Stochastic networks in the presence of heavy tails: a general introduction

Karl Sigman, Department of Industrial Engineering and Operations Research, Columbia University, New York, USA

Stochastic networks are *queueing* models used for various kinds of complex systems in which randomness and congestion play leading roles. Simple examples include banks, post offices and supermarkets but also include large telecommunication systems involving telephones, Internet/Web, and manufacturing processes and computer networks[1].

Customers arriving to such systems compete for service forming lines (queues) that in turn cause delays, and system performance can be measured from two perspectives: customers ("What is my delay likely to be?"), and the system ("How much work remains to complete?").

Randomness enters in two main ways: through the arrival streams of customers, and the sizes of the service requests.

Here we give an overview of the distinction between two cases of randomness and their effects on system performance: when component probability distributions such as service times S are *light-tailed* and *heavy-tailed*. Light-tailed refers to the case when the tail $P(S > x),\ x \geq 0$ decays approximately exponential, like $ce^{-\mu x}$, yielding finite moment generating functions, $E(e^{uS}) < \infty$, for example. Heavy-tailed is the case when the tail decays slower than any exponential yielding infinite moment generating function for example. In the early days of queueing, the light-tailed case was assumed, but now days data gathered from the Internet (and other areas) supports the existence of heavy-tailed phenomena; the heavy-tailed case has become increasingly important. In the context of Internet and World Wide Web services, risks can dramatically increase without much warning, and service level agreements (SLA) can incur heavy penalties because of the presence of heavy tails.

References

[1] B. von Bar (1975). Asymptotic ruin probabilities when exponential moments do not exist. *Scand. Act. J.* 6-10. A theorem on sums of independent, positive random variables and its applications to branching processes. *Theory Probab. Appl.* **9**, 640–648.

[2] V.P. Chistyakov (1964). A theorem on sums of independent, positive random variables and its applications to branching processes. *Theory Probab. Appl.* **9**, 640–648.

[3] J.W. Cohen (1973). Some results on regular variation for distributions in queueing and fluctuation theory. *QUESTA*, **33**, 205–232.

[4] P. Embrechts & N. Veraverbeke (1982). Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance: Mathematics and Economics* **1**, 55–72.

[5] A. Pakes (1975). On the tails of waiting time distributions. *J. Appl. Probab.* **12**, 555–564.

[6] K. Sigman (1999). (Guest Editor). Special Volume on Queues with Heavy-tailed Distributions, *QUESTA*, **33**.

[7] K. Sigman (1999). A primer on heavy-tailed distributions, *QUESTA*, **33**, 261–275.

[8] W.L. Smith (1972). On the tails of queueing time distributions. (Technical Report, MIMEO Series No. 830, Dept. Statistics, U. North Carolina, Chapel Hill.)

[1] Queueing theory – the mathematical theory of congestion – arose from the study of simple such examples; the early development of the field was motivated by telephone call systems and ways in which they could be made to operate more efficiently.