INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Asymptotics and Scalings for Large Closed Product-form Networks via the Central Limit Theorem

Guy FAYOLLE, Jean-Marc LASGOUTTES

N 2754

Février 1996

PROGRAMME 1

Rapport de recherche

ISSN 0249-6399

Asymptotics and Scalings for Large Closed Product-form Networks via the Central Limit Theorem

Guy FAYOLLE, Jean-Marc LASGOUTTES

Programme 1 — Architectures parallèles, bases de données, réseaux
et systèmes distribués
Projet MEVAL

Rapport de recherche n ° 2754 — Février 1996 — 40 pages

Abstract: The asymptotic behaviour of a closed BCMP network, with n queues and m_n clients, is analyzed when n and m_n become simultaneously large. Our method relies on Berry-Esseen type approximations coming in the Central Limit Theorem. We construct *critical sequences* m_n^0, which are necessary and sufficient to distinguish between saturated and non-saturated regimes for the network. Several applications of these results are presented. It is shown that some queues can act as bottlenecks, limiting thus the global efficiency of the system.

This report contains and extends the results obtained in [2].

(Résumé : tsvp)

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : (33 1) 39 63 55 11 – Télécopie : (33 1) 39 63 53 30

Étude asymptotique et charge critique pour les grands réseaux fermés à forme produit via le Théorème de la Limite Centrale

Résumé : On analyse le comportement asymptotique d'un réseau BCMP fermé comportant n files et m_n clients, lorsque n et m_n tendent vers l'infini simultanément. Notre méthode s'apparente aux approximations de type Berry-Esseen, liées au Théorème de la Limite Centrale. On construit des *séquences critiques* m_n^0 qui séparent de façon nécessaire et suffisante les régimes de saturation et de non saturation du réseau. Plusieurs applications sont présentées, incluant des cas d'engorgement local.

Ce travail reprend et étend les résultats présentés dans [2].

1 Introduction

In many applications (telecommunications, transportation, etc.), it is desirable to understand the behaviour and performance of stochastic networks as their size increases. From an engineering point of view, the problem can be roughly formulated as follows:

> *Consider a closed network with* n *nodes and exactly* m_n *customers circulating inside. Find a function* f, *such that* $m = f(n)$ *yields an* interesting *performance of the system as* n *increases.*

In this study, we start from the so-called *product-form* networks, which play an important role in quantitative analysis of systems. Although the equilibrium state probabilities have then a simple expression (see for example Kelly [4]), non-trivial problems remain, due to an intrinsic combinatorial explosion in the formulas, especially in those involving the famous normalizing constant. To circumvent these drawbacks, the idea is to compute asymptotic expansions of the characteristic values of the network, when m and n both tend to infinity.

This approach has been considered by Knessl and Tier [5], Kogan and Birman [6, 7, 1] and Malyshev and Yakovlev [10]. However, it relies on purely analytical tools, which are difficult to use in a more general setting and, in our opinion, do not really give a structural explanation of the phenomena involved.

The method proposed hereafter has direct connections with the *Central Limit Theorem*: instead of representing the values of interest as complex integrals, we express them in terms of distributions of scaled sums of independent random variables. Besides giving a clear interpretation of the computations, this allows to handle directly the general case of single-chain closed networks. We show by construction the existence of *critical sequences* m_n^o in the following sense: the network saturates if, and only if, $m_n \gg m_n^o$. These results can also be interpreted as *insensitivity* properties: as the number of stations n and the number of customers m_n go to infinity, the network is shown to be equivalent to an *open* network of n independent queues (having a total mean number of customers m_n), in the sense that both systems have asymptotically the same finite-dimensional distributions.

The paper is organized as follows. The model is introduced in Section 2, together with a presentation of the method. In Section 3, asymptotics of the marginal distribution of the queue lengths are given under normal conditions and also when some queues become overloaded. Section 4 unifies the results and contains the main theorems about scaling. Section 5 and 6 are devoted to concrete applications of these results, in particular to service vehicle networks (like the *Praxitèle* project, now developed at INRIA). Section 7 contains some conclusive remarks. Most of the technical proofs are postponed in Appendix.

2 Mathematical model and view of the main results

Consider a closed BCMP network $\mathcal{C}_n$ with n queues and m_n clients. The number of clients at queue k at steady state is a random variable $Q_{k,n}$. The service rate at queue k when there are q_k customers is $\mu_{k,n}(q_k)$. The routing probability from queue k to queue ℓ is $p_{k,\ell,n}$ and P_n denotes the transition matrix supposed to be ergodic, with invariant measure $\pi_n = (\pi_{1,n}, \ldots, \pi_{n,n})$, defined by:

$$\pi_n P_n = \pi_n \text{ and } \pi_{1,n} + \cdots + \pi_{n,n} = 1. \tag{2.1}$$

Then it is known that, for any $q_1, \ldots, q_n \geq 0$ such that $q_1 + \cdots + q_n = m_n$,

$$P_n(Q_{1,n} = q_1, \ldots, Q_{n,n} = q_n) = Z_{m_n,n}^{-1} \prod_{k=1}^{n} \frac{\pi_{k,n}^{q_k}}{\mu_{k,n}(1) \cdots \mu_{k,n}(q_k)}, \tag{2.2}$$

with the normalizing condition

$$Z_{m,n} = \sum_{q_1+\cdots+q_n=m} \prod_{k=1}^{n} \frac{\pi_{k,n}^{q_k}}{\mu_{k,n}(1) \cdots \mu_{k,n}(q_k)}. \tag{2.3}$$

It is worth noting that our analysis applies to any network which has a product form equilibrium distribution like (2.2). It includes for example, as soon as the matrix P_n is reversible, all systems having transition rates of the form $p_{k,\ell,n}\alpha_{k,n}(q_k)\beta_{\ell,n}(q_\ell)$, in which case finite capacity situations can be covered (e.g. $\beta_{\ell,n}(q_\ell) = \mathbb{1}_{\{q_\ell \leq \bar{q}_\ell\}}$). See Serfozo [11] for further examples.

To avoid hiding global results with tedious technicalities, we suppose throughout the study that, for all n, $\mathcal{C}_n$ contains at least one queue which, taken in isolation, can be saturated with a finite input flow (e.g. a $M/M/c/\infty$ queue).

The overall presentation requires definitions and an intermediate lemma, given in Section 2.1. The informal presentation of the central results appears in Section 2.2.

2.1 Preliminaries

Define, for each k, $1 \leq k \leq n$, the generating function

$$f_{k,n}(z) \stackrel{\text{def}}{=} \sum_{q=0}^{\infty} \frac{z^q}{\mu_{k,n}(1)\cdots\mu_{k,n}(q)}.$$

Note that for each n, $f_{k,n}$ has a singularity at finite distance for at least one $1 \leq k \leq n$.

To the original closed network $\mathcal{C}_n$, we let correspond a new system $\mathcal{O}_n(\lambda)$ which is open and consists of n parallel queues, with service rates $\mu_{k,n}(x)$ and arrival intensity $\lambda\pi_{k,n}$ at queue k, where the choice of λ will be made more precise later. The queue length $X_{k,n}(\lambda)$ of the k-th queue of $\mathcal{O}_n(\lambda)$ has a distribution given by

$$P(X_{k,n}(\lambda) = x) = \frac{1}{f_{k,n}(\lambda\pi_{k,n})} \frac{(\lambda\pi_{k,n})^x}{\mu_{k,n}(1)\cdots\mu_{k,n}(x)},$$

and $X_{1,n}(\lambda), \ldots, X_{n,n}(\lambda)$ are independent variables. We assume that $X_{k,n}(\lambda)$ has some finite moments of order $r \geq 2$ and introduce the following notation:

$$\begin{array}{rclrcl}
m_{k,n}(\lambda) & \stackrel{\text{def}}{=} & \mathbb{E}X_{k,n}(\lambda), & S_n(\lambda) & \stackrel{\text{def}}{=} & \sum_{k=1}^{n} X_{k,n}(\lambda), \\
\beta_{k,n}^{(r)}(\lambda) & \stackrel{\text{def}}{=} & \mathbb{E}|X_{k,n}(\lambda) - m_{k,n}(\lambda)|^r, & \beta_n^{(r)}(\lambda) & \stackrel{\text{def}}{=} & \sum_{k=1}^{n} \beta_{k,n}^{(r)}(\lambda), \\
\sigma_{k,n}^2(\lambda) & \stackrel{\text{def}}{=} & \beta_{k,n}^{(2)}(\lambda), & \sigma_n^2(\lambda) & \stackrel{\text{def}}{=} & \beta_n^{(2)}(\lambda), \\
\bar{\beta}_{k,n}^{(3)}(\lambda) & \stackrel{\text{def}}{=} & \mathbb{E}[X_{k,n}(\lambda) - m_{k,n}(\lambda)]^3, & \bar{\beta}_n^{(3)}(\lambda) & \stackrel{\text{def}}{=} & \sum_{k=1}^{n} \bar{\beta}_{k,n}^{(3)}(\lambda).
\end{array}$$

Let $\varphi_{k,n}(\theta;\lambda)$ be the characteristic function of $X_{k,n}(\lambda) - m_{k,n}(\lambda)$. Then, for any real θ,

$$\varphi_{k,n}(\theta;\lambda) \stackrel{\text{def}}{=} \mathbb{E}e^{i(X_{k,n}(\lambda)-m_{k,n}(\lambda))\theta} = \frac{f_{k,n}(\pi_{k,n}\lambda e^{i\theta})}{f_{k,n}(\pi_{k,n}\lambda)} e^{-im_{k,n}(\lambda)\theta}, \tag{2.4}$$

and

$$\varphi_n(\theta\,;\lambda) \stackrel{\text{def}}{=} \mathbb{E}e^{i(S_n(\lambda)-\mathbb{E}S_n(\lambda))\theta} = \varphi_{1,n}(\theta\,;\lambda)\cdots\varphi_{n,n}(\theta\,;\lambda) \tag{2.5}$$

The reason why $\mathcal{O}_n(\lambda)$ has been introduced is that the main performance characteristics of the network $\mathcal{C}_n$ can be expressed simply in terms of the distribution of $X_{1,n}(\lambda), \ldots, X_{n,n}(\lambda)$:

Lemma 2.1 (*i*) *For any choice of m_n, there exists a unique λ_n such that*

$$\mathbb{E}S_n(\lambda_n) = \mathbb{E}[X_{1,n}(\lambda_n) + \cdots + X_{n,n}(\lambda_n)] = m_n. \tag{2.6}$$

From now on, unless otherwise stated, all quantities will pertain to the network $\mathcal{O}_n(\lambda_n)$ and λ_n will be omitted.

(*ii*) *Equations (2.2) and (2.3) can be rewritten as*

$$P(Q_{1,n} = q_1, \ldots, Q_{n,n} = q_n) = \frac{1}{P(S_n = m_n)} \prod_{k=1}^{n} P(X_{k,n} = q_k). \tag{2.7}$$

(*iii*) *For any $\ell > 0$ and $q_1, \ldots, q_\ell \geq 0$, the joint distribution of the number of customers in the queues $1, \ldots, \ell$ of $\mathcal{C}_n$ is*

$$\begin{aligned} &P(Q_{1,n} = q_1, \ldots, Q_{\ell,n} = q_\ell) \qquad (2.8)\\ &\quad = \frac{P(S_n - \sum_{k=1}^{\ell} X_{k,n} = m_n - \sum_{k=1}^{\ell} q_k)}{P(S_n = m_n)} \prod_{k=1}^{\ell} P(X_{k,n} = q_k) \\ &\quad = P(X_{1,n} = q_1, \ldots, X_{\ell,n} = q_\ell | S_n = m_n), \end{aligned}$$

and, consequently, $\mathbb{E}Q_{\ell,n} = \mathbb{E}[X_{\ell,n}|S_n = m_n]$.

(*iv*) *For any $1 \leq \ell \leq n$,*

$$\mathbb{E}Q_{\ell,n} = m_{\ell,n}\frac{P(S_n - X_\ell + \widetilde{X}_\ell = m_n)}{P(S_n = m_n)}, \tag{2.9}$$

where $\widetilde{X}_{\ell,n}$ is an integer-valued r.v., independent from everything else and having distribution

$$P(\widetilde{X}_{\ell,n} = x) = \frac{x\,P(X_{\ell,n} = x)}{m_{\ell,n}}.$$

Note that λ_n can be obtained as the unique solution of the equation

$$m_n = \sum_{k=1}^{n} \frac{\pi_{k,n}\lambda_n f'_{k,n}(\pi_{k,n}\lambda_n)}{f_{k,n}(\pi_{k,n}\lambda_n)}. \tag{2.10}$$

While this equation is in general impossible to solve explicitly, λ_n can be computed numerically using classical methods.

Proof A straightforward computation yields, for all $1 \le k \le n$,

$$\frac{\partial m_{k,n}(\lambda)}{\partial \lambda} = \frac{\sigma^2_{k,n}(\lambda)}{\lambda} \ > \ 0. \tag{2.11}$$

The mean number of clients in $\mathcal{O}_n(\lambda)$ is thus a strictly increasing function of λ, which equals zero when $\lambda = 0$ and goes to infinity with λ. This proves the first assertion of the lemma.

Define

$$Y_n \stackrel{\text{def}}{=} \frac{Z_{m_n,n}\lambda_n^{m_n}}{\prod_{k=1}^{n} f_{k,n}(\pi_{k,n}\lambda_n)}.$$

Then (2.2) reads

$$P_n(Q_{1,n} = q_1, \dots, Q_{n,n} = q_n) = \frac{1}{Y_n}\prod_{k=1}^{n} P(X_{k,n} = q_k),$$

which yields (2.7), since

$$Y_n = \sum_{q_1+\cdots+q_n=m_n} \prod_{k=1}^{n} P(X_{k,n} = q_k) = P(X_{1,n} + \cdots + X_{n,n} = m_n).$$

Equation (2.9) and the first part of (2.8) are derived similarly. For the second part of (2.8), we simply note that

$$\begin{aligned}
&P(S_n - \textstyle\sum_{k=1}^{\ell} X_{k,n} = m_n - \sum_{k=1}^{\ell} q_k) \displaystyle\prod_{k=1}^{\ell} P(X_{k,n} = q_k) \\
&\quad = P(S_n = m_n | X_{1,n} = q_1, \dots, X_{\ell,n} = q_\ell) \prod_{k=1}^{\ell} P(X_{k,n} = q_k) \\
&\quad = P(X_{1,n} = q_1, \dots, X_{\ell,n} = q_\ell | S_n = m_n)\, P(S_n = m_n).
\end{aligned}$$

■

2.2 Informal description of the method

Most of the derivations obtained in the paper are based on the various representations given in Lemma 2.1. Whereas the studies [6, 7, 1, 10] use mainly saddle-point methods, our approach relies on direct limit theorems for the distribution of S_n.

For example, assume that $S_n - m_n$ satisfies a local limit theorem such as:

> *Under "suitable" conditions, there exists a distribution with density* f *and a sequence* a_n *such that, for any integer* x,
>
> $$\lim_{n\to\infty} a_n \, P(S_n - m_n = x) - f\Big(\frac{x}{a_n}\Big) = 0. \qquad (2.12)$$

Then Lemma 2.1 will yield

$$P(Q_{1,n} = q_1, \ldots, Q_{n,n} = q_n) \;\approx\; \frac{a_n}{f(0)} \prod_{k=1}^{n} P(X_{k,n} = q_k),$$

and, for any finite ℓ,

$$P(Q_{1,n} = q_1, \ldots, Q_{\ell,n} = q_\ell) \;\approx\; \prod_{k=1}^{\ell} P(X_{k,n} = q_k).$$

This amounts to say that the joint distribution of any *finite* number of queues in the BCMP network $\mathcal{C}_n$ is, at steady state, asymptotically equivalent to the product distribution of the corresponding queues in the system $\mathcal{O}_n$.

It is at this moment important to emphasize that we *do not* require any "smooth" limiting behaviour for $\mathcal{O}_n$, which is somehow an instrumental network, computationally easier to evaluate.

To prove local limit theorems like (2.12), it is necessary to investigate carefully the behaviour of the variables $X_{k,n}$. In particular, since $\mathbb{E}S_n = m_n < \infty$, all queues in $\mathcal{O}_n$ are ergodic, which reads, for any $1 \le k \le n$,

$$\lambda_n \pi_{k,n} < \mu_{k,n} \le \infty,$$

or, equivalently,

$$\rho_n^o \stackrel{\text{def}}{=} \lambda_n \max_{1\le k\le n} \frac{\pi_{k,n}}{\mu_{k,n}} < 1, \tag{2.13}$$

where typically

$$\mu_{k,n} = \varliminf_{q\to\infty} \sqrt[q]{\mu_{k,n}(1)\cdots\mu_{k,n}(q)}.$$

Three main situations have been analyzed:

(*i*) ρ_n^o *is bounded away from* 1: then S_n/σ_n satisfies a local Central Limit Theorem and tends in distribution to a *normal* law (see Theorem 4.2);

(*ii*) $\rho_n^o \to 1$ *and the supremum in (2.13) is attained for a finite number of queues*: then the network subdivides into two subsets, the "saturated" queues and the rest of the network. As shown in Theorem 4.3, under mild regularity assumptions, there exists a sequence α_n such that S_n/α_n tends to a *gamma* law;

(*iii*) $\rho_n^o \to 1$ *and the supremum in (2.13) is attained for an unbounded number of queues*: S_n/σ_n again tends in distribution to a *normal* law (see Theorem 4.4).

In fact, Theorems 4.2, 4.3 and 4.4 quoted above are general, in the sense that they provide a construction of efficient scalings in terms of m_n, the number of customers: the existence of *critical sequences* m_n^o for the network $\mathcal{C}_n$ is shown by explicit construction. Under reasonable assumptions, these sequences are *necessary and sufficient* to discriminate between saturated and non saturated regimes. This is similar to phase transition phenomena observed in [10], where it was assumed that $m_n/n \to \lambda > 0$ (see Section 6.1). Clearly, for a non-saturated regime to exist as $n \to \infty$, it is necessary to have $m_n = O(n)$; this condition is not sufficient (see Section 6.2).

Condition (2.13) can be used to determine an upper bound for λ_n and to exhibit queues which act as bottlenecks in the network $\mathcal{C}_n$ (see Section 4).

Remark Rather than simple limit theorems, the results in Sections 3 and 4 are given in terms of asymptotic expansions, using the operators O and Ω defined as follows:

$$\begin{aligned} a(\eta) &= O(b(\eta)), \quad \text{iff} \quad \exists K>0,\ \forall \eta,\ |a(\eta)| \le K|b(\eta)|, \\ a(\eta) &= \Omega(b(\eta)), \quad \text{iff} \quad a(\eta) = O(b(\eta)) \text{ and } b(\eta) = O(a(\eta)), \end{aligned}$$

where η is some unspecified argument. Unless otherwise stated, all these bounds are *uniform* with respect to n and all queue indexes.

3 Local limit theorems and asymptotic expansions

In this section, we compute estimates of several performance measures of $\mathcal{C}_n$ by means of local limit theorems on sums of independent random variables. The two series of results presented here are of somewhat different nature: whereas the conditions of Proposition 3.1 depend on moments, Proposition 3.3 relies on analytic properties of the generating function of some queues.

3.1 Normal traffic case

When the queues are not saturated (in a sense made more precise in Proposition 3.1), it is possible to prove local Central Limit Theorems, relying more exactly on Berry-Esseen type expansions (see for instance Feller [3]).

Define $\gamma_{k,n}^2$ from $X_{k,n}$ as in Lemma A.1 of the appendix, and let

$$\gamma_n^2 \stackrel{\text{def}}{=} \gamma_{1,n}^2 + \cdots + \gamma_{n,n}^2 \leq \sigma_n^2.$$

Proposition 3.1 (*i*) *Let, for any* $0 < r \leq 1$ *such that* $\beta_n^{(2+r)}$ *exists,*

$$\delta_n^r \stackrel{\text{def}}{=} \frac{1}{2} \frac{\sigma_n^2}{\beta_n^{(2+r)}}.$$

Let $\gamma_n \delta_n \to \infty$ *as* $n \to \infty$. *Then, for any integer* x, *the following approximation holds* uniformly in x*:*

$$\begin{aligned} \sigma_n P(S_n - m_n = x) - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_n^2}} & \qquad (3.1) \\ = O\Big(\frac{\beta_n^{(2+r)}}{\sigma_n^{2+r}}\Big) + O\Big(\frac{\sigma_n}{\gamma_n^2 \delta_n} \exp\Big(-\frac{\gamma_n^2 \delta_n^2}{5}\Big)\Big). \end{aligned}$$

(*ii*) *Let, for any* $0 < r \leq 1$ *such that* $\beta_n^{(3+r)}$ *exists,*

$$\delta_n \stackrel{\text{def}}{=} \frac{1}{2}\frac{\sigma_n^2}{\beta_n^{(3)}}.$$

Let $\gamma_n \delta_n \to \infty$ *as* $n \to \infty$. *Then, for any integer* x, *the following approximation holds* uniformly in x*:*

$$\begin{aligned}\sigma_n P(S_n - m_n = x) - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_n^2}}\left[1 + \frac{\bar{\beta}_n^{(3)}}{6\sigma_n^3}\Big(\frac{x^3}{\sigma_n^3} - 3\frac{x}{\sigma_n}\Big)\right] \\ = O\Big(\frac{\beta_n^{(3+r)}}{\sigma_n^{3+r}}\Big) + O\Big(\frac{\sigma_n}{\gamma_n^2 \delta_n}\exp\Big(-\frac{\gamma_n^2\delta_n^2}{5}\Big)\Big).\end{aligned} \tag{3.2}$$

Proof See Appendix A.2 ■

The main assumption of the previous proposition is classical, since it is nothing else but Lyapounov's condition, popular in the Central Limit Problem:

$$\textit{for some } r > 0, \ \lim_{n\to\infty} \frac{\beta_n^{(2+r)}}{\sigma_n^{2+r}} = 0. \tag{3.3}$$

This condition yields in particular (see e.g. Loève [9])

$$\lim_{n\to\infty} \max_{1\leq k\leq n} \frac{\sigma_{k,n}}{\sigma_n} = 0, \tag{3.4}$$

which in turn implies the *uniform asymptotic negligibility* of the $X_{k,n}$'s. Note that it would be possible by truncation methods to prove similar results without requiring the existence of moments.

We are now in a position to present some basic estimates when the size of the network increases.

Theorem 3.2 *Let* r *be a real number such that* $0 < r \leq 2$. *Assume that* $\sigma_n = O(\gamma_n)$*, that* $\beta_n^{(2+r)}$ *exists and* $\beta_n^{(2+r)}/\sigma_n^{2+r} \to 0$ *as* $n \to \infty$. *Then the following asymptotic expansions hold.*

(*i*)

$$\begin{aligned}P(Q_{1,n} = q_1, \ldots, Q_{n,n} = q_n) \qquad\qquad (3.5)\\ = \sqrt{2\pi}\sigma_n \prod_{k=1}^{n} P(X_{k,n} = q_k)\Big[1 + O\Big(\frac{\beta_n^{(2+r)}}{\sigma_n^{2+r}}\Big)\Big].\end{aligned}$$

(*ii*) *For any finite* ℓ, *if* $[\sum_{j=1}^{\ell} m_{j,n} - q_j]/\sigma_n \to 0$,

$$P(Q_{1,n} = q_1, \ldots, Q_{\ell,n} = q_\ell) = \prod_{k=1}^{\ell} P(X_{k,n} = q_k)\Big[1 + O(\varepsilon_{1,n})\Big], \qquad (3.6)$$

$$\begin{aligned}\varepsilon_{1,n} &= \frac{\beta_n^{(2+r)}}{\sigma_n^{2+r}} + \frac{\sum_{j=1}^{\ell}\sigma_{j,n}^2 + (\sum_{j=1}^{\ell} m_{j,n} - q_j)^2}{\sigma_n^2}\\ &\quad + 1\!\!1_{\{r>1\}}\frac{(\sum_{j=1}^{\ell} m_{j,n} - q_j)\bar\beta_n^{(3)}}{\sigma_n^4}.\end{aligned}$$

(*iii*) *For any* j,

$$\mathbb{E}Q_{j,n} = \mathbb{E}X_{j,n}\Big[1 + O(\varepsilon_{2,n})\Big], \qquad (3.7)$$

$$\varepsilon_{2,n} = \frac{\beta_n^{(2+r)}}{\sigma_n^{2+r}} + \frac{\sigma_{j,n}^2}{\sigma_n^2} + \frac{\beta_{j,n}^{(2+r)}}{m_{j,n}\sigma_n^{1+r}} + 1\!\!1_{\{r>1\}}\frac{\bar\beta_n^{(3)}}{\sigma_n^4}\sigma_{j,n}\Big(1 + \frac{\sigma_{j,n}}{m_{j,n}}\Big).$$

Proof Equation (3.5) is a simple application of Proposition 3.1 to (2.7).

To prove (3.6) from (2.8) when $r \le 1$, we simply write

$$\begin{aligned}&\frac{P(S_n - \sum_{k=1}^{\ell} X_{k,n} = m_n - \sum_{k=1}^{\ell} q_k)}{P(S_n = m_n)}\\ &= \Big(1 - \frac{\sum_{j=1}^{\ell}\sigma_{j,n}^2}{\sigma_n^2}\Big)^{-\frac{1}{2}}\Big[1 + O\Big(\frac{\beta_n^{(2+r)}}{\sigma_n^{2+r}} + e^{-\frac{\left(\sum_{j=1}^{\ell} m_{j,n} - q_j\right)^2}{2\sigma_n^2}} - 1\Big)\Big],\end{aligned}$$

and use the relation $|e^{-u^2/2} - 1| \le u^2$. When $1 < r \le 2$, it suffices to take into account the inequality

$$(u^3 - 3u)e^{-u^2/2} = O(u).$$

Relation (3.7) is also derived from (2.8). ∎

3.2 Heavy traffic case

We proceed now to analyze the behavior of the network $\mathcal{C}_n$ when some queues saturate, as $n \to \infty$. This, in particular, implies that the Lyapounov condition (3.3) is no more valid. In fact, after a suitable normalization, $S_n - m_n$ will be shown to converge in distribution to a random variable having a gamma distribution, under the broad assumption that the first singularities of the relevant generating functions are algebraic.

Let, for some $\rho_n^o \in [0, 1[$ (to be specified in Section 4),

$$\omega_n(\theta) \stackrel{\text{def}}{=} \frac{1-\rho_n^o}{1-\rho_n^o e^{i\theta}},$$

and assume

Assumption A1 *There exists a set $\mathcal{F}_n^o$ of "saturable" queues, for which*

$$\varphi_{k,n}(\theta) = e^{-im_{k,n}\theta}\omega_n^{\xi_{k,n}}(\theta)\psi_{k,n}(\theta),$$

with $\psi'_{k,n}(\theta) = O(1)$, uniformly in k and n. Moreover, there exist a constant $\xi_{\max}$ such that

$$1 \le \xi_{k,n} < \xi_{\max} < \infty.$$

Clearly, the term $\omega_n^{\xi_{k,n}}(\theta)$ coming in the definition of $\varphi_{k,n}(\theta)$ emphasizes the fact that the generating function $f_{k,n}(z)$ pertaining to queue $k \in \mathcal{F}_n^o$ has its first singularity which is algebraic of order $\xi_{k,n}$. If, in addition, $\rho_n^o \to 1$ as $n \to \infty$, the working conditions of the system ensure all queues in $\mathcal{F}_n^o$ saturate so that, in particular, $\mathbb{E}X_{k,n} \sim \xi_{k,n}\alpha_n$, where

$$\alpha_n \stackrel{\text{def}}{=} \frac{\rho_n^o}{1-\rho_n^o}.$$

While this assumption covers a wide range of known queues, it is clear that other types of singularities could be handled via the same method.

Let

$$\xi_n \stackrel{\text{def}}{=} \sum_{k\in\mathcal{F}_n^o} \xi_{k,n}.$$

and define the total characteristic function of the queues in $\mathcal{F}_n \setminus \mathcal{F}_n^o$ by

$$\widehat{\varphi}_n(\theta) \stackrel{\text{def}}{=} \prod_{k \notin \mathcal{F}_n^0} \varphi_{k,n}(\theta).$$

Let r be a real number, $0 < r \leq 1$. Hereafter, $\hat{\sigma}_n$, $\hat{\beta}_n^{(2+r)}$, $\hat{\gamma}_n$ and $\hat{\delta}_n$ will denote quantities having the same meaning as in Proposition 3.1, but related to $\widehat{\varphi}_n(\theta)$.

The counterpart of Theorem 3.2 now reads, in the case of heavy operating conditions:

Proposition 3.3 *Let $\rho_n^o \to 1$. If ξ_n is bounded, $\hat{\sigma}_n/\alpha_n \to 0$ and $\hat{\delta}_n\hat{\gamma}_n \to \infty$ as $n \to \infty$, then the following estimate holds:*

$$\begin{aligned} \alpha_n P(S_n - m_n = x) - \frac{(\xi_n + \frac{x}{\alpha_n})^{\xi_n - 1} e^{-\xi_n + \frac{x}{\alpha_n}}}{\Gamma(\xi_n)} \\ = O\Big(\Big(\frac{\hat{\sigma}_n}{\alpha_n}\Big)^2 + \frac{1}{\alpha_n} + \frac{\hat{\beta}_n^{(2+r)}}{\hat{\sigma}_n^{2+r}}\Big(\frac{\hat{\sigma}_n}{\alpha_n}\Big)^{2+r} + \frac{\hat{\beta}_n^{(2+r)}}{\hat{\sigma}_n^{2+r}}\Big(\frac{\hat{\sigma}_n}{\alpha_n}\Big)^{\xi_n - 1}\Big) \\ + O\Big(\frac{e^{-\frac{\hat{\gamma}_n^2\hat{\delta}_n^2}{5}}}{\hat{\gamma}_n^2\hat{\delta}_n^{\xi_n+1}\alpha_n^{\xi_n-1}}\Big). \end{aligned} \tag{3.8}$$

Proof See Appendix A.2 ■

The estimates of Proposition 3.3 allow to establish the main result of this section.

Theorem 3.4 *Let $\hat{\sigma}_n/\alpha_n \to 0$, $\hat{\beta}_n^{(2+r)}/\hat{\sigma}_n^{2+r} \to 0$ and $\hat{\sigma}_n = O(\hat{\gamma}_n)$ as $n \to \infty$. Then the following expansions hold when ξ_n is uniformly bounded:*

(*i*) *for any $q_1, \ldots, q_n \geq 0$,*

$$\begin{aligned} P(Q_{1,n} = q_1, \ldots, Q_{n,n} = q_n) \\ = \frac{\alpha_n \Gamma(\xi_n)}{e^{-\xi_n}\xi_n^{\xi_n - 1}} \prod_{k=1}^{n} P(X_{k,n} = q_k)\big[1 + O(\varepsilon_n)\big], \end{aligned} \tag{3.9}$$

with

$$\varepsilon_n = \Big(\frac{\hat{\sigma}_n}{\alpha_n}\Big)^2 + \frac{1}{\alpha_n} + \frac{\hat{\beta}_n^{(2+r)}}{\hat{\sigma}_n^{2+r}}\Big(\frac{\hat{\sigma}_n}{\alpha_n}\Big)^{2+r} + \frac{\hat{\beta}_n^{(2+r)}}{\hat{\sigma}_n^{2+r}}\Big(\frac{\hat{\sigma}_n}{\alpha_n}\Big)^{\xi_n - 1};$$

(*ii*) *for any finite* ℓ*, such that* $\mathcal{F}_n^o \cap [1,\ell] = \emptyset$*,*

$$P(Q_{1,n} = q_1, \ldots, Q_{\ell,n} = q_\ell) \tag{3.10}$$
$$= \prod_{k=1}^{\ell} P(X_{k,n} = q_k)\Big[1 + O(\varepsilon_n) + O\Big(\frac{\sum_{k=1}^{\ell} m_{k,n} - q_k}{\alpha_n}\Big)\Big].$$

(*iii*) *for any* $j \notin \mathcal{F}_n^o$*,*

$$\mathbb{E}Q_{j,n} = \mathbb{E}X_{j,n}\Big[1 + O(\varepsilon_n) + O\Big(\frac{\sigma_{j,n}^2 + m_{j,n}^2}{m_{j,n}\alpha_n}\Big)\Big], \tag{3.11}$$

(*iv*) *for any* $j \in \mathcal{F}_n^o$*,*

$$\mathbb{E}Q_{j,n} = \mathbb{E}X_{j,n}\Big[1 + O(\varepsilon_n)\Big], \tag{3.12}$$

Proof The proof of (*i*) -(*iii*) follows essentially along the same lines as for Theorem 3.2, while (*iv*) depends on Equation (2.9) of Lemma 2.1. ∎

4 Scaling

As said in the introduction, this section provides guidelines for using the above technical results in two ways.

- *Quantitative* estimates for the error terms (w.r.t. some limiting distribution), explicitly obtained from the original data (e.g. the total number of customers m_n).
- *Qualitative* understanding of the "critical" values for m_n which, in some sense, induce phase transitions of interest.

The queues are partitioned as follows:

$$\mathcal{F}_n \stackrel{\text{def}}{=} \Big\{0 \le k \le n \ : \ \varliminf_{q\to\infty} \sqrt[q]{\mu_{k,n}(1)\cdots\mu_{k,n}(q)} < \infty\Big\},$$
$$\mathcal{I}_n \stackrel{\text{def}}{=} \Big\{0 \le k \le n \ : \ \varliminf_{q\to\infty} \sqrt[q]{\mu_{k,n}(1)\cdots\mu_{k,n}(q)} = \infty\Big\}.$$

From the general discussion at the beginning of Section 2, $\mathcal{F}_n$ is never empty. Let also

$$\mu_{k,n} \stackrel{\text{def}}{=} \begin{cases} \displaystyle\lim_{q\to\infty} \sqrt[q]{\mu_{k,n}(1)\cdots\mu_{k,n}(q)}, & \text{if } k \in \mathcal{F}_n, \\ \mu_{k,n}(1), & \text{if } k \in \mathcal{I}_n, \end{cases}$$

$$\rho_{k,n} \stackrel{\text{def}}{=} \frac{\lambda_n \pi_{k,n}}{\mu_{k,n}}, \qquad \lambda_n^0 \stackrel{\text{def}}{=} \min_{k\in\mathcal{F}_n} \frac{\mu_{k,n}}{\pi_{k,n}}, \qquad \rho_n^0 \stackrel{\text{def}}{=} \max_{k\in\mathcal{F}_n} \rho_{k,n} = \frac{\lambda_n}{\lambda_n^0}.$$

We shall also need the following subset of $\mathcal{F}_n$:

$$\mathcal{F}_n^0 \stackrel{\text{def}}{=} \{k \in \mathcal{F}_n \ : \ \rho_{k,n} = \rho_n^0\}.$$

Note that the definitions of $\mu_{k,n}$ and ρ_n^0 are consistent with the discussion which lead to (2.13). Moreover, in most practical cases, $\mu_{k,n}(q) \to \mu_{k,n}$ as $q \to \infty$, provided that this limit exists and is finite.

To avoid uninteresting technicalities, it will be convenient to introduce Assumptions **A2** and **A3**, but it should be pointed out that the results of Section 3 are valid in a more general setting. Simple conditions ensuring **A1** and **A3** are discussed in Section 5.

Assumption A2 *The following limit holds:*

$$\lim_{n\to\infty} \max_{1\le k\le n} \frac{\frac{\pi_{k,n}}{\mu_{k,n}}}{\frac{\pi_{1,n}}{\mu_{1,n}} + \cdots + \frac{\pi_{n,n}}{\mu_{n,n}}} = 0.$$

Assumption **A2** is somehow unavoidable to obtain a meaningful asymptotic behaviour of the network. It says that it is possible to let $m_n \to \infty$ as $n \to \infty$, without saturating the network and, under the forthcoming Assumption **A3**, it amounts to Lyapounov's condition (3.3). Note that, when $\mu_{k,n} = \Omega(1)$ uniformly in k and n, **A2** is simply equivalent to

$$\lim_{n\to\infty} \max_{1\le k\le n} \pi_{k,n} = 0.$$

Assumption A3 (*i*) *For any real* $A < 1$ *and any integer* $r \le 4$, *and for any* $k \in \mathcal{F}_n$ *such that* $\rho_{k,n} \le A$,

$$m_{k,n} = \Omega(\rho_{k,n}), \quad \beta_{k,n}^{(r)} = \Omega(\rho_{k,n}), \quad \gamma_{k,n}^2 = \Omega(\rho_{k,n}) \tag{4.1}$$

uniformly in k *and* n.

(*ii*) *(4.1) also holds for all* $k \in \mathcal{I}_n$.

The derivation of the most general results of the section is done in Lemma 4.1 and Theorem 4.2. Further insight, under some additional assumptions, is presented in Theorems 4.3 and 4.4.

Definition *A sequence* m_n^o *is said to be* weakly critical *for* $\mathcal{C}_n$ *if, for any* $0 < t < 1$,

$$g(t) \stackrel{\text{def}}{=} \overline{\lim_{n\to\infty}} \frac{m_n(t\lambda_n^o)}{m_n^o} \tag{4.2}$$

exists and $\lim_{t\to 1-} g(t)$ *be either* 1 *or* ∞.

If, in addition, the relation

$$\lim_{t\to 1-} \underline{\lim_{n\to\infty}} \frac{m_n(t\lambda_n^o)}{m_n^o} = \lim_{t\to 1-} \overline{\lim_{n\to\infty}} \frac{m_n(t\lambda_n^o)}{m_n^o},$$

holds, then the sequence is said to be strongly critical *for* $\mathcal{C}_n$.

Before seeing how such critical sequences can be used, the next lemma proves their existence.

Lemma 4.1 *Under assumption* **A3**, *a convenient weakly critical sequence for* $\mathcal{C}_n$ *is, for some fixed* $0 < u < 1$,

$$m_n^o(u) \stackrel{\text{def}}{=} h_u m_n(u\lambda_n^o), \tag{4.3}$$

where h_u *is correctly chosen.*

Proof Choose $(t, u) \in]0, 1[\times]0, 1[$. From **A3**,

$$m_n(t\lambda_n^o) = \Omega(m_n(u\lambda_n^o)) = \Omega\Big(\sum_{k=1}^{n} \frac{t\lambda_n^o \pi_{k,n}}{\mu_{k,n}}\Big),$$

and the application $t \mapsto m_n(t\lambda_n^o)/m_n(u\lambda_n^o)$ is increasing and locally bounded. Therefore,

$$\hat{g}_u(t) \stackrel{\text{def}}{=} \overline{\lim_{n\to\infty}} \frac{m_n(t\lambda_n^o)}{m_n(u\lambda_n^o)}$$

exists and is increasing. To conclude the proof, take

$$h_u = \begin{cases} \lim\limits_{t\to 1-} \hat{g}_u(t), & \text{if the limit is finite,} \\ 1, & \text{otherwise.} \end{cases}$$

It is interesting to note that, if the above limit is finite for some u, it is finite for all $u \in]0, 1[$. The proof of the lemma is concluded. ■

In fact, as shown in Theorem 4.2, any critical sequence m_n^o acts as a threshold parameter for m_n. Under **A2** and **A3**, which are satisfied by a wide variety of networks, we provide a nearly complete classification in terms of *necessary and sufficient* scaling. It is worth to emphasize that any m_n^o chosen from (4.3) has a *pseudo-explicit* form, given in terms of the data of the original network.

The second step is to enumerate in a consistent way the desirable properties of the distribution of $Q_{1,n}, \dots, Q_{n,n}$: for some finite j and some unspecified ε_n, such that $\varepsilon_n \to 0$ as $n \to \infty$, we have

$$\mathbb{E}Q_{j,n} = \mathbb{E}X_{j,n}\Big[1 + O(\varepsilon_n)\Big], \tag{4.4}$$

$$P(Q_{1,n} = q_1, \dots, Q_{j,n} = q_j) = \prod_{k=1}^{j} P(X_{k,n} = q_k)\Big[1 + O(\varepsilon_n)\Big], \tag{4.5}$$

and also, when Theorem 3.2 [resp. Theorem 3.4] holds, the following equation (4.6) [resp. (4.7)]:

$$\begin{aligned}
P(Q_{1,n} = q_1, \dots, Q_{n,n} = q_n) & \\
&= \sqrt{2\pi}\sigma_n \prod_{k=1}^{n} P(X_{k,n} = q_k)\Big[1 + O(\varepsilon_n)\Big], \qquad (4.6)\\
P(Q_{1,n} = q_1, \dots, Q_{n,n} = q_n) & \\
&= \frac{\alpha_n \Gamma(\xi_n)}{\xi_n^{\xi_n - 1} e^{-\xi_n}} \prod_{k=1}^{n} P(X_{k,n} = q_k)\Big[1 + O(\varepsilon_n)\Big]. \qquad (4.7)
\end{aligned}$$

Theorem 4.2 *Let* **A2** *and* **A3** *hold and* m_n^o *be a weakly critical sequence for* $\mathcal{C}_n$*, with the associated function* $g(t)$*.*

Assume first that $\lim_{t\to 1-} g(t) = 1$*. Then the following classification holds:*

(*i*) *If*

$$\overline{\lim_{n\to\infty}} \frac{m_n}{m_n^o} < 1,$$

then (4.4), (4.5) and (4.6) hold with $\varepsilon_n = 1/m_n$*. In particular, for all* $k \in \{1, \dots, n\}$*,* $\mathbb{E}Q_{k,n}$ *is uniformly bounded in* n*.*

(*ii*) *If*

$$\overline{\lim_{n\to\infty}} \frac{m_n}{m_n^o} > 1,$$

then, for any sequence of queues k_n *in* $\mathcal{F}_n^o$*, we have* $\overline{\lim_{n\to\infty}} \mathbb{E}Q_{k_n,n} = \infty$*.*

(*iii*) *If* m_n^o *is a strongly critical sequence and*

$$\underline{\lim_{n\to\infty}} \frac{m_n}{m_n^o} > 1,$$

then, for any sequence of queues k_n *in* $\mathcal{F}_n^o$*, we have* $\lim_{n\to\infty} \mathbb{E}Q_{k_n,n} = \infty$*.*

In the situation $\lim_{t\to 1-} g(t) = \infty$*, the same results hold, just replacing* "< 1" *(resp.* "> 1"*) in the r.h.s. of the inequalities by* "$< \infty$" *(resp.* "$= \infty$"*).*

Proof To prove (*i*), note that $m_n = m_n(\lambda_n) = m_n(\rho_n^o \lambda_n^o)$. Since $m_n(t\lambda_n^o)$ is increasing in t, this implies that, when $\overline{\lim}_{n\to\infty} \rho_n^o = 1$, we have also $\overline{\lim}_{n\to\infty} m_n/m_n^o \geq 1$. Therefore, in case (*i*), there exists $\tau < 1$ such that $\rho_n^o \leq \tau$ for any $n \in \mathbb{N}$. Using **A3**, we can estimate all error terms coming in Theorem 3.2 and the result is proved.

Similarly in case (*ii*) [resp. (*iii*)], we have necessarily $\overline{\lim}_{n\to\infty} \rho_n^o = 1$ [resp. $\lim_{n\to\infty} \rho_n^o = 1$], and the result follows from the monotonicity of the function $t \mapsto m_{k_n,n}(t\lambda_n^o)$.

The case $\lim_{t\to 1-} g(t) = \infty$ is handled with the same method. ■

Direct applications of Theorem 4.2 are proposed farther on in sections 6.1 and 6.2.

In order to get finer results, the next assumption ensures that the queues not belonging to $\mathcal{F}_n^o$ stay uniformly away from saturation conditions.

Assumption A4 *There exists a constant* $A < 1$ *such that,*

$$\lambda_n^o \frac{\pi_{k,n}}{\mu_{k,n}} \leq A, \quad \textit{for all } k \in \mathcal{F}_n \setminus \mathcal{F}_n^o, \tag{4.8}$$

In order to properly reformulate the results of Section 3, let us define

$$\hat{m}_n(\lambda) \stackrel{\text{def}}{=} \sum_{k \notin \mathcal{F}_n^o} m_{k,n}(\lambda), \tag{4.9}$$

$$\hat{m}_n^o \stackrel{\text{def}}{=} \hat{m}_n(\lambda_n^o). \tag{4.10}$$

Using (2.11), it is not difficult to see that $\hat{m}_n^o$ defined (4.10) is a strongly critical sequence for $\mathcal{C}_n$ under **A1**, **A2**, **A3** and **A4**. Therefore, all results of Theorem 4.2 hold, as well as the following:

Theorem 4.3 *Let* **A1**, **A2**, **A3** *and* **A4** *hold. If* ξ_n *is uniformly bounded, then the following results hold:*

(*i*) *If there exists* $\theta_n > 0$, *such that, for all* $n \in \mathbb{N}$,

$$\frac{m_n}{\hat{m}_n^o} \leq 1 - \theta_n,$$

and $\lim_{n\to\infty} \theta_n^2 m_n = \infty$*, then (4.4), (4.5) and (4.6) hold with*

$$\varepsilon_n \stackrel{\text{def}}{=} \frac{1}{m_n} + \frac{1}{m_n^2\theta_n^4},$$

except when a queue in $\mathcal{F}_n^o$ *is concerned, in which case (4.4) and (4.5) hold with*

$$\varepsilon_n = \frac{1}{\theta_n^2 m_n}.$$

(*ii*) *If there exists* $\theta_n > 0$*, such that, for all* $n \in \mathbb{N}$*,*

$$\frac{m_n}{\hat{m}_n^o} \geq 1 + \theta_n,$$

and $\lim_{n\to\infty} \theta_n \hat{m}_n^o = \lim_{n\to\infty} \theta_n^2 \hat{m}_n^o = \infty$*, then (4.4) and (4.7) hold with*

$$\varepsilon_n \stackrel{\text{def}}{=} \frac{1}{\theta_n^2 \hat{m}_n^o} + \frac{1}{\theta_n \hat{m}_n^o} + \frac{1}{\sqrt{\hat{m}_n^o}}\left[\frac{1}{\hat{m}_n^o \theta_n^2}\right]^{\frac{\varepsilon_n - 1}{2}}.$$

Moreover, if in Equation (4.5), $[1, j] \cap \mathcal{F}_n^o = \emptyset$*, then the latter also holds, with* ε_n *having the above value.*

Proof To prove (*i*), note that when $m_n \leq (1 - \theta_n)\hat{m}_n^o$,

$$\hat{m}_n(\lambda_n) \leq m_n(\lambda_n) \leq (1 - \theta_n)\hat{m}_n(\lambda_n^o).$$

Moreover, using **A3**, **A4** and (2.11), Taylor's formula yields, for some $\lambda \in]\lambda_n, \lambda_n^o[$,

$$\begin{aligned}
\hat{m}_n^o - \hat{m}_n(\lambda_n) &= \hat{m}_n(\lambda_n^o) - \hat{m}_n(\lambda_n) \\
&= (\lambda_n^o - \lambda_n)\frac{\hat{\sigma}_n^2(\lambda)}{\lambda} \\
&= (\lambda_n^o - \lambda_n)\Omega\Big(\sum_{k\in\mathcal{F}_n\setminus\mathcal{F}_n^o} \frac{\pi_{k,n}}{\mu_{k,n}}\Big),
\end{aligned}$$

which implies
$$1 - \frac{\hat{m}_n(\lambda_n)}{\hat{m}_n^o} = \Omega(1-\rho_n^o) \geq \theta_n.$$

Hence,
$$\frac{1}{1-\rho_n^o} = O\Big(\frac{1}{\theta_n}\Big)$$
and, using $\hat{m}_n(\lambda_n) = \Omega(\hat{\beta}_n^{(r)}) = \Omega(\hat{\sigma}_n^2)$, a direct but tedious computation shows that Theorem 3.2 applies with appropriate error terms.

Let us now prove assertion (*ii*) . It follows from

$$m_n - \hat{m}_n(\lambda_n) = \Omega\Big(\frac{\xi_n \rho_n^o}{1-\rho_n^o}\Big) \geq \theta_n \hat{m}_n^o \to \infty,$$

that $\rho_n^o \to 1$ and

$$\begin{aligned}
\frac{\hat{\sigma}_n^2}{\alpha_n^2} &= \Omega(\hat{m}_n(\lambda_n)(1-\rho_n^o)^2) \\
&= O\Big(\frac{\hat{m}_n(\lambda_n)}{(m_n - \hat{m}_n(\lambda_n))^2}\Big) = O\Big(\frac{\hat{m}_n(\lambda_n)}{\theta_n^2[\hat{m}_n^o]^2}\Big) = O\Big(\frac{1}{\theta_n^2 \hat{m}_n^o}\Big).
\end{aligned}$$

Thus, Theorem 3.4 applies and (*ii*) is proved. ■

It remains to state what happens when $\xi_n \to \infty$ as $n \to \infty$. As shown below, this behaviour does not depend on the saturation of the queues in $\mathcal{F}_n^o$.

Theorem 4.4 *Let* $\xi_n \to \infty$ *as* $n \to \infty$*. Let also* **A1***,* **A2***,* **A3** *and* **A4** *hold. Then, under the uniformity assumption*

$$\beta_{k,n}^{(4)} = O\left(\frac{\xi_{k,n}\rho_n^o}{(1-\rho_n^o)^4}\right), \textit{ for all } k \in \mathcal{F}_n^o, \tag{4.11}$$

the results (4.4), (4.5) and (4.6) are again valid, with

$$\varepsilon_n \stackrel{\text{def}}{=} \frac{1}{(1-\rho_n^o)m_n}.$$

Proof The statement relies on Theorem 3.2, taking $r = 2$. First, from classical weak compactness and moment convergence theorems (see e.g. [9]), it follows that, for $k \in \mathcal{F}_n^o$ and all $0 < s \leq 4$

$$\beta_{k,n}^{(s)} = \Omega\left(\frac{\xi_{k,n}\rho_n^o}{(1-\rho_n^o)^s}\right).$$

Thus, the term coming in Lyapounov's condition (3.3) is equal to

$$\begin{aligned}
\frac{\beta_n^{(4)}}{\sigma_n^4} &= \Omega\left(\frac{\hat{m}_n + \frac{\rho_n^o\xi_n}{(1-\rho_n^o)^4}}{\left[\hat{m}_n + \frac{\rho_n^o\xi_n}{(1-\rho_n^o)^2}\right]^2}\right) \\
&= \Omega\left(\frac{(1-\rho_n^o)^4\hat{m}_n + \rho_n^o\xi_n}{[(1-\rho_n^o)^2\hat{m}_n + \rho_n^o\xi_n]^2}\right) \\
&= O\left(\frac{1}{(1-\rho_n^o)\hat{m}_n + \rho_n^o\xi_n}\right) \\
&= O\left(\frac{1}{(1-\rho_n^o)m_n}\right),
\end{aligned}$$

which tends to 0 as $n \to \infty$. The other error terms given in Theorem 3.2 are estimated in the same way.

The only thing left to check is that $\sigma_n^2 = O(\gamma_n^2)$. In fact, since $\gamma_n^2 = \Omega(\hat{m}_n)$, this relation will only hold when ρ_n^o is uniformly bounded away from 1. However, for any $k \in \mathcal{F}_n^o$ and for any $\theta \in [-\pi, \pi]$,

$$\begin{aligned}
|\varphi_{k,n}(\theta)| &= |\omega_{k,n}(\theta)|^{\xi_{k,n}}\Big|1 + O(\theta)\Big| \\
&\leq \left[\frac{1}{1+\frac{\alpha_n^2\theta^2}{6}}\right]^{\frac{\xi_{k,n}}{2}}\Big|1 + O(\theta)\Big| \\
&\leq \left[\frac{1}{1+\frac{\alpha_n^2\theta^2}{6}}\right]^{\frac{\xi_{k,n}}{4}},
\end{aligned}$$

provided that $a < \rho_n^o < 1$, where a is some fixed constant. This bound can be used to replace Equation (A.2) in the proof of Proposition 3.1 by

$$\begin{aligned} \left| \int_{\delta_n \le |\theta| \le \pi} e^{-i\theta x} \varphi_n(\theta) d\theta \right| &\le \int_{|\theta| \ge \delta_n} \left[\frac{1}{1 + \frac{\alpha_n^2 \theta^2}{6}} \right]^{\frac{\xi_n}{4}} d\theta \\ &= O\left(\frac{1}{\delta_n \alpha_n^2 \xi_n} \frac{1}{(1 + \alpha_n^2 \delta_n^2)^{\frac{\xi_n}{4} - 1}} \right), \end{aligned}$$

which is exponentially small in ξ_n, since $\delta_n \alpha_n = \Omega(1)$. ∎

5 Towards more tangible assumptions

The assumptions used in the results of the previous section may seem difficult to check in practice. However, as shown hereafter, they can be replaced (at the expense of a loss in generality) by simpler properties directly related to the service mechanisms of the queues.

The next lemma provides a realistic context in which **A3** is satisfied.

Lemma 5.1 *Assume that*

(*i*) *there exist sequences* $R(q)$ *and* $T(q)$ *such that*

$$\lim_{q \to \infty} \sqrt[q]{R(1) \cdots R(q)} = 1,$$

$$\lim_{q \to \infty} T(q) = \infty,$$

and, for any $q > 0$*,*

$$\begin{aligned} \mu_{k,n}(q) &\ge R(q)\mu_{k,n}, \quad \textit{for } k \in \mathcal{F}_n, \\ \mu_{k,n}(q) &\ge T(q)\mu_{k,n}, \quad \textit{for } k \in \mathcal{I}_n; \end{aligned}$$

(*ii*) *there exists a constant* $B < \infty$ *such that*

$$\lambda_n^o \frac{\pi_{k,n}}{\mu_{k,n}} < B, \textit{ for all } k \in \mathcal{I}_n.$$

Then **A3** *holds.*

Remark This lemma can be applied in particular to any mixing of $M/M/\infty$ and multiple-server queues with at most c servers, with

$$R(q) = \min\left[1, \frac{q}{c}\right], \quad T(q) = q.$$

Proof For each queue $k \in \mathcal{F}_n$ such that $\rho_{k,n} \le A$, and for all $r \in \mathbb{N}$, we have

$$\sum_{q=0}^{\infty} q^r \frac{(\lambda_n \pi_{k,n})^q}{\mu_{k,n}(1)\cdots\mu_{k,n}(q)} \le \sum_{q=0}^{\infty} \frac{q^r A^q}{R(1)\cdots R(q)} < \infty.$$

In particular, $f_{k,n}(\lambda_n\pi_{k,n}) = \Omega(1)$ and

$$m_{k,n} = \frac{\lambda_n \pi_{k,n}}{\mu_{k,n}(1) f_{k,n}(\lambda_n \pi_{k,n})} \sum_{q=1}^{\infty} q \frac{(\lambda_n \pi_{k,n})^{q-1}}{\mu_{k,n}(2)\cdots\mu_{k,n}(q)} = \Omega\left(\frac{\lambda_n \pi_{k,n}}{\mu_{k,n}(1)}\right).$$

Similarly, for any $r \in \mathbb{N}$,

$$\beta_{k,n}^{(r)} = \Omega\left(\frac{\lambda_n \pi_{k,n}}{\mu_{k,n}(1)}\right).$$

The same computations can be applied to $k \in \mathcal{I}_n$, thus proving **A3**-(*ii*). ■

The results of Section 4 can be easily generalized to a situation where some $M/M/\infty$ queues of $\mathcal{I}_n$ become saturated, in which case **A3**-(*ii*) is no longer satisfied. Indeed, the characteristic function of the number of clients X in an $M/M/\infty$ queue with parameter ρ can be written as

$$\mathbb{E}e^{i\theta X} = \exp\Big(\rho(e^{i\theta} - 1)\Big) = \left[\exp\Big(\frac{\rho}{\lfloor \rho \rfloor}(e^{i\theta} - 1)\Big)\right]^{\lfloor \rho \rfloor},$$

which means that a saturated infinite server queue can be replaced by several non-saturated infinite-server queues without changing the distribution of S_n. Therefore, the results of Section 4 still hold, except for marginal distributions containing one of the saturated queues.

Theorems 4.3 and 4.4 also required assumption **A1** on the service mechanisms of the so-called "saturable" queues. It is often enough to restrict ourselves to the following two categories of queues, which encompass the standard $M/M/c$ queue.

Lemma 5.2 *Assume that, for any* $k \in \mathcal{F}_n^o$, *either*

(i) there is a constant q_c, *independent of* k *and* n, *such that*

$$\frac{\mu_{k,n}(q)}{\mu_{k,n}} = \begin{cases} O(1), & \textit{if } q < q_c, \\ 1, & \textit{otherwise.} \end{cases} \tag{5.1}$$

or

(ii) for some finite constants $\xi_{\min}$ *and* $\xi_{\max}$,

$$\log \frac{\mu_{k,n}(q)}{\mu_{k,n}} = -\frac{\xi_{k,n} - 1}{q} + \Delta_{k,n}(q), \tag{5.2}$$

with

$$\Delta_{k,n}(q) = O\left(\frac{1}{q^2}\right), \qquad 1 < \xi_{\min} \le \xi_{k,n} \le \xi_{\max} \,,$$

uniformly in k *and* n. *(See also Section 7).*

Then **A1** *holds.*

Proof In view of Equation (2.4), for any fixed k and n, the quantity to estimate is related to

$$\begin{aligned} f_{k,n}(\lambda_n \pi_{k,n} e^{i\theta}) &= \sum_{q=0}^{\infty} \frac{(\lambda_n \pi_{k,n} e^{i\theta})^q}{\mu_{k,n}(1) \cdots \mu_{k,n}(q)} \\ &= \sum_{q=0}^{\infty} \prod_{p=1}^{q} \frac{\mu_{k,n}}{\mu_{k,n}(p)} (\rho_{k,n} e^{i\theta})^q. \end{aligned}$$

For the sake of brevity, let us omit the k and n subscripts and define, for any $z \in \mathbb{C}$, $|z| < 1$,

$$g(z) \stackrel{\text{def}}{=} \sum_{q=0}^{\infty} \prod_{p=1}^{q} \frac{\mu}{\mu(p)} z^q.$$

Thus, we have to estimate $g(\rho e^{i\theta})/g(\rho)$, for $\theta \in [-\pi, \pi]$ and $\rho < 1$. This proof proceeds in steps:

a) Assume first that (5.1) holds. Then, for

$$g(z) = \frac{1}{1-z}\left[(1-z)\sum_{q=0}^{q_c}\prod_{p=1}^{q}\frac{\mu}{\mu(p)}z^q + z^{q_c+1}\prod_{p=1}^{q_c}\frac{\mu}{\mu(p)}\right],$$

and Assumptions **A1** holds with $\xi = 1$.

b) Under (5.2), one obtains, for $q \geq 1$,

$$\begin{aligned}\prod_{p=1}^{q}\frac{\mu}{\mu(p)} &= \exp\left[(\xi-1)\sum_{p=1}^{q}\frac{1}{p} - \sum_{p=1}^{q}\Delta(p)\right] \\ &= \exp[(\xi-1)C - \Delta]\cdot q^{\xi-1}[1+\tfrac{a_q}{q}],\end{aligned}$$

where C is the Euler constant, $\Delta \stackrel{\text{def}}{=} \sum_{p=1}^{\infty}\Delta(p)$, and a_q is uniformly bounded. In the remainder of the proof, let

$$K \stackrel{\text{def}}{=} \exp[(\xi-1)C - \Delta].$$

c) Let, for $|z| < 1$ and $s \in \mathbb{C}$,

$$\phi(z,s) \stackrel{\text{def}}{=} \sum_{q=1}^{\infty}\frac{z^q}{q^s}.$$

Then, for $\mathrm{Re}(s) > 0$,

$$\phi(z,s) = \frac{z}{\Gamma(s)}\int_0^{\infty}\frac{t^{s-1}dt}{e^t - z}.$$

In fact, this integral representation can be used to get an analytic continuation with respect to s, by introducing the (classical) Hankel's contour. This yields, for all $|z| < 1$ and $\mathrm{Re}(s) > 0$,

$$\phi(z,s) = \frac{i\Gamma(1-s)}{2\pi}\int_{\mathcal{L}}\frac{(-t)^{s-1}dt}{e^t - z}.$$

Distorting $\mathcal{L}$ to include the zeros of $e^t - z$, the following expression holds, for $\mathrm{Re}(s) < 0$ and all values of z such that $|\arg(-\log z + 2in\pi)| \leq \pi$:

$$\phi(z,s) = \Gamma(1-s)\sum_{n\in\mathbb{Z}}(-\log z + 2in\pi)^{s-1}.$$

d) Using this expression, simple computations yield, when $\xi > 1$ and $|z| < 1$

$$\begin{aligned} g(z) &= 1 + K\left[\phi(z, 1-\xi) + \sum_{q=1}^{q} q^{\xi-2}a_q z^q\right] \\ &= \frac{K}{\log^{\xi} z}\left[\frac{\log^{\xi} z}{K} + 1 + \sum_{n\neq 0}\left(\frac{\log z}{\log z - 2in\pi}\right)^{\xi}\right. \\ &\qquad \left. + \log^{\xi} z \sum_{q=1}^{q} q^{\xi-2}a_q z^q\right], \end{aligned}$$

and, finally,

$$\frac{g(\rho e^{i\theta})}{g(\rho)} = \left[\frac{1-\rho}{1-\rho e^{i\theta}}\right]^{\xi}\left[1 + \xi\rho(e^{i\theta}-1)\right].$$

This concludes the proof of the lemma. ∎

6 Applications

6.1 A Jackson network with convergence properties

Consider the basic Jackson network (consisting of $M/M/1$ queues with constant service rates) analyzed in [10].

In this case,

$$m_n(t\lambda_n^o) = \sum_{k=1}^{n}\frac{tr_{k,n}}{1-tr_{k,n}}, \quad \text{with } r_{k,n} = \frac{\lambda_n^o \pi_{k,n}}{\mu_{k,n}}.$$

Under the assumption made in [10] that the counting measure

$$I_n(A) \stackrel{\text{def}}{=} \frac{1}{n}\mathrm{Card}(k : r_{k,n} \in A),$$

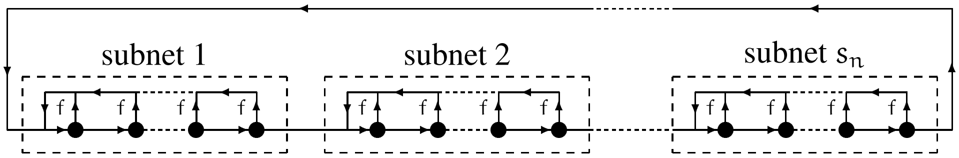

Figure 1: a compound network of tandem queues

defined for all Borel sets A, converges weakly to a probability measure I, we have

$$\lim_{n\to\infty} \frac{m_n(t\lambda_n^o)}{n} = \int_0^1 \frac{tr}{1-tr} dI(r),$$

and

$$\lim_{t\to 1-} \int_0^1 \frac{tr}{1-tr} dI(r) \stackrel{\text{def}}{=} \lambda_{cr} \leq \infty.$$

Thus, the results of [10] are contained in the theorems of Section 4, taking $m_n^o = n\lambda_{cr}$, which is then a *strongly critical sequence* for $\mathcal{C}_n$.

6.2 A network with tight bottlenecks

As pointed out in the introduction, there are cases of interest with $m_n = o(n)$. This will be illustrated in the next example.

Consider a closed network consisting of s_n subnetworks of $M/M/1$ queues having each a unique entry point, in which a fixed number m of tasks circulate. The queues are subject to failures, taking place with some probability $f < 1$. When a failure occurs, the task returns to the entry point of its current subnetwork. Tasks visit the various subnetworks according to some probability matrix.

This model exhibits tight bottlenecks, when the number and the size of the subnetworks grow. This fact, for the sake of simplicity, will be illustrated on a very simple topology, presented in Figure 1: all subnetworks are associated in tandem, and each of them consists itself of ℓ_n queues in tandem, with unit processing rates.

Here, the invariant measure of the routing matrix has the form

$$(\pi_{1,n}, \dots, \pi_{\ell_n,n}; \pi_{1,n}, \dots; \dots, \pi_{\ell_n,n}),$$

where $\pi_{k,n}$ is the invariant probability associated to the k-th queue of an arbitrary subnetwork. A straightforward computation, using symmetry properties, yields, for any $t \in]0,1[$,

$$\pi_{k,n} = \frac{1}{s_n}\frac{f(1-f)^{k-1}}{1-(1-f)^{\ell_n}} = (1-f)^{k-1}\pi_{1,n},$$

$$m_n(t\lambda_n^o) = s_n \sum_{k=1}^{\ell_n} \frac{t(1-f)^{k-1}}{1-t(1-f)^{k-1}}.$$

Choosing some fixed $u \in]0,1[$ and assuming that $\ell_n \to \infty$ as $n \to \infty$, we have

$$\lim_{n\to\infty} \frac{m_n(t\lambda_n^o)}{m_n(u\lambda_n^o)} = \frac{L_f(t)}{L_f(u)},$$

where L_f is defined on $]0,1[$ as

$$L_f(t) \stackrel{\text{def}}{=} \sum_{k=1}^{\infty} \frac{t(1-f)^{k-1}}{1-t(1-f)^{k-1}}$$

and $\lim_{t\to 1^-} L_f(t) = \infty$.

Therefore, $m_n(u\lambda_n^o)$ is a strongly critical sequence for the network and the size of the queues remain uniformly bounded if, and only if,

$$m_n = O\left(m_n(u\lambda_n^o)\right) = O(s_n) = o(n).$$

6.3 A service vehicle network

Consider a fleet of vehicles serving an area consisting of n stations forming a fully connected graph. These vehicles are used to transport goods or passengers. Vehicles wait at stations until they receive a request, in which case they go to an other station. The routing among stations is done according to some routing matrix P_n. When a request arrives to an empty station, it is immediately lost. The request arrivals form a Poisson stream at each queue.

We model this system as follows: for all $0 \le k \le n$, station k is represented as a single-server queue with service rate $\mu_{k,n}$ which is equal to the arrival rate at station k, since arrivals are lost when the station is empty. When a vehicle leaves station k, it chooses its destination according to the Markovian routing matrix $P_n = (p_{k\ell,n})$. The duration of the journey between two stations k and ℓ is represented by an infinite server queue placed on the edge between them. The service rate of this queue when there are q vehicles traveling between k and ℓ is $q\mu_{k\ell,n}$. Note that, contrary to the convention used throughout this paper, the total number of *queues* is $n^2 + n$. Let $(\pi_{1,n}, \dots, \pi_{n,n})$ be the invariant measure of P_n, defined as in (2.1). Then, with obvious notation, for all $k, \ell \in [1, n]$, for all $\theta \in [-\pi, \pi]$,

$$
\begin{aligned}
\rho_{k,n} &\stackrel{\text{def}}{=} \frac{\lambda_n \pi_{k,n}}{2\mu_{k,n}}, & \rho_{k\ell,n} &\stackrel{\text{def}}{=} \frac{\lambda_n \pi_{k,n} p_{k\ell,n}}{2\mu_{k\ell,n}}, \\
m_{k,n} &\stackrel{\text{def}}{=} \frac{\rho_{k,n}}{1-\rho_{k,n}}, & m_{k\ell,n} &\stackrel{\text{def}}{=} \rho_{k\ell,n}, \\
\varphi_{k,n}(\theta) &\stackrel{\text{def}}{=} \frac{(1-\rho_{k,n})e^{-im_{k,n}\theta}}{1-\rho_{k,n}e^{i\theta}}, & \varphi_{k\ell,n}(\theta) &\stackrel{\text{def}}{=} e^{\rho_{k\ell,n}(e^{i\theta}-1-i\theta)}.
\end{aligned}
$$

Define $\mathcal{F}_n^o$ as in Section 4 and assume that its cardinal is some fixed integer $K \geq 1$. Lemmas 5.1 and 5.2 apply, taking $R(q) = 1$, $T(q) = q$ and $\xi_{k,n} = 1$ for $q \geq 1$ and $k \in \mathcal{F}_n^o$. Thus, when **A4** holds, Theorem 4.3 can be used and estimates of many performance measures can be derived, with corresponding error terms.

Some questions of interest arise:

- which maximal efficiency can be expected from this system?
- how many vehicles should be provided?

To answer these questions, it is convenient to define the *loss probability* as

$$
\mathcal{P}_{\text{loss}}(n) \stackrel{\text{def}}{=} \frac{\sum_{k=1}^{n} \mu_{k,n} P(Q_{k,n} = 0)}{\sum_{k=1}^{n} \mu_{k,n}}.
$$

$\mathcal{P}_{\text{loss}}(n)$ is the proportion of customers that are lost because they arrive at an empty station. This is a good indicator of the quality of service provided by the network. Under appropriate conditions as $n \to \infty$:

$$
\begin{aligned}
\mathcal{P}_{\text{loss}}(n) &\sim \frac{\sum_{k=1}^{n} \mu_{k,n} P(X_{k,n} = 0)}{\sum_{k=1}^{n} \mu_{k,n}} \\
&\sim 1 - \frac{\lambda_n}{2\sum_{k=1}^{n} \mu_{k,n}}.
\end{aligned}
\tag{6.1}
$$

The last expression is a decreasing function of λ_n, which is itself bounded by λ_n^o. Therefore, the minimum loss probability is attained when $\lambda_n \to \lambda_n^o$; this happens with

$$m_n = (1 - \theta_n)\hat{m}_n^o, \quad \lim_{n\to\infty} \theta_n = 0,$$

where θ_n is chosen to satisfy the assumptions of Theorem 4.3-(*i*) . With this choice of m_n, (6.1) holds with

$$\lambda_n = \lambda_n^o(1 + O(\theta_n)),$$

which is asymptotically optimal. Consequently, a "good" value for m_n is $m_n = \hat{m}_n^o$, and having a number of vehicle proportional to the number of stations can be a poor choice, especially when some stations are more loaded than others. These stations act as *bottlenecks* of the system, which should be removed by altering the routing probabilities.

7 General remarks

First, a chief difficulty of the analysis is due to the need of dealing with rate of convergence and limits of *densities*: this is the field of Berry-Esseen theorems and large deviations.

Secondly, the results have been obtained under several technical assumptions (especially *uniformity*), which in some sense are unavoidable. This means precisely that the choice of conditions slightly different from **A1**, **A3** and **A4** would have led to different families of limit laws having infinitely divisible distributions.

In particular, from a physical point of view, it is worth commenting on equation (5.2). The inequality $\xi_{k,n} \geq 1$ implies that the maximum service rate of the queues in $\mathcal{F}_n^o$ is reached from below; this is not the case if $0 < \xi_{k,n} < 1$, and the analysis was omitted, since the technicalities involved would have made the text unnecessarily obscure. At last, the case $\xi_{k,n} \leq 0$ dealing with other types of singularities (for instance logarithmic), was not carried out, and would yield other limit laws.

The future class of problems of interest concerns some non-product form networks.

A Appendix

A.1 A bound on periodic characteristic functions

One of the problems arising in the computation of convergence rates in the Central Limit Theorem is to find upper bounds on the modulus of a characteristic function $\varphi(\theta)$ for θ away from 0. One typical property used can be stated as follows:

there exist $\theta_0 > 0$ *and* $a < 1$ *such that, for all* $|\theta| > \theta_0$, $|\varphi(\theta)| < a$.

It is pointed out in Feller [3] that this condition is usually easy to fulfill in practice, as long as X does not have a lattice distribution. Unfortunately, we are in the lattice case and thus must cope with the periodicity of φ.

Next lemma shows how a bound on $|\varphi(\theta)|$ can be derived for $|\theta| \leq \pi$.

Lemma A.1 *Let* X *be an integer-valued random variable with distribution* $P(X = k) = p_k$, $k \in \mathbb{N}$. *Define*

$$\gamma^2 \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \frac{p_{2k}p_{2k+1}}{p_{2k} + p_{2k+1}} \leq \min\Big(\operatorname{Var} X, \frac{1}{4}\Big),$$

where the summands are taken to be zero when $p_{2k} = p_{2k+1} = 0$. *Then, for any* $\theta \in [-\pi, \pi]$, *the characteristic function* φ *of* X *satisfies:*

$$|\varphi(\theta)| \leq \exp\Big(-\frac{\gamma^2}{5}\theta^2\Big). \tag{A.1}$$

Proof We have

$$|\varphi(\theta)| = \Big|\sum_{k=0}^{\infty} p_k e^{ik\theta}\Big| \leq \sum_{k=0}^{\infty} \Big|p_{2k} + p_{2k+1}e^{i\theta}\Big|.$$

Moreover,

$$\begin{aligned}
\Big|p_{2k} + p_{2k+1}e^{i\theta}\Big| &= \sqrt{(p_{2k} + p_{2k+1}\cos\theta)^2 + p_{2k+1}^2 \sin^2\theta} \\
&= \sqrt{(p_{2k} + p_{2k+1})^2 - 2p_{2k}p_{2k+1}(1 - \cos\theta)} \\
&\leq p_{2k} + p_{2k+1} - \frac{p_{2k}p_{2k+1}}{p_{2k} + p_{2k+1}}(1 - \cos\theta).
\end{aligned}$$

Hence, for $\theta \in [0, \pi]$,

$$\begin{aligned}
|\varphi(\theta)| &\leq 1 - (1-\cos\theta)\sum_{k=0}^{\infty}\frac{p_{2k}p_{2k+1}}{p_{2k}+p_{2k+1}} \\
&\leq 1 - \frac{2}{\pi^2}\theta^2\gamma^2 \\
&\leq \exp(-\frac{2\gamma^2}{\pi^2}\theta^2),
\end{aligned}$$

which yields (A.1). That $\gamma^2 \leq \operatorname{Var} X$ can be seen by a Taylor expansion of φ in the neighborhood of $\theta = 0$, while the relation $\gamma^2 \leq 1/4$ follows from the trivial inequality

$$\frac{p_{2k}p_{2k+1}}{p_{2k}+p_{2k+1}} \leq \frac{p_{2k}+p_{2k+1}}{4}.$$

∎

γ has the desirable property to be zero when X is an integer variable with a span strictly greater than 1, in which case the period of φ is less than 2π. Another desirable property would be that $\gamma \to \infty$ when the moments of X are unbounded; since $\gamma \leq 1/2$, this is obviously not possible here. That this “feature” is somehow unavoidable can be seen on the following example:

$$\begin{aligned}
\varphi(\theta) &\stackrel{\text{def}}{=} \frac{2+e^{i\theta}}{4} + \frac{1}{4}\sum_{k=2}^{\infty}\frac{e^{ik\theta}}{k(k-1)} \\
&= \frac{1+e^{i\theta}}{2} + (1-e^{i\theta})\ln(1-e^{i\theta}).
\end{aligned}$$

The random variable having φ as characteristic function admits no finite moment of order greater or equal to 1, but no bound on $|\varphi|$ is substantially better than (A.1).

A.2 Proof of Propositions 3.1 and 3.3

Proof of Proposition 3.1 Using a Fourier inversion formula, the left hand side of (3.1) can be rewritten as

$$\frac{\sigma_n}{2\pi}\int_{-\pi}^{\pi} e^{-i\theta x}\varphi_n(\theta)\,d\theta - \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\frac{x}{\sigma_n}u}e^{-\frac{u^2}{2}}\,du.$$

Thus, our goal is to evaluate the quantity

$$
\begin{aligned}
I_n &\stackrel{\text{def}}{=} \int_{-\pi}^{\pi} e^{-i\theta x}\varphi_n(\theta)\,d\theta - \int_{-\infty}^{\infty} e^{-i\theta x} e^{-\frac{\sigma_n^2\theta^2}{2}}\,d\theta \\
&= \int_{-\delta_n}^{\delta_n} e^{-i\theta x}\Big(\varphi_n(\theta) - e^{-\frac{\sigma_n^2\theta^2}{2}}\Big)\,d\theta \\
&\quad - \int_{|\theta|\geq\delta_n} e^{-i\theta x} e^{-\frac{\sigma_n^2\theta^2}{2}}\,d\theta + \int_{|\theta|\in[\delta_n,\pi]} e^{-i\theta x}\varphi_n(\theta)\,d\theta.
\end{aligned}
$$

It is known that

$$
\int_{|\theta|\geq\delta_n} e^{-\frac{\sigma_n^2\theta^2}{2}}\,d\theta \approx \frac{2}{\sigma_n^2\delta_n} e^{-\frac{\sigma_n^2\delta_n^2}{2}},
$$

applying Lemma A.1 to φ_n, we get

$$
\left|\int_{\delta_n\leq|\theta|\leq\pi} e^{-i\theta x}\varphi_n(\theta)\,d\theta\right| \leq \int_{|\theta|\geq\delta_n} e^{-\frac{\gamma_n^2\theta^2}{5}}\,d\theta = O\Big(\frac{1}{\gamma_n^2\delta_n} e^{-\frac{\gamma_n^2\delta_n^2}{5}}\Big). \tag{A.2}
$$

Finally, we obtain a bound on $|I_n|$ which is uniform in x:

$$
\begin{aligned}
|I_n| &\leq \int_{-\delta_n}^{\delta_n} \Big|\varphi_n(\theta) - e^{-\frac{\sigma_n^2\theta^2}{2}}\Big|\,d\theta \\
&\quad + O\Big(\frac{1}{\sigma_n^2\delta_n} e^{-\frac{\sigma_n^2\delta_n^2}{2}}\Big) + O\Big(\frac{1}{\gamma_n^2\delta_n} e^{-\frac{\gamma_n^2\delta_n^2}{5}}\Big).
\end{aligned} \tag{A.3}
$$

We proceed now to estimate the above integral, so that implicitly $|\theta| \leq \delta_n$. The derivation relies on the following simple inequality, valid for all complex numbers $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$:

$$
|x_1\cdots x_n - y_1\cdots y_n| \leq \sum_{k=1}^{n} |x_1\cdots x_{k-1}||x_k - y_k||y_{k+1}\cdots y_n|, \tag{A.4}
$$

which will be used with $x_k = \varphi_{k,n}(\theta)$ and $y_k = \exp(-\sigma_{k,n}^2\theta^2/2)$.

The characteristic function $\varphi_{k,n}$ of the random variable $X_{k,n}$ satisfies (see for example Loève [9])

$$
\Big|\varphi_{k,n}(\theta) - 1 + \sigma_{k,n}^2\frac{\theta^2}{2}\Big| \leq \beta_{k,n}^{(2+r)}\frac{|\theta|^{2+r}}{2}. \tag{A.5}
$$

Hence, using the inequality $|e^{-x} - 1 + x| \leq x^s/s$, valid for all $x \geq 0$ and $1 < s \leq 2$,

$$\begin{aligned}
\Big|\varphi_{k,n}(\theta) - e^{-\frac{\sigma_{k,n}^2\theta^2}{2}}\Big| &\leq \Big|\varphi_{k,n}(\theta) - 1 + \sigma_{k,n}^2\frac{\theta^2}{2}\Big| + \Big|e^{-\frac{\sigma_{k,n}^2\theta^2}{2}} - 1 + \sigma_{k,n}^2\frac{\theta^2}{2}\Big| \\
&\leq \beta_{k,n}^{(2+r)}\frac{|\theta|^{2+r}}{2} + \sigma_{k,n}^{2+r}\frac{|\theta|^{2+r}}{2} \leq \beta_{k,n}^{(2+r)}|\theta|^{2+r}. \qquad \text{(A.6)}
\end{aligned}$$

To find an upper bound for $|\varphi_{k,n}|$, assume first $\sigma_{k,n}\delta_n \leq 1$, so that

$$\begin{aligned}
|\varphi_{k,n}(\theta)| &\leq 1 - \sigma_{k,n}^2\frac{\theta^2}{2} + \beta_{k,n}^{(2+r)}\frac{|\theta|^{2+r}}{2} \\
&\leq \exp(-\sigma_{k,n}^2 + \beta_{k,n}^{(2+r)}\delta_n^r)\frac{\theta^2}{2}. \qquad \text{(A.7)}
\end{aligned}$$

In fact, (A.7) also holds when $\sigma_{k,n}\delta_n \geq 1$, since in this case

$$-\sigma_{k,n}^2 + \beta_{k,n}^{(2+r)}\delta_n^r \geq -\sigma_{k,n}^2 + \sigma_{k,n}^{2+r}\delta_n^r \geq 0.$$

From (3.4), we can choose n such that $\sigma_{k,n} \leq \sigma_n/2$ and, using (A.4), (A.6) and (A.7), we find

$$\begin{aligned}
\Big|\varphi_n(\theta) - e^{-\frac{\sigma_n^2\theta^2}{2}}\Big| &\leq \sum_{k=1}^{n}\beta_{k,n}^{(2+r)}|\theta|^{2+r}\exp\Big(-\sigma_n^2 + \sigma_{k,n}^2 + \beta_n^{(2+r)}\delta_n^r\Big)\frac{\theta^2}{2} \\
&\leq \beta_n^{(2+r)}|\theta|^{2+r}\exp\Big(-\sigma_n^2\frac{\theta^2}{8}\Big). \qquad \text{(A.8)}
\end{aligned}$$

Equation (3.1) follows, since the integral in (A.3) is bounded by

$$\begin{aligned}
\int_{-\delta_n}^{\delta_n}\Big|\varphi_n(\theta) - e^{-\frac{\sigma_n^2\theta^2}{2}}\Big|d\theta &\leq \beta_n^{(2+r)}\int_{-\infty}^{\infty}|\theta|^{2+r}\exp\Big(-\sigma_n^2\frac{\theta^2}{8}\Big)d\theta \\
&= O\Big(\frac{1}{\sigma_n}\frac{\beta_n^{(2+r)}}{\sigma_n^{2+r}}\Big).
\end{aligned}$$

The proof of (3.2) of the proposition is similar, although the computations be more involved. Redefine I_n as

$$I_n \stackrel{\text{def}}{=} \int_{-\pi}^{\pi} e^{-i\theta x}\varphi_n(\theta)d\theta - \int_{-\infty}^{\infty} e^{-i\theta x}\Big(1 - i\bar{\beta}_n^{(3)}\frac{\theta^3}{6}\Big)e^{-\frac{\sigma_n^2\theta^2}{2}}d\theta,$$

To find a bound for $|I_n|$, we have to estimate

$$\left|\varphi_n(\theta) - \Big(1 - i\bar{\beta}_n^{(3)}\frac{\theta^3}{6}\Big)e^{-\frac{\sigma_n^2\theta^2}{2}}\right| \tag{A.9}$$

$$\leq \quad \left|\varphi_n(\theta) - e^{-\frac{\sigma_n^2\theta^2}{2} - i\frac{\bar{\beta}_n^{(3)}\theta^3}{6}}\right| + \left|e^{-i\frac{\bar{\beta}_n^{(3)}\theta^3}{6}} - 1 + i\bar{\beta}_n^{(3)}\frac{\theta^3}{6}\right| e^{-\frac{\sigma_n^2\theta^2}{2}}.$$

The first part of the r.h.s. of (A.9) is evaluated as above with (A.4) and (A.7) replaced by

$$\varphi_{k,n}(\theta) \leq \exp(-\sigma_{k,n}^2 + \beta_{k,n}^{(3)}\delta_n)\frac{\theta^2}{2}.$$

For the second part, we use the following inequality, valid for $r \geq 0$ (see e.g. Loève [9])

$$\left[\frac{\beta_n^{(3)}}{\sigma_n^3}\right]^{1+\frac{r}{3}} \leq \frac{\beta_n^{(3+r)}}{\sigma_n^{3+r}},$$

which yields

$$\left|e^{-i\frac{\bar{\beta}_n^{(3)}\theta^3}{6}} - 1 + i\bar{\beta}_n^{(3)}\frac{\theta^3}{6}\right| \leq \left|\beta_n^{(3)}\frac{\theta^3}{6}\right|^{1+\frac{r}{3}} \leq \frac{\beta_n^{(3+r)}}{\sigma_n^{3+r}}\frac{\sigma_n^{3+r}|\theta|^{3+r}}{6},$$

and (3.2) follows. ∎

Proof of Proposition 3.3 The proof of this proposition is similar to the proof of Proposition 3.1 and is only sketched here. Define

$$\omega(u) \stackrel{\text{def}}{=} \frac{1}{1 - iu},$$

$$y_n \stackrel{\text{def}}{=} \frac{\sum_{j\in\mathcal{F}_n^0} m_{j,n} + x}{\alpha_n},$$

and

$$I_n \stackrel{\text{def}}{=} \alpha_n \int_{-\pi}^{\pi} e^{-i\theta\alpha_n y_n}\omega_n^{\xi_n}(\theta)\prod_{k\in\mathcal{F}_n^0}\psi_{k,n}(\theta)\widehat{\varphi}_n(\theta)d\theta$$

$$- \int_{-\infty}^{\infty} e^{-iuy_n}\omega^{\xi_n}(u)e^{-\frac{\hat{\sigma}_n^2}{\alpha_n^2}\frac{u^2}{2}}du$$

$$
\begin{aligned}
&= \int_{-\pi\alpha_n}^{\pi\alpha_n} e^{-iuy_n}\omega_n^{\xi_n}(u/\alpha_n)\Big[\prod_{k\in\mathcal{F}_n^0}\psi_{k,n}(u/\alpha_n)-1\Big]\widehat{\varphi}_n(u/\alpha_n)du \\
&\quad + \int_{-\pi\alpha_n}^{\pi\alpha_n} e^{-iuy_n}\omega_n^{\xi_n}(u/\alpha_n)\Big[\widehat{\varphi}_n(u/\alpha_n)-e^{-\frac{\hat{\sigma}_n^2}{\alpha_n^2}\frac{u^2}{2}}\Big]du \\
&\quad + \int_{-\pi\alpha_n}^{\pi\alpha_n} e^{-iuy_n}\Big[\omega_n^{\xi_n}(u/\alpha_n)-\omega^{\xi_n}(u)\Big]e^{-\frac{\hat{\sigma}_n^2}{\alpha_n^2}\frac{u^2}{2}}du \\
&\quad - \int_{|u|\geq\pi\alpha_n} e^{-iuy_n}\omega^{\xi_n}(u)e^{-\frac{\hat{\sigma}_n^2}{\alpha_n^2}\frac{u^2}{2}}du. \qquad \text{(A.10)}
\end{aligned}
$$

The evaluation of these integrals depends on the following straightforward estimations, valid for $|u| < \pi\alpha_n$,

$$
\begin{aligned}
|\omega_n^{\xi_n}(u/\alpha_n)| &= O\Big(\frac{1}{(1+u^2)^{\xi_n/2}}\Big), \\
|\omega_n^{\xi_n}(u/\alpha_n)-\omega^{\xi_n}(u)| &= O\Big(\frac{1}{\alpha_n}\frac{u^2}{(1+u^2)^{\xi_n}}\Big), \\
\Big|\prod_{k\in\mathcal{F}_n^0}\psi_{k,n}(u/\alpha_n)-1\Big| &= O\Big(\frac{1+|u|}{\alpha_n}\Big),
\end{aligned}
$$

and on (A.8), which yields for $|u| < \alpha_n\hat{\delta}_n$,

$$
\Big|\widehat{\varphi}_n(u/\alpha_n)-e^{-\frac{\hat{\sigma}_n^2}{\alpha_n^2}\frac{u^2}{2}}\Big| = O\Big(\frac{\hat{\beta}_n^{(2+r)}}{\alpha_n^{2+r}}\Big)u^{2+r}\exp\Big(-\frac{\hat{\sigma}_n^2}{\alpha_n^2}\frac{u^2}{8}\Big),
$$

$$
\Big|\widehat{\varphi}_n(u/\alpha_n)\Big| \leq \exp\Big(-\frac{\hat{\sigma}_n^2}{\alpha_n^2}\frac{u^2}{4}\Big).
$$

Moreover, we use the following approximation, valid for $a, b > 0$ and for sufficiently small z:

$$
J(a,b,z) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty}\frac{|u|^a}{(1+u^2)^b}e^{-z^2u^2}du = O(1)+O(z^{2b-a-1}).
$$

These relations, together with (A.10), yield:

$$
I_n = O\Big(\frac{1}{\alpha_n}\Big)J\Big(1,\frac{\xi_n}{2},\frac{\hat{\sigma}_n}{2\alpha_n}\Big)+O\Big(\frac{\hat{\beta}_n^{(2+r)}}{\alpha_n^{2+r}}\Big)J\Big(2+r,\frac{\xi_n}{2},\frac{\hat{\sigma}_n}{\sqrt{8}\alpha_n}\Big)
$$

$$
\begin{aligned}
&+ O\Big(\frac{1}{\alpha_n}\Big) J\Big(2, \xi_n, \frac{\hat{\sigma}_n}{\sqrt{2}\alpha_n}\Big) \\
&+ O\Big(\Big(\frac{\hat{\sigma}_n}{\alpha_n}\Big)^{\xi_n - 1}\Big) \int_{v \geq \pi \hat{\sigma}_n} v^{-\xi_n} e^{-\frac{v^2}{2}} dv \\
&+ O\Big(\Big(\frac{\hat{\gamma}_n}{\alpha_n}\Big)^{\xi_n - 1}\Big) \int_{v \geq \delta_n \hat{\gamma}_n} v^{-\xi_n} e^{-\frac{v^2}{2}} dv \\
= \;& O\Big(\frac{1}{\alpha_n} + \frac{\hat{\beta}_n^{(2+r)}}{\hat{\sigma}_n^{2+r}}\Big(\frac{\hat{\sigma}_n}{\alpha_n}\Big)^{2+r} + \frac{\hat{\beta}_n^{(2+r)}}{\hat{\sigma}_n^{2+r}}\Big(\frac{\hat{\sigma}_n}{\alpha_n}\Big)^{\xi_n - 1}\Big) \\
&+ O\Big(\frac{e^{-\frac{\hat{\gamma}_n^2 \hat{\delta}_n^2}{5}}}{\hat{\gamma}_n^2 \hat{\delta}_n^{\xi_n + 1} \alpha_n^{\xi_n - 1}}\Big).
\end{aligned}
$$

To conclude the proof of (3.8), the second term coming in the definition of I_n is evaluated using Parseval's identity and classical tools of complex analysis (see e.g. Lavrentiev and Chabat [8]). This yields

$$
\int_{-\infty}^{\infty} e^{-iy_n u} \omega^{\xi_n}(u) e^{-\frac{\hat{\sigma}_n^2}{\alpha_n^2}\frac{u^2}{2}} du = \frac{y_n^{\xi_n - 1} e^{-y_n}}{\Gamma(\xi_n)} \Big[1 + O\Big(\frac{\hat{\sigma}_n^2}{\alpha_n^2}\Big)\Big].
$$

■

References

[1] A. BIRMAN AND Y. KOGAN, *Asymptotic evaluation of closed queueing networks with many stations*, Communications in Statistics—Stochastic Models, 8 (1992), pp. 543–563.

[2] G. FAYOLLE AND J.-M. LASGOUTTES, *Limit laws for large product-form networks: connections with the Central Limit Theorem*, Rapport de Recherche 2513, INRIA, Mar. 1995.

[3] W. FELLER, *An Introduction to Probability Theory and its Applications, Vol. II*, John Wiley & Sons, 2 ed., 1971.

[4] F. P. KELLY, *Reversibility and Stochastic Networks*, Wiley, 1979.

[5] C. KNESSL AND C. TIER, *Asymptotic expansions for large closed queueing networks*, Journal of the ACM, 37 (1990), pp. 144–174.

[6] Y. KOGAN, *Another approach to asymptotic expansions for large closed queueing networks*, Operations Research Letters, 11 (1992), pp. 317–321.

[7] Y. KOGAN AND A. BIRMAN, *Asymptotic analysis of closed queueing networks with bottlenecks*, in Proceedings of the International Conference on Performance of Distributed Systems and Integrated Communication Networks, T. Hasegawa, H. Takagi, and Y. Takahashi, eds., Kyoto, 1991, pp. 237–252.

[8] M. LAVRENTIEV AND B. CHABAT, *Méthodes de la théorie des fonctions d'une variable complexe*, Éditions MIR, 1977.

[9] M. LOÈVE, *Probability Theory*, D. Van Nostrand Company, fourth ed., 1977.

[10] V. MALYSHEV AND A. YAKOVLEV, *Condensation in large closed Jackson networks*, Rapport de Recherche 1854, INRIA, 1993.

[11] R. F. SERFOZO, *Markovian network processes: Congestion-dependent routing and processing*, Queueing Systems, Theory and Applications, 5 (1989), pp. 5–36.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399