 16.5$ is greater than $\beta = 0.85$ (i.e., if the partial route $r' = \left(0, v_{(1)}, \ldots, v_{(i)}, \ldots, v_{(n_r)}\right)$ satisfies the chance constraint). In the example, since the probability of finishing the partial route $r'_1 = (0, 1)$ before the threshold is 0.96, r_1 is extensible and we proceed to evaluate r_2.

References

Latouche, G., V. Ramaswami. 1999. *Introduction to Matrix Analytic Methods in Stochastic Modeling*. SIAM, Philadelphia.

Moler, C., C. Loan. 2003. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review* **45** 3–49.

Resnick, S. 2002. *Adventures in Stochastic Processes*. Birkauser, Boston.