



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. 2021. hal-01894007v3

HAL Id: hal-01894007

<https://hal.science/hal-01894007v3>

Preprint submitted on 18 Jun 2021

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 with Markov switching, Poisson arrivals and where the service times are fat tailed with index $\alpha \in (0, 1)$. When the arrival rate is sped up by a factor n^γ , the transition probabilities of the underlying Markov chain are divided by n^γ and the service times are divided by n , we identify two regimes ("fast arrivals", when $\gamma > \alpha$, and "equilibrium", when $\gamma = \alpha$) in which we prove that a properly rescaled process converges pointwise in distribution to some limiting process. In a third "slow arrivals" regime, $\gamma < \alpha$, we show the convergence of the two first joint moments of the rescaled process.

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

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

1. Introduction and notation

1.1. Model and related work

The classical infinite server queue consists of a system where tasks or customers arrive according to a general arrival process and begin receiving service immediately. Such a model was studied extensively, under various assumptions on the interarrival and service time distributions, in [18, Chapter 3, Section 3]. Several variants or extensions have been considered, in particular where arrivals and service times are governed by an external background Markovian process [14, 5, 6, 4, 11, 3], or where customers arrive in batches [12]. An extension to a network of infinite-server queues where arrival and service rates are Markov modulated is studied in [9].

We consider yet another generalization of this model with Markov switching described as follows. Let $\{N_t, t \geq 0\}$ be a Poisson process with intensity $\lambda > 0$, corresponding jump times $(T_i)_{i \in \mathbb{N}}$ satisfying $T_0 = 0$, such that $(T_i - T_{i-1})_{i \geq 1}$ is a sequence of independent and identically distributed (iid) random variables with same exponential distribution with parameter $\lambda > 0$, denoted by $\mathcal{E}(\lambda)$. Let $(L_{ij})_{i \in \mathbb{N}, j=1, \dots, k}$ be a sequence of independent random variables such that the sequence of vectors $((L_{i1}, \dots, L_{ik}))_{i \in \mathbb{N}}$ is iid (with entries L_{i1}, \dots, L_{ik} having different distributions for each i). Finally, for some K and k in \mathbb{N}^* we consider the discrete set $\mathcal{S} = \{0, \dots, K\}^k$ and a stationary finite Markov chain $(X_i)_{i \in \mathbb{N}}$ with state space \mathcal{S} . Then, for all i , X_i is a vector of the form $X_i = (X_{i1}, \dots, X_{ik})$ with $X_{ij} \in \{0, \dots, K\}$, $j = 1, \dots, k$. We then define the following k dimensional process $\{Z(t) = (Z_1(t), \dots, Z_k(t)), t \geq 0\}$ with values in \mathbb{N}^k as

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

The process defined by (1) has many applications, of which we list two most important ones:

- *incurred but not reported correlated claims*: in an actuarial context, $Z(t) = (Z_1(t), \dots, Z_k(t))$ represents a set of branches where $Z_j(t)$ is the number of incurred but non reported (IBNR) claims in the j th branch of an insurance company. Here 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 the claim j is reported. 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 undeclared claims which have occurred by time t . Another application is when $K = 1$, in which case $X_{ij} = 0$ means that the claim in branch j occurring at time T_i is reported and dealt with immediately by the policyholder, whereas $X_{ij} = 1$ means that some effective lag in the report is observed. The Markovian nature of $(X_i)_{i \in \mathbb{N}}$ here is important from a practical point of view, as a claim amount at time T_i may impact the one at time T_{i+1} , or because a policyholder may decide to grant a long report delay for the claim at time T_{i+1} with high probability if the claim at time T_i is reported immediately.
- *infinite server queues with batch arrivals and Markov switching*: $Z(t) = (Z_1(t), \dots, Z_k(t))$ represents a set of k correlated queues with an infinite number of servers, such that customers arrive at each time T_i , with X_{ij} customers arriving in queue $j \in \{1, \dots, k\}$, with corresponding (same) service times L_{ij} (as an example, the basic case where $X_{ij} = 1$

for all $i \in \mathbb{N}$ and $j = 1, \dots, k$ corresponds to k customers arriving simultaneously at each instant T_i). $Z_j(t)$ can also be seen as the number of customers of class j in a (single) infinite-server queue, as illustrated in [15, Figure 1]. Other infinite-server queue, such as one where the customers within a batch arriving at time T_i have different service times, may be inferred from the model (1) by choosing an appropriate value of k and Markov chain $(X_i)_{i \in \mathbb{N}}$, see [15, Section 6]. Here, the Markov switching is a major novelty in the present model because it allows for some dependence between the successive number of incoming customers. One simple example is when $K = 1$, so that $X_{ij} = 0$ means that an incoming customer at time T_i in queue j is rejected from the system, whereas $X_{ij} = 1$ means that it is accepted: a classical situation would then be that if a customer is rejected at time T_i then the next one could be accepted with high probability, at time T_{i+1} . In other words, this Markov switching can help model traffic regulation mechanisms.

The present paper follows [16], which studies the transient or limiting distribution of a discounted version of $Z(t)$ of the form

$$Z_j(t) = \sum_{i=1}^{N_t} X_{ij} e^{-a(L_{ij}+T_i)} \mathbb{1}_{[t < L_{ij}+T_i]} = \sum_{i=1}^{\infty} X_{ij} e^{-a(L_{ij}+T_i)} \mathbb{1}_{[T_i \leq t < L_{ij}+T_i]}, \quad j = 1, \dots, k, \quad (2)$$

for $t \geq 0$. The main difference with [16] is that the latter has more general assumptions on the interarrival and service distributions, whereas we focus here on Poisson arrivals. Even though the assumptions are more restrictive than in [16], the goal here is different in that we are trying to exhibit different behaviours for the limiting models when the arrival rate is increased and the service times are decreased by suitable factors, whereas [16] is more focused on analytical stochastic properties such as the moments of $Z(t)$ or its limiting distribution as $t \rightarrow \infty$. The discounting factor $a \geq 0$ in (2) is important in situations e.g. where, in an actuarial context, $X_{ij} e^{-a(L_{ij}+T_i)}$ represents the value of the claim amount at the actual realization time $L_{ij} + T_i$. Furthermore, the state space $\mathcal{S} = \{0, \dots, K\}^k$, although seemingly artificially complex, allows in fact for some flexibility and enables us to retrieve some known models. In particular, consider a Markov-modulated infinite-server queue, i.e. a queueing process $\{\mathcal{Z}(t), t \geq 0\}$ of which interarrivals and service times are modulated by a background continuous time Markov chain $\{Y(t), t \geq 0\}$ with state space say $\{1, \dots, \kappa\}$, i.e. such that customers arrive on the switching times of the Markov chain, with service times depending on the state of the background process (see [12], [11, Model II]). Then [16, Section 6] explains how this process $\{\mathcal{Z}(t), t \geq 0\}$ can

be embedded in a process $\{Z(t), t \geq 0\}$ defined by (1) with an appropriate choice of k , K in function of κ , as well as of the Markov chain $(X_i)_{i \in \mathbb{N}}$ and the sequence $(L_{ij})_{i \in \mathbb{N}, j=1, \dots, k}$ of service times. Thus, studying a general process $\{Z(t), t \geq 0\}$ in (1) allows to study a broad class of infinite server queue models in a similar Markov modulated context.

We now proceed with some notation related to the model and used throughout the paper. 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 $s = (s_1, \dots, s_k) \in (-\infty, 0]^k$,

$$\tilde{\pi}(s, r) := \text{diag} \left[\mathbb{E} \left(\exp \left\{ \sum_{j=1}^k s_j x_j \mathbb{1}_{[L_j > r]} \right\} \right), x = (x_1, \dots, x_k) \in \mathcal{S} \right], \quad (3)$$

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

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 Laplace Transform (LT) of the process $Z(t)$ 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 Euclidian inner 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 finish this section with the following notation. For two sequences of random variables $(A_n)_{n \in \mathbb{N}}$ and $(B_n)_{n \in \mathbb{N}}$ and two random variables A and B , the notation $\mathcal{D}(A_n | B_n) \xrightarrow{n \rightarrow \infty} \mathcal{D}(A | B)$ indicates that, as $n \rightarrow \infty$, the conditional distribution of A_n given B_n converges weakly to the conditional distribution of A given B .

1.2. Rescaling

We arrive at the main topic of the paper, which is to be able to provide some information on the distribution of $Z(t)$ in (1). In the particular case of Poisson arrivals, and since $Z(t)$ in (1) is a particular case of the process in (2) with discount factor $a = 0$, the LT $\psi(s, t)$ defined in (5) is characterized by [16, Proposition 4], which we rewrite here:

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*

$$\partial_t \psi(s, t) = [\lambda(P - I) + \lambda P(\tilde{\pi}(s, t) - I)]\psi(s, t) \quad (6)$$

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

Unfortunately, the first order ordinary differential equation (6) does not have an explicit expression in general, so that studying the (transient or stationary) distribution of the couple $(Z(t), X_{N_t})$ is difficult. In that case, as in [3, 4, 11], 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. Similarly to those papers, 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 service time distributions. For this we will suppose that the rescalings of the parameters, denoted by **(S1)**, **(S2)** and **(S3)** hereafter, are performed as follows:

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

$$\text{(S1)} \quad \lambda_n = \lambda n^\gamma,$$

with associated Poisson process $\{N_t^{(n)}, t \geq 0\}$ and jump times $(T_i^n)_{i \in \mathbb{N}}$,

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

$$\text{(S2)} \quad P_n = P/n^\gamma + (1 - 1/n^\gamma)I,$$

with corresponding stationary Markov chain $(X_i^{(n)})_{i \in \mathbb{N}}$, having the same distribution π as $(X_i)_{i \in \mathbb{N}}$.

Since the transition matrix P_n verifies $P_n \xrightarrow{n \rightarrow \infty} I$, such normalizing assumptions imply that, as $n \rightarrow \infty$, one is close to a model where the arriving customers or claims come in the k queues in batches with same fixed size: those queues are nonetheless correlated because the customers arrive according to the same Poisson process. Also, observe that $\lambda_n(P_n - I)$ is the infinitesimal generator of the continuous time Markov chain $\left\{ Y^{(n)}(t) = X_{N_t^{(n)}}^{(n)}, t \geq 0 \right\}$ of which embedded Markov chain is the underlying Markov chain i.e. $(Y^{(n)}(T_i^n))_{i \in \mathbb{N}} = (X_i^{(n)})_{i \in \mathbb{N}}$.

Thus, since the rescaling is such that

$$\lambda_n(P_n - I) = \lambda(P - I)$$

for all n (a property which will be extensively used in the paper), we remark that the rescalings **(S1)** and **(S2)** for the arrival rate and the transition probabilities are such that the transition rates between the states of \mathcal{S} of $\{Y^{(n)}(t), t \geq 0\}$ are independent from n , which allows for enough dynamics in the model that compensates the fact that P_n tends to I , and yielding non trivial asymptotics in the convergence results in this paper 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

$$\mathbb{P}(L_j > t) \sim 1/t^\alpha, \quad t \rightarrow \infty,$$

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". In particular, those service times have *infinite expectation*. Furthermore, the rescaling for the service times is such that they are divided by n , denoted by

$$\textbf{(S3)} \quad L_j^{(n)} = L_j/n.$$

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 [4, 11], where the authors identified three regimes for different scalings in a Markov modulating context and obtained a Central Limit Theorem for a renormalized process, when the service times have general distribution with finite expectation or are exponentially distributed. [3] provides some precise asymptotics on the tail probability of the queue content for exponentially distributed service times. [9] provides a diffusion approximation for a model with exponentially distributed service times. 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 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 [8, Section 6.6.1]. This motivates the convergence results in this paper, which feature the exponent α as the only information required on those delays. This in itself is a noticeable difference from the Central Limit Theorems obtained in [4, Section 4], where the normalization and limiting distribution require the explicit cumulative distribution function of the service times. Not only that, but the scaling is rather done in those references [4, 11] 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. When the service times are heavy tailed, this particular model can also be seen as a generalization of the *infinite source model*, see [13, Section 2.2]. Since the class of fat tailed distributions is a sub-class of the set of heavy tailed distributions, the normalizations **(S1)** and **(S3)** can be directly compared to [13, Section 3.1], which studies limiting distributions of such normalized processes, and where the authors introduce the notion of so-called Slow and Fast Growth conditions when the arrival rate of customers is respectively negligible or dominant, compared to the service times. The reader is also referred to [7] for a similar model where the interarrivals are heavy tailed. 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 random vector

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

under rescaling **(S1)**, **(S2)** and **(S3)**, or of a renormalized version of it in the "fast" or "slow" arriving customers case. Note that that the convergence is proved on the interval $[0, 1]$, but all proofs can be adapted to show the convergence on any interval $[0, M]$ for $M > 0$. The corresponding joint Laplace Transform is given by

$$\psi^{(n)}(s, t) = \left[\mathbb{E} \left(e^{\langle s, Z^{(n)}(t) \rangle} \mathbb{1}_{\left[X_{N_t^{(n)}}^{(n)} = y \right]} \middle| X_0^{(n)} = x \right) \right]_{(x,y) \in \mathcal{S}^2}, \quad s = (s_1, \dots, s_j) \in (-\infty, 0]^k, \quad (7)$$

where we recall that $(X_i^{(n)})_{i \in \mathbb{N}}$ 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 . We also introduce the first and second joint matrix moments defined by

$$\begin{aligned} M_j^{(n)}(t) &:= \left[\mathbb{E} \left(Z_j^{(n)}(t) \mathbb{1}_{\left[X_{N_t^{(n)}}^{(n)}=y \right]} \middle| X_0^{(n)} = x \right) \right]_{(x,y) \in \mathcal{S}^2}, \quad j = 1, \dots, k, \\ M_{jj'}^{(n)}(t) &:= \left[\mathbb{E} \left(Z_j^{(n)}(t) Z_{j'}^{(n)}(t) \mathbb{1}_{\left[X_{N_t^{(n)}}^{(n)}=y \right]} \middle| X_0^{(n)} = x \right) \right]_{(x,y) \in \mathcal{S}^2}, \quad j, j' = 1, \dots, k. \end{aligned} \quad (8)$$

2. Statement of results and organization of paper

The core results of the paper concerning the different regimes are given in the following two Theorems 2 and 3:

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

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

$$\begin{aligned} \mathcal{D} \left(\left(\frac{Z^{(n)}(t)}{n^{\gamma-\alpha}}, X_{N_t^{(n)}}^{(n)} \right) \middle| X_0^{(n)} \right) \\ \rightarrow \mathcal{D} \left(\left(\frac{\lambda}{1-\alpha} \int_{1-t^{1-\alpha}}^1 \mathcal{X}^\alpha(v) dv, \mathcal{X}^\alpha(1) \right) \middle| \mathcal{X}^\alpha(1-t^{1-\alpha}) \right), \end{aligned} \quad (9)$$

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

$$\begin{aligned} \mathcal{D} \left(\left(Z^{(n)}(t), X_{N_t^{(n)}}^{(n)} \right) \middle| X_0^{(n)} \right) \\ \rightarrow \mathcal{D} \left(\left(\left(\int_{1-t^{1-\alpha}}^1 \mathcal{X}_j^\alpha(v) \nu_j^\alpha(dv) \right)_{j=1, \dots, k}, \mathcal{X}^\alpha(1) \right) \middle| \mathcal{X}^\alpha(1-t^{1-\alpha}) \right). \end{aligned} \quad (10)$$

We note that the terms in the limits on the right hand side of (9) and (10) feature simple objects (in regards to the complexity of the original model) where the only characteristic parameters needed are λ , P and α ; in particular, and apart from α , characteristics of the service times L_j , $j = 1, \dots, k$, such as their cumulative distribution functions, do not show up in the limiting distributions (9) and (10). The convergences in distribution (9) and (10) give some information on convergence of the (possibly renormalized) joint distribution of

the couple $\left(Z^{(n)}(t), X_{N_t^{(n)}}^{(n)} \right)$, $t \in [0, 1]$ given the initial state $X_0^{(n)}$. Intuitively, for fixed $t \in [0, 1]$, in the right hand sides of (9) and (10) we may interpret the inhomogeneous continuous time Markov chain $\{ \mathcal{X}^\alpha(v), v \in [1 - t^{1-\alpha}, 1] \}$ as the limiting counterpart of the modulating process $\left\{ X_{N_v^{(n)}}^{(n)}, v \in [0, t] \right\}$. On an even cruder level, we observe in the Fast arrivals case from (9) that each entry of $Z^{(n)}(t) = (Z_1^{(n)}(t), \dots, Z_k^{(n)}(t))$ behaves roughly like $n^{\gamma-\alpha}t^{1-\alpha}$. The intuition behind this behaviour may be explained as follows. Within queue $j = 1, \dots, k$, there are approximately $\lambda n^\gamma t$ arrivals in the interval $[0, t]$, each arriving customer with service time distributed as $L_j^{(n)}$, so that we may very grossly consider that a customer is still present at time t with probability $\mathbb{P}(L_j^{(n)} > t) = \mathbb{P}(L_j/n > t)$. Hence the number of customers in queue j is approximately

$$Z_j^{(n)}(t) \approx \lambda n^\gamma t \times \mathbb{P}(L^{(j)}/n > t) = \lambda n^\gamma t \times \mathbb{P}(L^{(j)} > nt) \approx \lambda n^\gamma t \times \frac{1}{(nt)^\alpha} = \lambda n^{\gamma-\alpha} t^{1-\alpha}$$

which is the expected order of growth $n^{\gamma-\alpha}t^{1-\alpha}$. Of course, such approximations are very crude, however this enables us to justify the presence of the normalizing factor $n^{\gamma-\alpha}$ as well as the time dilated factor $t^{1-\alpha}$ in (9).

In the case when $\gamma < \alpha$, proving the convergence in distribution of an adequate normalization of $Z^{(n)}(t)$ seems more difficult. The following result show that the two first moments of $Z^{(n)}(t)$ converge under respective normalization $n^{\alpha-\gamma}$ and $n^{(\alpha-\gamma)/2}$:

Theorem 3 (Slow arrivals). *If $\gamma < \alpha$ then the following convergences of the two joint moments hold as $n \rightarrow \infty$*

$$n^{\alpha-\gamma} M_j^{(n)}(t) \longrightarrow \lambda e^{\lambda t(P-I)} \int_0^t \frac{1}{v^\alpha} e^{-\lambda v(P-I)} \Delta_j e^{\lambda v(P-I)} dv, \tag{11}$$

$$n^{\alpha-\gamma} M_{jj}^{(n)}(t) \longrightarrow \lambda e^{\lambda t(P-I)} \int_0^t \frac{1}{v^\alpha} e^{-\lambda v(P-I)} \Delta_j^2 e^{\lambda v(P-I)} dv, \tag{12}$$

$$n^{\alpha-\gamma} M_{jj'}^{(n)}(t) \longrightarrow 0 \quad j \neq j', \tag{13}$$

for all $j, j' \neq j$, in $1, \dots, k$, $t \in [0, 1]$, where we recall that Δ_j is defined in (4).

One interesting by-product of Theorem 2 is that it in particular gives some insight on the (non conditional) limiting distribution of $Z^{(n)}(t)$, with a limit in a somewhat simpler form. More precisely, the following corollary follows from the proofs of (9) and (10):

Corollary 4. *In the Fast arrivals case $\gamma > \alpha$, the following convergence holds*

$$\frac{Z^{(n)}(t)}{n^{\gamma-\alpha}} \xrightarrow{\mathcal{D}} \lambda \int_0^t \frac{\mathcal{Y}(v)}{v^\alpha} dv, \quad n \rightarrow \infty, \quad t \in [0, 1], \tag{14}$$

where $\{\mathcal{Y}(t) = (\mathcal{Y}_1(t), \dots, \mathcal{Y}_k(t)), t \in [0, 1]\}$ is a (time homogeneous) stationary continuous time Markov chain on the state space \mathcal{S} , with infinitesimal generator matrix defined by

$$\lambda(\Delta_\pi^{-1}P'\Delta_\pi - I), \quad \Delta_\pi := \text{diag}(\pi(x), x \in \mathcal{S}). \quad (15)$$

In the Equilibrium case $\gamma = \alpha$, one has

$$Z^{(n)}(t) \xrightarrow{\mathcal{D}} \left(\int_0^t \mathcal{Y}_j(v) \tilde{\nu}_j^\alpha(dv) \right)_{j=1, \dots, k}, \quad n \rightarrow \infty, \quad t \in [0, 1]. \quad (16)$$

Here $\{\tilde{\nu}_j^\alpha(t), t \in [0, 1]\}$, $j = 1, \dots, k$, are the inhomogeneous Poisson processes defined by $\tilde{\nu}_j^\alpha(t) = \nu_j^\alpha(t^{1-\alpha})$, $t \in [0, 1]$, where $\{\nu_j^\alpha(t), t \in [0, 1]\}$, $j = 1, \dots, k$, are defined in Theorem 2.

As mentioned in Section 1.2, [13, 7] introduced a notion of Fast and Slow growth similar to the Fast and Slow arrivals presented in Theorems 2 and 3, for a process of interest which is either a superposition of renewal processes with heavy tailed interarrivals or the cumulative input of an infinite source Poisson model with heavy tailed services. In those references, the process is shown to converge weakly or in finite dimensional distributions towards specific limit processes under appropriate scaling, see [13, Theorem 1] and [7, Theorem 1]. Here, the outline of the proof of Theorem 2 is the following:

- We will first expand the LT of the left hand side of (9) and (10) as $n \rightarrow \infty$ and prove that the limit satisfies a particular ODE thanks to Proposition 1, which will be referred as Steps 1 and 2 in the proofs in the forthcoming Sections 3.2 and 3.3.
- Then, we will identify this limit as the LT of the right hand side of (9) and (10) thanks to a proper use of the Feynman-Kac or Campbell formula. This step will be referred as Step 3 in the proofs.

This is to be compared with the approach in [4, 11], where the authors derive ODEs for the limiting moment generating function and identify a gaussian limiting distribution for the normalized process.

The paper is organized in the following way. Section 3 is dedicated to the proofs of the main results, with Subsections 3.2, 3.3 and 3.4 giving the proofs of the convergences in distribution of Theorem 2 in fast arrivals and equilibrium cases, and of Theorem 3 in the slow arrivals case. The proof of Corollary 4 is included in Subsections 3.2, for the convergence (14), and 3.3, for the convergence (16). As a concluding remark, we will discuss in Section 4 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 $(0, 1)$.

3. Proofs of Theorems 2, 3 and Corollary 4

3.1. Preliminary results

We will repeatedly use the following general lemma in the proofs:

Lemma 5. *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 $n \rightarrow \infty$ for any matrix norm $\|\cdot\|$. 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*

$$\begin{cases} \frac{d}{dt} Y_n(t) = A_n(t) Y_n(t), & t \in [0, 1], \\ Y_n(0) = y \in \mathbb{R}^{\mathcal{S} \times \mathcal{S}}, \end{cases} \quad n \in \mathbb{N}. \quad (17)$$

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

$$\begin{cases} \frac{d}{dt} Y(t) = A(t) Y(t), & t \in [0, 1], \\ Y(0) = y. \end{cases} \quad (18)$$

Proof. We first observe that, because of continuity of $t \in [0, 1] \mapsto A_n(t)$ and $t \in [0, 1] \mapsto A(t)$, (17) and (18) read in integral form

$$Y_n(t) = y + \int_0^t A_n(v) Y_n(v) dv, \quad Y(t) = y + \int_0^t A(v) Y(v) dv \quad (19)$$

for all $t \in [0, 1]$. Since the norm $\|\cdot\|$ may be arbitrary, we pick a submultiplicative one on the set of $\mathcal{S} \times \mathcal{S}$ matrices. (19) implies the following inequality

$$\|Y_n(t)\| \leq \|y\| + \int_0^t \|A_n(v)\| \|Y_n(v)\| dv, \quad \forall t \in [0, 1].$$

Gronwall's lemma thus implies that $\|Y_n(t)\| \leq \|y\| \exp\left(\int_0^t \|A_n(v)\| dv\right)$ for all $t \in [0, 1]$. Since by assumption $\int_0^1 \|A_n(v) - A(v)\| dv \rightarrow 0$ as $n \rightarrow \infty$, one has that $\left(\int_0^1 \|A_n(v)\| dv\right)_{n \in \mathbb{N}}$ is a bounded sequence. We deduce the following finiteness

$$M_Y := \sup_{n \in \mathbb{N}} \sup_{t \in [0, 1]} \|Y_n(t)\| \leq \sup_{n \in \mathbb{N}} \sup_{t \in [0, 1]} \|y\| \exp\left(\int_0^t \|A_n(v)\| dv\right)$$

$$\leq \|y\| \exp \left(\int_0^1 \sup_{n \in \mathbb{N}} \|A_n(v)\| dv \right) < +\infty.$$

Let us then introduce $M_A := \sup_{v \in [0,1]} \|A(v)\|$, which is a finite quantity. Then one obtains that

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

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

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

Since the right hand side of the above inequality is independent from $t \in [0, 1]$, this proves the uniform convergence result. \square

We finish this subsection by stating the differential equation satisfied by the Laplace Transform $\psi^{(n)}(s, t)$ of $\left(Z^{(n)}(t), X_{N_t^{(n)}}^{(n)} \right)$ defined in (7), which will be the central object studied in Subsections 3.2 and 3.3. Thanks to equation (6) with the new parameters λ_n, P_n instead of λ and P (and remembering that $\lambda_n(P_n - I) = \lambda(P - I)$), this reads here

$$\begin{cases} \partial_t \psi^{(n)}(s, t) &= [\lambda(P - I) + \lambda n^\gamma P_n (\tilde{\pi}_n(s, t) - I)] \psi^{(n)}(s, t), \quad t \geq 0, \\ \psi^{(n)}(s, 0) &= I, \end{cases} \quad (20)$$

for all $s = (s_1, \dots, s_k) \in (-\infty, 0]^k$. And, from (3), 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 numbers a_1, \dots, a_k , we have the following expansion which will be useful later on:

$$\begin{aligned} \tilde{\pi}_n(s, t) - I &= \text{diag} \left(\prod_{j=1}^k \left((e^{s_j x_j} - 1) \mathbb{P} [L_j^{(n)} > t] + 1 \right) - 1, \quad x = (x_1, \dots, x_k) \in \mathcal{S} \right) \\ &= \text{diag} \left(\sum_{I \subset \{1, \dots, k\}} \prod_{\ell \in I} \left[(e^{s_\ell x_\ell} - 1) \mathbb{P} [L_\ell^{(n)} > t] \right], \quad x = (x_1, \dots, x_k) \in \mathcal{S} \right). \end{aligned} \quad (21)$$

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

We now proceed to show convergence (9) in Theorem 2. In the present case, it is sensible to guess that $Z^{(n)}(t)$ converges towards infinity as $n \rightarrow \infty$, hence it is natural to find a

normalization such that a convergence towards a proper distribution occurs. We renormalize the queue content by dividing it by $n^{\gamma-\alpha}$, i.e. we are here interested in $\left(Z^{(n)}(t)/n^{\gamma-\alpha}, X_{N_t^{(n)}}^{(n)}\right)$, of which Laplace transform is given by $\psi^{(n)}(s/n^{\gamma-\alpha}, t)$, $s = (s_1, \dots, s_k) \in (-\infty, 0]^k$. In order to avoid cumbersome notation, we introduce the quantity

$$\beta := \frac{1}{1-\alpha} \in (1, +\infty).$$

We observe then that

$$t \in [0, 1] \mapsto t^\beta \in [0, 1] \quad (22)$$

is a one to one mapping. Hence, studying the limiting distribution of $\left(Z^{(n)}(t)/n^{\gamma-\alpha}, X_{N_t^{(n)}}^{(n)}\right)$ for all $t \in [0, 1]$ amounts to study the limiting distribution of

$$\left(Z^{(n)}(t^\beta)/n^{\gamma-\alpha}, X_{N_{t^\beta}^{(n)}}^{(n)}\right) \quad (23)$$

for all $t \in [0, 1]$, then changing variable $t := t^{1/\beta}$. The time transformation (22) may at this point look artificial, but this is a key step which will later on enable us to use the convergence result in Lemma 5. The LT of (23) is given by

$$\chi^{(n)}(s, t) := \psi^{(n)}(s/n^{\gamma-\alpha}, t^\beta), \quad t \in [0, 1].$$

From (20), $\chi^{(n)}(s, t)$ satisfies

$$\begin{cases} \partial_t \chi^{(n)}(s, t) &= \beta t^{\beta-1} [\lambda(P - I) + \lambda n^\gamma P_n (\tilde{\pi}_n(s/n^{\gamma-\alpha}, t^\beta) - I)] \chi^{(n)}(s, t), \quad t \in [0, 1], \\ \chi^{(n)}(s, 0) &= I. \end{cases} \quad (24)$$

The starting point for proving (9) is the following: we will set to prove that

$$A_n(s, t) = \beta t^{\beta-1} [\lambda(P - I) + \lambda n^\gamma P_n (\tilde{\pi}_n(s/n^{\gamma-\alpha}, t^\beta) - I)] \quad (25)$$

converges to some limit $A(s, t)$ as $n \rightarrow \infty$, use Lemma 5, 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: Determining $A(s, t)$. This step is dedicated to finding the limit function $t \in [0, 1] \mapsto A(s, t)$ of (25). Recalling that $\lim_{n \rightarrow \infty} P_n = I$, studying the limit of (25) amounts to studying that of $\beta t^{\beta-1} \lambda n^\gamma (\tilde{\pi}_n(s/n^{\gamma-\alpha}, t^\beta) - I)$ as $n \rightarrow \infty$. In view of (21), the x th diagonal element of this latter term is

$$\beta t^{\beta-1} \lambda n^\gamma \sum_{I \subset \{1, \dots, k\}} \prod_{\ell \in I} \left[(e^{s_\ell x_\ell / n^{\gamma-\alpha}} - 1) \mathbb{P} \left[L_\ell^{(n)} > t^\beta \right] \right] \quad (26)$$

of which we proceed to find the limit as $n \rightarrow \infty$. In order to study its convergence, we are going to isolate the terms in the sum (26) 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 (26) as

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

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

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

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

$$\mathbb{P} \left[L_\ell^{(n)} > t^\beta \right] = \mathbb{P} \left[L_\ell > n t^\beta \right] \sim \frac{1}{n^\alpha t^{\beta\alpha}} \quad (29)$$

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 $t = 0$. Next we show that $J_n^2(s, t)$ tends to 0 by showing that each term on the right hand side of (28) tend to 0. So, if $I \subset \{1, \dots, k\}$ is such that $I = \{\ell_1, \ell_2\}$, i.e. $\text{Card}(I) = 2$, then

$$\begin{aligned} \left| \beta t^{\beta-1} \lambda n^\gamma \prod_{\ell \in I} \left[(e^{s_\ell x_\ell / n^{\gamma-\alpha}} - 1) \mathbb{P} \left[L_\ell^{(n)} > t^\beta \right] \right] \right| &= \beta t^{\beta-1} \lambda n^\gamma \left| e^{s_{\ell_1} x_{\ell_1} / n^{\gamma-\alpha}} - 1 \right| \mathbb{P} \left[L_{\ell_1} > n t^\beta \right] \\ &\cdot \left| e^{s_{\ell_2} x_{\ell_2} / n^{\gamma-\alpha}} - 1 \right| \mathbb{P} \left[L_{\ell_2} > n t^\beta \right] \leq \beta t^{\beta-1} \lambda n^\gamma |s_{\ell_1} x_{\ell_1}| \cdot |s_{\ell_2} x_{\ell_2}| \frac{1}{n^{2(\gamma-\alpha)}} \mathbb{P} \left[L_{\ell_1} > n t^\beta \right] \\ &= \beta t^{\beta-1} \lambda |s_{\ell_1} x_{\ell_1}| \cdot |s_{\ell_2} x_{\ell_2}| \frac{1}{n^{\gamma-\alpha}} n^\alpha \mathbb{P} \left[L_{\ell_1} > n t^\beta \right], \quad (30) \end{aligned}$$

where we used the inequality $|e^u - 1| \leq |u|$ for $u \leq 0$ and $\mathbb{P} \left[L_{\ell_2} > n t^\beta \right] \leq 1$. Thanks to (29), the right hand side of (30) thus tends to zero when $t \in (0, 1]$. The case $\text{Card}(I) > 2$ is dealt with similarly. Finally, all terms on the right hand side of (28) tend to 0 as $n \rightarrow \infty$, i.e. $\lim_{n \rightarrow \infty} J_n^2(s, t) = 0$ for all $t \in (0, 1]$. When $t = 0$ then $J_n^2(s, t) = 0$, so that the limit holds for all $t \in [0, 1]$.

Hence we have that (26) tends to $\lim_{n \rightarrow \infty} J_n^1(s, t) + \lim_{n \rightarrow \infty} J_n^2(s, t)$, i.e. 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

$$t \in [0, 1] \mapsto A(s, t) := \beta t^{\beta-1} \lambda (P - I) + \beta \lambda \sum_{\ell=1}^k s_\ell \Delta_\ell \quad (31)$$

where we recall from (4) that $\Delta_\ell = \text{diag}[x_\ell, x = (x_1, \dots, x_k) \in \mathcal{S}]$. This is where the time transformation (22) described previously is important, as without it it would not have been possible to exhibit the limit (31) for $A_n(s, t)$. Note that the limit when $t = 0$ for $A_n(s, t)$ in (25) differs from $A(s, 0) = \beta\lambda \sum_{\ell=1}^k s_\ell \Delta_\ell$, 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 $\lim_{n \rightarrow \infty} A_n(s, 0)$ should rather be the 0 matrix. This is due to the fact that the limit $t \in [0, 1] \mapsto A(s, t)$ in Lemma 5 has to be *continuous* so that the lemma holds.

Step 2: Determining $\chi(s, t) = \lim_{n \rightarrow \infty} \chi_n(s, t)$. We now need to prove that $\int_0^1 \|A_n(s, v) - A(s, v)\| dv \rightarrow 0$ as $n \rightarrow \infty$ in order to apply Lemma 5. Thanks to (25) and (31), and by the definitions (27) and (28) of $J_n^1(s, t, x)$ and $J_n^2(s, t, x)$, we observe that $A_n(s, t)$ can be decomposed as

$$\begin{aligned} A_n(s, t) &= A(s, t) + P_n \text{diag} \left(J_n^1(s, t, x) - \beta\lambda \sum_{\ell=1}^k s_\ell x_\ell, x \in \mathcal{S} \right) \\ &\quad + P_n \text{diag} (J_n^2(s, t, x), x \in \mathcal{S}) + (P_n - I)\beta\lambda \sum_{\ell=1}^k s_\ell \Delta_\ell, \quad t \in [0, 1]. \end{aligned}$$

Hence, since $\lim_{n \rightarrow \infty} P_n = I$, proving $\lim_{n \rightarrow \infty} \int_0^1 \|A_n(s, v) - A(s, v)\| dv = 0$ amounts to proving that

$$\begin{aligned} \int_0^1 \left| J_n^1(s, v, x) - \beta\lambda \sum_{\ell=1}^k s_\ell x_\ell \right| dv &\rightarrow 0, \text{ and} \\ \int_0^1 |J_n^2(s, v, x)| dv &= \int_0^1 J_n^2(s, v, x) dv \rightarrow 0, \end{aligned} \tag{32}$$

as $n \rightarrow \infty$, for each fixed $x \in \mathcal{S}$. Let us first focus on $\int_0^1 \left| J_n^1(s, v, x) - \beta\lambda \sum_{\ell=1}^k s_\ell x_\ell \right| dv$. We have

$$\begin{aligned} \int_0^1 \left| J_n^1(s, v, x) - \beta\lambda \sum_{\ell=1}^k s_\ell x_\ell \right| dv &\leq \sum_{\ell=1}^k (I_n^1(\ell) + I_n^2(\ell)), \text{ where, for all } \ell = 1, \dots, k, \\ I_n^1(\ell) &:= \int_0^1 \lambda \beta v^{\beta-1} \left| n^\gamma (e^{s_\ell x_\ell / n^{\gamma-\alpha}} - 1) - n^\alpha s_\ell x_\ell \right| \mathbb{P} [L_\ell^{(n)} > v^\beta] dv \\ I_n^2(\ell) &:= |s_\ell x_\ell| \int_0^1 \lambda \left| \beta v^{\beta-1} n^\alpha \mathbb{P} [L_\ell^{(n)} > v^\beta] - \beta \right| dv. \end{aligned} \tag{33}$$

Expanding the exponential function, one has that $|e^{s_\ell x_\ell / n^{\gamma-\alpha}} - 1 - s_\ell x_\ell / n^{\gamma-\alpha}| \leq M_\ell / n^{2(\gamma-\alpha)}$ where $M_\ell > 0$ only depends on s_ℓ and x_ℓ . Thus, one deduces the following upper bounds for $I_n^1(\ell)$, $\ell = 1, \dots, k$:

$$I_n^1(\ell) = \int_0^1 \lambda \beta v^{\beta-1} \left| n^\gamma (e^{s_\ell x_\ell / n^{\gamma-\alpha}} - 1) - n^\alpha s_\ell x_\ell \right| \mathbb{P} [L_\ell > n v^\beta] dv$$

$$\begin{aligned}
&\leq n^\gamma \frac{M_\ell}{n^{2(\gamma-\alpha)}} \lambda \int_0^1 \beta v^{\beta-1} \mathbb{P} [L_\ell > nv^\beta] dv = \frac{M_\ell}{n^{\gamma-\alpha}} \beta \lambda \int_0^1 n^\alpha v^{\beta-1} \mathbb{P} [L_\ell > nv^\beta] dv \\
&= \frac{M_\ell}{n^{\gamma-\alpha}} \beta \lambda \int_0^1 (nv^\beta)^\alpha \mathbb{P} [L_\ell > nv^\beta] dv, \tag{34}
\end{aligned}$$

the last equality holding because $\beta - 1 = \beta\alpha$ implies that the integrand verifies $n^\alpha v^{\beta-1} = (nv^\beta)^\alpha$. 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

$$\sup_{j \in \mathbb{N}, v \in [0,1]} (jv^\beta)^\alpha \mathbb{P} [L_\ell > jv^\beta] < +\infty \tag{35}$$

(note that those two latter suprema are in fact equal). One then gets from (34) that

$$I_n^1(\ell) \leq \frac{M_\ell}{n^{\gamma-\alpha}} \beta \lambda \left[\sup_{j \in \mathbb{N}, v \in [0,1]} (jv^\beta)^\alpha \mathbb{P} [L_\ell > jv^\beta] \right] \longrightarrow 0, \quad n \rightarrow \infty. \tag{36}$$

We now turn to $I_n^2(\ell)$, $\ell = 1, \dots, k$. Using again $\beta - 1 = \beta\alpha$, one may write in the integrand of $I_n^2(\ell)$ that $v^{\beta-1} n^\alpha = (nv^\beta)^\alpha$: hence

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

Since L_ℓ is fat tailed with index α , estimates similar to the ones leading to the upper bound (36) for $I_n^1(\ell)$ yield that

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

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

$$I_n^2(\ell) \longrightarrow 0, \quad n \rightarrow \infty. \tag{37}$$

Gathering (33), (36) and (37), we thus deduce finally that $\int_0^1 \left| J_n^1(s, v, x) - \beta \lambda \sum_{\ell=1}^k s_\ell x_\ell \right| dv$ tends to 0 as $n \rightarrow \infty$ for each $x \in \mathcal{S}$.

We now prove that $\int_0^1 J_n^2(s, v, x) dv \rightarrow 0$ as $n \rightarrow \infty$. In view of the definition (28), it suffices to prove that

$$\int_0^1 \beta v^{\beta-1} \lambda n^\gamma \prod_{\ell \in I} \left[|e^{s_\ell x_\ell / n^{\gamma-\alpha}} - 1| \mathbb{P} [L_\ell^{(n)} > v^\beta] \right] dv \tag{38}$$

tends to 0 as $n \rightarrow \infty$ for $I \subset \{1, \dots, k\}$ such that $\text{Card}(I) \geq 2$. Let us prove the convergence for $\text{Card}(I) = 2$, i.e. for $I = \{\ell_1, \ell_2\}$ for some $\ell_1 \neq \ell_2$ in $1, \dots, k$, the case $\text{Card}(I) > 2$

being dealt with similarly. By the basic inequality $|e^u - 1| \leq |u|$ for $u \leq 0$ we deduce that $|e^{s_{\ell_i} x_{\ell_i}/n^{\gamma-\alpha}} - 1| \leq |s_{\ell_i} x_{\ell_i}|/n^{\gamma-\alpha}$, $i = 1, 2$. Since $\mathbb{P}\left[L_{\ell_1}^{(n)} > v^\beta\right] \leq 1$ for all $v \in [0, 1]$, we then deduce that (38) is upper bounded by

$$\frac{|s_{\ell_1} x_{\ell_1}|}{n^{\gamma-\alpha}} |s_{\ell_2} x_{\ell_2}| \int_0^1 \beta v^{\beta-1} \lambda n^\alpha \mathbb{P}\left[L_{\ell_2}^{(n)} > v^\beta\right] dv.$$

As $v^{\beta-1} n^\alpha = (nv^\beta)^\alpha$, and thanks to (35), the latter quantity is in turn written then bounded as follows

$$\begin{aligned} \frac{|s_{\ell_1} x_{\ell_1}|}{n^{\gamma-\alpha}} |s_{\ell_2} x_{\ell_2}| \int_0^1 \beta \lambda (nv^\beta)^\alpha \mathbb{P}\left[L_{\ell_2}^{(n)} > v^\beta\right] dv &= \frac{|s_{\ell_1} x_{\ell_1}|}{n^{\gamma-\alpha}} |s_{\ell_2} x_{\ell_2}| \int_0^1 \beta \lambda (nv^\beta)^\alpha \mathbb{P}\left[L_{\ell_2} > nv^\beta\right] dv \\ &\leq \frac{|s_{\ell_1} x_{\ell_1}|}{n^{\gamma-\alpha}} |s_{\ell_2} x_{\ell_2}| \beta \lambda \left[\sup_{j \in \mathbb{N}, v \in [0,1]} (jv^\beta)^\alpha \mathbb{P}\left[L_{\ell_2} > jv^\beta\right] \right] \longrightarrow 0, \quad n \rightarrow \infty, \end{aligned}$$

proving that (38) tends to 0 as $n \rightarrow \infty$ when $I = \{\ell_1, \ell_2\}$.

Hence we just proved (32), which implies $\int_0^1 \|A_n(s, v) - A(s, v)\| dv \longrightarrow 0$. We may then use Lemma 5 to deduce that $\chi^{(n)}(s, t)$ converges to $\chi(s, t)$ which satisfies

$$\begin{cases} \partial_t \chi(s, t) &= A(s, t) \chi(s, t) = \left[\beta t^{\beta-1} \lambda (P - I) + \beta \lambda \sum_{\ell=1}^k s_\ell \Delta_\ell \right] \chi(s, t), \quad t \in [0, 1], \\ \chi(s, 0) &= I. \end{cases} \quad (39)$$

Step 3: Identifying the limit in distribution. Let us note that (39) 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 $P^{(r)}$ 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

$$\{U(t) = (U_j(t))_{j=1, \dots, k} \in \mathcal{S}, t \in [0, 1]\} \quad (40)$$

with values in \mathcal{S} , and 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 $t \in [0, 1]$, i.e. that $\{U(t), t \in [0, 1]\}$ is stationary. Since $\sum_{j=1}^k s_j \Delta_j$ is diagonal, one checks easily that $A(s, t) = \Delta_\pi^{-1} \left[\beta t^{\beta-1} \lambda (P^{(r)'} - I) + \sum_{j=1}^k s_j \Delta_j \right] \Delta_\pi$ and that $Y(t) = Y(s, t) := \Delta_\pi^{-1} \chi(s, t)' \Delta_\pi$ satisfies the differential equation

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

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 [17, 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 reversed process $\{U(1-t), t \in [0, 1]\}$ admits $\Delta_\pi^{-1} \beta (1-t)^{\beta-1} \lambda (P^{(r)} - I) \Delta_\pi = \beta (1-t)^{\beta-1} \lambda (P - I)$ as infinitesimal generator matrix, which is the generator of the process $\{\mathcal{X}^\alpha(t) = (\mathcal{X}_1^\alpha(t), \dots, \mathcal{X}_k^\alpha(t)) \in \mathcal{S}, t \in [0, 1]\}$ introduced in the statement of Theorem 2, so that $\{\mathcal{X}^\alpha(t), t \in [0, 1]\} \stackrel{\mathcal{D}}{=} \{U(1-t), t \in [0, 1]\}$ pathwise. Hence, one obtains for all x and y in \mathcal{S} that

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1}_{[U(t)=y]} \exp \left(\sum_{j=1}^k s_j \beta \lambda \int_0^t U_j(v) dv \right) \middle| U(0) = x \right] \\ &= \mathbb{E} \left[\mathbf{1}_{[\mathcal{X}^\alpha(1-t)=y]} \exp \left(\sum_{j=1}^k s_j \beta \lambda \int_{1-t}^1 \mathcal{X}_j^\alpha(v) dv \right) \middle| \mathcal{X}^\alpha(1) = x \right] \\ &= \mathbb{E} \left[\mathbf{1}_{[\mathcal{X}^\alpha(1)=x]} \exp \left(\sum_{j=1}^k s_j \beta \lambda \int_{1-t}^1 \mathcal{X}_j^\alpha(v) dv \right) \middle| \mathcal{X}^\alpha(1-t) = y \right] \frac{\pi(y)}{\pi(x)}, \end{aligned} \quad (41)$$

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

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

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

$$\mathcal{D} \left(\left(Z^{(n)}(t^\beta) / n^{\gamma-\alpha}, X_{N_{t^\beta}^{(n)}}^{(n)} \right) \middle| X_0^{(n)} \right) \longrightarrow \mathcal{D} \left(\left(\beta \lambda \int_{1-t}^1 \mathcal{X}^\alpha(v) dv, \mathcal{X}^\alpha(1) \right) \middle| \mathcal{X}^\alpha(1-t) \right) \quad (42)$$

as $n \rightarrow \infty$ for all $t \in [0, 1]$. Changing t into $t^{1/\beta}$ yields (9).

Proof of the convergence (14) in Corollary 4. With the previous definitions of processes $\{U(t), t \in [0, 1]\}$ in (40) and $\{\mathcal{X}^\alpha(t), t \in [0, 1]\}$, (41) implies the following matrix equality

$$\begin{aligned} & \left[\mathbb{E} \left[\mathbf{1}_{[U(t)=y]} \exp \left(\sum_{j=1}^k s_j \beta \lambda \int_0^t U_j(v) dv \right) \middle| U(0) = x \right] \right]_{(x,y) \in \mathcal{S}^2} \\ &= \left[\mathbb{E} \left[\mathbf{1}_{[\mathcal{X}^\alpha(1)=x]} \exp \left(\sum_{j=1}^k s_j \beta \lambda \int_{1-t}^1 \mathcal{X}_j^\alpha(v) dv \right) \middle| \mathcal{X}^\alpha(1-t) = y \right] \frac{\pi(y)}{\pi(x)} \right]_{(x,y) \in \mathcal{S}^2} \end{aligned} \quad (43)$$

Left-multiplying and right-multiplying (43) respectively by the row vector $(\pi(x))_{x \in \mathcal{S}}$ and the column vector $\mathbf{1}$ results in in the following equality of LT

$$\begin{aligned} & \mathbb{E} \left[\exp \left(\sum_{j=1}^k s_j \beta \lambda \int_0^t U_j(v) dv \right) \right] \\ &= \mathbb{E} \left[\exp \left(\sum_{j=1}^k s_j \beta \lambda \int_{1-t}^1 \mathcal{X}_j^\alpha(v) dv \right) \right], \quad s = (s_1, \dots, s_k) \in (-\infty, 0]^k, \end{aligned} \quad (44)$$

which, combined with (42), yields the convergence $\frac{Z^{(n)}(t^\beta)}{n^{\gamma-\alpha}} \xrightarrow{\mathcal{D}} \beta \lambda \int_0^t U(v) dv$ as $n \rightarrow \infty$. Changing t into $t^{1/\beta}$ and performing the change of variable $v := v^{1/\beta} = v^{1-\alpha}$, we obtain

$$\frac{Z^{(n)}(t)}{n^{\gamma-\alpha}} \xrightarrow{\mathcal{D}} \lambda \int_0^t \frac{U(v^{1-\alpha})}{v^\alpha} dv, \quad n \rightarrow \infty, \quad t \in [0, 1]. \quad (45)$$

Since the \mathcal{S} valued Markov process $\{U(t), t \in [0, 1]\}$ admits $\beta t^{\beta-1} \lambda (P^{(r)} - I)$ as the infinitesimal generator matrix, the time changed Markov process $\{\mathcal{Y}(t) := U(t^{1-\alpha}) = U(t^{1/\beta}), t \in [0, 1]\}$ admits (15) as generator, so that (14) follows from (45).

3.3. Equilibrium case $\gamma = \alpha$

We now proceed to show convergence (10) in Theorem 2. Intuitively, we are in the critical case where customers should arrive just fast enough such that 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 3.2, we first consider t^β instead of t and let

$$\chi^{(n)}(s, t) := \psi^{(n)}(s, t^\beta), \quad t \in [0, 1],$$

the corresponding Laplace transform, where $s = (s_1, \dots, s_k) \in (-\infty, 0]^k$. $t \in [0, 1] \mapsto \chi^{(n)}(s, t)$ then satisfies, thanks to (20), the following differential equation

$$\begin{cases} \partial_t \chi^{(n)}(s, t) &= \beta t^{\beta-1} [\lambda(P - I) + \lambda n^\gamma P_n(\tilde{\pi}_n(s, t^\beta) - I)] \chi^{(n)}(s, t), & t \in [0, 1], \\ \chi^{(n)}(s, 0) &= I. \end{cases} \quad (46)$$

The present case has the same roadmap as Subsection 3.2: 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

$$A_n(s, t) = \beta t^{\beta-1} [\lambda(P - I) + \lambda n^\gamma P_n(\tilde{\pi}_n(s, t^\beta) - I)] \quad (47)$$

then getting a limiting matrix differential equation with solution $\chi(s, t) = \lim_{n \rightarrow \infty} \chi^{(n)}(s, t)$. Then we will identify $\chi(s, t)$ as the Laplace transform of a (conditional) distribution, yielding (10).

Step 1: Determining $A(s, t) = \lim_{n \rightarrow \infty} A_n(s, t)$. We recall that the (x, x) th diagonal element of $\lambda n^\gamma (\tilde{\pi}_n(s, t^\beta) - I)$ is (from (21)) $\sum_{I \subset \{1, \dots, k\}} \prod_{\ell \in I} [(e^{s_\ell x_\ell} - 1) \mathbb{P}[L_\ell > nt]]$, which we decompose as in Section 3.2 as $K_n^1(s, t) + K_n^2(s, t)$ with

$$K_n^1(s, t) = K_n^1(s, t, x) := \beta t^{\beta-1} \lambda n^\gamma \sum_{\ell=1}^k (e^{s_\ell x_\ell} - 1) \mathbb{P}[L_\ell > nt^\beta], \quad (48)$$

$$K_n^2(s, t) = K_n^2(s, t, x) := \beta t^{\beta-1} \lambda n^\gamma \sum_{\text{Card}(I) \geq 2} \prod_{\ell \in I} [(e^{s_\ell x_\ell} - 1) \mathbb{P}[L_\ell > nt^\beta]]. \quad (49)$$

The important point here is that, throughout this subsection, we have $\gamma = \alpha$ in the expressions (47), (48) and (49), which will impact on the convergences and limiting results we are going to prove. 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 (49) is over $\text{Card}(I) \geq 2$, and using the fat tailed property of the service times. The candidate for the continuous function is thus

$$t \in [0, 1] \mapsto A(s, t) := \beta t^{\beta-1} \lambda(P - I) + \beta \lambda \sum_{\ell=1}^k \text{diag}(e^{s_\ell x_\ell} - 1, x = (x_1, \dots, x_k) \in \mathcal{S}). \quad (50)$$

Step 2: Determining $\chi(s, t) = \lim_{n \rightarrow \infty} \chi_n(s, t)$. We now wish to apply Lemma 5 and prove that $\int_0^1 \|A_n(s, v) - A(s, v)\| dv \rightarrow 0$ where $A_n(s, t)$ and $A(s, t)$ are defined in (47) and (50).

The method is very similar as to proving (32) in Step 2 of Section 3.2, as this is equivalent to proving for all $x \in \mathcal{S}$ that

$$\int_0^1 \left| K_n^1(s, v, x) - \beta \lambda \sum_{\ell=1}^k (e^{s_\ell x_\ell} - 1) \right| dv \longrightarrow 0, \quad (51)$$

$$\int_0^1 K_n^2(s, v, x) dv \longrightarrow 0 \quad (52)$$

as $n \rightarrow \infty$. In view of the expression (49) of $K_n^2(s, t)$, (52) 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 Step 2 of Section 3.2. More precisely, it suffices from (49) to prove that

$$\lim_{n \rightarrow \infty} \int_0^1 \beta t^{\beta-1} \lambda n^\alpha \prod_{\ell \in I} \mathbb{P} [L_\ell > nt^\beta] dt = 0 \quad (53)$$

for all $I \subset \{1, \dots, k\}$, $\text{Card}(I) \geq 2$. We prove it for $I = \{\ell_1, \ell_2\}$, $\ell_1 \neq \ell_2$, the proof for $\text{Card}(I) > 2$ being very similar. The trick is again to use that $v^{\beta-1} n^\alpha = (nv^\beta)^\alpha$ as well as the previously established upper bound (35), resulting in

$$\begin{aligned} & \int_0^1 \beta t^{\beta-1} \lambda n^\alpha \mathbb{P} [L_{\ell_1} > nt^\beta] \mathbb{P} [L_{\ell_2} > nt^\beta] dt \\ & \leq \beta \lambda \left[\sup_{j \in \mathbb{N}, v \in [0,1]} (jv^\beta)^\alpha \mathbb{P} [L_{\ell_1} > jv^\beta] \right] \int_0^1 \mathbb{P} [L_{\ell_2} > nt^\beta] dt, \end{aligned}$$

which converges to zero as $n \rightarrow \infty$ by the dominated convergence theorem, proving (53) when $\text{Card}(I) = 2$. As to (51), this is proved, in view of the expression (48) of $K_n^1(s, t)$, by showing that $\int_0^1 \lambda |\beta v^{\beta-1} n^\alpha \mathbb{P} [L_\ell > nv^\beta] - \beta| dv$ 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 3.2 when proving that $\lim_{n \rightarrow \infty} I_n^2(\ell) = 0$, $\ell = 1, \dots, k$, see the arguments leading to the convergence (37). All in all, one has the convergence $\int_0^1 \|A_n(s, v) - A(s, v)\| dv \longrightarrow 0$, and Lemma 5 is applicable so that $\chi^{(n)}(s, t)$ converges to $\chi(s, t)$ which satisfies

$$\begin{cases} \partial_t \chi(s, t) &= A(s, t) \chi(s, t) = [\beta t^{\beta-1} \lambda (P - I) \\ &+ \beta \lambda \sum_{\ell=1}^k \text{diag}(e^{s_\ell x_\ell} - 1, x = (x_1, \dots, x_k) \in \mathcal{S})] \chi(s, t), \quad t \in [0, 1], \\ \chi(s, 0) &= I. \end{cases} \quad (54)$$

Step 3: Identifying the limit in distribution. With the same notation as in Step 3 of Section 3.2 for process $\{\mathcal{X}^\alpha(t) = (\mathcal{X}_1^\beta(t), \dots, \mathcal{X}_k^\beta(t)) \in \mathcal{S}, t \in [0, 1]\}$, one finds this time that

$$\chi(s, t) = \left[\mathbb{E} \left[\mathbf{1}_{[\mathcal{X}^\alpha(1)=y]} \exp \left(\sum_{j=1}^k \beta \lambda \int_{1-t}^1 (e^{s_j \mathcal{X}_j^\alpha(v)} - 1) dv \right) \middle| \mathcal{X}^\alpha(1-t) = x \right] \right]_{(x,y) \in \mathcal{S}^2} \quad (55)$$

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

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

where $\{\nu(x), x \geq 0\}$ is a Poisson process with intensity ξ , see [10, Section 3.2]. Conditioning on $\{\mathcal{X}^\alpha(v), v \in [0, 1]\}$, this results in (55) being written as

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

where $\{\nu_j^\alpha(t), t \geq 0\}$, $j = 1, \dots, k$, are independent Poisson processes with intensities $\beta\lambda = \lambda/(1-\alpha)$, and independent from $\{\mathcal{X}^\alpha(t), t \in [0, 1]\}$. Identifying Laplace Transforms, we obtain in conclusion that

$$\begin{aligned} \mathcal{D} \left(\left(Z^{(n)}(t^\beta), X_{N_{t^\beta}^{(n)}}^{(n)} \right) \middle| X_0^{(n)} \right) \\ \longrightarrow \mathcal{D} \left(\left(\left(\int_{1-t}^1 \mathcal{X}_j^\alpha(v) \, d\nu_j^\alpha(v) \right)_{j=1, \dots, k}, \mathcal{X}^\alpha(1) \right) \middle| \mathcal{X}^\alpha(1-t) \right) \end{aligned} \quad (56)$$

as $n \rightarrow \infty$ for all $t \in [0, 1]$. Changing t into $t^{1/\beta}$ completes the proof of (10).

Proof of the convergence (16) in Corollary 4. This follows the same pattern as the proof of (14), to which we refer here. More precisely, one verifies this time that, from (55), the analog of (44) in the Fast arrival case is here

$$\begin{aligned} \mathbb{E} \left[\exp \left(\sum_{j=1}^k s_j \int_0^t U_j(v) \nu_j^\alpha(dv) \right) \right] \\ = \mathbb{E} \left[\exp \left(\sum_{j=1}^k s_j \int_{1-t}^1 \mathcal{X}_j^\alpha(v) \, d\nu_j^\alpha(v) \right) \right], \quad s = (s_1, \dots, s_k) \in (-\infty, 0]^k, \end{aligned} \quad (57)$$

which, combined with (56), yields the convergence $Z^{(n)}(t^\beta) \xrightarrow{\mathcal{D}} \left(\int_0^t U_j(v) \nu_j^\alpha(dv) \right)_{j=1, \dots, k}$ as $n \rightarrow \infty$. Changing t into $t^{1/\beta}$ and performing the change of variable $v := v^{1/\beta} = v^{1-\alpha}$, we obtain

$$Z^{(n)}(t) \xrightarrow{\mathcal{D}} \left(\int_0^t U_j(v^{1-\alpha}) \tilde{\nu}_j^\alpha(dv) \right)_{j=1, \dots, k}, \quad n \rightarrow \infty, \quad t \in [0, 1], \quad (58)$$

where $\{\tilde{\nu}_j^\alpha(t), t \in [0, 1]\}$, $j = 1, \dots, k$, are the inhomogeneous independent Poisson processes given by $\tilde{\nu}_j^\alpha(t) = \nu_j^\alpha(t^{1/\beta}) = \nu_j^\alpha(t^{1-\alpha})$, i.e. Poisson processes with non constant intensity $\lambda t^{-\alpha}$.

Arguing, as in the Fast arrival case, the time changed Markov process $\{\mathcal{Y}(t) := U(t^{1-\alpha}) = U(t^{1/\beta}), t \in [0, 1]\}$ admits (15) as generator, hence (16) follows from (58).

3.4. Proof of Theorem 3: Slow arriving customers.

We now consider the case $\gamma < \alpha$. Section 4.2 of [16] provide the first two joint moments of the $Z_j^{(n)}(t)$, $j = 1, \dots, k$, with a particular discount factor $a \geq 0$ (recall the notation (2) in Section 1 for the discounted counterpart of the queueing process (1)). Recalling that the rescaling implies that $\lambda_n(P_n - I) = \lambda(P - I)$, we get from [16, Theorems 14 and 15 with $a = 0$ discount factor], that those moments are given by

$$M_j^{(n)}(t) = \lambda_n e^{\lambda t(P-I)} \int_0^t \mathbb{P}(L_j^{(n)} > v) e^{-\lambda v(P-I)} \Delta_j P_n e^{\lambda v(P-I)} dv, \quad (59)$$

$$\begin{aligned} M_{jj}^{(n)}(t) &= \lambda_n e^{\lambda t(P-I)} \int_0^t \mathbb{P}(L_j^{(n)} > v) e^{-\lambda v(P-I)} \Delta_j^2 P_n e^{\lambda v(P-I)} \\ &\quad + 2\mathbb{P}(L_j^{(n)} > v) \Delta_j P_n M_j^{(n)}(v) dv, \end{aligned} \quad (60)$$

$$\begin{aligned} M_{jj'}^{(n)}(t) &= \lambda_n e^{\lambda t(P-I)} \int_0^t \mathbb{P}(L_j^{(n)} > v) \mathbb{P}(L_{j'}^{(n)} > v) e^{-\lambda v(P-I)} \Delta_j \Delta_{j'} P_n e^{\lambda v(P-I)} \\ &\quad + \mathbb{P}(L_j^{(n)} > v) \Delta_j P_n M_{j'}^{(n)}(v) + \mathbb{P}(L_{j'}^{(n)} > v) \Delta_{j'} P_n M_j^{(n)}(v) dv, \end{aligned} \quad (61)$$

for all $t \geq 0$ and $j \neq j'$, j and j' in $\{1, \dots, k\}$. We first show (11). Since $\lambda_n = \lambda n^\gamma$, multiplying (59) by $n^{\alpha-\gamma}$ yields for $j = 1, \dots, k$

$$n^{\alpha-\gamma} M_j^{(n)}(t) = \lambda e^{\lambda t(P-I)} \int_0^t n^\alpha \mathbb{P}(L_j^{(n)} > v) e^{-\lambda v(P-I)} \Delta_j P_n e^{\lambda v(P-I)} dv. \quad (62)$$

By definition of $L_j^{(n)}$ and the fat tail property of L_j :

$$n^\alpha \mathbb{P}(L_j^{(n)} > v) = n^\alpha \mathbb{P}(L_j > nv) \sim n^\alpha \frac{1}{(nv)^\alpha} = \frac{1}{v^\alpha}, \quad v \in (0, t), \quad n \rightarrow \infty.$$

Now, since $\lim_{n \rightarrow \infty} P_n = P$ and

$$\begin{aligned} \sup_{n \in \mathbb{N}} n^\alpha \mathbb{P}(L_j^{(n)} > v) &= \sup_{n \in \mathbb{N}} n^\alpha \mathbb{P}(L_j > nv) = \frac{\sup_{n \in \mathbb{N}} (nv)^\alpha \mathbb{P}(L_j > nv)}{v^\alpha} \\ &\leq \frac{\sup_{u \geq 0} u^\alpha \mathbb{P}(L_j > u)}{v^\alpha}, \quad v \in (0, 1), \end{aligned} \quad (63)$$

the dominated convergence enables us to let $n \rightarrow \infty$ in (62) to get (11). We now turn to (12).

Multiplying (60) by $n^{\alpha-\gamma}$ yields

$$\begin{aligned}
 n^{\alpha-\gamma}M_{jj}^{(n)}(t) &= \lambda e^{\lambda t(P-I)} \int_0^t n^\alpha \mathbb{P}\left(L_j^{(n)} > v\right) e^{-\lambda v(P-I)} \Delta_j^2 P_n e^{\lambda v(P-I)} dv \\
 &\quad + 2\lambda e^{\lambda t(P-I)} \int_0^t n^\alpha \mathbb{P}\left(L_j^{(n)} > v\right) \Delta_j P_n M_j^{(n)}(v) dv. \quad (64)
 \end{aligned}$$

Since (11) in particular implies that $\lim_{n \rightarrow \infty} M_j^{(n)}(v) = 0$ for all $v \in (0, 1)$, and thanks to the upper bound (63), a dominated convergence argument entails that the second integral on the right hand side of (64) tends to 0 as $n \rightarrow \infty$. We also conclude by a dominated convergence argument that the first integral on the right hand side of (64) tends to the right hand side of (12), and we are done. As to (61), we have for $j \neq j'$

$$\begin{aligned}
 n^{\alpha-\gamma}M_{jj'}^{(n)}(t) &= \lambda e^{\lambda t(P-I)} \int_0^t n^\alpha \mathbb{P}\left(L_j^{(n)} > v\right) \mathbb{P}\left(L_{j'}^{(n)} > v\right) e^{-\lambda v(P-I)} \Delta_j \Delta_{j'} P_n e^{\lambda v(P-I)} dv \\
 &\quad + \lambda e^{\lambda t(P-I)} \int_0^t \left\{ n^\alpha \mathbb{P}\left(L_j^{(n)} > v\right) \Delta_j P_n M_{j'}^{(n)}(v) + n^\alpha \mathbb{P}\left(L_{j'}^{(n)} > v\right) \Delta_{j'} P_n M_j^{(n)}(v) \right\} dv. \quad (65)
 \end{aligned}$$

Similarly to the second integral on the right hand side of (64), we show that the second integral on the right hand side of (65) converges to 0 as $n \rightarrow \infty$. As to the first integral, the fact that $\mathbb{P}\left(L_{j'}^{(n)} > v\right) = \mathbb{P}\left(L_{j'} > nv\right) \rightarrow 0$ as $n \rightarrow \infty$, combined with the upper bound (63), yields by the dominated convergence theorem that it tends to 0 as $n \rightarrow \infty$, achieving the proof of (61) and of the theorem.

4. 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 larger or equal to α by obtaining the corresponding limiting joint Laplace transform $\chi(s, t)$ in each case. Even though the distributional limits (9) and (10) involve simple processes $\{\mathcal{X}^\alpha(t), t \in [0, 1]\}$ and $\{\nu_j^\alpha(t), t \geq 0\}$, $j = 1, \dots, k$, it turns out that the Laplace transforms $\chi(s, t)$, which are solutions to the differential equations (39) and (54), 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 $t \in [0, 1] \mapsto \chi(s, t) \in \mathbb{R}^{S \times S}$ into a power series with matrix coefficients, as explained in [1, Section 1.1]. Let us focus on the fast arrival case

in Section 3.2, 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 (39) that $t \in [0, 1] \mapsto \check{\chi}(s, t)$ verifies the matrix differential equation

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

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

$$\check{\chi}(s, t) = \sum_{j=0}^{\infty} U_j(s) t^j, \quad t \in [0, 1], \quad (66)$$

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

$$U_j(s) = \begin{cases} 0, & 1 \leq j < p, \\ Q_2(s) U_{j-p}(s) / j, & p \leq j < q+1, \\ [Q_2(s) U_{j-p}(s) + Q_1 U_{j-q-1}(s)] / j, & j \geq q+1, \end{cases} \quad (67)$$

and that (66) converges for all t , as 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 (67), this expansion for $\chi(s, t)$ is then easy to handle as it can be e.g. approximated by truncation.

References

- [1] Balsler 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] Gaigalas R. and Kaj I. Convergence of scaled renewal processes and a packet arrival model. *Bernoulli*, **9(4)**: 671–703, 2003.
- [8] Hindley D. *Claims Reserving in General Insurance*. Cambridge University Press, 2017.
- [9] Jansen H.M., Mandjes M., De Turck K. and Wittevrongel S. Diffusion limits for networks of Markov-modulated infinite-server queues. *Performance Evaluation*, **135**: 102039, 2019.
- [10] Kingman J.F.C. *Poisson Processes*. Oxford studies in probability, 1993.
- [11] Mandjes M. and De Turck K. Markov-modulated infinite-server queues driven by a common background process. *Stochastic Models*, **32(2)**: 206–232, 2016.
- [12] 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.
- [13] Mikosch T., Resnick S., Rootzén H. and Stegeman A. Is network traffic approximated by stable Lévy motion or fractional brownian motion? *Annals of Applied Probability*, **12(1)**: 23–68, 2002.
- [14] O’Cinneide C., and Purdue P. The $M/M/\infty$ in a random environment. *Journal of Applied Probability*, **23**: 175–184, 1986.
- [15] 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*, **90(3)**: 307–350, 2018.
- [16] Rabehasaina L. and Woo J.-K. Analysis of the infinite server queues with semi-Markovian multivariate discounted inputs. *Queueing Systems, Theory and Applications*, **94(3)**: 393–420, 2020.
- [17] Rogers L.C.G. and Williams D. *Diffusions, Markov Processes, and Martingales: Volume 1, Foundations*. Cambridge mathematical library, 2000.
- [18] Takács L. *Introduction to the theory of queues*. Oxford University press, 1962.