Available online at www.sciencedirect.com

ScienceDirect

C. R. Acad. Sci. Paris, Ser. I 345 (2007) 233–237

COMPTES RENDUS MATHEMATIQUE

http://france.elsevier.com/direct/CRASS1/

Probability Theory

Stationarity of measure-valued stochastic recursions: applications to the pure delay system and the SRPT queue

Pascal Moyal

CEREMADE, université Paris IX Dauphine, place du-Marechal-de-Lattre-de-Tassigny, 75775 Paris cedex 16, France

Received 25 February 2007; accepted after revision 14 June 2007

Available online 1 August 2007

Presented by Marc Yor

Abstract

In this Note we present a stability criterion for finite measure-valued stochastic recursions, generalizing Loynes's Theorem to spaces of measures. This result, developed in detail elsewhere, provides conditions for reaching a 'total stationary state' for the queue with an infinity of servers and the single-server SRPT queue. Indeed, we give in both cases a condition of existence of a stationary measure-valued recursive sequence characterizing the queueing system exhaustively. ***To cite this article: P. Moyal, C. R. Acad. Sci. Paris, Ser. I 345 (2007).***

Résumé

Stationnarité des suites récurrentes stochastiques à valeurs mesures. Applications aux files d'attente « pur délai » et SRPT. Nous présentons un critère général de stabilité pour des suites récurrentes stochastiques (SRS) à valeurs mesures finies positives, qui généralise le Théorème de Loynes aux espaces de mesures. Ce résultat, développé en détail ailleurs, donne des conditions d'atteinte d'un « régime stationnaire total » pour la file d'attente à une infinité de serveurs et la file d'attente à un serveur travaillant sous la discipline SRPT. En effet, nous donnons dans chaque cas une condition d'existence d'une SRS stationnaire à valeurs mesures représentant exhaustivement la file d'attente. ***Pour citer cet article : P. Moyal, C. R. Acad. Sci. Paris, Ser. I 345 (2007).***

Version française abrégée

Nous étudions l'existence d'une version stationnaire pour la suite récurrente stochastique (SRS) $\{\mu_n\}_{n\in\mathbb{N}}$ définie sur l'espace probabilisé $(\Omega,\mathcal{F},\mathbf{P}^0)$ (muni de l'endomorphisme θ sous lequel $\mathbf{P}^0$ est stationnaire et ergodique), à valeurs dans l'espace $\mathbf{M}_f^+$ des mesures finies positives sur $\mathbb{R}_+^*$, et porté par l'équation de récurrence

$$\mu_{n+1}^{\kappa}=\Phi\circ\theta^n(\mu_n^{\kappa}),$$

où l'exposant dénote la variable aléatoire initiale de la suite. L'application aléatoire Φ est à valeurs dans l'espace $\mathcal{C}(\mathbf{M}_f^+)$ des fonctions continues de $\mathbf{M}_f^+$ dans lui-même, $\mathbf{M}_f^+$ étant muni de la topologie faible. L'espace $\mathbf{M}_f^+$ est

E-mail address: moyal@ceremade.dauphine.fr.

1631-073X/$ – see front matter © 2007 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.crma.2007.06.014

partiellement ordonné suivant l'*ordre croissant* $\preceq$ ($\mu \preceq \nu$ ssi $\mu([a,\infty)) \leqslant \nu([a,\infty))$ pour tout $a \geqslant 0$) et on note ζ la mesure nulle. Soit $\mathcal{M}$ le sous-espace des mesures de comptage. On note pour tout $\mu \in \mathcal{M} \setminus \{\zeta\}$, $m(\mu)$ et $M(\mu)$ le plus petit atome de μ et le plus grand atome de μ, respectivement. L'existence d'une version stationnaire (i.e., compatible avec θ) de $\{\mu_n\}_{n\in\mathbb{N}}$ revient à celle d'une solution pour l'équation (1). En remarquant que toute suite $\preceq$-croissante et bornée converge pour la topologie faible, nous appliquons le théorème de Loynes (voir [4], [1] p. 107), fondé sur un schéma de récurrence arrière, pour montrer le résultat suivant :

Théorème 0.1. *L'équation* (1) *admet une solution lorsque Φ est p.s. croissante et lorsque la suite $\{\mu_n^{\zeta} \circ \theta^{-n}\}_{n\in\mathbb{N}}$ est p.s. $\preceq$-bornée.*

Nous donnons une première application de ce résultat au cas du système à une infinité de serveurs G/G/∞. On note N le processus des arrivées, ξ la v.a. générique des interarrivées, σ la v.a. générique des temps de service et pour tout $n \in \mathbb{Z}$, ξ_n et σ_n la n-ième interarrivée et le temps de service du n-ième client, respectivement. Le *temps de service résiduel* du client n à l'instant t pourvu que celui-ci soit dans le système à t est donné par $T_n + \sigma_n - t$. Nous considérons le processus $(\mu(t))_{t\geqslant 0}$ à valeurs $\mathcal{M}$ dont la valeur $\mu(t)$ à t place ses unités de masse aux temps de service résiduels des clients présents à t. L'existence d'une version de $(\mu(t))_{t\geqslant 0}$ jointement stationnaire avec le processus des arrivées N revient à l'existence d'une version stationnaire pour la SRS $\{\mu_n\}_{n\in\mathbb{N}} := \{\mu(T_n-)\}_{n\in\mathbb{N}}$. Puisque (2) tient pour tout n, ce problème revient à la résolution de l'équation (3) dans l'espace de Palm de N et pour son flot stationnaire ergodique naturel θ. La fonctionnelle de récurrence étant croissante, nous nous fondons sur le Théorème 0.1 (existence) et le Théorème 2 de [5] pour montrer le résultat suivant :

Théorème 0.2. *L'équation* (3) *admet une solution $\mu_\infty \in \mathcal{M}$, $\mathbf{P}^0$-p.s., donnée par* (4)*. Cette solution est unique si* (5) *est vérifiée.*

Notons W_n et X_n, la charge de travail et le nombre de clients dans le système, respectivement, à l'arrivée du client n. On a alors $W_n = \int_0^{+\infty} x \,\mathrm{d}\mu_n(x)$ et $X_n = \int_0^{+\infty} \mathrm{d}\mu_n(x)$. Le corollaire (i) est une conséquence naturelle du schéma de récurrence arrière, et les suivants découlent de (i) en vertu du théorème de l'application continue (l'exposant des suites représente leur v.a. initiale).

Corollaire 0.3. *On a les convergences en loi suivantes* :

(i) $\mu_n^{\zeta} \xrightarrow{\mathcal{L}} \mu_\infty$,
(ii) $X_n^0 \xrightarrow{\mathcal{L}} \mu_\infty(\mathbb{R}_+^*)$,
(iii) $W_n^0 \xrightarrow{\mathcal{L}} \int x \,\mathrm{d}\mu_\infty(x)$.

Avec les mêmes notations, considérons maintenant la file d'attente G/G/1, régie par la discipline de service *Shortest Remaining Processing Time* (SRPT), qui donne priorité préemptive au client ayant le temps de service résiduel le plus court. Le processus des profils $(\nu(t))_{t\geqslant 0}$ à valeurs $\mathcal{M}$ marquant les temps de services résiduels des clients du système est régi par la dynamique décrite par (6). Ce processus admet donc une version stationnaire pourvu que l'Éq. (7) admette une solution. La fonctionnelle de récurrence est croissante, et il est possible de borner la suite des profils en vertu de la stabilité de la congestion et de la charge de travail de la file. On a donc le résultat suivant :

Théorème 0.4. *Si* $\mathbf{E}^0[\sigma] < \mathbf{E}^0[\xi]$, (7) *admet une solution appartenant à $\mathcal{M}$, $\mathbf{P}^0$-p.s.*

1. Stationarity of measure-valued SRS

Let $\mathbf{M}_f^+$ and $\mathcal{C}_b$ denote respectively the set of positive finite measures on $\mathbb{R}_+^*$ and the set of bounded continuous functions from $\mathbb{R}$ to $\mathbb{R}$. Equipped with the *weak topology* $\sigma(\mathbf{M}_f^+, \mathcal{C}_b)$, $\mathbf{M}_f^+$ is Polish (see [2]). Let ζ be the zero measure on $\mathbb{R}$ (i.e., such that $\zeta(\mathfrak{B}) = 0$ for any Borel set $\mathfrak{B}$ on $\mathbb{R}$). For any $\mu \in \mathbf{M}_f^+$ and for any measurable $f:\mathbb{R} \to \mathbb{R}$, we classically write $\langle \mu, f\rangle := \int f \,\mathrm{d}\mu$. We denote $\mathcal{C}(\mathbf{M}_f^+)$, the space of continuous mappings from $\mathbf{M}_f^+$ into itself. Let the set $\mathbf{M}_f^+$ be endowed with the *increasing partial integral order* $\preceq$: for any two $\mu, \nu \in \mathbf{M}_f^+$, $\mu \preceq \nu$ if $\mu([a,\infty)) \leqslant \nu([a,\infty))$ for any $a \geqslant 0$. Remark, that $\zeta \preceq \mu$ for any $\mu \in \mathbf{M}_f^+$. Let us denote for any $y \in \mathbb{R}$ and any measurable $f : \mathbb{R} \to \mathbb{R}$, $\tau_y f(\cdot) = f(\cdot - y)\mathbf{1}_{\{\cdot > y\}}$. Then, for any $\mu \in \mathbf{M}_f^+$, $\tau_y \mu$ denotes the only element of $\mathbf{M}_f^+$ s.t. $\langle \tau_y \mu, f\rangle = \langle \mu, \tau_y f\rangle$.

Let now $\mathcal{M} \subset \mathbf{M}_f^+$ be the subset of finite counting measures on $\mathbb{R}_+^*$. Any $\mu \in \mathcal{M}\backslash\{\zeta\}$ reads $\mu = \sum_{i=1}^{\mu(\mathbb{R}_+^*)} \delta_{\alpha_i(\mu)}$, where δ_x is the Dirac measure at $x \in \mathbb{R}_+^*$ and $\alpha_1(\mu) < \alpha_2(\mu) < \cdots < \alpha_{\mu(\mathbb{R}_+^*)}(\mu)$. Then,

$$\tau_y(\mu) = \sum_{i=1}^{\mu(\mathbb{R}_+^*)} \delta_{\alpha_i(\mu) - y}\mathbf{1}_{\{\alpha_i(\mu) > y\}}$$

and for any two $\mu, \nu \in \mathcal{M}\backslash\{\zeta\}$,$\mu \preceq \nu$ whenever

$$\begin{cases} \text{(i)} \quad \mu(\mathbb{R}_+^*) \leqslant \nu(\mathbb{R}_+^*), \\ \text{(ii)} \quad \text{for all } i = 0, \ldots, \mu(\mathbb{R}_+^*) - 1, \quad \alpha_{\mu(\mathbb{R}_+^*)-i}(\mu) \leqslant \alpha_{\nu(\mathbb{R}_+^*)-i}(\nu). \end{cases}$$

Let us finally denote for any $\mu \in \mathcal{M}\backslash\{\zeta\}$, $m(\mu) = \alpha_1(\mu)$ and $M(\mu) = \alpha_{\mu(\mathbb{R}_+^*)}(\mu)$, the smallest and the largest atom of μ, respectively.

Let $(\Omega, \mathcal{F}, \mathbf{P}^0, \theta)$ be a probability space furnished with a discrete bijective flow θ, under which $\mathbf{P}^0$ is stationary and ergodic. Let Φ be a $\mathcal{C}(\mathbf{M}_f^+)$-valued r.v., κ be an $\mathbf{M}_f^+$-valued random variable (r.v. for short), and define the following sequence of $\mathbf{M}_f^+$-valued r.v.:

$$\begin{cases} \mu_0^\kappa = \kappa, \\ \mu_n^\kappa = \Phi \circ \theta^n(\mu_n^\kappa), \quad n \in \mathbb{N}. \end{cases}$$

Hence, the existence of a stationary (i.e., compatible with θ) version of $\{\mu_n^\kappa\}_{n\in\mathbb{N}}$ amounts to that of a solution of the following equation in the space of $\mathbf{M}_f^+$-valued r.v.:

$$\kappa \circ \theta = \Phi(\kappa). \tag{1}$$

We have the following result:

Theorem 1.1. *Eq.* (1) *admits a solution whenever Φ is a.s. $\preceq$-non decreasing and the sequence $\{\mu_n^\zeta \circ \theta^{-n}\}_{n\in\mathbb{N}}$ is a.s. $\preceq$-bounded.*

Idea of the proof. (See details in [5].) We first prove that any bounded $\preceq$-increasing sequence of $\mathbf{M}_f^+$ converges for the weak topology. The result then follows from the classical Loynes's backwards scheme ([4], [1] p. 107): the sequence $\{\mu_n^\zeta \circ \theta^{-n}\}_{n\in\mathbb{N}}$ is $\mathbf{P}^0$-a.s. non-decreasing and bounded, and thus converges a.s. to the $\preceq$-minimal solution of (1), which belongs $\mathbf{P}^0$-a.s. to $\mathbf{M}_f^+$. □

2. The G/G/∞ system

Let $(\Omega, \mathcal{F}, \mathbf{P}, \theta_t)$ be a probability space furnished with a bijective flow $(\theta_t)_{t\geqslant 0}$, under which $\mathbf{P}$ is stationary and ergodic. Define on Ω the θ_t-compatible simple point process $(N_t)_{t\geqslant 0}$ of points

$$\cdots < T_{-2} < T_{-1} < T_0 \leqslant 0 < T_1 < T_2 < \cdots,$$

marked by a sequence $\{\sigma_n\}_{n\in\mathbb{Z}}$. Also denote for all $n \in \mathbb{Z}$, $\xi_n = T_{n+1} - T_n$, and suppose that the generic r.v. σ and ξ are integrable. Consider the G/G/∞ queue fed by $N(\sigma)$: there is an infinity of servers, thus e.g., the customer C_n arrived at time T_n is immediately attended and spend in the system a time equal to the duration σ_n of his service. Assume that C_n is present in the system at time t, i.e., $T_n \leqslant t < T_n + \sigma_n$. The *remaining processing time* (i.e., time before the service completion) of C_n at t is hence given by $T_n + \sigma_n - t$. The *profile* of the system at t is the collection of the remaining processing times of all the customers in the system at t, and can thus be represented by the following measure, which $\mathbf{P}^0$-a.s. belongs to $\mathcal{M}$:

$$\mu(t) = \sum_{n\in\mathbb{Z}} \delta_{T_n+\sigma_n-t}\mathbf{1}_{\{T_n \leqslant t < T_n+\sigma_n\}} = \begin{cases} \zeta & \text{if the system is empty at } t, \\ \displaystyle\sum_{i=1}^{X(t)} \delta_{\alpha_i(t)} & \text{if not,} \end{cases}$$

where $0 < m(\mu(t)) = \alpha_1(t) < \alpha_2(t) < \cdots < \alpha_{X(t)}(t) = M(\mu(t))$ are the remaining service times of the $X(t)$ customers present in the system at t. Then $(\mu(t))_{t\geqslant 0}$ provides an exhaustive representation of the system in that it keeps track of all present customers at a given time. Note, that a diffusion approximation of a scaled sequence of such profile processes under Poissonian assumptions is presented in [3]. In particular, for any $t \geqslant 0$, the *congestion* $X(t)$ (i.e., the number of customers in the system at time t) and the *workload* $W(t)$ (i.e., the quantity of work in the system at time t, in time units) read

$$\begin{cases} X(t) = \langle \mu(t), 1\rangle, \\ W(t) = \displaystyle\sum_{i=1}^{X(t)} \alpha_i(t) = \big\langle \mu(t), I\big\rangle, \end{cases}$$

where I is the identity function. The processes $(\mu(t))_{t\geqslant 0}$, $(X(t))_{t\geqslant 0}$ and $(W(t))_{t\geqslant 0}$ have rcll (i.e. right continuous with limit on the left) paths and denoting for any $t \geqslant 0$, $\mu(t-) = \lim_{s\uparrow\uparrow t} \mu(s)$ (and accordingly, $X(t-)$ and $W(t-)$), we have for all $n \in \mathbb{Z}$

$$\mu(t) = \tau_{t-T_n}\big(\mu(T_n-) + \delta_{\sigma_n}\big), \quad t \in [T_n, T_{n+1}), \ \mathbf{P}\text{-a.s.} \tag{2}$$

Let $(\Omega, \mathcal{F}, \mathbf{P}^0, \theta)$ be the Palm space of N. Denote $\theta := \theta_{T_1}$ and for all $n \in \mathbb{Z}$, $\theta^n := \theta \circ \theta \circ \cdots \circ \theta$. Denoting $\xi := \xi_0$ and $\sigma := \sigma_0$, we have for all $n \in \mathbb{Z}$, $\xi_n := \xi \circ \theta^n$ and $\sigma_n := \sigma \circ \theta^n$. For all $\mathbf{M}_f^+$-valued r.v. κ, denote for all $n \in \mathbb{N}$, $\mu_n^\kappa = \mu(T_n-)$ provided that $\mu^\kappa(T_0-) = \kappa$ (and accordingly, X_n^X and W_n^W for any integer-valued r.v. X and any real-valued r.v. W).

The problem of finding a stationary version of the process $(\mu(t))_{t\geqslant 0}$ amounts to that of finding a sequence $\{\mu_n^\kappa\}_{n\in\mathbb{N}}$ compatible with the discrete flow θ, which amounts in turn in view of (2) to finding a $\mathbf{P}^0$-a.s. finite random variable κ of $\mathbf{M}_f^+$ such that

$$\kappa \circ \theta = \tau_\xi(\kappa + \delta_\sigma). \tag{3}$$

Theorem 2.1. *Eq.* (3) *admits a solution* $\mu_\infty \in \mathcal{M}$, $\mathbf{P}^0$*-a.s., given by*

$$\mu_\infty = \sum_{i=1}^{\infty} \delta_{(\sigma_{-i} - \sum_{j=1}^{i} \xi_{-j})} \mathbf{1}_{\{\sigma_{-i} \geqslant \sum_{j=1}^{i} \xi_{-j}\}}. \tag{4}$$

This solution is unique if

$$\mathbf{P}^0\left[\sup_{i\in\mathbb{N}^*}\left(\sigma_{-i} - \sum_{j=1}^{i} \xi_{-j}\right) \leqslant 0\right] > 0. \tag{5}$$

Idea of the proof. (See details in [5].) That $\mu_\infty \in \mathcal{M}$ (or equivalently, that $\mu_\infty(\mathbb{R}_+^*) < \infty$), $\mathbf{P}^0$-a.s., is a consequence of Birkhoff's ergodic theorem. The first statement follows from Theorem 1.1: the mapping $\mu \mapsto \tau_\xi(\mu + \delta_\sigma)$ is $\mathbf{P}^0$-a.s. continuous and non-decreasing, and for all $n \in \mathbb{N}$,

$$\mu_n^\zeta \circ \theta^{-n} = \sum_{i=1}^{n} \delta_{(\sigma_{-i} - \sum_{j=1}^{i} \xi_{-j})} \mathbf{1}_{\{\sigma_{-i} \geqslant \sum_{j=1}^{i} \xi_{-j}\}}.$$

Thus μ_∞ is the $\preceq$-smallest upper-bound of $\{\mu_n^\zeta\}_{n\in\mathbb{N}}$ and hence the $\preceq$-minimal solution of (3). For the second statement, note that any solution κ of (3) satisfies

$$M(\kappa) \circ \theta = \big[\max\big\{M(\kappa), \sigma\big\} - \xi\big]^+,$$

which is a recursive equation of the type (2) of [5]. Hence (Theorem 2 of [5], uniqueness result), $M(\kappa) = M(\mu_\infty)$, which implies uniqueness whenever (5) holds in view of the minimality of μ_∞. □

Corollary 1. *The following convergences in distribution hold.*

(i) $\mu_n^\zeta \xrightarrow{\mathcal{L}} \mu_\infty$,

(ii) $X_n^0 \xrightarrow{\mathcal{L}} \langle \mu_\infty, 1\rangle = \mu_\infty(\mathbb{R}_+^*)$,

(iii) $W_n^0 \xrightarrow{\mathcal{L}} \langle \mu_\infty, I\rangle = \int x \,\mathrm{d}\mu_\infty(x)$.

Idea of the proof. (See details in [5].) (i) classically follows from Loynes's construction, whereas (ii) and (iii) are consequences of (i) and the Continuous Mapping Theorem. □

3. The SRPT queue

Under the settings and notations of the previous section, consider now a G/G/1 queue fed by the input $N(\sigma)$. We assume furthermore that the single server obeys the *Shortest Remaining Processing Time* (SRPT) service discipline, i.e., it always gives a preemptive priority to the customer having the least remaining service time. Let $(\nu(t))_{t\geqslant 0}$ denote the profile process of the queue, keeping track of the remaining service times of the customers in the system at current time. This process is governed by the following dynamics: for all $n \in \mathbb{Z}$,

$$\nu(t) = \Upsilon_{t-T_n}\big(\nu(T_n-) + \delta_{\sigma_n}\big), \quad t \in [T_n, T_{n+1}), \ \mathbf{P}\text{-a.s.} \tag{6}$$

where for any $x \in \mathbb{R}_+^*$ and any $\mu \in \mathcal{M}$,

$$\Upsilon_x(\mu) = \sum_{i=1}^{\mu(\mathbb{R}_+^*)} \tau_{\{(x-\sum_{k=1}^{i-1}\alpha_k(\mu))^+\}}\delta_{\alpha_i(\mu)},$$

setting $\sum_{k=1}^{0} . \equiv 0$. Hence there exists a stationary version of $\{\nu_n\}_{n\in\mathbb{N}}$ provided that for some $\mathcal{M}$-valued r.v. κ,

$$\kappa \circ \theta = \Upsilon_\xi(\kappa + \delta_\sigma). \tag{7}$$

Theorem 3.1. *If* $\mathbf{E}^0[\sigma] < \mathbf{E}^0[\xi]$, (7) *admits a solution* $\nu_\infty \in \mathcal{M}$, $\mathbf{P}^0$*-a.s.*

Idea of the proof. (See details in [5].) Denote as above $\{\nu_n^\kappa\}_{n\in\mathbb{N}}$, the profile sequence provided that $\nu_0^\kappa = \kappa$. The result is a consequence of Theorem 1.1 since the mapping $\mu \mapsto \Upsilon_\xi(\mu + \delta_\sigma)$ is $\mathbf{P}^0$-a.s. $\preceq$-non-decreasing and continuous, and since $\{\nu_n^\zeta \circ \theta^{-n}\}_{n\in\mathbb{N}}$ is $\preceq$-unbounded only if the congestion sequence or the workload sequence initiated at 0 is $\leqslant$-unbounded, which is of probability 0 in view of Loynes' theorem for G/G/1 queues. □

References

[1] F. Baccelli, P. Brémaud, Elements of Queueing Theory, second ed., Springer, 2002.
[2] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.
[3] L. Decreusefond, P. Moyal, A Functional Central Limit Theorem for the M/GI/∞ queue, Preprint, math.PR/0608256, 2006.
[4] R.M. Loynes, The stability of queues with non-independent interarrivals and service times, Proceedings of the Cambridge Philosophical Society 58 (1962) 497–520.
[5] P. Moyal, Measure-valued stochastic recurrences and the stability of queues, Preprint, math.PR/0706.0817, 2007, in preparation.