Skorokhod Mapping Theorem. Reflected Brownian Motion

Content.

1. G/G/1 queueing system
2. One dimensional reflection mapping
3. Reflected Brownian Motion (RBM)

1 Technical preliminaries

We use the following technical facts. If x is a Lipschitz continuous, then it also has derivative everywhere except measure zero set of points on $[0, \infty)$ and $x(t) = \int_0^t \dot{x}(u)du$. Note, however that if x is a realization of a Brownian motion, then it is nowhere differentiable and we cannot write this integral representation. Also if x is a continuous and non-decreasing, then it has derivatives almost everywhere and again $x(t) = \int_0^t \dot{x}(u)du$.

2 G/G/1 queueing system

We consider the following G/G/1 queueing system consisting of a single server which works at a unit speed. The G/G/1 notation stands for general arrival/general service times/single server type queueing system (as opposed to for example M/M/1 which stands for Poisson arrival/Exponential service time/single server type systems). Initially we assume there are $Q(0)$ jobs in the system. There is an arrival process into the system. The times between successive arrivals is an i.i.d. sequence $u_n \geq 0, n \geq 1$ of random variables. We let $\lambda = 1/\mathbb{E}[u_1]$ denote the arrival rate. Namely, $1/\lambda$ is the mean interarrival time. We let $U(n) = \sum_{1 \leq i \leq n} u_i$ and let

$$A(t) = \sup\{n : U(n) \leq t\}$$

denote the counting process describing the arrival process.

The time it takes to process the i job arriving to the server (including the first $Q(0)$ jobs) is also assumed to be an i.i.d. sequence $v_n \geq 0, n \geq 1$ of random variables with service rate $\mu = 1/\mathbb{E}[v_1]$. Also $V(n) = \sum_{1 \leq i \leq n} v_i$ and let

$$S(t) = \sup\{n : U(n) \leq t\}$$

denote the counting process describing the service process.

Denote by $D(t)$ the cumulative departure process from the queueing system – the number of jobs that departed up to time t. If the server was working all the time, then we would have $D(t) = S(t)$, but occasionally, the server runs out of jobs. So let $B(t)$ denote the total cumulative time that the server was working on jobs during $[0, t]$. Naturally $B \geq 0$, B is non-decreasing and $B(t) \leq t$. Then $D(t) = S(B(t))$. Note that we can also write

$$B(t) = \int_0^t 1\{Q(s) > 0\} ds.$$

The jobs which have not been processed yet form a queue. Denote by $Q(t)$ the length of the queue at time t. Then we naturally obtain

$$Q(t) = Q(0) + A(t) - D(t) = Q(0) + A(t) - S(B(t)). \tag{1}$$

In addition we consider *workload process* $Z(t)$ – this represents the amount of time it takes to process all the jobs present in the system at time t (queue plus possibly a job being served), assuming no further jobs arrive after time t. Loosely speaking this is the time it takes to clear the system after we ”shut the door”.

Note that

$$Z(t) = V(Q(0) + A(t)) - B(t). \tag{2}$$

(convince yourself that this is indeed the case).

Finally, it is also useful consider the cumulative idling process

$$I(t) = t - B(t) = \int_0^t 1\{Q(s) = 0\} ds. \tag{3}$$

Recall our earlier notations $C[0, \infty)$ is the space of continuous functions on $[0, \infty)$ and $D[0, \infty)$ is the space of right-continuous function with left limits on the same domain.

Problem 1. *Show that* $Q, Z \in D[0, \infty); Y, Z \in C[0, \infty)$

The main question we ask for G/G/1 system is to characterize the behavior of the queue length $Q(t)$ and workload $Z(t)$ as a process or in steady state ($t = \infty$), when steady-state exists, assuming that we know the distribution of the interarrival and service times. Ideally, we would like to compute the distribution of $Q(t), Z(t)$. But this is, with some exceptions, infeasible. Thus the focus will be on obtaining approximations. As we will see we can get good approximations when the system is in so called *heavy traffic* regime. This is the regime when $\lambda \approx \mu$. Then the typically observed queue length will be very large and, as we will see, can be approximated by a certain *reflected Brownian motion*, for which both time dependent and steady-state distribution is known.

3 One dimensional reflection (Skorohod) mapping

Introduce

$$X(t) = V(Q(0) + A(t)) - t \tag{4}$$

Then $X(t)$ represents the total work that arrived into the system up to time t minus t. Then we can rewrite (2) as

$$Z(t) = X(t) + I(t) \geq 0. \tag{5}$$

Observe that $I(t)$ is a non-decreasing piece-wise linear continuous process (it increases at rate 1 when there are no jobs in the system, and stays constant when there are jobs in the system). In particular,

$$dY(t) \geq 0 \tag{6}$$

whenever derivative is defined. Moreover, since the cumulative idling time can increase *only* when there are no jobs in the system, then

$$\int_0^\infty Z(t) dI(t) = 0. \tag{7}$$

There is a good reason to write equations (5),(6),(7) – these equations turn out to be defining Q, Y uniquely from X and we now establish this fact in a more general setting. We say that $x \in D[0,\infty)$ has no downward jumps, if at every point of discontinuity t_0 we have $\lim_{t\uparrow t_0} x(t) \leq x(t_0)$.

Problem 2. *Show that X has no downward jumps.*

Theorem 1 (Reflection (Skorohod) Mapping Theorem)**.** *Given $x \in D[0,\infty), x(0) \geq 0$, with no downward jumps, there exists a* unique *pair $y \in C[0,\infty), z \in D[0,\infty)$ such that*

1. $z(t) = x(t) + y(t) \geq 0,\ t \in \mathbb{R}_+$

2. $dy(t) \geq 0$

3. $z(t)dy(t) = 0$ *for all* $t \in \mathbb{R}_+$.

This pair is uniquely defined by

$$y(t) = \sup_{0 \leq s \leq t} (-x(s))^+ \tag{8}$$

$$z(t) = x(t) + \sup_{0 \leq s \leq t} (-x(s))^+ \tag{9}$$

Moreover, the mappings $\Psi : D[0,\infty) \to C[0,\infty),\ \Phi : D[0,\infty) \to D[0,\infty)$ *given by* $y = \Psi(x), z = \Phi(x)$ *are Lipschitz continuous with respect to the uniform norm* $\|\cdot\|_T$ *for every* $T > 0$*. We call* z *the* reflected *process of* x *and* y *the regulator of* x

Intuitively, the "role" of the regulator process y is to provide enough "pushing" to ensure that the modified process $z = x + y$ remains non-negative, given a (possibly negative) process x.

Proof. We begin by verifying that the solutions given in the theorem are valid. Clearly, $z \geq 0$ and y is non-decreasing, implying $dy \geq 0$. Let us show that $y \in C[0,\infty)$. Fix t. If x is continuous in t then continuity of y in t is immediate, as supremum of a continuous function is a continuous function. Suppose x is discontinuous in t, namely it has an upward jump. Let $x^- = \lim_{t' \uparrow t} x(t)$. Then $x(t) > x^-$ and, since x is right-continuous, $x(t') > x^-$ for all t' in some interval $[t, t+\delta]$. This means $y(t') = y(t)$ on $[t, t+\delta]$. Also since x^- is the left limit of x, then for every $\epsilon > 0$ there exists δ, such that $|x(t') - x^-| < \epsilon$ for $t' \in [t-\delta, t]$. This means $|y(t') - y(t)| < \epsilon$ for $t' \in [t-\delta, t]$ and the proof of continuity is complete. Finally, we check that $zdy = 0$. Suppose $dy > 0$. Then we claim that $x(t) = \inf_{0 \leq s \leq t} x(s)$ (inf is achieved in t). Indeed, otherwise $x(t) > \inf_{0 \leq s \leq t} x(s)$ and since x can only have upward jumps, $x(t+\delta) > \inf_{0 \leq s \leq t} x(s)$ for all sufficiently small δ. As a result $y(t+\delta) = y(t)$ for all sufficiently small δ, contradicting $dy > 0$. We conclude $x(t) = \inf_{0 \leq s \leq t} x(s)$, which implies $z(t) = 0$ and $z(t)dy(t) = 0$.

We now establish the uniqueness of the solution. Suppose there exists an additional solution $y' \in C[0,\infty), z' \in D[0,\infty)$ satisfying the properties of the reflection mapping, for the same input process x. Note that $z' - z = y' - y$ is a difference between two continuous non-decreasing processes. As a result it is differentiable almost everywhere and

$$\begin{aligned}\frac{1}{2}(z(t)-z'(t))^2 &= \int_0^t (z(s)-z'(s))\frac{d}{ds}(z(s)-z'(s))ds \\ &= \int_0^t (z(s)-z'(s))(\frac{d}{ds}y(s)-\frac{d}{ds}y'(s))ds\end{aligned}$$

But, by assumption $zdy = z'dy' = 0$, and we obtain an expression

$$\int_0^t (-z(s)\frac{d}{ds}y'(s) - z'(s)\frac{d}{ds}y(s)ds) \leq 0$$

We conclude $z(t) = z'(t)$. It is then immediate that $y(t) = y'(t)$.

It remains to show that the mappings $y = \Psi(x)$ and $z = \Phi(x)$ are Lipschitz continuous. That is, we need to show that for some constants $C_1, C_2 > 0$ and every pair $x, x' \in D[0,T]$, we have

$$\|\Psi(x) - \Psi(x')\|_T \leq \|x - x'\|_T, \quad \|\Phi(x) - \Phi(x')\|_T \leq \|x - x'\|_T$$

Denote $\|x - x'\|_T$ by $V \geq 0$. For any $t \in [0,T]$ we have

$$\begin{aligned}\Psi(x)(t) - \Psi(x')(t) &= \sup_{0\leq s\leq t}(-x(s))^+ - \sup_{0\leq s\leq t}(-x'(s))^+ \\ &\leq \sup_{0\leq s\leq t}(-x'(s)+V)^+ - \sup_{0\leq s\leq t}(-x'(s))^+ \\ &\leq \sup_{0\leq s\leq t}(-x'(s))^+ + V - \sup_{0\leq s\leq t}(-x'(s))^+ \\ &= V\end{aligned}$$

Similarly, we show $\Psi(x')(t) - \Psi(x)(t) \leq V$. This proves Lipschitz continuity of Ψ with constant $C_1 = 1$. The Lipschitz continuity of Φ follows then immediately with constant $C_2 = 2$. $\square$

The reflection mapping also satisfies the following important "memoryless" property: if we consider the process starting from some time t_0 its reflection is the same as if we started at time 0:

Proposition 1. *Given $t_0 > 0, x \in D[0,\infty)$ and the reflection $y = \Psi(x), z = \Phi(x)$ of x, consider a modified process $\hat{x}(t) = z(t_0) + x(t_0 + t) - x(t_0), t \geq 0$. Then $\Phi(\hat{x})(t) = z(t_0 + t), \Psi(\hat{x})(t) = y(t_0 + t) - y(t_0), t \geq 0$.*

Problem 3. *Establish this proposition.*

4 Reflected Brownian motion

We will see later on that when the G/G/1 queueing system is in heavy-traffic, the process $X(t)$ defined in (4) is approximated well by a Brownian motion. For now assume that it is in fact a Brownian motion. Note, that $X(t)$ is a process, whose distribution we now in principle, since it is directly linked the arrival and service processes. Thus if we can find a reflection of X with respect to the Skorohod mapping, we obtain an approximation of the workload process $Z(t)$.

Definition 1. *A (one-dimensional) Reflected Brownian Motion (RBM) is the process $Z = \Phi(B)$ obtained by Skorohod mapping (Ψ, Φ), when the input process is a Brownian motion $B(t), B(0) \geq 0$. When B has drift θ and variance σ^2, we also write $Z = RBM(\theta, \sigma^2)$.*

Knowing the forms (8),(9) of the reflection mapping allows us to obtain say something about the distribution of the reflected process Z when the input process is a Brownian motion.

Theorem 2. *A Reflected Brownian Motion $Z(t) = RBM(\theta, \sigma^2)$, converges in distribution to some limiting random variable $Z(\infty)$ iff $\theta < 0$. In this case $Z(\infty)$ is exponentially distributed with parameter $-2\theta/\sigma^2$.*

It should not be surprising that we get limiting exponential distribution when the drift is negative and no limiting distribution when drift is non-negative. After all we have established these facts for a maximum of a Brownian motion $M(t) = \sup_{0\leq s\leq t} B(s)$. We just do the appropriate adjustments.

Proof. First assume $B(0) = 0$. Observe that in this case $\sup_{0\leq s\leq t}(-B(s))^+ = \sup_{0\leq s\leq t}(-B(s))$. Then

$$\mathbb{P}(Z(t) \geq z) = \mathbb{P}(B(t) + \sup_{0\leq s\leq t}(-B(s)) \geq z).$$

Consider the process $\hat{B}_t(u) = B(t) - B(t-u), 0 \leq u \leq t$ and $\hat{B}_t(u) = B(u), u \geq t$. Observe that it is also a Brownian motion with drift θ and variance σ^2 (this is similar to a problem on midterm and some of the hw problems). Therefore

$$\begin{aligned}
\mathbb{P}(B(t) + \sup_{0\leq s\leq t}(-B(s)) \geq z) &= \mathbb{P}(\sup_{0\leq s\leq t}(B(t) - B(s)) \geq z) \\
&= \mathbb{P}(\sup_{0\leq s\leq t}\hat{B}(s) \geq z) \\
&= \mathbb{P}(\sup_{0\leq s\leq t}B(s) \geq z)
\end{aligned}$$

Now we just recall basic properties of a Brownian motion. When $\theta \geq 0$ we have $\sup_{s\geq 0} B(s) = \infty$ with probability one. Therefore $\mathbb{P}(\sup_{0\leq s\leq t} B(s) \geq z) \to 1$ as $t \to \infty$, and limiting distribution does not exist.

When $\theta < 0$, we know that $\mathbb{P}(\sup_{s\geq 0} B(s) \geq z) = \exp(-2\theta/\sigma^2)$, thus

$$\lim_{t\to\infty} \mathbb{P}(\sup_{0\leq s\leq t} B(s) \geq z) = \exp(-2\theta/\sigma^2).$$

This concludes the proof for the case $B(0) = 0$.

Suppose now $B(0) = w > 0$. We let $\tau = \inf\{t : B(t) = 0\}$, including the possibility $\tau = \infty$. When $\theta \leq 0$ we know that $\tau < \infty$ a.s. Then from strong Markov property of a Brownian motion we know that $B(\tau + t)$ is a Brownian motion which starts with origin. Applying Proposition 1, $Z(\tau+t)$ is the reflected Brownian motion implying

$$\lim_{t\to\infty} \mathbb{P}(Z(t) \geq z) = \lim_{t\to\infty} \mathbb{P}(Z(\tau + t) \geq z)$$

which is unity when $\theta = 0$ and $\exp(-2\theta/\sigma^2)$ when $\theta < 0$.

Suppose now $\theta > 0$. Then we simply use the fact $Z(t) \geq B(t)$, which follows from non-negativity of $Y(t)$ and $\lim_{t\to\infty} \mathbb{P}(B(t) \geq z) = 1$ when $\theta > 0$, to conclude $\lim_{t\to\infty} \mathbb{P}(Z(t) \geq z) = 1$.

This concludes the proof. $\square$

5 Additional reading materials

- Chapter 6 from Chen & Yao book "Fundamentals of Queueing Networks" [1]

References

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

15.070J / 6.265J Advanced Stochastic Processes
Fall 2013