Queueing Networks Simulation: Artificial Regeneration and Heavy Tail Phenomena

Irina V. Aminova

Department of Applied Mathematics and Cybernetics,
University of Petrozavodsk

Lenin St., 33, Petrozavodsk, Republic of Karelia, 185640, Russia

E-mail: iaminova@mainpgu.karelia.ru

Abstract

Many queueing and network stochastic models do not guarantee classical assumptions for the existence of regeneration points. In particular, the infinite second moment of regenerative cycle's length (it means that they have so-called *heavy-tailed* distribution) induce long-range dependence, which does not except the existence of regenerative points, but such events are very rare. At that we discuss distributions with heavy tails in detail. To increase the frequency of regeneration points one can use the so-called *artificial classical regeneration* method. In this paper, we introduce artificial regeneration based on exponential splitting.

1 Introduction

The regenerative approach is one of the basic and efficient methods to simulation queueing stochastic processes. it is appliance is due to that property that many stochastic systems can "probabilistically" restart at some random times called regeneration points. It means that the paths of the regeneration process can be separated into independent groups of data belonging to one regeneration cycle (the distance between any pair of successive regeneration points).

First of all we give basic definitions.

Let (Z, β) be a random pair, where $\beta = (\beta_1, \beta_2, \dots)$ is an increasing sequence of nonnegative finite random times, $Z = \{Z_t, t \geq 0\}$— random process with state space $(\mathfrak{Z}, \mathfrak{B})$ and the shift operator Θ is defined as $\Theta_{\beta_k}(Z, \beta) = \{(Z_{\beta_i+t})_{t\geq 0}, (\beta_k - \beta_i)_{k\geq i}\}$.

A random pair (Z, β) is called *a (classic) regenerative process*, if for any $k \geq 0$:

(1) Distribution of the process $\Theta_{\beta_k}(Z, \beta)$ does not depend on k;

(2) Distribution of $\Theta_{\beta_k}(Z, \beta)$ does not depend on the “pre-history” $\{(Z_t)_{t<\beta_k}, \beta_0, \dots, \beta_k\}$.

The sequence β is called *the renewal process embedded into the original process* $\{Z, \beta\}$. Variables β_k are called *(strong) regeneration points (r.p.s)*, G_k=$\{Z_t, 0 < t < \beta_1\}$ are independent and identically distributed (i.i.d.) parts of the trajectory of the process Z between r.p.s called *regeneration cycles (r.c.s)*, $\alpha_k = \beta_{k+1} - \beta_k, k \geq 1$ are i.i.d. lengths of r.c.s.

Under some conditions, the regenerative structure guarantees the existence of a limiting distribution. Usually, interarrival and service times are assumed to be i.i.d. with finite first moments.

Classical approaches (for example, central limit theorem (CLT) in the case of classic regeneration or an extended CLT for the so-called m-dependent random variables) allow us to obtain statistical estimates for various characteristics of the limiting distribution: waiting time, sojourn time, queue-size, and so on, [2].

Now let us consider a queueing system with an infinite second moment of r.c.s lengths, $\mathbf{E}\,\alpha^2 = \infty$ (α is a generic variable for cycle lengths). In this case we say that the random variable α has *heavy-tailed distribution.* Since $\mathbf{E}\,\alpha^2 < \infty$ regeneration points generally exist, but such events turn out to be too rare. One of the ways to increase the efficiency of the regeneration approach is to increase the frequency of r.p.s. To do this one can use, for instance, the so-called *artificial classical regeneration method.*

In this paper, we present artificial regeneration based on exponential splitting. The main idea of such splitting is to plot the exponential function into a given density that is possible in the case of heavy-tailed distribution. One can expect that the heavier the tail of the given distribution the more frequent regeneration points we can obtain. At that we discuss in detail heavy-tailed distributions.

The paper is organized as follows. In section 2, we give mathematical definitions of long-range dependent and self-similar processes. In section 3, we consider distributions with heavy tails. Section 4 describes an artificial regeneration method in the case of exponential splitting. In section 5, we conclude with a discussion of statistical methods of *heavy-tailed index* estimation.

2 Self-similarity

The network arrival process is often considered in accordance with Poisson processes for analytical simplicity. But increasing instrumentation of local- and wide-area networks (LAN and WAN) has made possible the observation of large amounts of data [13, 14, 16, 17]. Analysis of this data shows that classical queueing assumptions of *light tails* (see section 3) and independence are inapplicable for them [19]. Huge data sets in LAN traffic appear to exhibit long-range dependence (i.e., the process is *self-similar* or has *fractal* behavior with scaling parameter greater then $1/2$, while a number of file size appear to be generated by heavy-tailed distributions. It means that traffic shows noticeable bursts at all timescales.

We note that self-similarity is a significant factor in network modeling. For example, delays in data transferring can be more severe when traffic is self-similar than would be predicted by Poisson models [16, 17]

The theory of self-similar stochastic processes is not so well studied as the theory of Poisson processes. Empirical observation of fractal network behavior makes it important to develop tools for understanding self-similarity [11, 15].

Let us now dwell on the definitions of self-similarity and long-range dependence.

Definition 1. ([15]) A discrete time process $Z = \{Z_n : n = 0, 1, \dots\}$ is *second order (covariance) stationary*, if it has

(1) constant mean $\mu = \mathsf{E}\, Z_n$ for all $n \geq 0$;

(2) finite variance $v = \mathsf{E}(Z_n - \mu)^2$.

Let $\gamma_k = \text{cov}(Z_n, Z_{n+k})$, $k = 0, 1, \ldots$ be an auto-covariance function, which depends only on k and does not depend on n. Denote $\rho_k = \frac{\gamma_k}{\gamma_0}$— auto-correlation function and note that $v = \gamma_0$.

Let us construct a new second order stationary process $Z^{(m)} = \{Z_j^{(m)}, j = 1, 2, \ldots\}$, obtained by averaging the given process Z over blocks of size m—the so-called *aggregated process*. That is, for all $m \geq 1$,

$$Z_j^{(m)} = \frac{Z_{jm-m+1} + Z_{jm-m+2} + \ldots Z_{jm}}{m} , \quad j \geq 1 .$$

Let $v^{(m)} = \textbf{Var}\, Z_k^{(m)}$ be a variance of aggregated process, which is independent of subscript k. Denote $\gamma_k^{(m)}$, $\rho_k^{(m)}$ auto-covariance and auto-correlation functions for aggregated process $Z^{(m)}$, respectively.

It is easy to check that

$$v^{(m)} = \frac{v}{m} + \frac{2}{m^2} \sum_{k=1}^{m} (m-k)\gamma_k \tag{2.1}$$

or

$$v^{(m)} = \frac{v}{m} + \frac{2}{m^2} \sum_{s=1}^{m-1} \sum_{k=1}^{s} \gamma_k . \tag{2.2}$$

Definition 2. The process Z is called *short-range dependent* if

$$\sum_k \gamma_k < \infty . \tag{2.3}$$

It follows from (2.2) that for short-range dependent process

$$v^{(m)} \sim \frac{c}{m} , \quad \text{as} \quad m \to \infty , \tag{2.4}$$

where c is positive finite constant. (For example, for the so-called k-dependent variables, $c = \textbf{Var}\, Z_1$. So, in this case, $v^{(m)} \to 0$ as $m \to \infty$.)

For instance, if a process Z has exponentially decaying autocovariation function, i.e. $\gamma_k \sim ca^k$ for large k, where $0 < c < \infty$ and $a \in (0, 1)$ are constants, then process is short-range dependent.

Remark. If relation (2.4) holds for all $k = 1, 2, 3, \ldots$, then $\gamma_k^{(m)} \to 0$ as $m \to \infty$.

Definition 3. The process Z is called *long-range dependent (LRD)* if

$$\sum_k \gamma_k = \infty \ . \tag{2.5}$$

Since for each k

$$\gamma_k \sim (1 - \frac{k}{m})\gamma_k \ \text{ as } \ m \to \infty \ ,$$

then divergence in (2.5) implies that series

$$\frac{1}{m^2} \sum\nolimits_k (m-k)\gamma_k$$

is also diverges. So,

$$mv^{(m)} \to \infty \ , \quad \text{as} \ \ m \to \infty \ .$$

For instance, if a process Z has hyperbolically decaying autocovariance function $\gamma_k \sim ck^{-\beta}$, and variance $v^{(m)} = cm^{-\beta}$, $0 < \beta < 1$, $0 < c < \infty$, then it is LRD.

One can show [15] that for LRD processes

$$\rho_k^{(m)} \to \rho_k \ , \quad \text{as} \quad m \to \infty \ , \tag{2.6}$$

that means, that for large m, ρ_k does not depend on m. In this case the process is called *asymptotically second order self-similar*.

Definition 4. ([15, 13]) The process Z is called *exactly second order self-similar* if

$$\rho_k^{(m)} = \rho_k \quad \forall m, \ k \geq 0 \ , \tag{2.7}$$

$$v^{(m)} = vm^{-\beta} \ , \quad m > 0 \ , \quad 0 < \beta < 2 \ . \tag{2.8}$$

In other words, the aggregated process $Z^{(m)}$ and given process Z have the same correlation structure.

From definition 4 it is not obvious at first glance that self-similar processes actually exist, but in fact a number of families of self-similar processes are known. The most widely-studied self-similar processes are fractional Gaussian noise (FGN) and fractional ARIMA processes [13, 20]. Associated with FGN is Fractional Browian motion (FBM), which is simply the integrated version of FGN (that is, a FBM process is simply the sum of FGN increments).

Remarks.

1. Usually instead of parameter β one can use index $H = 1 - \beta/2$ which is called *the Hurst index*, $0 < H < 1$. For example, BM is $1/2$ self-similar.
2. (2) If $1/2 < H < 1$, then the process is LRD, otherwise the process is short-range dependent or has independent increments.

Recent measurements [16, 17, 21] of LAN and WAN traffic characteristics show that long-range dependence and heavy tails are tightly connected. For example, Samorodnitsky and Resnick [19] suggest that long-range dependent inputs to queueing system can induce heavy tailed outputs. In [19] Resnick notes that heavy tails can induce long-range dependence. Asmussen [1, 8], Boxma [6] show that heavy-tailed service times in $M/M/1$ and in $M/G/1$ (FIFO discipline of service and Processor Sharing service) induce heavy-tailed outputs and therefore long-range dependence.

3 Distributions with heavy tails

This section is important for understanding the splitting procedure (see section 4) which is based on the existence of heavy tails for density functions. Now we start a discussion on main definitions and the properties of heavy-tailed distributions.

Let F be the distribution function (d.f.) of nonnegative r.v. X.

Definition 5. ([5]) D.f. F has *heavy tail* if

$$m(\lambda) = \mathbf{E}\{\exp(\lambda X)\} = \infty \quad \forall\, \lambda > 0\ . \tag{3.1}$$

Otherwise, d.f. F is called *light-tailed.*

Heavy-tailed distributions have the following property:

$$\lim_{x\to\infty} e^{\lambda x}(1 - F(x)) = \infty\ , \quad \forall\, \lambda > 0\ . \tag{3.2}$$

Let us consider examples of heavy-tailed distributions.

Example 1. Among distribution functions with heavy tails subclass S of *subexponential distributions* is quite important; let X_1, X_2, ... be nonnegative, i.i.d. r.v.s with common d.f. F.

Definition 6. ([1]) F is called *subexponential distributed* ($F \in S$) if the following relation holds

$$\lim_{x\to\infty} \frac{P(X_1 + \cdots + X_n > x)}{P(X_1 > x)} = \frac{1 - F^{(n)}(x)}{1 - F(x)} = n \quad \forall\, n\,, \tag{3.3}$$

where $F^{(n)}(x)$ is an n-multiple convolution of d.f. F.

We note, that for any distribution F,

$$P(\max(X_1, \ldots, X_n) > x) \sim nP(X_1 > x)$$

(here $\sim$ means that the ratio converges to one as $x \to \infty$). So, this is the only way that superposition of subexponential d.f.s is large when one of the summands is so.

For a distribution F on $(-\infty, +\infty)$, $F \in S$, if $1 - F(x) \sim 1 - G(x)$, $x \to +\infty$ for some $G \in S$ which is concentrated on $(0, \infty)$.

The following properties are often used [5]:

Proposition 2. *Let $F \in S$. Then*

(1) Property (3.2) holds;

(2) $\forall \epsilon > 0$ there is a constant $\alpha = \alpha(\epsilon) > 0$ such that

$$1 - F^{(n)}(x) \le \alpha(1+\epsilon)^n (1 - F(x)) \quad \forall\, x \ge 0\,, \quad n \ge 2\,;$$

(3) $\displaystyle\lim_{x\to\infty} \frac{1 - F(x+u)}{1 - F(x)} = 1$ uniformly in u on compact sets.

Example 2. *Regularly varying d.f.'s* (which were defined by Karamata, see [19]).

Definition 7. ([3]) A positive function L (not necessarily monotone) is called *slowly varying* if

$$\frac{L(tx)}{L(t)} \to 1\,, \quad t \to \infty\,, \quad \forall\, x > 0\,. \tag{3.4}$$

Example 3. Each almost everywhere positive measurable function is slowly varying, if it has a positive limit as $x \to \infty$. Typical slowly varying functions are [19]

$$\text{const} + o(1)\ , \quad o(1) \to 0\ , \quad \text{as} \quad x \to \infty\ ;$$

$$\log(x)\ , \quad x > 1\ ;$$

$$\log(\log(x))\ , \quad \text{for large } x\ ;$$

$$\frac{1}{log(x)}\ , \quad x > 1\ ,$$

On the other hand, exponents e^x and e^{-x} are not slowly varying. These examples intuitively give the feeling of this concept. Before giving the definition of regularly varying functions, let us consider properties of slowly varying functions [18, 19].

Proposition 3. ([19]) *Let $L(x)$, $L_1(x)$, $L_2(x)$ be slowly varying functions. Then*

(1) for all $\nu > 0$, $x \to \infty$

$$x^\nu L(x) \to \infty\ , \quad x^{-\nu} L(x) \to 0\ ;$$

(2) $L_1(x) \cdot L_2(x)$ and $L_1(x) + L_2(x)$ are slowly varying;

(3) for all fixed $\epsilon > 0$ and large enough x

$$x^{-\epsilon} < L(x) < x^\epsilon\ .$$

Definition 8. ([3]) A positive function u is called *regularly varying* with rate ν $((-\infty < \nu < \infty))$ if

$$u(tx) \to x^{-\nu} u(t)\ , \quad t \to \infty\ , \quad \forall\, x\ . \tag{3.5}$$

Proposition 4. ([3]) *Let $1 - F_1(x)$, $1 - F_2(x)$ be regularly varying tails of d.f.'s. $1 - F_i(x) = x^{-\nu} L_i(x)$. Then the convolution $F^{(2)} = F_1 * F_2$ has regularly varying tail*

$$1 - F^{(2)}(x) \sim x^{-\nu}(L_1(x) + L_2(x))\ .$$

This proposition is useful because of the following corollary.

Corollary 4.1. *If tail of d.f. F is regularly varying, i.e.* $1 - F(x) \sim x^{-\nu}L(x)$*, then*

$$1 - F^{(r)}(x) \sim r \cdot x^{-\nu}L(x) \ . \tag{3.6}$$

Thus, if d.f. F is regularly varying then F is subexponential.

Definition 9. ([19]) A positive function u is called *second order regularly* with first order parameter $-\nu$ and second order parameter ρ, if there is a function $A(t) \to 0$, $t \to \infty$ which ultimately has a constant sign such that the following refinement of (3.5) holds:

$$\lim_{t\to\infty} \frac{\frac{u(tx)}{u(t)} - x^{-\nu}}{A(t)} = H(x) = cx^{-\nu}\int_1^x v^{\rho-1}dv \ , \quad x > 0 \ , \tag{3.7}$$

for $c \neq 0$.

Note that for $x > 0$

$$H(x) = \begin{cases} cx^{-\nu}\log x \ , & \text{if } \rho = 0 \ , \\ cx^{-\nu}\dfrac{x^{\rho} - 1}{\rho} \ , & \text{if } \rho < 0 \ . \end{cases}$$

To integrate regularly varying functions one can use Karamata's theorem [19]. This theorem roughly says that when one integrates a regularly varying function, one may treat the slowly varying function as a constant.

Example 4. *Pareto, Lognormal and Weibull d.f.s* belong to subexponential distributions.

Definition 10. D.f. F with tail $1 - F(x) = \left(\frac{x_0}{x}\right)^{\alpha}$, $\alpha > 0$, $x_0 > 0$, $x \geq x_0$ is called *Pareto d.f.*

Random variable $e^{a+\sigma\eta}$ has *lognormal distribution* if random variable η is standard normal.

D.f. F with tail $1 - F(x) = e^{-x^{\beta}}$, $0 < \beta < 1$ is called *Weibull.*

4 Exponential splitting

Let T be a positive random variable (r.v.) with density function $f(x)$.

Definition 11. ([4]) Random variable T is called *exponentially splitted* if there are constants $\lambda > 0$, $\tau > 0$ and $0 < q < 1$ such that

$$f(t) \geq q\lambda e^{-\lambda(t-\tau)} , \quad \forall\, t \geq \tau .$$

The exponentially splitted random variable can be represented in the following way:

$$T = \begin{cases} T_0 , & \text{with probability } q , \\ T_1 , & \text{with probability } 1-q , \end{cases}$$

where T_1 has the density

$$f_1(t) = \begin{cases} \dfrac{f(t)}{1-q} , & \text{for } t \leq \tau , \\ \dfrac{f(t) - q\lambda e^{-\lambda(t-\tau)}}{1-q} , & \text{for } t > \tau , \end{cases}$$

and T_0 is a left hand truncated exponentially splitted random variable with parameter λ. The density of r.v. T_0 is defined as follows

$$f_0(t) = \begin{cases} 0 , & \text{for } t \leq \tau , \\ \lambda e^{-\lambda(t-\tau)} , & \text{for } t > \tau . \end{cases}$$

So, density f is splitted into two density functions f_1 and f_0, where f_0 is the shifted exponent:

$$f(t) = qf_0(t) + (1-q)f_1(t) , \quad t \geq 0 .$$

Remarks.

1. Distributions with heavy tails always satisfy the above mentioned conditions of exponential splitting. So, any stochastic queueing system with heavy-tailed interarrivals and service times is obliged to apply artificial regeneration.
2. Note that the heavier the tail of d.f., the bigger the area we can plot into graphics of d.f. f and therefore the more frequent the artificial regeneration points we can obtain.

5 How to estimate index ν of heavy-tailed function?

The material of this section is based on works [19, 10]. There are many different techniques to estimate index ν of d.f.'s tails in (3.5). We dwell on the Hill estimator, which is widely used.

Suppose $X_1, \dots, X_n$ be i.i.d. r.v.'s with d.f. F. Let

$$X_{(1)} > X_{(2)} > \cdots > X_{(n)}$$

be order statistics. Assume that F has h.t.:

$$\bar{F}(x) = x^{-\nu} L(x) \;, \quad x \to \infty \;.$$

Definition 12. ([10]) Function

$$H_{k,n} = \frac{1}{k} \sum_{i=1}^{k} \log \frac{X_{(i)}}{X_{(k+1)}} \;, \quad k < n \;, \tag{5.1}$$

is called the *Hill estimator* (here k is the number of upper statistics used in the estimation).

We use only k upper statistics, because it is better to sample from that part the distribution what looks most Pareto-like. A more detailed explanation is that sample

$$\frac{X_{(1)}}{X_{(k+1)}} \;, \quad \dots \;, \quad \frac{X_{(k)}}{X_{(k+1)}}$$

is distributed like the order statistics from sample size k and has distribution of the tail

$$\frac{1 - F(xX_{(k+1)})}{1 - F(X_{(k+1)})} \;, \quad x \geq 1 \;.$$

Because of (3.5) if $X_{(k+1)}$ is large, then

$$\frac{1 - F(xX_{(k+1)})}{1 - F(X_{(k+1)})} \sim x^{-\nu} \;.$$

Proposition 5. ([19]) *If* $n \to \infty$*,* $k \to \infty$*, but* $\frac{k}{n} \to 0$ *we have*

$$H_{k,n} \to \nu^{-1} \ . \tag{5.2}$$

In practice the Hill estimator is used in the following way: we graph $\left\{k\,,\ H_{k,n}^{-1}\,,\ 1 \le k \le n\right\}$.

6 Conclusion

Now we make some important conclusion comments.

When classic regeneration is ineffective (for instance, in the case of rare events, that is when $\mathsf{E}\,\alpha_1^2 = \infty$), one can apply artificial regeneration based on exponential splitting. The existence of heavy-tailed interarrivals and service times in queues allow us to apply artificial regeneration. Moreover, the heavier the tails, the more frequent the regeneration points we can obtain. This property sometimes may be very useful to increase the efficiency of regenerative simulation. In this connection, we note that it seems to be very promising to study the conservation property of heavy tails for output in network context when interarrivals and service times have heavy-tailed distributions and to develop the method of construction r.p.s based on artificial regeneration for queueing networks.

Acknowledgments

I want to thank my adviser professor Evsei Morozov for essential help in setting the problem and preparation of the paper. I would also like to thank professor Vladimir Kalashnikov and professor Alexander Andronov for helpful discussions.

References

[1] S. Asmussen, K. Binswanger, B. Hojgaard, *Rare events simulation for heavy-tailed distributions.* 1998. (to be appeared)

[2] E. V. Morozov, S. G. Sigovtsev, *Simulation of Queueing processes based on weak regeneration.* J. of Math. Sciences, vol. 89, no 5, 1998.

[3] W. Feller, *An introduction to probability theory and its applications.* Vol. 2, 1984.

[4] A. Andronov, *Artificial regeneration points for stochastic simulation of complex system.* Riga Aviation Univ., 1998.

[5] V. Kalashnikov, *Geometric sums.*

[6] A. P. Zwart, O. J. Boxma, *Sojourn time asymptotic in the $M/G/1$ processor sharing queue.* CWI, 1998.

[7] C. Kluppelberg, *Subexponential distributions and integrated tails.* J. Appl. Prob., vol. 25, 1998.

[8] S. Asmussen, J. Møller, *Tail asymptotic for $M/G/1$ type queueing processes with subexponential increments.* Lund University, 1998.

[9] S. Asmussen, J. Teugels, *Convergence rates for $M/G/1$ queues and ruin problems with heavy tails.* Aalborg Univ., 1995.

[10] H. Drees, L. Haan, S. Resnick, *How to make a Hill plot.* AMS, 1991.

[11] B. K. Ryu, *Fractal network traffic: from understanding to implications.* Columbia Univ., 1996.

[12] Y. Chen, Z. Deng, C. Williamson, *A model for self-similar Ethernet LAN traffic: design, implementations, and performance implications.* Univer. of Saskatchewan.

[13] V. Paxson, *Fast, Approximate synthesis of fractional Gaussian noise for generating self-similar network traffic.* Comp. Commun. Review, vol. 27, 1997.

[14] V. Paxson, S. Floyd, *Wide-area traffic: the failure of Poisson modeling.* IEEE ACM Transactions on Networking, vol. 3, no 3, pp. 226-244, 1995.

[15] B. K. Ryu, *Fractal network traffic modeling: past, present, and future.* 1997.

[16] W. Willinger, V. Paxson, M.Taqqu, *Self-similarity and heavy tails: structural modeling of networks traffic.* In “A practical Guide to heavy tails: statistical techniques and applications”, Boston, pp. 27-53., 1998.

[17] M. Crovella, M. Taqqu, A. Bestavros, *Heavy-tailed probability distributions in the World Wide Web.* In “A practical Guide to heavy tails: statistical techniques and applications”, Boston, pp. 3-25, 1998.

[18] E. Seneta, *Regularly varying functions.* Springer-Verlag, Berlin-Heidelberg-New York, 1976.

[19] S. I. Resnick, *Heavy tail modeling and teletraffic data.* Cornell University, 1995.

[20] I. Norros, P. Mannersalo, S. L. Wang, *Simulation of Fractional Brownian Motion with midpoint displacement.* A journal of Teletraffic theory and performance analysis of communication systems and networks, vol. 2, no 1, 1998.

[21] I. Kaj, *Stochastic modeling broadband communications systems.* Uppsala, 1999.