

Note sur la simulation d'une file $M/G/1$ selon la distribution du temps de service

Thomas Begin ⁽¹⁾ et Alexandre Brandwajn ⁽²⁾

(1) Université Lyon 1 / LIP UMR CNRS - ENS Lyon - UCB Lyon 1 - INRIA 5668.

(2) University of California Santa Cruz / Jack Baskin School of Engineering.

Résumé

Dans cette note nous souhaitons mettre en lumière une difficulté particulière de la simulation liée aux types de distributions supposées dans le modèle. Nous appuyons notre étude sur un modèle simple de type file d'attente pour lequel une résolution analytique existe et peut donc servir de base de comparaison.

Nos résultats suggèrent que l'usage de certaines distributions (dont celle de Pareto) dans les modèles tend à accentuer la complexité de leur résolution par simulation. Cet effet est d'autant plus prononcé que le coefficient de variation de la distribution est élevé. Ainsi, si dans un modèle une loi n'est connue que par ses deux (ou n -premiers) moments, il convient de choisir la distribution la plus simple à simuler.

Keywords: Distribution du temps de service, Simulation à événements discrets, File $M/G/1$, Performances.

1 Introduction

La simulation à événements discrets est un outil efficace et simple en apparence pour évaluer le comportement de systèmes informatiques [BRA87]. Elle peut servir à vérifier des résultats théoriques ou expérimentaux ou bien à résoudre un modèle dont les caractéristiques empêchent la résolution par voie analytique.

Vu la complexité, la taille et l'hétérogénéité des réseaux informatiques, l'évaluation de leurs performances se réalise souvent par simulation. Ainsi, comme le montre l'étude de Pawlikowski [PAW03], une majorité des articles publiés récemment dans les conférences réseaux ou dans des revues renommées s'appuient sur un simulateur pour obtenir leur résultat.

Dans cette note nous considérons un modèle simple de type file d'attente à solution analytique connue. Notre objectif est de mettre en lumière une difficulté particulière liée aux types de distributions supposées dans le modèle. Nous montrons à travers une série d'exemples que, même pour ce modèle simple, une résolution par simulation peut s'avérer difficile. Ce type de difficultés est connu dans la littérature spécialisée [GRO02, ASM00] mais semble être relativement mal connu des chercheurs dans d'autres domaines.

2 Formulation du problème

Nous étudions le comportement stationnaire d'une file $M/G/1$. Le serveur de la file a un taux d'utilisation égal à ρ ($\rho < 1$) et un temps de service caractérisé par sa moyenne, m , et par son coefficient de variation, cv . Soit \bar{n} le nombre moyen de clients présents dans la file. D'après la formule de Pollaczek-Khintchine [ALL90], on a

$$\bar{n} = \rho + \rho^2(1 + cv^2)/2(1 - \rho).$$
 Notons que ce résultat est valable quels que soient les moments d'ordre supérieur à 2 et la forme de la distribution du temps de service (sous réserve que sa moyenne et sa variance soient finies). Ainsi, si deux files $M/G/1$ ont une même utilisation de leur serveur et des temps de service distribués différemment mais de même moyenne et de même variance, alors le nombre moyen de clients dans ces deux files sera identique.

Nous évaluons par simulation le comportement moyen d'une file $M/G/1$ pour différents types de distributions du temps de service ayant les deux premiers moments identiques. D'après la propriété précédente, toutes les simulations doivent converger asymptotiquement à la même valeur de \bar{n} . Comme nous le montrons dans la section suivante, cette convergence peut se faire avec des vitesses très différentes selon les distributions du temps de service.

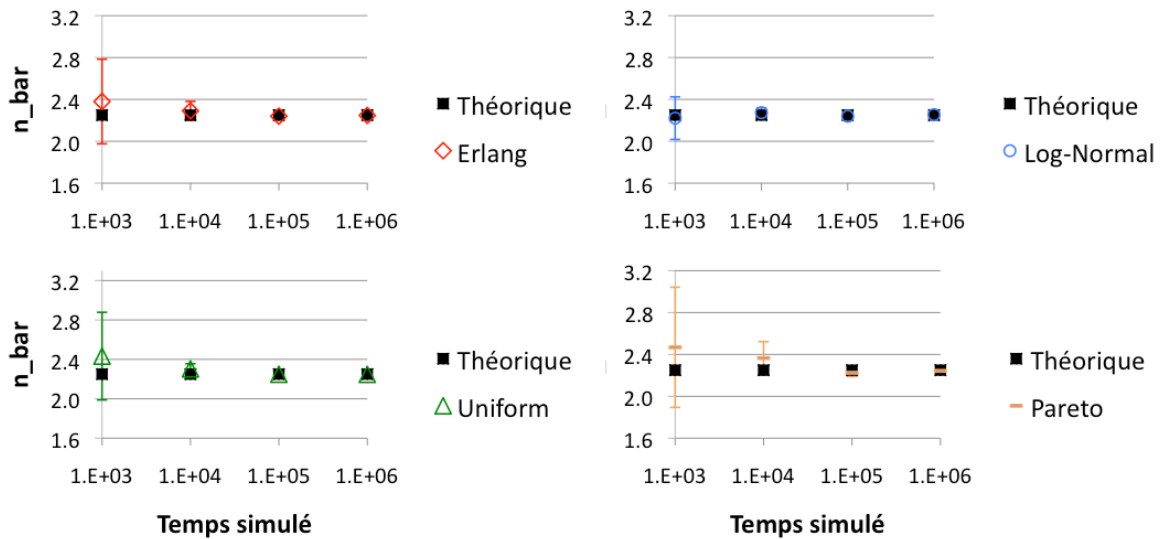


Fig. 1 – Convergence du comportement simulé d'une file $M/G/1$ vers la valeur exacte pour différents types de distributions avec $\rho=0.75$ et $cv < 1$

	Temps simulé			
	1.E+03	1.E+04	1.E+05	1.E+06
Erlang	11.652	2.077	0.301	0.163
Uniforme	0.763	2.598	0.968	0.467
Log-Normal	9.513	0.529	2.776	0.049
Pareto	19.108	5.526	1.014	0.490

Table 1 – Ecarts relatifs en pourcentage entre la valeur exacte du nombre moyen de clients dans la file $M/G/1$ et la valeur fournie par la simulation pour différents types de distributions du temps de service avec $cv < 1$

3 Résultats numériques

Dans cette note nous considérons un niveau de charge modéré du serveur, i.e. $\rho=0.75$. Selon notre expérience, pour des niveaux de charges inférieurs (resp. supérieurs), la vitesse de convergence des simulations tend à se réduire (resp. s'allonger) quelle que soit distribution du temps de service considérée. Le temps de simulation est exprimé en nombre de services accomplis par le serveur de la file. Les intervalles de confiance représentés sur les figures ont été calculés à partir de 7 répliquations indépendantes pour un degré de confiance de 95%. Bien que d'autres techniques existent pour construire les intervalles de confiance, les travaux de Law et Kelton [LAW84] sur la file $M/M/1$ ont montré qu'en général les autres approches ne produisaient pas de meilleurs résultats.

La simulation de certaines distributions, notamment celle de Pareto avec $cv > 1$, ne permet pas, même après un grand nombre d'échantillons ($2.E+9$), de reproduire correctement son coefficient de variation (que ce soit en utilisant le générateur pseudo aléatoire de Matlab, NS-2 ou SMPL). Par conséquent, le cas échéant, nous avons indiqué sur les figures la valeur de \bar{n} calculée à partir des moments théoriques de la distribution et celle calculée à partir des moments effectivement simulés. La première valeur est dénotée *Théorique* et la seconde *Référence*. Notons que ces deux valeurs sont obtenues à partir de Pollaczek-Khintchine. Les écarts relatifs reportés dans les tables correspondent à l'écart entre la valeur de \bar{n} obtenue par simulation et celle de Référence.

3.1 Cas pour $\rho=0.75$ et $cv < 1.0$

Ici, nous considérons quatre lois pour la distribution du temps de service : Erlang, Uniforme, Log-Normale et Pareto. Les paramètres de ces distributions, reportés en Annexe, ont été choisis tels que les quatre distributions aient les deux mêmes premiers moments : $m=1$ et $cv=1/\sqrt{3}$. Nous avons simulé le comportement de la file $M/G/1$ pour chacune de ces lois et reporté dans la Fig. 1 la valeur obtenue pour le nombre moyen de clients dans la file, \bar{n} , de même que sa valeur théorique exacte. Dans la Table 1, nous avons reporté l'écart relatif entre ces deux valeurs. La Fig. 1 montre qu'après un temps de simulation d'environ 100 000 fins de service, les valeurs

simulées de \bar{n} sont toutes très proches de celle calculée exactement (moins de 3% d'écart, cf. Table 1) et qu'asymptotiquement elles convergent vers cette valeur.

	Temps simulé						
	1.E+03	1.E+04	1.E+05	1.E+06	1.E+07	1.E+08	1.E+09
Hyper-Exponentielle	16.527	5.322	1.547	0.852	<0.5%		
Cox-2 (type 1)	1.424	0.684	0.881	0.629	<0.5%		
Cox-2 (type 2)	7.587	1.792	4.159	0.418	<0.5%		
Cox-2 (type 3)	32.729	38.740	4.846	4.846	<0.5%		
Log-Normal	35.062	1.568	5.773	0.087	<0.5%		
Pareto	10.357	0.463	6.890	2.273	3.020	2.650	<0.5%

Table 2 – *Ecarts relatifs en pourcentage entre la valeur exacte du nombre moyen de clients dans la file M/G/1 et la valeur fournie par la simulation pour différents types de distributions du temps de service avec $cv = 1.2$*

	Temps simulé						
	1.E+03	1.E+04	1.E+05	1.E+06	1.E+07	1.E+08	1.E+09
Hyper-Exponentielle	57.346	9.472	0.708	0.334	<0.5%		
Cox-2 (type 1)	24.449	22.804	2.923	0.326	<0.5%		
Cox-2 (type 2)	51.045	15.451	6.366	0.262	<0.5%		
Cox-3 (type 3)	95.841	17.241	71.504	22.990	3.116	3.751	<0.5%
Log-Normal	57.957	0.533	12.242	1.815	<0.5%		
Pareto	1.214	9.119	5.269	5.794	9.308	7.435	3.480

Table 3 – *Ecarts relatifs en pourcentage entre la valeur exacte du nombre moyen de clients dans la file M/G/1 et la valeur fournie par la simulation pour différents types de distributions du temps de service avec $cv = 2.0$*

3.2 Cas avec $\rho=0.75$ et $cv > 1.0$

A présent nous considérons le cas où le coefficient de variation du temps de service est supérieur à 1 pour les lois suivantes de la distribution du temps de service: Hyper-Exponentielle selon la décomposition proposée par Morse [MOR], plusieurs Cox-2, Log-Normale et Pareto. Cet ensemble de distributions diffère de celui de la Section 3.1 car une distribution en Erlang (resp. Uniforme à valeurs positives) ne peut pas avoir un coefficient de variation supérieur à 1 (resp. $1/\sqrt{3}$). Inversement, le coefficient de variation d'une distribution Hyper-exponentielle (resp. Cox-2) est nécessairement supérieur à 1 (resp. $1/\sqrt{2}$). Les paramètres des lois ont été choisis de façon à ce que le temps de service ait une moyenne de 1 et un coefficient de variation de 1.2 (cf. Annexe).

Dans la Table 2, nous avons représenté l'écart relatif dans les valeurs de \bar{n} entre la simulation et le calcul exact. Nous remarquons, qu'en dehors du cas d'une loi de Pareto, la valeur de \bar{n} obtenue par simulation converge rapidement vers la valeur théorique attendue. Pour le cas où le temps de service est distribué selon une loi de Pareto, cette convergence s'avère beaucoup plus lente. La durée de la simulation doit être allongée d'un facteur d'environ 1000 afin que la valeur obtenue pour \bar{n} soit proche de la valeur exacte (écart inférieur à 1%).

A présent nous considérons le même ensemble de distributions mais avec un coefficient de variation égal à 2.0 (leurs paramètres figurent en Annexe). Nous reportons dans la Table 3 et sur la Fig. 2 les résultats obtenus. Comme l'illustre la comparaison des Tables 2 et 3, d'une manière générale les temps de convergence des simulations tendent à s'allonger lorsque le cv du temps de service augmente. Ces résultats suggèrent également la difficulté de simuler le comportement d'une file M/G/1 avec un temps de service distribué selon une loi de Pareto (cf. [GRO02]). Dans ce dernier cas, le temps de convergence dépasse de loin celui des autres exemples. Ainsi, même pour une durée de 1 milliard de fins de service, la valeur de \bar{n} obtenue par simulation demeure éloignée de plus de 3% de la valeur de Référence (cf. Fig. 2 et Table 3). Par ailleurs, ces résultats montrent que la taille de l'intervalle de confiance ne décroît pas toujours de façon monotone avec la durée de la simulation. Plus surprenant peut-être, nous observons sur cet exemple que l'intervalle de confiance englobe la valeur de Référence de \bar{n} pour certaines durées de simulation (1.E+03; 1.E+04; 1.E+05; 1.E+06; 1.E+07), mais que pour des durées plus longues (1.E+08; 1.E+09) cette valeur se situe à l'extérieur de l'intervalle de confiance. Ce dernier point illustre les risques potentiels liés à la présence de certaines distributions dans un modèle résolu par simulation (Law et Kelton en ont étudié [LAW84] les causes majeures).

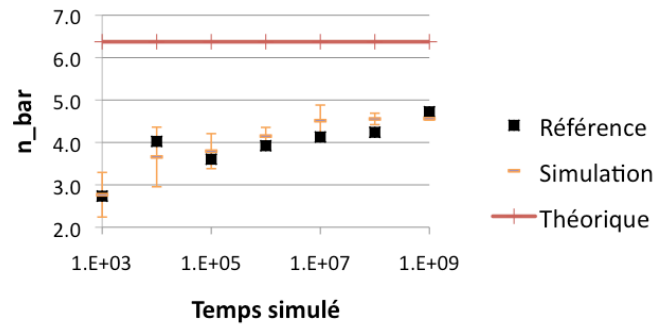


Fig. 2 – Convergence du comportement simulé d'une file $M/G/1$ vers la valeur exacte pour une distribution de Pareto avec $\rho=0.75$ et $cv = 2$

4 Conclusion

Dans cette note nous avons voulu rappeler que l'usage de certaines distributions dans les modèles tend à accentuer la complexité de leur résolution par simulation. Pour cela, nous avons considéré le nombre moyen de clients dans une file $M/G/1$ en régime stationnaire dont la valeur théorique est connue. Nous avons étudié l'impact des lois de la distribution du temps de service sur la précision des résultats des simulations. Nos résultats confirment que la résolution par simulation d'une file $M/G/1$ nécessite des durées de simulations très variables (facteur 1000) selon la loi du temps de service. L'effet de la distribution est d'autant plus prononcé que le coefficient de variation du temps de service et le taux d'utilisation du serveur sont élevés. Ainsi, si dans un modèle une loi n'est connue que par ses deux (ou n -premiers) moments, il convient de choisir la distribution la plus simple à simuler car un autre choix peut affecter sensiblement la précision et le temps de convergence de la simulation. Ainsi, il est intéressant de noter que c'est la loi de Pareto qui nous a posée le plus de difficultés alors même que l'usage de distributions à queue lourde (dont celle de Pareto) dans la littérature scientifique a décuplé (taille des flots, tailles des objets Web, etc).

Références

- [ALL90] ALLEN A. O., Probability, statistics and queuing theory with computer science applications, Vol. 2, Academic Press, 1990.
- [ASM00] ASMUSSEN K., BINSWANGER K., HOJGAARD B. "Rare events simulation for heavy-tailed distributions", Bernoulli, vol. 6, No. 2, 2000, pp. 303-322.
- [BRA87] BRATLEY P., FOX, B. L., SCHRAGE, L. E., A Guide to Simulation (2nd Ed.), Springer-Verlag New York, Inc, 1987.
- [GRO02] GROSS D., SHORTLE J.F., FISCHER M.J., MASI D.M. "Simulation input analysis: difficulties in simulating queues with Pareto service". In Proceedings of the 34th Winter Simulation Conference, pp. 407-415, 2002.
- [LAW84] LAW A.M., KELTON W.D., "Confidence Intervals for Steady-State Simulations: I. A Survey of Fixed Sample Size Procedures". Operations Research, vol. 32, No. 6, November-December 1984, pp. 1221-1239.
- [MOR04] MORSE P. M., Queues, Inventories and Maintenance: The Analysis of Operational Systems with Variable Demand and Supply, Dover Phoenix Editions.
- [PAW03] PAWLIKOWSKI K., "Do Not Trust All Simulation Studies of Telecommunication Networks". Lecture Notes in Computer Science, vol. 2662, 2003, pp. 899-908. Published also in Proc. Int. Conf. on Information Networking, ICOIN'03 (Jeju Island, Korea, February 2003).

Annexe

	$m=1 ; cv = \sqrt{3}$	$m=1 ; cv=1.2$	$m=1 ; cv=2.0$
Erlang	3 étages de taux 3.0	-	-
Uniforme	Min=0.0 ; Max=2.0	-	-
Hyper-Exponentielle	-	$\mu_1=1.404 ; p_1=0.288 ; \mu_2=3.476$	$\mu_1=1.127 ; p_1=0.1127 ; \mu_2=8.873$
Cox-2 (type 1)	-	$\mu_1=10.0 ; p_2=0.723 ; \mu_2=0.806$	$\mu_1=10.0 ; p_2=0.337 ; \mu_2=0.3745$
Cox-2 (type 2)	-	$\mu_1=1.111 ; p_2=3.12E-2 ; \mu_2=0.303$	$\mu_1=1.111 ; p_2=6.25E-3 ; \mu_2=6.25E-2$
Cox-2 (type 3)	-	$\mu_1=1.01 ; p_2=4.35E-4 ; \mu_2=4.348E-2$	$\mu_1=1.01 ; p_2=6.62E-5 ; \mu_2=6.623E-3$
Log-Normale (paramètres de la Normale associée)	Moyenne=-0.1438; Ecart-type=0.536	Moyenne=-0.446; Ecart-type=0.945	Moyenne=-0.8047; Ecart-type = 1.269
Pareto	Echelle=2/3; Forme=3.0	Echelle=0.5655; Forme=2.302	Echelle=0.5279; Forme=2.118

Table 4 – Paramètres des distributions considérées