Computers & Operations Research 65 (2016) 111–124

Contents lists available at ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/caor

Computational analysis of a Markovian queueing system with geometric mean-reverting arrival process

Daniel Wei-Chung Miao [a,b,*], Xenos Chang-Shuo Lin [c], Wan-Ling Chao [a]

[a] National Taiwan University of Science and Technology, Taipei, 106, Taiwan
[b] Imperial College Business School, South Kensington Campus, London SW7 2AZ, UK
[c] Aletheia University, No. 32, Zhenli St., Danshui Dist., New Taipei City 25103, Taiwan

ARTICLE INFO

Available online 30 June 2015

Keywords:
Markov-modulated Poisson process (MMPP)
Cox process
Diffusion process
Geometric mean-reverting (GMR)
Markovian queueing systems
Matrix geometric method

ABSTRACT

In the queueing literature, an arrival process with random arrival rate is usually modeled by a Markov-modulated Poisson process (MMPP). Such a process has discrete states in its intensity and is able to capture the abrupt changes among different regimes of the traffic source. However, it may not be suitable for modeling traffic sources with smoothly (or continuously) changing intensity. Moreover, it is less parsimonious in that many parameters are involved but some are lack of interpretation. To cope with these issues, this paper proposes to model traffic intensity by a geometric mean-reverting (GMR) diffusion process and provides an analysis for the Markovian queueing system fed by this source. In our treatment, the discrete counterpart of the GMR arrival process is used as an approximation such that the matrix geometric method is applicable. A conjecture on the error of this approximation is developed out of a recent theoretical result, and is subsequently validated in our numerical analysis. This enables us to calculate the performance measures with high efficiency and precision. With these numerical techniques, the effects from the GMR parameters on the queueing performance are studied and shown to have significant influences.

1. Introduction

One approach to model a time-varying and volatile arrival process is to consider a Poisson process whose intensity $\lambda(t)$ is also a random process. This is known as a doubly stochastic Poisson process, or a Cox process (see, e.g. [7,12,2,22]), in which the randomness in $\lambda(t)$ may bring in some desired features. The well-known MMPP (see, e.g. [8,9,13]) falls into this category in that the intensity process is modulated by a continuous-time Markov chain. With proper assumptions (e.g. phase-type distributed service time), a queueing system fed by an MMPP can be modeled by a large Markov chain for which some numerical techniques are readily available (e.g. the matrix geometric method of Neuts [18]). This makes MMPP a popular model for the traffic sources with random intensity.

Despite the convenience in modeling and the tractability in computation, there are some drawbacks of using MMPP as an arrival process. As commented in [10], there may be too many parameters involved in an MMPP. Consider the simplest case of a two-state MMPP, there are four parameters (transition rates between two states, and arrival rates at these states) to be fit (see, e.g. the fitting algorithm in [13]). But the number of parameters will grow dramatically (at an exponential rate) as the number of states increases and this limits the use of an MMPP with more states. In addition to the explosive growth in parameters, the lack of intuitive interpretations on these parameters (e.g. transition rates between states) also makes it difficult to understand their physical meanings. Furthermore, the abrupt change in its intensity makes the MMPP unsuitable for modeling the traffic sources with continuously changing random intensity or those without contrasting regimes (states).

From the perspective of a parsimonious model with interpretable parameters, the MMPP seems less competitive. Some new models are proposed to address these issues. One example is the discrete autoregressive (DAR) model (see, e.g. [10,11,15]) which is able to generate the geometric decaying autocorrelation with fewer parameters. The present study considers an alternative way to address the problems with MMPPs. We propose to use a geometric (GMR) mean-reverting diffusion process to model the intensity process. Such a process is continuous, positive, and mean-reverting, in that there is a long term mean and the process will be pulled back toward this mean when it deviates away. It is commonly used in finance for modeling asset prices with such features. For example, in [3,19,20], the GMR is used to model fuel and electricity prices. Here in the context of queueing, mean-reversion is motivated by the traffic control mechanism which will

* Corresponding author.
E-mail address: miao@mail.ntust.edu.tw (D.-C. Miao).

http://dx.doi.org/10.1016/j.cor.2015.06.015
0305-0548/

divert part of incoming traffic away when a line becomes heavily loaded, and will route some traffic back when it is less busy. Consider a queue in the middle stage of this line. When the traffic intensity goes up (down), the mechanism will bring an opposite effect against this upward (downward) trend. This causes the traffic intensity to be continuously changing and mean-reverting. Other motivations of using mean-reverting traffic sources can be found in recent studies. For example, [24] proposed to use a source of this kind to model "overdispersion" and "autocorrelation" as observed in call center traffic arrivals. Compared to modeling the intensity by a Markov chain, using GMR is able to reflect the smooth (instead of abrupt) changes in the intensity process. More importantly, the GMR is more parsimonious since fewer parameters are required to characterize the main features and each of them has clear physical interpretation. When the intensity is modulated by a GMR process, the arrival process is termed a GMR–MPP. The differences between Poisson processes with discrete and continuous intensities are depicted in Fig. 1.

To understand how such a continuous-state process influences the queueing performance, we consider the Markovian queueing system GMR–MPP/M/s. Because of the continuous nature in intensity, this system cannot be directly formulated as a large Markov chain. To make the existing numerical techniques applicable, we consider its counterpart system as an approximation, where the arrival process is replaced by DGMR(m)-MPP, a discrete-state (but still continuous-time) version of GMR-MPP. This system can be analyzed by the matrix-geometric method, but it remains to investigate how good the approximation is. Motivated by a theoretical convergence result, we establish a conjecture on the relation between the performance measures of the original and approximate systems. This conjecture is validated in the subsequent numerical analysis where the errors on the first fourth moments of queue length and waiting time are examined. Our results demonstrate that using the matrix geometric method together with Richardson extrapolation is able to provide very accurate and efficient estimates on the queueing performance. These numerical techniques are then applied to investigate the influences from the key traffic parameters on the queueing performance measures.

This rest of this paper is organized as follows. Section 2 introduces the GMR arrival process as well as its discrete-state counterpart. Section 3 formulates the counterpart queueing system as a large Markov chain and applies the matrix geometric method to analyze its performance. Section 4 presents the numerical results which validate our conjecture and investigate the influences from the GMR parameters. Finally Section 5 gives conclusions.

2. The geometric mean-reverting arrival process

This section introduces the geometric mean-reverting (GMR) process and defines the GMR–MPP arrival process. We also introduce its discrete counterpart in order to apply the existing numerical techniques. For the convenience of the subsequent queueing analysis, a conjecture is proposed and will be validated in our numerical study.

A process $\lambda(t)$ is said to follow a GMR diffusion process if it solves the following stochastic differential equation (see [19,20])

$$d\ \ln\lambda(t) = \kappa(\theta - \ln\lambda(t))dt + \sigma dW(t), \quad \lambda(0) = \lambda. \tag{1}$$

The process $\lambda(t)$ is driven by the standard Brownian motion $W(t)$. Each of the three parameters κ, θ, and σ has its own physical meaning: κ is the speed of mean reversion, θ is the long-term mean, and σ is the volatility. If we define $X(t) = \ln\lambda(t)$, then $X(t)$ is the well known Ornstein–Uhlenbeck process (originally proposed in the physical literature [21]) which satisfies the following stochastic differential equation:

$$dX(t) = \kappa(\theta - X(t))dt + \sigma dW(t), \quad X(0) = \ln\lambda. \tag{2}$$

The mean reversion property is seen in its drift term: it will be pulled up if $X(t) < \theta$ and vice versa. When the OU process is at steady state, $X(t)$ follows a normal distribution with mean $\mathsf{E}[X(t)] = \theta$ and variance $\mathsf{Var}[X(t)] = \sigma^2/2\kappa$. Its moment generating function is given by

$$\mathsf{E}[e^{\alpha X(t)}] = \exp\left(\alpha\theta + \frac{\alpha^2\sigma^2}{4\kappa}\right). \tag{3}$$

We are now in a position to define the Poisson process modulated by the above GMR process and discuss its basic properties.

2.1. The GMR–MPP and its basic properties

The GMR modulated Poisson process (GMR–MPP) is a Poisson process whose intensity follows a GMR process as in (1). For convenience, we assume the stationarity of the underlying processes $\lambda(t)$ and $X(t)$ (i.e. they are already at steady state). For such a doubly stochastic Poisson process, conditional on the intensity over the interval $[0,t]$ (i.e. given $\lambda(s), 0 \le s \le t$), it can be seen as a time-inhomogeneous Poisson process and the number of arrivals $A(t)$ during $[0,t]$ has the following distribution:

$$\mathsf{P}(A(t) = n \mid \lambda(s), 0 \le s \le t) = \frac{e^{-\int_0^t \lambda(s)\, ds}}{n!}\left(\int_0^t \lambda(s)\, ds\right)^n, \quad n = 0, 1, 2, \ldots.$$

Taking expected value over all the random paths, the unconditional probability distribution of $A(t)$ can be expressed as

$$\mathsf{P}(A(t) = n) = \mathsf{E}\left[\frac{e^{-\int_0^t \lambda(s)\, ds}}{n!}\left(\int_0^t \lambda(s)\, ds\right)^n\right], \quad n = 0, 1, 2, \ldots.$$

This distribution is crucial to the queueing performance when the system is fed by such an arrival process. The above formula also reveals the complication caused by the randomness in $\lambda(t)$. It is clear that if $\sigma = 0$ or $\kappa \to \infty$, this randomness in $\lambda(t)$ will disappear (i.e. $\lambda(t) = e^\theta$ is constant) and the GMR–MPP will degenerate to a usual Poisson process.

Note that under stationarity $X(t)$ is normally distributed (with mean and variance given above), the intensity $\lambda(t)$ of the GMR–MPP at any given time t follows a log-normal distribution. By using (3) with $\alpha = 1$, its mean is be obtained as

$$\bar{\lambda} = \mathsf{E}[\lambda(t)] = \exp\left(\theta + \frac{\sigma^2}{4k}\right). \tag{4}$$

Its variance can also be obtained from (3) with $\alpha = 2$ as

$$\mathsf{Var}[\lambda(t)] = \exp\left(2\theta + \frac{\sigma^2}{2\kappa}\right)\left[\exp\left(\frac{\sigma^2}{2\kappa}\right) - 1\right].$$

We see that the parameters σ and κ have a major influence on the variance of intensity. Larger σ ($\lambda(t)$ is more volatile) and smaller κ ($\lambda(t)$ is pulled back to its mean more slowly when it moves away) will lead to larger variance, i.e. the distribution of $\lambda(t)$ and in turn $A(t)$ are more widely spread. These observations suggest that the triplet $(\bar{\lambda}, \sigma, \kappa)$ is a more meaningful parameter set which is interchangeable with (κ, θ, σ). As expected, when $\sigma \to 0$ or $\kappa \to \infty$, the variance tends to zero and the GMR–MPP reduces to a Poisson process.

The focus of this study is on a Markovian queueing system with such a GMR–MPP traffic source. The influence from the mean intensity $\bar{\lambda}$ is better known and more straightforward. We will mainly investigate the influences from σ and κ on the distribution of $A(t)$ and in turn on the queueing performance.