Large deviations and statistical analysis for queueing systems with regenerative input flow

Larisa G. Afanasyeva
Lomonosov Moscow State University, Moscow, Russia - afanas@mech.math.msu.su

Elena Bashtova
Lomonosov Moscow State University, Moscow, Russia - bashtovaelena@rambler.ru

Andrey Tkachenko
National Research University Higher School of Economics, Moscow, Russia - tkachenko_av@hse.ru

1 Introduction

Studying of queueing models is an appealing part of applied mathematics because queues are familiar and intuitively clear and they can be need to model many real systems. We consider a single-server queueing system specified by regenerative input flow and independent identically distributed service times. A stochastic flow is regenerative if there are moments where it starts anew independently of the past. The majority of flows used in queueing theory are regenerative therefore our model extend classical models. Regeneration usually is not a primary assumption in applications. Rather regeneration is a key property that makes processes amenable for analysis. Given the input flow and service requirements, we estimate the probability that queue length exceeds a high level. The basic tool for studying these rare events is large deviation theory. There is an extensive literature deals with corresponding limit theorems (see, e.g. [1] and reference there). We establish conditions providing the existence of logarithmic limit $x^{-1}(1-F(x)) = -q$ as $x \to \infty$ for distribution function $F(x)$ of waiting time in stationary regime. It turns out that the parameter q is defined by the Laplace-Stieltjes transforms (LST) of the service time distribution and LST of the joint distribution of the regeneration period and the number of customers arrived during this period. For applications of this result we need to estimate parameter q with the help of statistical data.

If we observe n regeneration periods of the input flow and service times of all arrived customers, then an unbiased and consistent (as n goes to infinity) estimate for the parameter q is proposed. The confidence interval is also can be constructed. Another statistical problem is the estimation of parameters of the input flow. We consider two situations. Firstly, it is supposed that regeneration points of the input flow and the times of customer arrivals are observed. Then classical statistical methods can be applied. In the second case the times of customer arrivals are only observed. Then we need a new approach. Due to asymptotic stationarity of increments and other properties of regenerative flows as well as results from renewal theory we propose consistent and asymptotically unbiased estimates for basic parameters of the input flow.

2 Model description

Suppose an integer-valued stochastic process $\{X(t), t \geq 0\}$ is defined on some probability space $(\Omega, \mathcal{F}, P)$ and $X(t)$ has nondecreasing right-continuous sample pathes and $X(0) = 0$. Assume that there exists a filtration $\{\mathcal{F}_t\}_{t\geq 0}$, $(\mathcal{F}_t \in \mathcal{F})$ such that $X(t)$ is measurable with respect to $\mathcal{F}_t$ $(t \geq 0)$.

Definition 1. *The stochastic flow $X(t)$ is called regenerative if there is an increasing sequence of Markov moments $\{\theta_i,\ i \geq 0\}, \theta_0 = 0$ (with respect to $\{\mathcal{F}_t\}_{t\geq 0}$) such that the sequence*

$$\{\varkappa_i\}_{i=1}^{\infty} = \{X(\theta_{i-1} + t) - X(\theta_{i-1}), \theta_i - \theta_{i-1}, t \in [0, \theta_i - \theta_{i-1})\}_{i=1}^{\infty}$$

consists of independent identically distributed random elements on $(\Omega, \mathcal{F}, P)$.

This definition is close to the one in [2]. So we define zero-delayed regenerative flow, but all our results are true if random element $\varkappa_1$ has a distribution different from distribution of other elements. The random variable θ_i is said to be the ith regeneration moment of $X(t)$ and $\tau_i = \theta_i - \theta_{i-1}$ is the ith regeneration period ($i = 1, 2, \ldots$). Let $\xi_i = X(\theta_i) - X(\theta_{i-1})$ be the number of arrived customers during the ith regeneration period. Assume that $\mu = E\tau_1 < \infty$, $a = E\xi_1 < \infty$. The limit $\lambda = \lim_{t\to\infty} \frac{X(t)}{t}$ a.s. is called the *intensity* of $X(t)$. It is easy to prove that $\lambda = \frac{a}{\mu}$ (e.g., see [3]).

We consider a single-server queueing system with regenerative input flow $X(t)$ and independent identically distributed service times $\{\eta_n\}_{n\geq 1}$ with distribution function $B(x) = \mathbf{P}\{\eta_n \leq x\}$ and $b = \mathbf{E}\eta_n < \infty$.

We focus on model with a regenerative flow for two reasons. Firstly, a process describing the system under some natural conditions turns out to be a classical regenerative process and renewal theory gives very efficient tools for its asymptotic analysis. Secondly, the class of regenerative flows is rather wide and includes the most usually used in queueing theory. We establish conditions providing the existence of logarithmic limit $x^{-1}(1 - F(x)) = -q$ as $x \to \infty$ for distribution function $F(x)$ of waiting time in stationary regime. Our analysis is the generalization of the well-known characterization of the $GI|G|1$ queue obtained using classical probabilistic technique of the exponential change of measure and renewal theory (see, e.g. [4]). We also employ the properties of regenerative flows [3]. To prove our results we need additional assumption.

Condition 1. *The greatest common devisor of numbers* $\{i = 1, 2, \ldots\}$ *such that* $\mathbf{P}\{\xi_1 = i\} > 0$ *is equal to one.*

Let us note that the distribution of stochastic flow $X(t)$ is defined by the distribution of the random elements $\{\varkappa_i\}_{i\geq 1}$. However, the asymptotic behaviour of the limit distribution of waiting time is defined by functions

$$G(z, s) = \mathbf{E}z^{\xi_1}e^{-s\tau_1},\ b(s) = \mathbf{E}e^{-s\eta_1},\ Re\ s \geq 0,\ |z| < 1.$$

We use the following property of regenerative flows established in [3]. Let t_n be the time of the nth customer arrival ($n = 1, 2, \ldots$) and $\{\zeta_n = t_n - t_{n-1}\}_{n=1}^{\infty}$ ($t_0 = 0$) be a sequence of interarrival times. If Condition 1 holds, then the sequence $\{\zeta_{n+k}\}_{k=1}^{\infty}$ weakly converges (as $n \to \infty$) to a stationary metrically transitive sequence $\{\widetilde{\zeta}_k\}_{k=1}^{\infty}$. Let $W(t)$ be the virtual waiting time process and

$$w_n = W(t_n - 0), \quad W_n = W(\theta_n - 0).$$

It is shown in [3] that these processes have proper limit distributions if traffic coefficient $\rho = \lambda b < 1$. We further assume this condition to be fulfilled. Denote

$$\Phi(x) = \lim_{n\to\infty} \mathbf{P}(W_n \leq x),\ F(x) = \lim_{n\to\infty} \mathbf{P}(w_n \leq x). \tag{1}$$

The first result concerns the asymptotic behaviour of these functions as $x \to \infty$.

3 Large deviations

Theorem 1. *Let Condition 1 be fulfilled,*

$$\delta_0 = \sup\{s : G(b(-s), s) < \infty\} > 0 \tag{2}$$

and

$$G(b(-\delta_0), \delta_0) > 1. \tag{3}$$

Then there are limits

$$\lim_{x\to\infty} x^{-1} ln(1 - F(x)) = -q, \tag{4}$$

$$\lim_{x\to\infty} x^{-1} ln(1 - \Phi(x)) = -q, \tag{5}$$

where q is a unique positive solution of the equation

$$G(b(-s), s) = 1. \tag{6}$$

Proof. Firstly, we consider $F(x)$. In view of weak convergence $\{\zeta_{n+k}\}_{k=1}^{\infty}$ to a stationary sequence as $n \to \infty$ our analysis is based on results from [1][Th. 3.1]. We have to establish the existence

$$\lim_{n\to\infty} n^{-1}\Lambda_n(s) = \Lambda(s), \tag{7}$$

where $\Lambda_n(s) = ln\mathbf{E}e^{s\sum_{j=1}^{n}(\eta_j-\zeta_j)}$ and $\Lambda(s)$ is a finite and differentiable function in some neighborhood of zero.

Lemma 1. *Under conditions of Theorem 1 the limit* (7) *exists and*

$$\Lambda(s) = ln\frac{b(-s)}{z_1(s)}, \tag{8}$$

where $z_1(s)$ is a unique solution of the equation

$$G(z, s) = 1 \tag{9}$$

with respect to z and $z_1(s) > 1$ as $s > 0$.

Proof of the Lemma. Since $\Lambda_n(s) = n\ln b(-s) + \ln \mathbf{E}e^{-st_n}$ it is sufficient to prove that

$$\lim_{n\to\infty} n^{-1}\ln \mathbf{E}e^{-st_n} = -\ln z_1(s). \tag{10}$$

Let $\nu(n) = \min\{k : \xi_1 + \ldots + \xi_k \geq n\}$. Then

$$\theta_{\nu(n)-1} < t_n \leq \theta_{\nu(n)} \tag{11}$$

and $\varphi_n(s) = \mathbf{E}e^{-s\theta_{\nu(n)}}$ satisfies the equation

$$\varphi_n(s) = \Psi_n(s) + \sum_{j=0}^{n-1}\varphi_{n-j}(s)g_j(s), \quad (s > 0) \tag{12}$$

where $\Psi_n(s) = \mathbf{E}e^{-s\tau_1}\mathbb{I}(\xi_1 \geq n)$, $g_j(s) = \mathbf{E}e^{-s\tau_1}\mathbb{I}(\xi_1 = j)$. Since $\sum_{j=0}^{\infty} g_j(s) < 1$ employing Theorem 2 from [5][Ch. 11, par. 6] we get

$$\varphi_n(s) \sim \frac{C}{(z_1(s))^n} \quad (n \to \infty), \tag{13}$$

where $z_1(s)$ is a unique solution of (9), $z_1(s) > 1$ $(s \geq 0)$ and C ia a constant. In view of (11) we get (10) from (12). ■

Proof of the Theorem (continuous). From Lemma 1 it follows that $\Lambda(s)$ is defined by (8). Therefore, conditions of Theorem 3.1 from [1] are fulfilled and we get (4), where q is a unique positive solution of the equation $b(-s) = z_1(s)$.

Consider the process W_n. Let γ_n be the total service time of customers arrived during the nth regenerative period. We introduce the auxiliary process W_n^- by the recurrent relation

$$W_n^- = \left[W_{n-1}^- + \gamma_n - \tau_n\right]^+, W_0^- = 0.$$

Then $W_n^- \geq W_n$ with probability 1 (w.p.1) if $W_0 = 0$. It is well known (see, e.g. [6]) that function $\Phi^-(x) = \lim_{n\to\infty}\mathbf{P}(W_n^- \leq x)$ has the asymptotics

$$1 - \Phi^-(x) \sim Ce^{-qx}, \text{ as } x \to \infty, \tag{14}$$

where q is a positive solution of (13) and C is a constant. Therefore

$$\liminf_{x\to\infty} x^{-1}\ln(1 - \Phi(x)) \geq -q. \tag{15}$$

The estimate of the upper limit is based on the inequality $W_n \le w_{\mu(n)} + \eta_{\mu(n)}$ w.p.1., where $\mu(n) = \xi_1 + \ldots + \xi_n$. Here random variables $w_{\mu(n)}$ and $\eta_{\mu(n)}$ are independent. One may easily see that for any $\varepsilon > 0$ and some constant C_ε

$$1 - \Phi(x) \le Ce^{-(q-\varepsilon)x}, \ x \to \infty. \tag{16}$$

Now the proof of (5) follows from (15) and (16). ■ For the sake of brevity let us assume that the distribution of regenerative period τ_1 of the input flow $X(t)$ has an absolutely continuous component. Then there exists $\lim_{t\to\infty} \mathbf{P}\{W(t) \le x\} = \Psi(t)$ and $\Psi(x)$ is a distribution function if $\rho < 1$ (see e.g. [7] and [8]).

Corollary 1. *Under conditions of Theorem 1 there exists*

$$\lim_{x\to\infty} x^{-1} \ln(1 - \Psi(x)) = -q, \tag{17}$$

where q is a unique positive solution of the equation (6).

Proof. Without loss of generality we may assume that $\mathbf{P}(\xi_1 > 0) = 1$. Define $l(t) = \sup\{n \ge 1 : \theta_n \le t\}$, $n(t) = \sum_{j=1}^{l(t)} \xi_j$. Then

$$\mathbf{P}(W_{l(t)-1} - \tau_{l(t)} > x) \le \mathbf{P}(W(t) > x) \le \mathbf{P}(w_{n(t)} + \eta_{n(t)} > x) \tag{18}$$

and random variables $W_{l(t)-1}, \tau_{l(t)}$ as well as $w_{n(t)}, \eta_{n(t)}$ are independent. Employing the basic result from the renewal theory and (18) we get

$$\mu^{-1} \int_0^\infty y(1 - \Phi(x+y)) d\mathbf{P}(\tau_1 \le y) \le 1 - \Psi(x) \le \mu^{-1} \int_0^x (1 - F(x-y)) dB(y) + 1 - B(x).$$

Now (17) follows from these inequalities and (4) and (5). ■

4 Statistical analysis

4.1 Statistical estimation for q

Assume that we observe n cycles of the regenerative input flow $X(t)$ that is $\{\tau_j, \ \xi_j\}_{j=1}^n$ and distribution function of service times $B(x)$ is known. Let $b(s) = \int_0^\infty e^{-sx} dB(x)$ and

$$\beta = \sup\{s : b(-s) < \infty\}. \tag{19}$$

We need the following conditions to be fulfilled

$$0 < \beta < \infty; \quad b(-\beta) = \infty. \tag{20}$$

Let $f(s) = G(b(-s), s) = \mathbf{E}(b(-s)^{\xi_1} e^{-s\tau_1})$. It follows from (20) and ergodicity condition that $f(\beta) = \infty$, $f(0) = 1$, $f'(0) = \lambda b - \boldsymbol{\tau_1} < 0$. Besides, there is a unique positive solution q of the equation $f(s) = 1$ and $0 < q < \beta$. As the estimate $f_n(s)$ of the function $f(s)$ we take

$$f_n(s) = n^{-1} \sum_{j=1}^n (b(-s))^{\xi_j} e^{-s\tau_j}. \tag{21}$$

For any integer $k > 0$ we put $\Delta_k = \beta 2^{-k}$ and define $j_{n,k} = \min\{i : f_n(i\Delta_k) > 1\}$.

Theorem 2. *Let conditions* (20) *be fulfilled and*

$$q^*_{n,k} = j_{n,k}\Delta_k. \tag{22}$$

Then for any integer $k > 0$

$$\mathbf{P}\{|q^*_{n,k} - q| > \Delta_k\} \xrightarrow[n\to\infty]{} 0. \tag{23}$$

Proof. Firstly, we note that in accordance with strong law of large numbers for any $k = 1, 2, \ldots$

$$\mathbf{P}\{f_n(i\Delta_k) \xrightarrow[n\to\infty]{} f(i\Delta_k),\ i = 0, 1, \ldots 2^{k-1}\} = 1. \tag{24}$$

Let $j_k = \min\{i : f(i\Delta_k) > 1\}$ so that $q \in [(j_k - 1)\Delta_k, j_k\Delta_k)$. Since $j_{n,k} \to j_k$ as $n \to \infty$ w.p.1 (in view of (24)) we get (23). ■

Basing on Theorem 2 for any $\varepsilon > 0$ and $\delta > 0$ one may find n_ε and construct the estimate q_n^* for the parameter q so that

$$\mathbf{P}\{|q_n^* - q| > \varepsilon\} < \delta \text{ as } n > n_0.$$

Of course, to do this we need to estimate $Var\ q_{n,k}^*$.

Now we consider the case when $B(x)$ is unknown but conditions (20) hold for some known β. Let $\{\eta_{i,j}, i = \overline{1, \xi_j}\}$, be service times of customers arrived during the jth cycle ($j = \overline{1, n}$). As estimate $\widetilde{f}_n(s)$ for $f(s)$ we take

$$\widetilde{f}_n(s) = n^{-1} \sum_{j=1}^{n} e^{s \sum_{i=1}^{\xi_j} \eta_{i,j} - s\tau_j}.$$

One may easily verify that Theorem 2 is true if we take $\widetilde{f}_n(s)$ instead of $f_n(s)$.

4.2 Statistical estimates for basic parameters of a regenerative flow.

For queueing systems with rather complicated input flows it is impossible (with rare exception) to obtain explicit expression of their operating characteristics. Therefore, the study of the limit behaviour of systems in heavy traffic situation is a very important problem. There is a extensive literature dealing with functional limit theorems for processes describing queueing systems in this situation (see [9], [6], [7], [10], [8] and references there). For application of the results for system with a regenerative input flow $X(t)$ we need to estimate the parameters $\lambda_X = \frac{a}{\mu}$ and

$$\sigma_X^2 = \lim_{t\to\infty} \frac{Var\ X(t)}{t} = \frac{Var\ \xi_1}{\mu} + \frac{a^2 Var\ \tau_1}{\mu^3} - \frac{2a\ cov(\xi_1, \tau_1)}{\mu^2}. \tag{25}$$

This section is devoted to this problem. Thus, let $\{X(t),\ t \ge 0\}$ be a regenerative flow and $N(t)$ the number of regenerative points in $[0, t]$, i.e. $N(t) = \max\{n \ge 0 : \theta_n < t\}$. If both processes $\{X(u), N(u), u \in [0, t]\}$ are observed we may estimate λ_X and σ_X^2 basing on (25) and traditional statistical methods. By means of the renewal theory it is possible to prove that these estimates are consistent, asymptotically unbiased and asymptotically normal. It is not difficult to construct the confidence interval as well.

Problems arise if only process $X(t)$ is observed. The estimate of intensity λ_X has the form $\widehat{\lambda}_X(t) = t^{-1}X(t)$ and $\widehat{\lambda}_X(t) \to \lambda_X$ w.p.1 as $t \to \infty$. If $\lambda_X(t)$ is known we put $\widehat{\sigma}_X^2(t) = t^{-1}(X(t) - \lambda_X t)^2$ and under some natural assumptions $\widehat{\sigma}_X^2(t) \to \sigma_X^2$ w.p.1 as $t \to \infty$. If λ_X is unknown we need to use another approach. Choosing $A > 0$ we use the following notation

$$Z_k(A) = X(kA) - X((k-1)A),\ k = 1, 2, \ldots;$$

$$\Delta_n(A) = (nA)^{-1} \sum_{k=1}^{n} (Z_k(A) - \lambda_X A)^2,\ \widehat{\Delta}_n(A) = (nA)^{-1} \sum_{k=1}^{n} (Z_k(A) - \widehat{\lambda}_X(nA)A)^2.$$

Theorem 3. *Let $\{X(t), t \ge 0\}$ be a regenerative flow and $\mathbf{E}\xi^4 < \infty$, $\mathbf{E}\tau^4 < \infty$. Then $\widehat{\Delta}_n(A)$ converges in probability to σ_X^2 as $n \to \infty$, $A \to \infty$.*

Proof The proof is based on the convergence $Cov(X(t), X(t + A) - X(t)) \to C_1$ as $t \to \infty$, $A \to \infty$ (see [11], [12]). We show that $Cov(\widehat{Z}_k^2(A),\ \widehat{Z}_j^2(A)) < C_2 < \infty$ ($k \ne j$). Here C_1 and C_2 are constants. This means that $(nA)^{-2} \sum_{j\ne k} Cov(\widehat{Z}_k^2(A),\ \widehat{Z}_j^2(A)) \to 0$ as $n \to \infty$, $A \to \infty$. Since $Var\ \widehat{Z}_k^2(A) \sim C_3 A$ as $A \to \infty$ (see [11]) then $Var\ \Delta_n(A) \to 0$ as $n \to \infty$, $A \to \infty$. In view of convergence $\widehat{\lambda}_X(t) \to \lambda_X$ w.p.1 as $t \to \infty$ we have

$Var\, \widehat{\Delta}_n(A) \to 0$ as $n \to \infty$, $A \to \infty$. Note that we take $t = nA$. Finally, we get convergence $\widehat{\Delta}_n(A) \to \sigma_X^2$ in probability as $n \to \infty$, $A \to \infty$.■

Now let the process $X(t)$ be observed on interval $(0, t)$. Set $A = t^\alpha$, $n = [t^{1-\alpha}]$, $\alpha \in (0, 1)$. Then estimate for σ_X^2 has the form

$$\widehat{\sigma}_X^2(t) = t^{-1} \sum_{k=1}^{[t^{1-\alpha}]} (\widehat{Z}_k(t^\alpha) - X(t)t^{-(1-\alpha)})^2.$$

According to Theorem 3 we get a consistent estimate for σ_X^2. This estimate satisfies the following asymptotic relation as $t \to \infty$

$$Var\, \widehat{\sigma}_X^2(t) \sim C_1 t^{-1} + C_2 t^{-2\alpha},$$

where C_1 and C_2 are some constants. For $\alpha \geq 0.5$ the main term of $Var\, \widehat{\sigma}_X^2(t)$ has order t^{-1} whereas $\alpha < 0.5$ the order is $t^{-2\alpha}$. Therefore, we take $\alpha \geq 0.5$. Clearly, for $\alpha = 0.5$ we have $Var\, \widehat{\sigma}_X^2(t) \sim t^{-1}(C_1 + C_2)$ and $Var\, \widehat{\sigma}_X^2(t) \sim t^{-1} C_1$ for $\alpha > 0.5$.

Acknowledgement This paper is partially supported by grant of Russian Foundation For Basic Research 13-01-00653.

References

[1] A. Ganesh, N. O'Connell, D. Wischik, Big queues, Springer, 2004.

[2] W. Smith, Regenerative stochastic processes, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 232 (1188) (1955) 6–31.

[3] L. Afanasyeva, E. Bashtova, Coupling method for asymptotic analysis of queues with regenerative input and unreliable server, Queueing Systems 76 (2) (2014) 125–147.

[4] A. Borovkov, Asymptotic methods in theory of queues, Nauka, Moscow, 1980.

[5] W. Feller, An introduction to probability theory and its applications, 2nd Edition, Wiley, New York, NY, USA, 1957.

[6] A. A. Borovkov, Stochastic processes in queueing theory, Vol. 4, Springer Verlag, 1976.

[7] L. G. Afanaseva, Queuing systems with cyclic control processes, Cybernetics and Systems Analysis 41 (1) (2005) 43–55. doi:10.1007/s10559-005-0040-9.

[8] L. Afanasyeva, A. Tkachenko, Multichannel queueing systems with regenerative input flow, Theory of Probability & Its Applications 58 (2) (2014) 174–192.

[9] W. Whitt, Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues, Springer, 2002.

[10] L. Afanasyeva, E. Bashtova, E. Bulinskaya, Limit theorems for semi-markov queues and their applications, Communications in Statistics-Simulation and Computation 41 (6) (2012) 688–709.

[11] D. R. Cox, Renewal theory, Vol. 4, Methuen London, 1962.

[12] W. L. Smith, Renewal theory and its ramifications, Journal of the Royal Statistical Society. Series B (Methodological) (1958) 243–302.