



HAL
open science

Asymptotics for infinite server queues with fast/slow Markov switching and fat tailed service times

Landy Rabehasaina

► **To cite this version:**

Landy Rabehasaina. Asymptotics for infinite server queues with fast/slow Markov switching and fat tailed service times. 2018. hal-01894007v1

HAL Id: hal-01894007

<https://hal.science/hal-01894007v1>

Preprint submitted on 12 Oct 2018 (v1), last revised 18 Jun 2021 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Asymptotics for infinite server queues with fast/slow Markov switching and fat tailed service times

Landy Rabehasaina

*Laboratory of Mathematics, University Bourgogne Franche-Comté,
16 route de Gray, 25030 Besançon cedex, France.*

e-mail: lrabehas@univ-fcomte.fr

Abstract: We study a general k dimensional infinite server queues process. When the service times are fat tailed, we prove that the properly rescaled process converges to some limiting process: in particular we identify three regimes including slow arrivals, fast arrivals, and equilibrium, which lead to different limits in distribution.

AMS 2000 subject classifications: Primary 60G50, 60K30, 62P05, 60K25.

Keywords and phrases: Multivariate discounted rewards, Incurred But Not Reported (IBNR) claims, Infinite server queues, Markov modulation, Rescaled process.

1. Introduction and statement of result

We consider the following k dimensional process $\{Z(t) = (Z_1(t), \dots, Z_k(t)), t \geq 0\}$ defined by

$$Z_j(t) = \sum_{i=1}^{N_t} X_{ij} \mathbb{1}_{[t < L_{ij} + T_i]} = \sum_{i=1}^{\infty} X_{ij} \mathbb{1}_{[T_i \leq t < L_{ij} + T_i]}, \quad j = 1, \dots, k, \quad (1)$$

with $\{N_t, t \geq 0\}$ a Poisson process with intensity $\lambda > 0$, $(T_i)_{i \in \mathbb{N}}$ satisfying $T_0 = 0$, such that $(T_i - T_{i-1})_{i \geq 1}$ is independent and identically distributed (iid) with same distribution $\mathcal{E}(\lambda)$. $(L_{ij})_{i \in \mathbb{N}, j=1, \dots, k}$ is a sequence of independent random variables (rvs) such that $(L_{i1}, \dots, L_{ik})_{i \in \mathbb{N}}$ is iid (although L_{i1}, \dots, L_{ik} may have different distributions). $(X_i)_{i \in \mathbb{N}}$ is a finite Markov chain with state space $\mathcal{S} = \{0, \dots, K\}^k$ for some K and k in \mathbb{N}^* , so that X_i is for all i of the form $X_i = (X_{i1}, \dots, X_{ik})$ with $X_{ij} \in \{0, \dots, K\}$ for $j = 1, \dots, k$.

Process defined by (1) has many applications. It may represent a set of k correlated queues with infinite servers, such that customers arrive at each time T_i , with X_{ij} customers arriving in queue $j \in \{1, \dots, k\}$, with corresponding service time L_{ij} . $Z_j(t)$ could also be seen as the number of customers of class j in a (global) infinite server queue, see [12, Figure 1]; other infinite queues applications may be found in [12, Section 6]. In an actuarial context, $Z_j(t)$ may represent the number of incurred but non reported (IBNR) claims in the j th branch of an insurance company, where X_{ij} is the number of such claims arriving in that branch at time T_i , and L_{ij} is the related delay time before declaring the claim; from another point of view, $X_{ij} \in [0, +\infty)$ may also represent the amount (say, in euros) of the claim occurring at time T_i in the j th branch, in which case $Z_j(t)$ is the total amount of not declared yet claims which have nonetheless occurred by time t . Getting back to the queueing interpretation, and since arrivals are modelled by a Poisson process and $(X_i)_{i \in \mathbb{N}}$ is a Markov chain, the process $\{Z(t), t \geq 0\}$ may thus be seen as (multidimensional) Markov modulated infinite queue process, i.e. a queueing process governed by an external continuous time Markov chain

$\{Y(t), t \geq 0\}$ with state space \mathcal{S} , associated switching times $(T_i)_{i \in \mathbb{N}}$ and embedded Markov chain $(Y(T_i))_{i \in \mathbb{N}} = (X_i)_{i \in \mathbb{N}}$, see [10], [9, Model II]. Such a model was studied in a similar Markov modulated context analytically in [11, 5, 10] for the steady and/or transient state of the queue, but this usually concerns partial information such as the moments, and when the services or interarrivals are matrix exponential distributed. The case of semi-Markov arrivals and (modulated) arrival and service was studied in [5, 6].

Let $P = (p(x, x'))_{(x, x') \in \mathcal{S}^2}$ and $\pi = (\pi(x))_{x \in \mathcal{S}}$ (written as a row vector) be respectively the transition matrix and stationary distribution of the Markov chain. We next define for all $r \geq 0$ and $s = (s_1, \dots, s_k) \in (-\infty, 0]^k$,

$$\tilde{\pi}(s, r) := \text{diag} \left[\mathbb{E} \left(\exp \left\{ \sum_{j=1}^k s_j x_j \mathbb{1}_{[L_j > r]} \right\} \right), x = (x_1, \dots, x_k) \in \mathcal{S} \right], \quad (2)$$

$$\tilde{Q}(s, r) := \tilde{\pi}(s, r)P', \quad (3)$$

$$\Delta_i := \text{diag}[x_i, x = (x_1, \dots, x_k) \in \mathcal{S}], \quad i = 1, \dots, k, \quad (4)$$

where P' denotes the transpose of matrix P . I is the identity matrix, $\mathbf{0}$ is a column vector with zeroes, and $\mathbf{1}$ is a column vector with 1's, of appropriate dimensions. The moment generating function/Laplace Transform (mgf/LT) of the process $\tilde{Z}(t)$ jointly to the state of X_{N_t} given the initial state of X_0 is denoted by

$$\psi(s, t) := \left[\mathbb{E} \left(e^{\langle s, Z(t) \rangle} \mathbb{1}_{[X_{N_t} = y]} \middle| X_0 = x \right) \right]_{(x, y) \in \mathcal{S}^2}, \quad t \geq 0, \quad s = (s_1, \dots, s_k) \in (-\infty, 0]^k \quad (5)$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product on \mathbb{R}^k . Note that X_0 has no direct physical interpretation here, as the claims sizes/customer batches are given by $X_i, i \geq 1$, and is rather introduced for technical purpose. We also remark that (5) is defined for $s_j \leq 0$ because we will only need non positive arguments, although this mgf may in fact be defined for $s_j \in \mathbb{R}$, as one can prove easily that the expectation would then be finite.

The transient distribution of $Z(t)$ for $t \geq 0$ is studied under various assumptions on the interarrival and service distributions in [13]. In the particular case of Poisson arrivals, it is described by the following proposition, which is [13, Proposition 4] with $\delta = 0$.

Proposition 1. *When $\{N_t, t \geq 0\}$ is a Poisson process with intensity $\lambda > 0$, then $\psi(s, t)$ is the unique solution to the first order linear (matrix) differential equation*

$$\partial_t \psi(s, t) = [-\lambda I + \lambda \tilde{Q}(s, t)'] \psi(s, t) = [\lambda(P - I) + \lambda P(\tilde{\pi}(s, t) - I)] \psi(s, t) \quad (6)$$

with the initial condition $\psi(s, 0) = I$.

Unfortunately, such a first order ordinary differential equation does not have an explicit expression in general. In that case, it is appealing to study the process when the intensity of the Poisson process is sped up and the switching rates of the Markov chain are modified, as in [3, 4, 9]. Similarly to those paper, the goal of this paper is thus to study the behaviour of the queue/IBNR process in "extreme conditions" for the arrival rates, transition rates and delays, while trying to maintain minimal assumptions on the distributions. For this we will suppose that the rescaling is performed as follows:

- the arrival rate is multiplied by n^γ for some $\gamma > 0$, denoted by

$$\lambda_n = \lambda n^\gamma,$$

- the transition probabilities $p(x, y)$ are slowed down by dividing them by n^γ when $x \neq y$, x, y in \mathcal{S} , i.e. the new transition matrix is given by

$$P_n = P/n^\gamma + (1 - 1/n^\gamma)I.$$

Such normalizing assumptions imply that, as $n \rightarrow \infty$, one is close to a model where the k queues almost evolve independently (in the sense that arriving customers or claims come in batches with same fixed size) and are correlated only through the sequence $(T_i)_{i \in \mathbb{N}}$, as indeed the transition matrix P_n verifies $P_n \rightarrow I$. The arrival rate λ is however sped up accordingly such that $\lambda_n(P_n - I) = \lambda(P - I)$ i.e. the transition rates between the states of \mathcal{S} are independent from n , which allows for enough dynamics in the model that compensates the fact that P_n tends to I , and yielding no trivial asymptotics in the following as $n \rightarrow \infty$.

The assumptions for the service times/delays distribution are the following. We first suppose that the base model features fat tailed distributed service times with same index $\alpha \in (0, 1)$, i.e. such that $\mathbb{P}(L_j > t) \sim 1/t^\alpha$ as $t \rightarrow \infty$ for all $j = 1, \dots, k$. This kind of distribution (included in the wider class of heavy tailed distributions) mean that the service times are "large". Furthermore, the scaling is such that the service times are divided by n , denoted by

$$L_j^{(n)} = L_j/n.$$

Hence, the situation is the following: the arrivals are sped up by factor n^γ , but this is compensated by the fact that the delay times are diminished with factor n , so that one expects one of the three phenomena to occur at time t for the limiting model: the arrivals occur faster than it takes time for customers to be served and the corresponding queue content $Z^{(n)}(t)$ grows large as $n \rightarrow \infty$, the arrivals occur slower and services are completed fast so that $Z^{(n)}(t)$ tends to 0 as $n \rightarrow \infty$, or an equilibrium is reached. Those three cases will be studied in the forthcoming sections. Some limiting behaviour was studied in [3, 4, 9], where the authors identified three regimes for different scalings in a Markov modulating context, and where the limiting distribution is normally distributed (obtaining a Central Limit Theorem), and depends on t as well as on the different parameters, when the service times have general distribution with finite expectation or are exponentially distributed in [4, 9], or where some precise information on the tail probability of the queue content is given for exponential service times in [3]. Quite a novelty in this paper is that we restrict here the class of distributions to that of fat tailed distributions in order to exhibit (under slightly different scalings) a different behaviour and different limiting distribution which is not gaussian. Also note that the class of fat tailed distributions is interesting in itself as, in actuarial practice, this corresponds to *latent claims*, i.e. very long delays which are incidentally in practice often not observed (as the case $\alpha \in (0, 1)$ corresponds to the L_j 's having infinite expectation), see [7, Section 6.6.1]; This motivates the following results which feature the exponent α as the only information required on those delays. Also note that those service times have infinite mean, which may explain why the limits in distribution in the main Theorem 2 below fall out of the cases studied in [4, Section 4], and are significantly different from the ones in this latter reference. Not only that, but the scaling is rather done in those references [4, 9] on the transition rates of the underlying continuous time Markov chain modulating the arrival and service rates, whereas here these are constant, as we saw that $\lambda_n(P_n - I) = \lambda(P - I)$ is independent of n , and the scaling is rather done on the service times instead. All in all, what is going to be studied hereafter is, when t is fixed in say $[0, 1]$ w.l.o.g., the limiting distribution as $n \rightarrow \infty$ of the $\mathbb{N}^k \times \mathcal{S}$ valued r.v.

$$\left(Z^{(n)}(t), X_{N_t^{(n)}}^{(n)} \right)$$

or of a renormalized version of it in the "fast arriving customers" case. The corresponding joint Laplace Transform is given by

$$\psi^{(n)}(s, t) = \left[\mathbb{E} \left(e^{\langle s, Z^{(n)}(t) \rangle} \mathbb{1}_{\left[X_{N_t^{(n)}}^{(n)} = y \right]} \middle| X_0^{(n)} = x \right) \right]_{(x,y) \in \mathcal{S}^2}, \quad s = (s_1, \dots, s_j) \in (-\infty, 0]^k,$$

where $(X_i^{(n)})_{i \in \mathbb{N}}$ is the underlying Markov chain with generating matrix P_n , stationary distribution π , and $\{N_t^{(n)}, t \geq 0\}$ is a Poisson process representing the arrivals, with scaled intensity λ_n . The core result of the paper concerning the different regimes is given in the following theorem:

Theorem 2. *Let $\beta := 1/(1-\alpha)$ and $t \in [0, 1]$ fixed. Let $\{\mathcal{X}(t) = (\mathcal{X}_1(t), \dots, \mathcal{X}_k(t)), t \in [0, 1]\}$ be a $\{0, \dots, K\}^k$ valued continuous time Markov chain with infinitesimal generating matrix $\lambda(P - I)$ with $\mathcal{X}(0) \sim \pi$, $\{\mathcal{X}^\beta(t) = (\mathcal{X}_1^\beta(t), \dots, \mathcal{X}_k^\beta(t)), t \in [0, 1]\}$ be a $\{0, \dots, K\}^k$ valued continuous time inhomogeneous Markov chain with infinitesimal generating matrix $\beta(1-t)^{\beta-1}\lambda(P - I)$ with $\mathcal{X}^\beta(0) \sim \pi$, and $\{\nu_j^\beta(t), t \in [0, 1]\}$, $j = 1, \dots, k$, be k independent Poisson processes with same intensity $\beta\lambda$, independent from $\{\mathcal{X}^\beta(t), t \in [0, 1]\}$. Then one has one of the three limiting behaviours:*

- **Slow arrivals:** *If $\gamma < \alpha$ then*

$$\mathcal{D} \left(\left(Z^{(n)}(t), X_{N_t^{(n)}}^{(n)} \right) \middle| X_0^{(n)} \right) \rightarrow \mathcal{D} \left((\mathbf{0}, \mathcal{X}(t)) \middle| \mathcal{X}(0) \right) = \delta_{\mathbf{0}} \otimes \mathcal{D}(\mathcal{X}(t) \middle| \mathcal{X}(0)), \quad (7)$$

$n \rightarrow \infty$, where $\mathbf{0} = (0, \dots, 0) \in \mathbb{R}^k$,

- **Fast arrivals:** *If $\gamma > \alpha$ then, as $n \rightarrow \infty$,*

$$\mathcal{D} \left(\left(\frac{Z^{(n)}(t)}{n^{\gamma-\alpha}}, X_{N_t^{(n)}}^{(n)} \right) \middle| X_0^{(n)} \right) \rightarrow \mathcal{D} \left(\left(\beta\lambda \int_{1-t^{1/\beta}}^1 \mathcal{X}^\beta(v) dv, \mathcal{X}^\beta(1) \right) \middle| \mathcal{X}^\beta(1-t^{1/\beta}) \right), \quad (8)$$

- **Equilibrium:** *If $\gamma = \alpha$ then, as $n \rightarrow \infty$,*

$$\begin{aligned} & \mathcal{D} \left(\left(Z^{(n)}(t), X_{N_t^{(n)}}^{(n)} \right) \middle| X_0^{(n)} \right) \\ & \rightarrow \mathcal{D} \left(\left(\left(\int_{1-t^{1/\beta}}^1 \mathcal{X}_j^\beta(v) \nu_j^\beta(dv) \right)_{j=1, \dots, k}, \mathcal{X}^\beta(1) \right) \middle| \mathcal{X}^\beta(1-t^{1/\beta}) \right), \quad (9) \end{aligned}$$

Sections 2, 3 and 4 below are dedicated to the proofs of the convergences in distribution of this theorem in the slow arrivals, fast arrivals and equilibrium cases. As said earlier, we note that the terms in the limits on the righthandside of (7), (8) and (9) feature simple objects (in regards to the complexity of the original model) where the only characteristic parameters needed are λ , P and α . As a concluding remark, we will discuss in Section 5 some computational aspect for the limiting distributions mentioned in those different regimes in Theorem 2 in the particular case when α is a rational number lying in $(0, 1)$.

In all three cases we will repeatedly use the following general lemma, of which proof is not really relevant in this paper and is given in the Appendix.

Lemma 3. Let $(t \in [0, 1] \mapsto A_n(t))_{n \in \mathbb{N}}$ be a sequence of continuous functions with values in $\mathbb{R}^{\mathcal{S} \times \mathcal{S}}$, and let us assume that there exists some continuous function $t \in [0, 1] \mapsto A(t) \in \mathbb{R}^{\mathcal{S} \times \mathcal{S}}$ such that $\int_0^1 \|A_n(v) - A(v)\| dv \rightarrow 0$ as $n \rightarrow \infty$. Let $y \in \mathbb{R}^{\mathcal{S} \times \mathcal{S}}$ and $t \in [0, 1] \mapsto Y_n(t) \in \mathbb{R}^{\mathcal{S} \times \mathcal{S}}$ be the solution to the following differential equation

$$\begin{cases} \frac{d}{dt} Y_n(t) = A_n(t) Y_n(t), & t \in [0, 1], \\ Y_n(0) = y, \end{cases} \quad n \in \mathbb{N}. \quad (10)$$

Then one has $Y_n(t) \rightarrow Y(t)$ uniformly in $t \in [0, 1]$, as $n \rightarrow \infty$, where $t \in [0, 1] \mapsto Y(t) \in \mathbb{R}^{\mathcal{S} \times \mathcal{S}}$ is the solution to the following differential equation

$$\begin{cases} \frac{d}{dt} Y(t) = A(t) Y(t), & t \in [0, 1], \\ Y(0) = y. \end{cases} \quad (11)$$

Finally, let us recall that the differential equation (6) reads here, with the new parameters λ_n, P_n being such that $\lambda_n(P_n - I) = \lambda(P - I)$:

$$\begin{cases} \partial_t \psi^{(n)}(s, t) = [\lambda(P - I) + \lambda n^\gamma P_n(\tilde{\pi}_n(s, t) - I)] \psi^{(n)}(s, t), & t \geq 0, \\ \psi^{(n)}(s, 0) = I, \end{cases} \quad (12)$$

for all $s = (s_1, \dots, s_k) \in (-\infty, 0]^k$, where, from (2), and using the expansion $\prod_{j=1}^k (a_j + 1) = 1 + \sum_{I \subset \{1, \dots, k\}} \prod_{\ell \in I} a_\ell$ for all real or complex numbers a_1, \dots, a_k ,

$$\begin{aligned} \tilde{\pi}_n(s, t) - I &= \text{diag} \left(\prod_{j=1}^k \left((e^{s_j x_j} - 1) \mathbb{P} [L_j^{(n)} > t] + 1 \right) - 1, x = (x_1, \dots, x_k) \in \mathcal{S} \right) \\ &= \text{diag} \left(\sum_{I \subset \{1, \dots, k\}} \prod_{\ell \in I} \left[(e^{s_\ell x_\ell} - 1) \mathbb{P} [L_\ell^{(n)} > t] \right], x = (x_1, \dots, x_k) \in \mathcal{S} \right). \end{aligned} \quad (13)$$

2. Case $\gamma < \alpha$: Slow arriving customers

This case corresponds to slow arriving customers, compared to short service times. Intuitively, one guesses that the normalized queue content does tend to 0. The following transformation will also be useful: Since one wants to study $\left(Z^{(n)}(t), X_{N_t^{(n)}}^{(n)} \right)$ when n is large, and since t is fixed say in $[0, 1]$, one may as well study $\left(Z^{(n)}(t^\beta), X_{N_{t^\beta}^{(n)}}^{(n)} \right)$ where $\beta := 1/(1 - \alpha) > 1$, as indeed $t \mapsto t^\beta$ is a bijective function mapping $[0, 1]$ to $[0, 1]$. Using this transformation will turn out to be useful here as well as in the following sections so as to avoid pathological cases in the forthcoming involved differential equations, and in order to properly apply Lemma 3. Let $s = (s_1, \dots, s_k) \in (-\infty, 0]^k$ be fixed. The corresponding joint mgf/Laplace transform is given by $\chi^{(n)}(s, t) := \psi^{(n)}(s, t^\beta)$ and satisfies, from (12):

$$\begin{cases} \partial_t \chi^{(n)}(s, t) = \beta t^{\beta-1} [\lambda(P - I) + \lambda n^\gamma P_n(\tilde{\pi}_n(s, t^\beta) - I)] \chi^{(n)}(s, t), & t \in [0, 1], \\ \chi^{(n)}(s, 0) = I. \end{cases} \quad (14)$$

Since the L_j 's are fat tailed, one has that $\mathbb{P}[L_\ell^{(n)} > t^\beta] = \mathbb{P}[L_\ell > nt^\beta] \sim \frac{1}{n^\alpha} \frac{1}{t^{\alpha\beta}}$, $n \rightarrow \infty$, when $t > 0$, $\ell = 1, \dots, k$. Since $\gamma < \alpha$ implies that $1/n^\alpha = o(1/n^\gamma)$ as $n \rightarrow \infty$, we deduce that

$$\beta t^{\beta-1} \prod_{\ell \in I} \left[(e^{s_\ell x_\ell} - 1) \mathbb{P}[L_\ell^{(n)} > t] \right] = \begin{cases} 0 & \text{if } t = 0, \\ o(1/n^\gamma) & \text{if } t > 0 \end{cases}$$

for all $x = (x_1, \dots, x_k) \in \mathcal{S}$ and $I \subset \{1, \dots, k\}$. Then, from (13), we deduce that $\beta t^{\beta-1} n^\gamma (\tilde{\pi}_n(s, t) - I) \rightarrow 0$ as $n \rightarrow \infty$ for all $t \in [0, 1]$, hence $A_n(t) := \beta t^{\beta-1} [\lambda(P - I) + \lambda n^\gamma P_n(\tilde{\pi}_n(s, t^\beta)) - I]$ converges to $A(t) := \beta t^{\beta-1} \lambda(P - I)$ for all $t \in [0, 1]$ (Remember that $\lim_{n \rightarrow \infty} P_n = I$). We now want to use Lemma 3 and prove that those functions satisfy $\int_0^1 \|A_n(v) - A(v)\| dv \rightarrow 0$ as $n \rightarrow \infty$, i.e. that

$$\int_0^1 \beta v^{\beta-1} n^\gamma \|\tilde{\pi}_n(s, v^\beta) - I\| dv \rightarrow 0, \quad n \rightarrow +\infty \quad (15)$$

(again because $\lim_{n \rightarrow \infty} P_n = I$). In view of (13), it suffices to prove that for all $x = (x_1, \dots, x_k) \in \mathcal{S}$ and all $I \subset \{1, \dots, k\}$, one has $\int_0^1 \beta v^{\beta-1} n^\gamma \left| \prod_{\ell \in I} \left[(e^{s_\ell x_\ell} - 1) \mathbb{P}[L_\ell^{(n)} > v^\beta] \right] \right| dv \rightarrow 0$. However, one just needs to prove this convergence when $\text{Card}(I) = 1$, that is $I = \{\ell\}$ for all $\ell = 1, \dots, k$, as indeed one checks that the convergence for $\text{Card}(I) \geq 2$ can be proved similarly, and actually converges faster than the case $\text{Card}(I) = 1$ towards 0. So, when $I = \{\ell\}$, $\ell = 1, \dots, k$, the change of variable $u = nv^\beta$ yields

$$\begin{aligned} \int_0^1 \beta v^{\beta-1} n^\gamma \left| (e^{s_\ell x_\ell} - 1) \mathbb{P}[L_\ell^{(n)} > v^\beta] \right| dv &= \int_0^1 \beta v^{\beta-1} n^\gamma \left| (e^{s_\ell x_\ell} - 1) \mathbb{P}[L_\ell > nv^\beta] \right| dv \\ &= |e^{s_\ell x_\ell} - 1| n^{\gamma-1} \int_0^1 \mathbb{P}[L_\ell > u] du \end{aligned}$$

which tends to 0 as $n \rightarrow \infty$, as indeed $\gamma < \alpha < 1 \implies \lim_{n \rightarrow \infty} n^{\gamma-1} = 0$.

All in all, one has then that $\lim_{n \rightarrow \infty} \int_0^1 \|A_n(v) - A(v)\| dv = 0$. Lemma 3 thus implies that $\chi^{(n)}(s, t)$ defined in (14) converges as $n \rightarrow \infty$ towards $\chi(s, t)$ that satisfies

$$\begin{cases} \partial_t \chi(s, t) &= \beta t^{\beta-1} \lambda(P - I) \chi(s, t), \quad t \in [0, 1], \\ \chi(s, 0) &= I. \end{cases} \quad (16)$$

This matrix linear differential equation admits the explicit solution $\chi(s, t) = \exp(t^\beta \lambda(P - I))$, $t \in [0, 1]$, which is the Laplace Transform of the random variable $(0, \dots, 0) \in \mathbb{R}^k$ jointly to $\mathcal{X}(t^\beta)$, where $\{\mathcal{X}(t), t \geq 0\}$ is the continuous time Markov chain defined in Theorem 2. Hence, the distribution of $\left(Z^{(n)}(t^\beta), X_{N_{t^\beta}^{(n)}}^{(n)} \right)$ given $X_0^{(n)}$ converges towards $\mathcal{D}(((0, \dots, 0), \mathcal{X}_{t^\beta}) | X_0)$. Changing t into $t^{1/\beta}$ yields (7).

3. Case $\gamma > \alpha$: Fast arriving customers

We now proceed to show convergence (8). Since we saw in the previous section that the queue content converged towards 0 in the case of slow arriving customers, it is sensible to guess that here it will converge towards infinity, hence it is natural to find a normalization such that a convergence towards a proper distribution occurs. As shown in (8), we renormalize the

queue content by dividing it by $n^{\gamma-\alpha}$, i.e. we are here interested in $\left(Z^{(n)}(t)/n^{\gamma-\alpha}, X_{N_t^{(n)}}^{(n)}\right)$. Its corresponding mgf/Laplace transform is given by $\psi^{(n)}(s/n^{\gamma-\alpha}, t)$, $s = (s_1, \dots, s_k) \in (-\infty, 0]^k$. As in the proof in the slow arrival case, thanks to $t \in [0, 1] \mapsto t^\beta \in [0, 1]$ being a one to one mapping, we are going to study $\left(Z^{(n)}(t^\beta)/n^{\gamma-\alpha}, X_{N_{t^\beta}^{(n)}}^{(n)}\right)$, of which mgf is given by $\chi^{(n)}(s, t) := \psi^{(n)}(s/n^{\gamma-\alpha}, t^\beta)$. From (12), $\chi^{(n)}(s, t)$ satisfies

$$\begin{cases} \partial_t \chi^{(n)}(s, t) &= \beta t^{\beta-1} [\lambda(P - I) + \lambda n^\gamma P_n(\tilde{\pi}_n(s/n^{\gamma-\alpha}, t^\beta)) - I] \chi^{(n)}(s, t), \quad t \in [0, 1], \\ \chi^{(n)}(s, 0) &= I. \end{cases} \quad (17)$$

The starting point is similar to that of Section 2: we will set to prove that

$$A_n(s, t) = \beta t^{\beta-1} [\lambda(P - I) + \lambda n^\gamma P_n(\tilde{\pi}_n(s/n^{\gamma-\alpha}, t^\beta)) - I]$$

converges to some limit $A(s, t)$ as $n \rightarrow \infty$, use Lemma 3, then identify the limit $\chi(s, t) := \lim_{n \rightarrow \infty} \chi^{(n)}(s, t)$ as the Laplace Transform of a known distribution.

Step 1: Finding $A(s, t)$. Let us dedicate to finding the limit function $t \in [0, 1] \mapsto A(s, t)$. In view of (13), the x th diagonal element of $\beta t^{\beta-1} \lambda n^\gamma (\tilde{\pi}_n(s/n^{\gamma-\alpha}, t^\beta) - I)$ is

$$\beta t^{\beta-1} \lambda n^\gamma \sum_{I \subset \{1, \dots, k\}} \prod_{\ell \in I} \left[(e^{s_\ell x_\ell / n^{\gamma-\alpha}} - 1) \mathbb{P} \left[L_\ell^{(n)} > t^\beta \right] \right] \quad (18)$$

of which we proceed to find the limit as $n \rightarrow \infty$. Contrarily to what happened in Section 2, we are going to isolate the terms in the sum for which $\text{Card}(I) = 1$ and $\text{Card}(I) \geq 2$, and show that the former admit a non zero limit and the latter tend to 0. We thus write (18) as

$$\beta t^{\beta-1} \lambda n^\gamma \sum_{I \subset \{1, \dots, k\}} \prod_{\ell \in I} \left[(e^{s_\ell x_\ell / n^{\gamma-\alpha}} - 1) \mathbb{P} \left[L_\ell^{(n)} > t^\beta \right] \right] = J_n^1(s, t) + J_n^2(s, t), \quad \text{where}$$

$$J_n^1(s, t) = J_n^1(s, t, x) := \beta t^{\beta-1} \lambda n^\gamma \sum_{\ell=1}^k (e^{s_\ell x_\ell / n^{\gamma-\alpha}} - 1) \mathbb{P} \left[L_\ell^{(n)} > t^\beta \right], \quad (19)$$

$$J_n^2(s, t) = J_n^2(s, t, x) := \beta t^{\beta-1} \lambda n^\gamma \sum_{\text{Card}(I) \geq 2} \prod_{\ell \in I} \left[(e^{s_\ell x_\ell / n^{\gamma-\alpha}} - 1) \mathbb{P} \left[L_\ell^{(n)} > t^\beta \right] \right]. \quad (20)$$

Both terms $J_n^1(s, t)$ and $J_n^2(s, t)$ are studied separately. Using that $e^{s_\ell x_\ell / n^{\gamma-\alpha}} - 1 \sim s_\ell x_\ell / n^{\gamma-\alpha}$ as $n \rightarrow \infty$ and $\mathbb{P} \left[L_\ell^{(n)} > t^\beta \right] = \mathbb{P} \left[L_\ell > n t^\beta \right] \sim \frac{1}{n^\alpha t^{\beta\alpha}}$ when $t > 0$, and since $\beta\alpha = \alpha/(1-\alpha) = \beta - 1$, we arrive at

$$J_n^1(s, t) \sim \beta \lambda \sum_{\ell=1}^k t^{\beta-1} n^\gamma \frac{s_\ell x_\ell}{n^{\gamma-\alpha}} \frac{1}{n^\alpha t^{\beta\alpha}} \sim \beta \lambda \sum_{\ell=1}^k s_\ell x_\ell, \quad n \rightarrow \infty,$$

when $t > 0$, and is 0 when $t = 0$. Using similar estimates, and exploiting the fact that the sum is over $\text{Card}(I) \geq 2$, one proves easily that $J_n^2(s, t) = O(1/n^\gamma)$ for all $t \geq 0$, which thus tends to 0 as $n \rightarrow \infty$. Hence we have that (18) tends to $\beta \lambda \sum_{\ell=1}^k s_\ell x_\ell$ when $t \in (0, 1]$, and to 0 when $t = 0$. The candidate for the continuous function $A(s, t)$ is then

$$t \in [0, 1] \mapsto A(s, t) := \beta t^{\beta-1} \lambda (P - I) + \beta \lambda \sum_{\ell=1}^k s_\ell \Delta_\ell \quad (21)$$

where we recall that $\Delta_\ell = \text{diag}[x_\ell, x = (x_1, \dots, x_k) \in \mathcal{S}]$. Note that the limit when $t = 0$ differs from what was expected, as indeed a closer look from the study of the limits of $J_n^1(s, t)$ and $J_n^2(s, t)$ would yield that the limiting function for $t = 0$ should rather be the 0 matrix, and not $\beta\lambda \sum_{\ell=1}^k s_\ell \Delta_\ell$ which is $A(s, 0)$ defined above. This is due to the fact that one requires a *continuous* function $t \in [0, 1] \mapsto A(s, t)$ in Lemma 3.

Step 2: Finding $\chi(s, t)$. So, in order to apply this Lemma we now need to prove that $\int_0^1 \|A_n(v) - A(v)\| dv \rightarrow 0$ as $n \rightarrow \infty$. Since $A_n(v) = A(v) + P_n \text{diag}(J_n^1(s, v, x) - \beta\lambda \sum_{\ell=1}^k s_\ell x_\ell, x \in \mathcal{S}) + P_n \text{diag}(J_n^2(s, v, x), x \in \mathcal{S})$ and since $\lim_{n \rightarrow \infty} P_n = I$, this thus amounts to prove that $\int_0^1 |J_n^1(s, v, x) - \beta\lambda \sum_{\ell=1}^k s_\ell x_\ell| dv$ and $\int_0^1 |J_n^2(s, v, x)| dv = \int_0^1 J_n^2(s, v, x) dv$ tend to 0 as $n \rightarrow \infty$, for each fixed $x \in \mathcal{S}$. Let us first focus on $\int_0^1 |J_n^1(s, v, x) - \beta\lambda \sum_{\ell=1}^k s_\ell x_\ell| dv$. We have

$$\begin{aligned} \int_0^1 \left| J_n^1(s, v, x) - \beta\lambda \sum_{\ell=1}^k s_\ell x_\ell \right| dv &\leq \sum_{\ell=1}^k (I_n^1(\ell) + I_n^2(\ell)), \text{ where, for all } \ell = 1, \dots, k, \quad (22) \\ I_n^1(\ell) &:= \int_0^1 \lambda \beta v^{\beta-1} \left| n^\gamma (e^{s_\ell x_\ell / n^{\gamma-\alpha}} - 1) - n^\alpha s_\ell x_\ell \right| \mathbb{P} \left[L_\ell^{(n)} > v^\beta \right] dv \\ I_n^2(\ell) &:= |s_\ell x_\ell| \int_0^1 \lambda \left| \beta v^{\beta-1} n^\alpha \mathbb{P} \left[L_\ell^{(n)} > v^\beta \right] - \beta \right| dv. \end{aligned}$$

Furthermore, one has, expanding the exponential function, that $|e^{s_\ell x_\ell / n^{\gamma-\alpha}} - 1 - s_\ell x_\ell / n^{\gamma-\alpha}| \leq M/n^{2(\gamma-\alpha)}$ for some constant M independent from everything. Thus, one deduces the following upper bounds for $I_n^1(\ell)$, $\ell = 1, \dots, k$:

$$\begin{aligned} I_n^1(\ell) &= \int_0^1 \lambda \beta v^{\beta-1} \left| n^\gamma (e^{s_\ell x_\ell / n^{\gamma-\alpha}} - 1) - n^\alpha s_\ell x_\ell \right| \mathbb{P} \left[L_\ell > nv^\beta \right] dv \\ &\leq n^\gamma \frac{M}{n^{2(\gamma-\alpha)}} \lambda \int_0^1 \beta v^{\beta-1} \mathbb{P} \left[L_\ell > nv^\beta \right] dv = \frac{M}{n^{\gamma-\alpha}} \beta \lambda \int_0^1 n^\alpha v^{\beta-1} \mathbb{P} \left[L_\ell > nv^\beta \right] dv \\ &= \frac{M}{n^{\gamma-\alpha}} \beta \lambda \int_0^1 (nv^\beta)^\alpha \mathbb{P} \left[L_\ell > nv^\beta \right] dv, \quad (23) \end{aligned}$$

the last equality holding because $\beta - 1 = \beta\alpha$. A consequence of the fact that L_ℓ is fat tailed with index α is that $\sup_{u \geq 0} u^\alpha \mathbb{P}(L_\ell > u) < +\infty$, from which one deduces immediately that $\sup_{j \in \mathbb{N}, v \in [0,1]} (jv^\beta)^\alpha \mathbb{P} \left[L_\ell > jv^\beta \right] < +\infty$ (those suprema being in fact equal). One then gets from (23) that

$$I_n^1(\ell) \leq \frac{M}{n^{\gamma-\alpha}} \beta \lambda \sup_{j \in \mathbb{N}, v \in [0,1]} (jv^\beta)^\alpha \mathbb{P} \left[L_\ell > jv^\beta \right] \rightarrow 0, \quad n \rightarrow \infty. \quad (24)$$

We now turn to $I_n^2(\ell)$, $\ell = 1, \dots, k$. Using again $\beta - 1 = \beta\alpha$, one way write

$$I_n^2(\ell) = |s_\ell x_\ell| \int_0^1 \lambda \left| \beta (nv^\beta)^\alpha \mathbb{P} \left[L_\ell > nv^\beta \right] - \beta \right| dv.$$

Similar estimates yield that

$$\sup_{n \in \mathbb{N}, v \in [0,1]} \left| \beta (nv^\beta)^\alpha \mathbb{P} \left[L_\ell > nv^\beta \right] - \beta \right| < +\infty.$$

Furthermore, since L_ℓ is fat tailed with index α , one has $\mathbb{P}[L_\ell > nv^\beta] \sim 1/(nv^\beta)^\alpha$ as $n \rightarrow \infty$ when $v > 0$, hence $|\beta(nv^\beta)^\alpha \mathbb{P}[L_\ell > nv^\beta] - \beta| \rightarrow 0$ as $n \rightarrow \infty$ when $v \in (0, 1]$, and is equal to β when $v = 0$. The dominated convergence theorem thus implies that

$$I_n^2(\ell) \rightarrow 0, \quad n \rightarrow \infty. \quad (25)$$

Gathering (22), (24) and (25), we thus deduce finally that $\int_0^1 |J_n^1(s, v, x) - \beta \lambda \sum_{\ell=1}^k s_\ell x_\ell| dv$ tends to 0 as $n \rightarrow \infty$ for each $x \in \mathcal{S}$. As to $\int_0^1 J_n^2(s, v, x) dv$, the fact that we are summing in (20) over $\text{Card}(I) \geq 2$, and using again the upper bound $\sup_{j \in \mathbb{N}, v \in [0, 1]} (jv^\beta)^\alpha \mathbb{P}[L_\ell > jv^\beta] < +\infty$, yields easily by the dominated convergence theorem that $\lim_{n \rightarrow \infty} \int_0^1 J_n^2(s, v, x) dv = 0$. Hence we just proved that $\int_0^1 \|A_n(v) - A(v)\| dv \rightarrow 0$, and we may then use Lemma 3 to deduce that $\chi^{(n)}(s, t)$ converges to $\chi(s, t)$ which satisfies

$$\begin{cases} \partial_t \chi(s, t) = A(s, t) \chi(s, t) = [\beta t^{\beta-1} \lambda (P - I) + \beta \lambda \sum_{\ell=1}^k s_\ell \Delta_\ell] \chi(s, t), & t \in [0, 1], \\ \chi(s, 0) = I. \end{cases} \quad (26)$$

Step 3: Identifying the limit in distribution. Contrarily to the differential equation (16) in the Slow arrival case, (26) does not admit an explicit expression. However, since we purposely chose $s = (s_1, \dots, s_k)$ with $s_j \leq 0$, $j = 1, \dots, k$, one has that $\sum_{j=1}^k s_j \Delta_j = \sum_{j=1}^k s_j \text{diag}(x_j, x \in \mathcal{S})$ is a diagonal matrix with non positive entries. Let $\Delta_\pi := \text{diag}(\pi(x), x \in \mathcal{S})$ and let us introduce the matrix $P^{(r)}$ defined by $P^{(r)} = \Delta_\pi^{-1} P' \Delta_\pi \iff P = \Delta_\pi^{-1} P^{(r)} \Delta_\pi$. It is standard that the $P^{(r)}$ is the transition matrix of the reversed version of the stationary Markov chain $\{X_i, i \in \mathbb{N}\}$ with distribution π , and that $\beta t^{\beta-1} \lambda (P^{(r)} - I)$ is the infinitesimal generator matrix of an inhomogeneous Markov process $\{U(t) = (U_j(t))_{j=1, \dots, k} \in \mathcal{S}, t \in [0, 1]\}$ with values in \mathcal{S} , with initial distribution $U(0) \sim \pi$. In fact, it turns out that the conditional distribution of $U(t)$ given $U(0)$ is given by $[\mathbb{P}(U(t) = y | U(0) = x)]_{(x, y) \in \mathcal{S}} = \exp(t^\beta \lambda (P^{(r)} - I))$, which results in $U(t) \sim \pi$ for all $t \in [0, 1]$. Since $\sum_{j=1}^k s_j \Delta_j$ is diagonal, one checks easily that $A(s, t) = \Delta_\pi^{-1} [\beta t^{\beta-1} \lambda (P^{(r)} - I) + \sum_{j=1}^k s_j \Delta_j] \Delta_\pi$ and that $Y(t) = Y(s, t) := \Delta_\pi^{-1} \chi(s, t)' \Delta_\pi$ satisfies the differential equation

$$\begin{cases} \partial_t Y(t) = Y(t) [\beta t^{\beta-1} \lambda (P^{(r)} - I) + \beta \lambda \sum_{\ell=1}^k s_\ell \Delta_\ell], & t \in [0, 1], \\ Y(0) = I. \end{cases}$$

The Feynman-Kac formula ensures that one has the representation

$$Y(t) = Y(s, t) = \left[\mathbb{E} \left[\mathbf{1}_{[U(t)=y]} \exp \left(\sum_{j=1}^k s_j \beta \lambda \int_0^t U_j(v) dv \right) \middle| U(0) = x \right] \right]_{(x, y) \in \mathcal{S}^2}, \quad \forall t \in [0, 1],$$

see [14, Chapter III, 19, p.272] for the general theorem on this formula, or [2, Section 5, Expression (5.2) and differential equation (5.3)] for the particular case of a finite Markov chain, adapted here to an inhomogeneous Markov process. Also, the process $\{\mathcal{X}^\beta(t) = (\mathcal{X}_1^\beta(t), \dots, \mathcal{X}_k^\beta(t)) \in \mathcal{S}, t \in [0, 1]\}$ introduced in the theorem statement is the reversed version of $\{U(t), t \in [0, 1]\}$, i.e. is such that $\{\mathcal{X}^\beta(t), t \in [0, 1]\} \stackrel{\mathcal{D}}{=} \{U(1-t), t \in [0, 1]\}$ pathwise. Hence, one obtains for

all x and y in \mathcal{S} that

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1}_{[U(t)=y]} \exp \left(\sum_{j=1}^k s_j \beta \lambda \int_0^t U_j(v) dv \right) \middle| U(0) = x \right] \\ &= \mathbb{E} \left[\mathbf{1}_{[\mathcal{X}^\beta(1-t)=y]} \exp \left(\sum_{j=1}^k s_j \beta \lambda \int_{1-t}^1 \mathcal{X}_j^\beta(v) dv \right) \middle| \mathcal{X}^\beta(1) = x \right] \\ &= \mathbb{E} \left[\mathbf{1}_{[\mathcal{X}^\beta(1)=x]} \exp \left(\sum_{j=1}^k s_j \beta \lambda \int_{1-t}^1 \mathcal{X}_j^\beta(v) dv \right) \middle| \mathcal{X}^\beta(1-t) = y \right] \frac{\pi(y)}{\pi(x)}, \end{aligned}$$

the last line coming from the fact that $U(0)$, $U(t)$, $\mathcal{X}^\beta(1-t)$ and $\mathcal{X}^\beta(1)$ all have same distribution π . Switching the role of x and y above results in the following relationship:

$$\begin{aligned} & \left[\mathbb{E} \left[\mathbf{1}_{[\mathcal{X}^\beta(1)=y]} \exp \left(\sum_{j=1}^k s_j \beta \lambda \int_{1-t}^1 \mathcal{X}_j^\beta(v) dv \right) \middle| \mathcal{X}^\beta(1-t) = x \right] \right]_{(x,y) \in \mathcal{S}^2} \\ &= \left[\mathbb{E} \left[\mathbf{1}_{[U(t)=x]} \exp \left(\sum_{j=1}^k s_j \beta \lambda \int_0^t U_j(v) dv \right) \middle| U(0) = y \right] \frac{\pi(y)}{\pi(x)} \right]_{(x,y) \in \mathcal{S}^2} \\ &= \Delta_\pi^{-1} Y(t)' \Delta_\pi = \chi(s, t). \end{aligned}$$

Since we just proved that $\chi^{(n)}(s, t) := \psi^{(n)}(s/n^{\gamma-\alpha}, t^\beta)$ converges as $n \rightarrow \infty$ towards $\chi(s, t)$, expressed above, for all $s = (s_1, \dots, s_k) \in (-\infty, 0]^k$, and identifying Laplace transforms, we obtained in conclusion that

$$\mathcal{D} \left(\left(Z^{(n)}(t^\beta)/n^{\gamma-\alpha}, X_{N_t^{(n)}}^{(n)} \right) \middle| X_0^{(n)} \right) \rightarrow \mathcal{D} \left(\left(\beta \lambda \int_{1-t}^1 \mathcal{X}^\beta(v) dv, \mathcal{X}^\beta(1) \right) \middle| \mathcal{X}^\beta(1-t) \right)$$

as $n \rightarrow \infty$ for all $t \in [0, 1]$. Changing t into $t^{1/\beta}$ yields (8).

4. Equilibrium case $\gamma = \alpha$

We now proceed to show convergence (9). Intuitively, we are in the critical case where customers should arrive just fast enough such the queue at time t converges as $n \rightarrow \infty$. We are here interested in the behaviour of $\mathcal{D} \left(\left(Z^{(n)}(t), X_{N_t^{(n)}}^{(n)} \right) \middle| X_0^{(n)} \right)$ as $n \rightarrow \infty$ when $t \in [0, 1]$ is fixed. As in Section 2, we first consider t^β instead of t and let $\chi^{(n)}(s, t) := \psi^{(n)}(s, t^\beta)$ the corresponding mgf/Laplace transform, where $s = (s_1, \dots, s_k) \in (-\infty, 0]^k$. The differential equation (14) still holds with $\gamma = \alpha$. The present case has the same roadmap as Subsection 3: We will study the behaviour as $n \rightarrow \infty$ of $\lambda n^\gamma (\tilde{\pi}_n(s, t^\beta)) - I$ in order to obtain a limit as $n \rightarrow \infty$ of

$$A_n(s, t) = \beta t^{\beta-1} [\lambda(P - I) + \lambda n^\gamma P_n(\tilde{\pi}_n(s, t^\beta)) - I]$$

then getting a limiting matrix differential equation for a candidate $\chi(s, t) = \lim_{n \rightarrow \infty} \chi^{(n)}(s, t)$. Then we will identify $\chi(s, t)$ as the Laplace transform of a (conditional) distribution, yielding (9).

Step 1: Finding $A(s, t) = \lim_{n \rightarrow \infty} A_n(s, t)$. We recall that the (x, x) th diagonal element of $\lambda n^\gamma (\tilde{\pi}_n(s, t^\beta) - I)$ is (from (13)) $\sum_{I \subset \{1, \dots, k\}} \prod_{\ell \in I} [(e^{s_\ell x_\ell} - 1) \mathbb{P}[L_\ell > nt]]$, which we decompose as in Section 3 as $K_n^1(s, t) + K_n^2(s, t)$ with

$$K_n^1(s, t) = K_n^1(s, t, x) := \beta t^{\beta-1} \lambda n^\gamma \sum_{\ell=1}^k (e^{s_\ell x_\ell} - 1) \mathbb{P}[L_\ell > nt^\beta], \quad (27)$$

$$K_n^2(s, t) = K_n^2(s, t, x) := \beta t^{\beta-1} \lambda n^\gamma \sum_{\text{Card}(I) \geq 2} \prod_{\ell \in I} [(e^{s_\ell x_\ell} - 1) \mathbb{P}[L_\ell > nt^\beta]]. \quad (28)$$

Using that $\mathbb{P}[L_\ell > nt^\beta] \sim \frac{1}{n^\alpha} \frac{1}{t^{\alpha\beta}}$, $n \rightarrow \infty$, when $t > 0$, and since $\alpha\beta = \beta - 1$, and $\gamma = \alpha$, one here finds that

$$K_n^1(s, t) = K_n^1(s, t, x) \longrightarrow \begin{cases} \beta \lambda \sum_{\ell=1}^k (e^{s_\ell x_\ell} - 1), & t > 0, \\ 0, & t = 0, \end{cases} \quad n \rightarrow \infty.$$

As to $K_n^2(s, t)$, one proves easily that it tends to 0 as $n \rightarrow \infty$ for all $t \in [0, 1]$, as the sum in (28) is over $\text{Card}(I) \geq 2$. The candidate for the continuous function is thus

$$t \in [0, 1] \mapsto A(s, t) := \beta t^{\beta-1} \lambda (P - I) + \beta \lambda \sum_{\ell=1}^k \text{diag}(e^{s_\ell x_\ell} - 1, x = (x_1, \dots, x_k) \in \mathcal{S}). \quad (29)$$

Step 2: Finding $\chi(s, t)$. We now wish to apply Lemma 3 and prove that $\int_0^1 \|A_n(v) - A(v)\| dv \rightarrow 0$ where $A(s, t)$ is defined in (29). The method is very similar as in Step 2 of Section 3, as this is equivalent to proving for all $x \in \mathcal{S}$ that $\int_0^1 |K_n^1(s, v, x) - \beta \lambda \sum_{\ell=1}^k (e^{s_\ell x_\ell} - 1)| dv$ and $\int_0^1 K_n^2(s, v, x) dv$ tend to 0 as $n \rightarrow \infty$. This latter limit is proved the same way as for proving that $\lim_{n \rightarrow \infty} \int_0^1 J_n^2(s, v, x) dv = 0$ in Section 3. As to the former limit, this is proved by showing that $\int_0^1 \lambda \left| \beta v^{\beta-1} n^\alpha \mathbb{P}[L_\ell^{(n)} > v^\beta] - \beta \right| dv$ tends to 0 as $n \rightarrow \infty$ for all $\ell = 1, \dots, k$, as again we have that $\gamma = \alpha$; however this was already proved in Step 2 of Section 3 when proving that $\lim_{n \rightarrow \infty} I_n^1(\ell) = 0$, $\ell = 1, \dots, k$. All in all, Lemma 3 is applicable and we get that $\chi^{(n)}(s, t)$ converges to $\chi(s, t)$ which satisfies

$$\begin{cases} \partial_t \chi(s, t) = A(s, t) \chi(s, t) = [\beta t^{\beta-1} \lambda (P - I) \\ \quad + \beta \lambda \sum_{\ell=1}^k \text{diag}(e^{s_\ell x_\ell} - 1, x = (x_1, \dots, x_k) \in \mathcal{S})] \chi(s, t), & t \in [0, 1], \\ \chi(s, 0) = I. \end{cases} \quad (30)$$

Step 3: Identifying the limit in distribution. With the same notation as in Step 3 of Section 3 for process $\{\mathcal{X}^\beta(t) = (\mathcal{X}_1^\beta(t), \dots, \mathcal{X}_k^\beta(t)) \in \mathcal{S}, t \in [0, 1]\}$, one finds this time that

$$\chi(s, t) = \left[\mathbb{E} \left[\mathbf{1}_{[\mathcal{X}^\beta(1)=y]} \exp \left(\sum_{j=1}^k \beta \lambda \int_{1-t}^1 (e^{s_j \mathcal{X}_j^\beta(v)} - 1) dv \right) \middle| \mathcal{X}^\beta(1-t) = x \right] \right]_{(x,y) \in \mathcal{S}^2} \quad (31)$$

for all $s = (s_1, \dots, s_k) \in (-\infty, 0]^k$. We recall the Campbell formula which states that for all measurable function $f : t \in [0, +\infty) \mapsto f(t) \in \mathbb{R}$ such that $\int_0^\infty (e^{f(v)} - 1) \xi dv$ is finite for some $\xi > 0$ then one has the identity

$$\int_0^\infty (e^{f(v)} - 1) \xi dv = \mathbb{E} \left[\exp \left(\int_0^\infty f(v) \nu(dv) \right) \right],$$

where $\{\nu(x), x \geq 0\}$ is a Poisson process with intensity ξ , see [8, Section 3.2]. The results in (31) being written as

$$\chi(s, t) = \left[\mathbb{E} \left[\mathbf{1}_{[\mathcal{X}^\beta(1)=y]} \exp \left(\sum_{j=1}^k s_j \int_{1-t}^1 \mathcal{X}_j^\beta(v) \nu_j^\beta(dv) \right) \middle| \mathcal{X}^\beta(1-t) = x \right] \right]_{(x,y) \in \mathcal{S}^2}$$

where $\{\nu_j^\beta(t), t \geq 0\}$, $j = 1, \dots, k$, are independent Poisson processes with intensities $\beta\lambda$, and independent from $\{\mathcal{X}^\beta(t), t \in [0, 1]\}$. Changing t into $t^{1/\beta}$ completes the proof of (9).

5. A remark on the computation of the limiting joint Laplace transform when $\alpha \in \mathbb{Q}$

We identified in Theorem 2 the different limiting regimes when γ is less, larger or equal to α by obtaining the corresponding limiting joint Laplace transform $\chi(s, t)$ in each case. Even though the dsitributional limits (8) and (9) involve simple processes $\{\mathcal{X}^\beta(t), t \in [0, 1]\}$ and $\{\nu_j^\beta(t), t \geq 0\}$, $j = 1, \dots, k$, it turns out that the Laplace transforms $\chi(s, t)$, which are solutions to the differential equations (26) and (30), are in general not explicit in the fast or equilibrium arriving cases. We suggest to show that things are much simpler when $\alpha \in (0, 1)$ is rational, say of the form

$$\alpha = 1 - p/q$$

for some p and $q \in \mathbb{N}^*$, with $p < q$. The idea here is quite simple and standard, and consists in expanding a transformation of the solution $t \in [0, 1] \mapsto \chi(s, t) \in \mathbb{R}^{\mathcal{S} \times \mathcal{S}}$ into a power series with matrix coefficients, as explained in [1, Section 1.1]. Let us focus on the fast arrival case in Section 3, although the method is of course applicable to the equilibrium case, and let us put $\check{\chi}(s, t) := \chi(s, t^p)$, $t \in [0, 1]$. In that case, we deduce from (26) that $t \in [0, 1] \mapsto \check{\chi}(s, t)$ verifies the matrix differential equation

$$\begin{cases} \partial_t \check{\chi}(s, t) &= \left[(p+q)t^q \lambda(P-I) + pt^{p-1} \beta \lambda \sum_{\ell=1}^k s_\ell \Delta_\ell \right] \check{\chi}(s, t), \quad t \in [0, 1], \\ &= [Q_1 t^q + Q_2(s) t^{p-1}] \check{\chi}(s, t), \\ \chi(s, 0) &= I. \end{cases}$$

where $Q_1 := (p+q)\lambda(P-I)$ and $Q_2(s) := p\beta\lambda \sum_{\ell=1}^k s_\ell \Delta_\ell$, $s = (s_1, \dots, s_k)$. It is quite simple to check that $\check{\chi}(s, t)$ can then be expanded as

$$\check{\chi}(s, t) = \sum_{j=0}^{\infty} U_j(s) t^j, \quad t \in [0, 1], \quad (32)$$

where the sequence of matrices $(U_j(s))_{j \in \mathbb{N}}$ is defined from [1, Relation (1.4)] by $U_0(s) = I$ and

$$U_j(s) = \begin{cases} 0, & 1 \leq j < p, \\ Q_2(s) U_{j-p}(s) / j, & p \leq j < q+1, \\ [Q_2(s) U_{j-p}(s) + Q_1 U_{j-q-1}(s)] / j, & j \geq q+1, \end{cases} \quad (33)$$

and that (32) converges for all t , as is proved in [1, Lemma 1 p.2]. The final solution is then expressed in that case as

$$\chi(s, t) = \check{\chi}(s, t^{1/p}) = \sum_{j=0}^{\infty} U_j(s) t^{j/p}, \quad t \in [0, 1].$$

The $U_j(s)$'s, $j \in \mathbb{N}$, being simply expressed with the simple linear recurrence (33), this expansion for $\chi(s, t)$ is then easy to handle as it can be e.g. approximated by truncation.

Appendix: Proof of Lemma 3.

We first observe that, because of continuity of $t \in [0, 1] \mapsto A_n(t)$ and $t \in [0, 1] \mapsto A(t)$, (10) and (11) read in integral form

$$Y_n(t) = y + \int_0^t A_n(v)Y_n(v)dv, \quad Y(t) = y + \int_0^t A(v)Y(v)dv$$

for all $t \in [0, 1]$. Let $\|\cdot\|$ be a submultiplicative norm on $\mathcal{S} \times \mathcal{S}$ matrices, and let us put $M_Y := \sup_{v \in [0, 1]} \|Y(v)\|$, $M_A := \sup_{v \in [0, 1]} \|A(v)\|$, which are finite quantities. Then one obtains that

$$\begin{aligned} \|Y_n(t) - Y(t)\| &\leq \int_0^t \|A_n(v) - A(v)\| \|Y(v)\| dv + \int_0^t \|A(v)\| \|Y_n(v) - Y(v)\| dv \\ &\leq M_Y \int_0^t \|A_n(v) - A(v)\| dv + M_A \int_0^t \|Y_n(v) - Y(v)\| dv, \quad \forall t \in [0, 1]. \end{aligned}$$

Gronwall's lemma thus implies that, for all $t \in [0, 1]$,

$$\begin{aligned} \|Y_n(t) - Y(t)\| &\leq M_Y \left[\int_0^t \|A_n(v) - A(v)\| dv \right] \cdot e^{M_A t} \\ &\leq M_Y \left[\int_0^1 \|A_n(v) - A(v)\| dv \right] \cdot e^{M_A} \longrightarrow 0 \text{ as } n \rightarrow \infty, \end{aligned}$$

proving the result.

References

- [1] Balser W. *Formal power series and linear systems of meromorphic ordinary differential equations*. Springer, 2000.
- [2] Berman S.H. and Halina F. Distributions associated with markov processes with killing. *Stochastic Models*, **12:3**: 367–388, 1996.
- [3] Blom J., De Turck K. and Mandjes M. Refined large deviations asymptotics for Markov-modulated infinite-server systems. *European Journal of Operation Research*, **259**: 1036–1044, 2017.
- [4] Blom J., Kella O., Mandjes M. and Thorsdottir H. Markov-modulated infinite-server queues with general service times. *Queueing Systems, Theory and Applications*, **76**: 403–424, 2014.
- [5] D'Auria B. $M/M/\infty$ queues in semi-Markovian random environment. *Queueing Systems, Theory and Applications*, **58**: 221–237, 2008.
- [6] Fralix B.H. and Adan I.J.B.F. An infinite-server queue influenced by a semi-Markovian environment. *Queueing Systems, Theory and Applications*, **61**: 65–84, 2009.
- [7] Hindley D. *Claims Reserving in General Insurance*. Cambridge University Press, 2017.
- [8] Kingman J.F.C. *Poisson Processes*. Oxford studies in probability, 1993.

- [9] Mandjes M. and De Turck K. Markov-modulated infinite-server queues driven by a common background process. *Stochastic Models*, **32(2)**: 206–232, 2016.
- [10] Masuyama H. and Takine T. Analysis of an infinite-server queue with batch markovian arrival streams. *Queueing Systems, Theory and Applications*, **42(3)**: 269–296, 2002.
- [11] O’Cinneide C., and Purdue P. The $M/M/\infty$ in a random environment. *Journal of Applied Probability*, **23**: 175–184, 1986.
- [12] Rabehasaina L. and Woo J.-K. On a multivariate renewal-reward process involving time delays and discounting: applications to IBNR processes and infinite server queues. *Queueing Systems, Theory and Applications*, to appear.
- [13] Rabehasaina L. and Woo J.-K. Analysis of the incurred but not reported/infinite server queue process with semi-Markovian multivariate discounted inputs. Submitted, available on [Hal](#) and [ArXiv](#).
- [14] Rogers L.C.G. and Williams D. *Diffusions, Markov Processes, and Martingales: Volume 1, Foundations*. Cambridge mathematical library, 2000.