Advanced Stochastic Processes.

David Gamarnik

LECTURE 24

G/G/1 in heavy-traffic. Introduction to queueing networks

Lecture outline

- G/G/1 in heavy-traffic (continued)
- Introduction to queueing networks. Generalized Jackson Network

24.1. G/G/1 queueing system in heavy-traffic. Generalized Jackson Network

We begin by proving Theorem 23.3.

Proof of Theorem 23.3. We have almost completed the proof when we did the analysis for the case $\rho < 1$ is a constant. Now we have $\rho_n = 1 - \frac{\theta}{\sqrt{n}}$ as well as interarrival times are rescaled by the same factor ρ_n^{-1}. We recall $Z^n(t) = V(Q^n(0) + A^n(t)) - B^n(t) = X^n(t) + I^n(t)$, where as usual $X^n(t) = V(Q^n(0)+A^n(t))-t$ and $I^n(t) = t-B^n(t)$ (recall that service times are uneffected by n). As before we consider time nt instead of t and write

$$
\begin{aligned}
\frac{X^n(nt)}{\sqrt{n}} &= \frac{V(Q^n(0) + A^n(nt)) - \mu^{-1}(Q^n(0) + A^n(nt))}{\sqrt{n}} \\
&\quad + \mu^{-1}\frac{Q^n(0) + A^n(nt) - \lambda^n nt}{\sqrt{n}} \\
&\quad + \frac{\rho^n nt - nt}{\sqrt{n}}
\end{aligned}
$$

Now we have $A^n(nt) = \rho^n A(nt)$, implying by FSLLN that $A^n(nt)/n \to \lambda t$, since $\lim_n \rho_n = 1$. Also $Q^n(0)/n \to 0$. Therefore

$$\begin{aligned}
&\frac{V(Q(0) + A^n(nt)) - \mu^{-1}(Q^n(0) + A^n(nt))}{\sqrt{n}} \\
&= \frac{V(Q^n(0) + A^n(nt)) - \mu^{-1}(Q^n(0) + A^n(nt))}{\sqrt{Q^n(0) + A^n(nt)}}\sqrt{\frac{Q^n(0) + A^n(nt)}{n}} \\
&\Rightarrow \lambda^{\frac{1}{2}}\mu^{-1}c_s W_1(t) \\
&= \lambda^{-\frac{1}{2}}c_s W_1(t)
\end{aligned}$$

where W_1 is a standard Brownian motion, where we again use Donsker's Theorem for V. Using FCLT for counting process $A(t)$ we obtain

$$\begin{aligned}
\mu^{-1}\frac{Q^n(0) + A^n(nt) - \lambda^n nt}{\sqrt{n}} &= \mu^{-1}\frac{Q^n(0) + \rho^n(A(nt) - \lambda nt)}{\sqrt{n}} \\
&\Rightarrow q\mu^{-1} + \mu^{-1}\lambda^{\frac{1}{2}}c_a W_2(t) \\
&= q\mu^{-1} + \lambda^{-\frac{1}{2}}c_a W_2(t)
\end{aligned}$$

where W_2 is a standard Brownian motion independent from W_1.

Finally

$$\frac{\rho^n nt - nt}{\sqrt{n}} = -\theta t.$$

Putting all this together, we conclude that $X^n(nt)/\sqrt{n}$ converges weakly to a Brownian motion W with $W(0) = q/\mu$, drift $-\theta$ and variance $\lambda^{-1}(c_a^2 + c_s^2)$. Also $X^n(nt), I^n(nt), Z^n(nt)$ solve jointly the Skorohod mapping problem. Moreover, recall that the mappings $\Psi(X^n) = I^n, \Phi(X^n) = Z^n$ are (Lipshitz) continuous mappings on $D[0,T]$. Using Mapping Theorem (Theorem 20.1 from Lecture 20), we conclude that $Z^n \Rightarrow RBM$ which starts from q/μ, has drift $-\theta$ and variance $\lambda^{-1}(c_a^2 + c_s^2)$, and $I^n(t) \Rightarrow \sup_{0\le s\le t}(-W(s))^+$, where W is a Brownian motion with the same drift, variance and starting value. This concludes the derivation for the workload process Z. The derivation for the queue length process is similar and involves "inverting" the process V. We skip these technical, but not insightful details. □

24.2. Generalized Jackson Network. Model description

A G/G/1 queueing system we considered thus far is limited in its modeling power. We now extend this model to a network case – Generalized Jackson Network (GJN). This is a queueing network consisting of J single servers. Each server $j = 1, 2, \ldots, J$ process jobs with i.i.d. service time $v_n^j, n \ge 1$, with service rate $\mu_j = 1/\mathbb{E}[v_1^j]$. There are J external arrival processes, each having i.i.d. interarrival times $u_n^j, n \ge 1$. Upon arrivals the jobs enter buffers corresponding to server j and form a queue, but upon service completion the jobs do not necessarily depart from the network but join a another queue corresponding to a different server k. The mechanism for this transitions is as follows. We fix a $J \times J$ matrix P with non-negative entries, such that each row sum is at most unity $\sum_k p_{jk} \le 1$. To each server $j = 1, 2, \ldots, J$ we associate an i.i.d. sequence of vectors $R_n^j, n \ge 1$, where each R_n^j takes value $e_k, 1 \le k \le J$ with probability p_{jk} and value 0 with probability $1 - \sum_k p_{jk}$. Here e_k is the k-th unit vector. The event $R_n^j = e_k$ corresponds to sending the n-job completing service in server j to the queue of server k. The event $R_n^j = 0$ corresponds to this job exiting the network. For each pair j, k we also let $R_j^k(n)$ denote the counting process – the number of times that $R_n^k = e_j$ out of the first n trials. That is the number of jobs which were sent to server j after service completion in k, out of the first n jobs served there. The jobs joining the queue corresponding to each server j are merged together, whether they arrive externally or internally, and they are served using the First-In-First-Out rule.

We use the same notations as for G/G/1 queueing system, but now the processes are indexed by j: $A_j, S_j, D_j, B_j, I_j, Z_j, Q_j$. The balance equation for the queue length are written as

$$\begin{aligned} Q_j(t) &= Q_j(0) + A_j(t) + \sum_{1 \le k \le J} R_j^k(S_k(B_k(t))) - S_j(B_j(t)), \\ B_j(t) &= \int_0^t 1\{Q_j(s) > 0\} ds \\ Z_j(t) &= V_j(Q_j(0) + A_j(t) + \sum_{1 \le k \le J} R_j^k(S_k(B_k(t)))) - B_j(t) \\ &= V_j(Q_j(t) + S_j(B_j(t))) - B_j(t). \end{aligned}$$

24.3. Additional reading materials

- Chapter 6 of Chen & Yao book [**1**] from the course packet.

BIBLIOGRAPHY

1. H. Chen and D. Yao, *Fundamentals of queueing networks: Performance, asymptotics and optimization*, Springer-Verlag, 2001.