ARTICLE IN PRESS

Performance Evaluation ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Performance Evaluation

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

Rare-event analysis of mixed Poisson random variables, and applications in staffing

Mariska Heemskerk[a,*], Julia Kuhn[a,b], Michel Mandjes[a,c,d]

[a] Korteweg–de Vries Institute for Mathematics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands
[b] The University of Queensland, St Lucia, Queensland, Australia
[c] Eurandom, Eindhoven University of Technology, Eindhoven, The Netherlands
[d] Amsterdam Business School, Faculty of Economics and Business, University of Amsterdam, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Queueing
Overdispersion
Rare events
Large deviations
Importance sampling
Staffing

ABSTRACT

A common assumption when modeling queueing systems is that arrivals behave like a Poisson process with constant parameter. In practice, however, call arrivals are often observed to be significantly overdispersed. This motivates that in this paper we consider a *mixed* Poisson arrival process with arrival rates that are resampled every $N^{-\alpha}$ time units, where $\alpha > 0$ and N a scaling parameter.

In the first part of the paper we analyze the asymptotic tail distribution of this doubly stochastic arrival process. That is, for large N and i.i.d. arrival rates $X_1, \ldots, X_N$, we focus on the evaluation of the probability that the scaled number of arrivals exceeds Na,

$$P_N(a) := \mathbb{P}\left(\text{Pois}\left(N\overline{X}_{N^\alpha}\right) \geqslant Na\right), \quad \text{with} \quad \overline{X}_N := \frac{1}{N}\sum_{i=1}^{N} X_i.$$

The logarithmic asymptotics of $P_N(a)$ are easily obtained from previous results; we find constants r_P and γ such that $N^{-\gamma}\log P_N(a) \to -r_P$ as $N \to \infty$. Relying on elementary techniques, we then derive the exact asymptotics of $P_N(a)$: For $\alpha < \frac{1}{3}$ and $\alpha > 3$ we identify (in closed-form) a function $\widetilde{P}_N(a)$ such that $P_N(a)/\widetilde{P}_N(a)$ tends to 1 as $N \to \infty$. For $\alpha \in [\frac{1}{3}, \frac{1}{2})$ and $\alpha \in [2, 3)$ we find a partial solution in terms of an asymptotic lower bound. For the special case that the X_is are gamma distributed, we establish the exact asymptotics across all $\alpha > 0$. In addition, we set up an asymptotically efficient importance sampling procedure that produces reliable estimates at low computational cost.

The second part of the paper considers an infinite-server queue assumed to be fed by such a mixed Poisson arrival process. Applying a scaling similar to the one in the definition of $P_N(a)$, we focus on the asymptotics of the probability that the number of clients in the system exceeds Na. The resulting approximations can be useful in the context of staffing. Our numerical experiments show that, astoundingly, the required staffing level can actually *decrease* when service times are more variable.

* Corresponding author.
E-mail addresses: j.m.a.heemskerk@uva.nl (M. Heemskerk), j.kuhn@uva.nl (J. Kuhn), m.r.h.mandjes@uva.nl (M. Mandjes).

http://dx.doi.org/10.1016/j.peva.2017.03.006
0166-5316/

Please cite this article in press as: M. Heemskerk, et al., Rare-event analysis of mixed Poisson random variables, and applications in staffing, Performance Evaluation (2017), http://dx.doi.org/10.1016/j.peva.2017.03.006.

1. Introduction

In communications engineering it is increasingly accepted that traditional Poisson processes do not succeed in capturing the variability that is typically observed in real call arrival processes [1,2]. This led to the idea to instead use Cox processes [3] to model arrivals, i.e., Poisson processes in which the arrival rate follows some (non-negative) stochastic process. Perhaps the simplest choice, advocated in [4], is to *resample* the arrival rate (in an i.i.d. manner) every Δ units of time; during the resulting time intervals the arrival rate is assumed constant. We denote these i.i.d. arrival rates by $(X_i)_{i\in\mathbb{N}}$. This paper studies two settings in which such an overdispersed arrival process is featured.

1. *Number of arrivals*. We start by studying the tail asymptotics of the total number of arrivals in a time interval of given length. We do so in a scaling regime that was proposed in [4], in which the arrival rates and sampling frequency are jointly inflated as follows. In the first place, it is natural to assume that arrival rates are large, as these represent the contributions of many potential clients; this can be achieved by letting these arrival rates be $NX_1, NX_2, \dots$ for i.i.d. $(X_i)_{i\in\mathbb{N}}$ and some large N. In addition, the sampling frequency is set to N^α (assumed to be integer) and hence the size of each time slot is assumed to be $\Delta = N^{-\alpha}$. Evidently, the larger α, the more frequently the arrival rate is resampled.

The focus is on the probabilities $P_N(a)$ and $p_N(a)$, where

$$P_N(a) := \mathbb{P}\left(\text{Pois}\left(N\overline{X}_{N^\alpha}\right) \geqslant Na\right), \quad \text{with} \quad \overline{X}_N := \frac{1}{N}\sum_{i=1}^{N} X_i,$$

and $p_N(a)$ denotes the corresponding probability that the mixed Poisson random variable equals Na (assumed to be integer). We consider the situation that a is larger than $\nu := \mathbb{E}X_i$, which entails that the event under consideration is rare and that we are in the framework of large deviations theory.

We would like to stress the important role that is played by the time-scale parameter $\alpha > 0$. One could imagine that in a rapidly changing environment, the inherent overdispersion of the arrival process hardly plays a role, whereas in a slowly changing random environment, overdispersion is expected to be more dominant. Hence the parameter α can be tweaked in order to match any real-world scenario in that sense. That is, if α is large, since the arrival rate is resampled relatively frequently, it is anticipated that the mixed Poisson random variable behaves Poissonian with parameter $N\nu$. If on the contrary α is small, one would expect that detailed characteristics of the distribution of the X_i matter. For $\alpha = 1$ both effects play a role. This intuition underlies nearly all results presented in this paper.

2. *Number of customers in an infinite-server queue.* In the second part of this paper we focus on a cornerstone model in the design and performance evaluation of communication networks: the *infinite-server queue*. This model can be used to produce approximations for many-server systems. In our paper, the arrival process is the overdispersed process we introduced above, and the service times are i.i.d. samples from a (non-negative) distribution with distribution function $F(\cdot)$. The number of clients in this infinite-server queue, under the arrival process described above, is studied in [4]. As it turns out, one can prove the (conceivable) property that the number of clients in the system at time t (which we, for simplicity, assume to be a multiple of Δ), has a *mixed Poisson* distribution, i.e., a Poisson distribution with random parameter. This parameter is given by

$$\sum_{i=1}^{t/\Delta} X_i\,\Delta\, f_i(t,\Delta),$$

where $f_i(t,\Delta)$ denotes the probability that a call arriving at a uniformly distributed epoch in the interval $[(i-1)\Delta, i\Delta)$ is still in the system at time t. Evidently, for small Δ this probability essentially behaves as $\overline{F}(t-i\Delta)$, with $\overline{F}(\cdot) := 1 - F(\cdot)$ denoting the complementary distribution function.

We renormalize time such that $t \equiv 1$ (which can be done without loss of generality), and again impose the scaling along the lines of [4]: the arrival rates are NX_i and the interval width $N^{-\alpha}$. Then the number of clients in the system is Poisson with random parameter

$$\sum_{i=1}^{N^\alpha}(NX_i)\,N^{-\alpha}\, f_i(1,N^{-\alpha}) = N^{1-\alpha}\sum_{i=1}^{N^\alpha} X_i\,\omega_i(N^\alpha), \tag{1}$$

where $\omega_i(N) := f_i(1,N^{-1}) \approx \overline{F}(1-i/N)$. A clearly relevant object of study concerns the probability that the number of clients in the system exceeds some threshold Na:

$$Q_N(a) := \mathbb{P}\left(\text{Pois}\left(N^{1-\alpha}\sum_{i=1}^{N^\alpha} X_i\,\omega_i(N^\alpha)\right) \geqslant Na\right); \tag{2}$$

$q_N(a)$ denotes the corresponding probability that the mixed Poisson random variable equals Na. To ensure that the event under consideration is rare, a is assumed to be larger than

$$\frac{\nu}{N^\alpha}\sum_{i=1}^{N^\alpha}\omega_i(N^\alpha) \approx \frac{\nu}{N^\alpha}\sum_{i=1}^{N^\alpha}\overline{F}(1-i/N^\alpha) \approx \nu\int_0^1 \overline{F}(x)\mathrm{d}x.$$

Please cite this article in press as: M. Heemskerk, et al., Rare-event analysis of mixed Poisson random variables, and applications in staffing, Performance Evaluation (2017), http://dx.doi.org/10.1016/j.peva.2017.03.006.