Solution for Problem Set 3

October 15, 2009

1.

(a)

In the following derivation for $P(S(n) = i | Y(n) = i, Y(n-1), Y(n-2), \ldots, Y(1))$, assume $0 \leq i \leq n$, and that $i \bmod 2 \equiv n \bmod 2$, otherwise the probability is clearly 0. Additionally, if $i = 0$, the probability is clearly 1. Then for $i > 0$,

$$
\begin{aligned}
& P(S(n) = i | Y(n) = i, Y(n-1), Y(n-2), \ldots, Y(1)) \\
= \; & \frac{\binom{n}{(n+i)/2} p^{(n+i)/2}(1-p)^{(n-i)/2}}{\binom{n}{(n+i)/2} p^{(n+i)/2}(1-p)^{(n-i)/2} + \binom{n}{(n-i)/2} p^{(n-i)/2}(1-p)^{(n+i)/2}} \\
= \; & \frac{p^{i/2}(1-p)^{-i/2}}{p^{i/2}(1-p)^{-i/2} + p^{-i/2}(1-p)^{i/2}} \\
= \; & \frac{p^i}{p^i + (1-p)^i}
\end{aligned}
$$

where we have used the fact $\binom{n}{(n+i)/2} = \binom{n}{n-(n+i)/2} = \binom{n}{(n-i)/2}$.

To elaborate, the numerator represents the probability of the random walk ending up in state i, whereas the denominator represents the probability of the random walk ending up in either state i or $-i$ (distinct, since $i \neq 0$).

(b)

Now we can proceed to calculate $P(Y(n+1) = i+1 | Y(n) = i, Y(n-1), Y(n-2), \ldots, Y(1))$.

Again we assume $0 \leq i \leq n$, and that $i \bmod 2 \equiv n \bmod 2$, otherwise the probability is clearly 0. Additionally, if $i = 0$, the probability is clearly 1. Then for $i > 0$,

$$
\begin{aligned}
& P(Y(n+1) = i+1 | Y(n) = i, Y(n-1), Y(n-2), \ldots, Y(1)) \\
= \; & P(Y(n+1) = i+1 | S(n) = i, Y(n) = i, Y(n-1), Y(n-2), \ldots, Y(1)) \\
& + P(Y(n+1) = i+1 | S(n) = -i, Y(n) = i, Y(n-1), Y(n-2), \ldots, Y(1)) \\
= \; & \{P(Y(n+1) = i+1 | S(n) = i) \cdot P(S(n) = i | Y(n) = i, Y(n-1), Y(n-2), \ldots, Y(1))\} \\
& + \{P(Y(n+1) = i+1 | S(n) = -i) \cdot P(S(n) = -i | Y(n) = i, Y(n-1), Y(n-2), \ldots, Y(1))\} \\
= \; & \left\{ p \cdot \frac{p^i}{p^i + (1-p)^i} \right\} + \left\{ (1-p) \cdot \frac{(1-p)^i}{p^i + (1-p)^i} \right\} = \frac{p^{i+1} + (1-p)^{i+1}}{p^i + (1-p)^i}
\end{aligned}
$$

where we have made use of the fact that $P(Y(n) = i) = \{P(S(n) = i)\} + \{P(S(n) = -i)\}$ when $i > 0$.

Now consider $P(Y(n+1) = i-1 | Y(n) = i, Y(n-1), Y(n-2), \ldots, Y(1))$.

Note that this probability must be 0 when $i = 0$, since Y is a non-negative random process.

With an exactly analogous derivation as before, under the same assumptions, we have

$P(Y(n+1) = i-1 | Y(n) = i, Y(n-1), Y(n-2), \ldots, Y(1)) = \dfrac{(1-p)p^i + p(1-p)^i}{p^i + (1-p)^i}$

Summarizing, we have

$$P(Y(n+1) = i+1 | Y(n) = i, Y(n-1), Y(n-2), \ldots, Y(1)) = \begin{cases} 1 & , \ i = 0, \ n \bmod 2 \equiv 0 \\ \dfrac{p^{i+1} + (1-p)^{i+1}}{p^i + (1-p)^i} & , \ i > 0, \ i \bmod 2 \equiv n \bmod 2 \\ 0 & , \text{ otherwise} \end{cases}$$

$$P(Y(n+1) = i-1 | Y(n) = i, Y(n-1), Y(n-2), \ldots, Y(1)) = \begin{cases} \dfrac{(1-p)p^i + p(1-p)^i}{p^i + (1-p)^i} & , \ i > 0, \ i \bmod 2 \equiv n \bmod 2 \\ 0 & , \text{ otherwise} \end{cases}$$

It is also clear that no other transitions are possible for Y; it can only increment or decrement at each time.

Since the RHS to these expressions are only functions of Y only through $Y(n) = i$, Y is a Markov chain, with initial state $Y(0) = 0$ and transition matrix P, with entries given by

$P_{0,1} = 1$

$P_{i,i+1} = \dfrac{p\nu^i + (1-p)}{\nu^i + 1} \ \forall i \geq 1$

$P_{i,i-1} = \dfrac{(1-p)p^i + p(1-p)^i}{p^i + (1-p)^i} \ \forall i \geq 1$

and all other entries 0.

2: Worst case Chernoff Bound

This problem deals with a scenario where the actual distribution of the source is not known, but a bound on the expectation of the source is known. Under this constraint we need to find which distribution represents the worst case for the Chernoff bound:

$$P(\sum_{i=1}^{n} X_i \geq nx) \quad \leq \quad e^{-nI(x)}.$$

We need to find the distribution that minimizes the rate function subject to the constraint $EX \leq \rho$ and $0 \leq X \leq M$. So the worst case rate function is

$$I(x) \quad = \quad \min_{p(x)} \sup_{\theta} \theta x - \Lambda(\theta).$$

The order of the max and min can be interchanged since the function $\theta x - \Lambda(\theta)$ is convex is $p(x)$ and concave in θ. Therefore,

$$I(x) \quad = \quad \sup_{\theta} \min_{p(x)} \theta x - \Lambda(\theta).$$

We can now focus on maximizing $Ee^{\theta X}$ over $p(x)$ satisfying the constraints. We claim that the following distribution minimizes this MGF:

$$\begin{aligned} p(X = x) &= \frac{\rho}{M}, \text{if x = M} \\ &= 1 - \frac{\rho}{M}, \text{if x = 0} \\ &= 0, \text{otherwise.} \end{aligned}$$

For this distribution, $Ee^{\theta X} = \frac{\rho}{M}e^{\theta M} + (1-\frac{\rho}{M})$. If we show that, for every $p(x)$ satisfying the constraints, $Ee^{\theta X}$ is lesser than this value, then we are done. We observe that, by the convexity of $e^{\theta x}$ and $0 \leq X \leq M$,

$$\begin{aligned} e^{\theta X} &\leq \frac{X}{M}e^{\theta M} + (1 - \frac{X}{M}), \\ \Rightarrow Ee^{\theta X} &\leq \frac{\rho}{M}e^{\theta M} + (1 - \frac{\rho}{M}) \end{aligned}$$

and we are done!

In this problem, since we already had a guess for the optimizing distribution, verifying that this is optimal was easy. If we did not have a guess, then we can start by observing that $Ee^{\theta X}$ is linear $p(x)$ and that the set of $p(x)$ satisfying the constraint is a convex, compact set. Therefore the maximum would occur on one of the corner points of this set. The corner points of this set of distributions can be shown to be the set of all two-point distributions satisfying the constraints, and from these, we can easily identify the optimizing distribution.

3

(a)

$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-(x-\mu)^2/(2\sigma^2)\right]$

$M(s) = E[e^{sX}] = \int_{-\infty}^{\infty} e^{sx} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-(x-\mu)^2/(2\sigma^2)\right] dx$

$$\begin{aligned} e^{sx} \cdot \exp\left[-(x-\mu)^2/(2\sigma^2)\right] &= \exp\left\{-\frac{1}{2\sigma^2} \cdot \left[x^2 - 2\mu x + \mu^2 - 2s\sigma^2 x\right]\right\} \\ &= \exp\left\{-\frac{1}{2\sigma^2} \cdot \left[x^2 - 2(\mu + s\sigma^2)x + \mu^2\right]\right\} \\ &= \exp\left\{-\frac{1}{2\sigma^2} \cdot \left[x^2 - 2(\mu + s\sigma^2)x + (\mu + s\sigma^2)^2 - 2\mu s\sigma^2 + \left(s\sigma^2\right)^2\right]\right\} \\ &= \exp\left\{-\frac{1}{2\sigma^2} \cdot \left[x^2 - 2(\mu + s\sigma^2)x + (\mu + s\sigma^2)^2\right]\right\} \cdot \exp\left(s\mu + \frac{1}{2}s^2\sigma^2\right) \\ &= \exp\left(s\mu + \frac{1}{2}s^2\sigma^2\right) \cdot \exp\left[-(x - \mu - s\sigma^2)^2/(2\sigma^2)\right] \end{aligned}$$

$$\begin{aligned} M(s) &= \int_{-\infty}^{\infty} e^{sx} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-(x-\mu)^2/(2\sigma^2)\right] dx \\ &= \exp\left(s\mu + \frac{1}{2}s^2\sigma^2\right) \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-(x - \mu - s\sigma^2)^2/(2\sigma^2)\right] dx \\ &= \exp\left(s\mu + \frac{1}{2}s^2\sigma^2\right) \end{aligned}$$

$I(x) = \sup_s \left[sx - s\mu - \frac{1}{2}s^2\sigma^2\right]$

$sx - s\mu$ is linear and $-\frac{1}{2}s^2\sigma^2$ is concave and continuous, so $\left[sx - s\mu - \frac{1}{2}s^2\sigma^2\right]$ is concave and continuous. Then we can solve for the optimal $s = s^*$ by setting the derivative to zero:

$x - \mu - s^*\sigma^2 = 0$

$s^* = \dfrac{x-\mu}{\sigma^2}$

$I(x) = s^*(x-\mu) - \dfrac{1}{2}(s^*)^2\sigma^2 = \dfrac{(x-\mu)^2 - \frac{1}{2}(x-\mu)^2}{\sigma^2} = \dfrac{(x-\mu)^2}{2\sigma^2}$

(b)

$f_X(x) = \lambda e^{-\lambda x}$

$M(s) = E[e^{sX}] = \int_0^\infty e^{sx}\lambda e^{-\lambda x}dx = \lambda \int_0^\infty e^{-(\lambda-s)x}dx = \dfrac{\lambda}{\lambda - s}\cdot \int_0^\infty (\lambda-s)e^{-(\lambda-s)x}dx = \begin{cases} \frac{\lambda}{\lambda-s} & , \ s<\lambda \\ \infty & , \ s \geq \lambda \end{cases}$

$I(x) = \sup_s\left[sx - \log M(s)\right] = \sup_{s<\lambda}\left[sx - \log\lambda + \log(\lambda-s)\right]$

sx is linear and $\log(\lambda - s)$ is concave and continuous for $s<\lambda$, so $[sx - \log\lambda + \log(\lambda - s)]$ is concave and continuous.

$\dfrac{d}{ds}\left[sx - \log M(s)\right] = x - \dfrac{1}{\lambda - s}$. This is negative for all $s<\lambda$ when $x \leq 0$, so the supremum occurs as $s \to -\infty$,

$I(x) = \lim_{s\to-\infty}\left[sx - \log\lambda + \log(\lambda - s)\right] = \infty \ \forall x \leq 0.$

If $x > 0$, we can solve for the optimal $s = s^*$ by setting the derivative to zero:

$x - \dfrac{1}{\lambda - s^*} = 0$

$s^* = \lambda - \dfrac{1}{x}$

$I(x) = s^*x - \log\lambda + \log(\lambda - s^*) = \lambda x - 1 - \log\left(\lambda x\right)$

To summarize,

$I(x) = \begin{cases} +\infty & , \ x \leq 0 \\ \lambda x - 1 - \log\left(\lambda x\right) & , \ x > 0 \end{cases}$

(c)

$p_X(k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$

$M(s) = E[e^{sX}] = \sum_{k=0}^\infty \left[e^{sk}\cdot\dfrac{\lambda^k e^{-\lambda}}{k!}\right] = \sum_{k=0}^\infty \dfrac{\left(e^s\lambda\right)^k e^{-\lambda}}{k!} = \dfrac{e^{e^s\lambda}}{e^\lambda}\cdot\sum_{k=0}^\infty \dfrac{\left(e^s\lambda\right)^k e^{-e^s\lambda}}{k!} = \exp(\lambda(e^s - 1))$

$I(x) = \sup_s\left[sx - \log M(s)\right] = \sup_s\left[sx - \lambda(e^s - 1)\right]$

sx is linear and $-\lambda(e^s-1)$ is concave and continuous, so $[sx - \lambda(e^s - 1)]$ is concave and continuous.

$\dfrac{d}{ds}\left[sx - \log M(s)\right] = x - \lambda e^s$. This is negative for all s when $x \leq 0$, so the supremum occurs as $s \to -\infty$,

$I(x) = \lim_{s\to-\infty}\left[sx - \lambda(e^s - 1)\right] = \infty, \ x < 0.$

Note that for the edge case $x = 0$, the supremum again occurs at $-\infty$, but the derivative there is zero. This allows the following optimization to work at the point $x = 0$.

If $x \geq 0$, we can solve for the optimal $s = s^*$ by setting the derivative to zero:

$x - \lambda e^{s^*} = 0$

$s^* = \log\left(\dfrac{x}{\lambda}\right)$

$I(x) = x\log\left(\dfrac{x}{\lambda}\right) - \lambda(\dfrac{x}{\lambda} - 1) = x\left[\log\left(\dfrac{x}{\lambda}\right) - 1\right] + \lambda$

To summarize,

$I(x) = \begin{cases} +\infty & , \ x < 0 \\ x\left[\log\left(\dfrac{x}{\lambda}\right) - 1\right] + \lambda & , \ x \geq 0 \end{cases}$

where we have used the convention $0\log 0 = 0$.

(d)

Assume $0 < p < 1$.

$p_X(k) = \begin{cases} p, & k = 1 \\ 1-p, & k = 0 \end{cases}$

$M(s) = E[e^{sX}] = e^{s1}\cdot p + e^{s0}\cdot(1-p) = 1 - p + pe^s$

$I(x) = \sup_s \{sx - \log[1-p+pe^s]\} = \sup_s \left\{\log\left[\frac{e^{sx}}{1-p+pe^s}\right]\right\}$

Note that $1-p+pe^s > 0\ \forall s$, so it is easy to see that $-\log M(s) = -\log[1-p+pe^s]$ is an analytic function of s.

Now we can verify $-\log M(s)$ is concave in s by checking the second derivative.

$$\frac{d}{ds}[-\log M(s)] = -\frac{pe^s}{1-p+pe^s}$$

$$\begin{aligned}\frac{d^2}{ds^2}[-\log M(s)] &= -\frac{[1-p+pe^s]\cdot pe^s - pe^s \cdot pe^s}{1-p+pe^s} \\ &= \frac{-(1-p)\cdot pe^s}{1-p+pe^s}\end{aligned}$$

Thus $\frac{d^2}{ds^2}[-\log M(s)] < 0\ \forall s$. Combined with the fact that $-\log M(s)$ is analytic, we have that $-\log M(s)$ is concave and continuous.

sx is linear and $-\log M(s)$ is concave and continuous, so $[sx - \log M(s)]$ is concave and continuous.

$\frac{d}{ds}[sx - \log M(s)] = x - \frac{pe^s}{1-p+pe^s}$. This is negative for all s when $x < 0$, so the supremum occurs as $s \to -\infty$,

$I(x) = \lim_{s\to-\infty} \log\left[\frac{e^{sx}}{1-p+pe^s}\right] = \infty\ \forall x < 0.$

Similarly, this is positive for all s when $x > 1$, so the supremum occurs as $s \to +\infty$,

$I(x) = \lim_{s\to+\infty} \log\left[\frac{e^{sx}}{1-p+pe^s}\right] = \infty\ \forall x > 0.$

Note that for the edge cases $x = 0$ and $x = 1$, the supremum again occurs at $-\infty$ and $+\infty$ respectively, but the derivative there is zero. This allows the following optimization to work at these two points.

If $0 \le x \le 1$, we can solve for the optimal $s = s^*$ by setting the derivative to zero:

$x - \frac{pe^{s^*}}{1-p+pe^{s^*}} = 0$

$x - px + pe^{s^*}x = pe^{s^*}$

$e^{s^*} = \frac{x(1-p)}{p(1-x)}$

$s^* = \log\left[\frac{x(1-p)}{p(1-x)}\right]$

$$\begin{aligned}I(x) &= s^*x - \log\left[1-p+pe^{s^*}\right] \\ &= x\log\left[\frac{x(1-p)}{p(1-x)}\right] - \log\left[1-p+\frac{x(1-p)}{1-x}\right] \\ &= x\log\left[\frac{x(1-p)}{p(1-x)}\right] - \log\left[\frac{1-p}{1-x}\right] \\ &= x\log\left[\frac{x}{p}\right] + (1-x)\log\left[\frac{1-x}{1-p}\right] \\ &= D(x\|p)\end{aligned}$$

where $D(x\|p)$ is the Kullback-Leibler divergence between Bernoulli distributions with parameters x and p.

To summarize,

$$I(x) = \begin{cases} +\infty & , \ x < 0 \\ (1-x)\log\left[\dfrac{1-x}{1-p}\right] + x\log\left[\dfrac{x}{p}\right] & , \ 0 \le x \le 1 \\ +\infty & , \ x > 1 \end{cases}$$

where we have used the convention $0 \log 0 = 0$.

4: Network Traffic Policing

This problem considers the insertion of a traffic device which regulates the number of packets inserted by a source x to $h(x)$, such that $Eh(X) = \mu$. Clearly $h(x)$ satisfies $h(x) \le x$. The objective of the problem is to design $h(x)$ such that the effective bandwidth of $h(X)$ is minimized. Minimizing the effective bandwidth is the same as minimizing the MGF (moment generating function), by definition of effective bandwidth.

Let $h(a)$ be any policy and $h^*(a) = min(a, M)$ be the given policy. We will show that

$$E(e^{\theta h(a)}) \quad \ge \quad E(e^{\theta h^*(a)}).$$

To begin with, observe

$$\begin{aligned} e^{\theta h(a)} - e^{\theta h^*(a)} &= e^{\theta h^*(a)}(e^{\theta(h(a)-h^*(a))} - 1) \\ &\ge e^{\theta h^*(a)}\theta(h(a) - h^*(a)), \end{aligned}$$

since $e^x - 1 \ge x$.

Also observe $h(a) \le a = h^*(a)$, for $a \le M$.

$$\begin{aligned} E(e^{\theta h(a)} - e^{\theta h^*(a)}) &\ge E(\theta e^{\theta h^*(a)}(h(a) - h^*(a))) \\ &= E(\theta e^{\theta h^*(a)}(h(a) - h^*(a))I_{a\le M}) + E(\theta e^{\theta h^*(a)}(h(a) - h^*(a))I_{a>M}) \\ &= E(\theta e^{\theta h^*(a)}(h(a) - h^*(a))I_{a\le M}) + E(\theta e^{\theta M}(h(a) - h^*(a))I_{a>M}) \\ &= E(\theta e^{\theta M} h(a) - h^*(a))I_{a\le M}) + E(\theta e^{\theta M}(h(a) - h^*(a))I_{a>M}) \\ &= E(\theta e^{\theta M}(h(a) - h^*(a))\{I_{a\le M} + I_{a>M}\}) \\ &= \theta e^{\theta M} E(h(a) - h^*(a)) \\ &= 0 \end{aligned}$$

Thus we have proved that the given h^* is optimal.

In case we were not given h^* to start with, we can use the following alternate (but less rigorous) approach to figure out the optimal h^*. To begin with, we can prove the following property: Let $f(x)$ be a function having a certain MGF and a given mean. Consider $x_1 \le x_2$, such that $f(x_1) < f(x_2)$. If we define $g(x)$ to be equal to $f(x)$ elsewhere, but increase the value of $g(x)$ at x_1 and decrease the value of $g(x)$ at x_2, then $g(x)$ has MGF no greater than $f(x)$.

Proof: Let $p(x_i)$ be the probability that source takes values x_i. Define

$$\begin{aligned} g(x_1) &= f(x_1) + \epsilon\frac{p(x_2)}{p(x_1)} \\ g(x_2) &= f(x_2) - \epsilon \\ g(x) &= f(x) \text{ elsewhere.} \end{aligned}$$

Then $Eg(X) = Ef(X)$ clearly. Now

$$Ee^{g(X)} - Ee^{f(X)} \quad = \quad p(x_1)(e^{\theta f(x_1)}e^{\epsilon\frac{p(x_2)}{p(x_1)}} - e^{\theta f(x_1)}) + p(x_2)(e^{\theta f(x_2)}e^{-\epsilon} - e^{\theta f(x_2)})$$

Now, $e^x = 1 + x + o(x)$,, and we have

$$
\begin{aligned}
Ee^{g(X)} - Ee^{f(X)} &= p(x_1)(e^{\theta f(x_1)}\{1+\epsilon\frac{p(x_2)}{p(x_1)}\} - e^{\theta f(x_1)}) + p(x_2)(e^{\theta f(x_2)}\{1-\epsilon\} - e^{\theta f(x_2)}) + o(\epsilon) \\
&= p(x_1)e^{\theta f(x_1)}\{\epsilon\frac{p(x_2)}{p(x_1)}\} + p(x_2)e^{\theta f(x_2)}(-\epsilon) + o(\epsilon) \\
&= \epsilon p(x_2)\{e^{\theta f(x_1)} - e^{\theta f(x_2)}\} + o(\epsilon) \\
&< 0, \text{ since } f(x_1) < f(x_2), \text{ and if } \epsilon \text{ is small.}
\end{aligned}
$$

The above property implies that the optimal $h(x)$ is monotone non-decreasing. Also, this property implies that we can keep increasing $f(x)$ at small x and decreasing $f(x)$ at large x as long as the constraints are not violated. Now, since $f(x) \leq x$, we have to stop when $f(x)$ reaches x for all small x. The above process terminates only when $f(x) \leq x, \forall x \leq x_t$ for a threshold x_t, and $f(x) = K, \forall x > x_t$. Clearly $K = f(x_t)$. Thus this function minimizes the effective bandwidth under the mean constraint.

5.

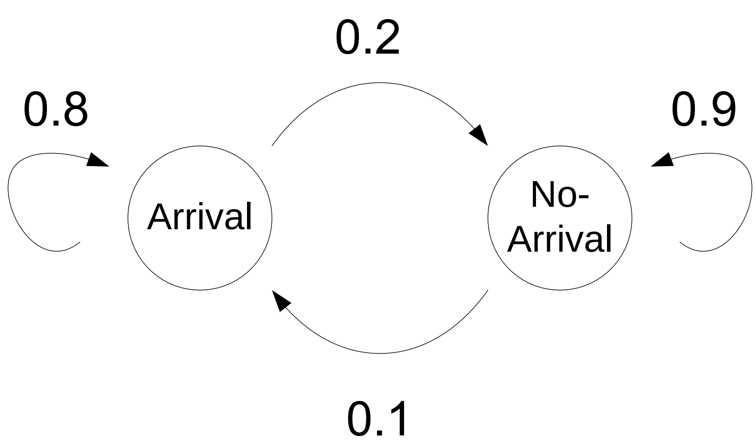

Figure 1: Markov chain describing the arrival process.

(i) The arrival process itself is a Markov chain as shown in Figure 1. Denote the steady state probability that an arrival happens by π_A and no arrival does by π_{NA}. From $\pi = \pi P$ and $\sum_i \pi_i = 1$,

$$
\begin{aligned}
\pi_A &= 0.8\pi_A + 0.1\pi_{NA} \\
\pi_{NA} &= 0.2\pi_A + 0.9\pi_{NA} \\
\pi_A + \pi_{NA} &= 1,
\end{aligned}
\tag{1}
$$

which can be solved as

$$
\begin{aligned}
\pi_A &= \frac{1}{2}\pi_{NA} = \frac{1}{3} \\
\pi_{NA} &= \frac{2}{3}.
\end{aligned}
\tag{2}
$$

Thus the mean arrival rate is $\lambda := \pi_A = \frac{1}{3}$. For this queueing system to be stable in a mean sense, $\mu > \frac{1}{3}$.

(ii) Queue dynamics is $q(k+1) = \big(q(k) + a(k) - s(k)\big)^+$ where $a(k), s(k) \in \{0,1\}$, $s(k)$ is iid over k and $a(k)$ is described by the Markov chain in Figure 1 in (i).

The queueing system can be described by a Markov chain with states represented by a pair $\big(q(k), a(k)\big)$. We can immediate see that this Markov chain is aperiodic and irreducible. To prove that this Markov chain is positive recurrent, we use Foster-Lyapunov theorem with a Lyapunov function $V\big(q(k), a(k)\big) = \frac{1}{2}q^2(k)$.

Given that we're in state (q, a), consider the drift

$$\mathrm{E}[V\big(q(k+M), a(k+M)\big) - V\big(q(k), a(k)\big)|q(k) = q, a(k) = q] \tag{3}$$

If $M = 1$, (3) $\leq C_1 + q\big(\mathrm{E}[a(k)|a(k) = a] - \mu\big)$.
If $M = 2$, (3) $\leq C_2 + q\big(\mathrm{E}[a(k+1) + a(k)|a(k) = a] - \mu\big)$.
For arbitrary $M = m$, (3)

$$\begin{aligned} &\leq C_m + q\big(\mathrm{E}[\sum_{i=0}^{m-1} a(k+i)|a(k) = a] - m \cdot \mu\big) \\ &= C_m + m \cdot q\big(\mathrm{E}[\frac{1}{m}\sum_{i=0}^{m-1} a(k+i)|a(k) = a] - \mu\big) \end{aligned} \tag{4}$$

for some constants C_l, $l = 1, 2, \cdots$.

Since $\lim_{m\to\infty} \frac{1}{m}\mathrm{E}[\sum_{i=0}^{m-1} a(k+i)|a(k) = a] = \mathrm{E}[a(k)] = \frac{1}{3}$, though we started from state (q, a), we can always find large m such that

$$\frac{1}{m}\mathrm{E}[\sum_{i=0}^{m-1} a(k+i)|a(k) = a] < \frac{1}{3}, \tag{5}$$

for any $a \in \{0, 1\}$. Therefore, for large q and m, the drift in (3) $< -\epsilon$, which means that the considered Markov chain is positive recurrent.

6.

In Problem 5 of Problem Set 2, S_{max} was used to upper-bound $\mathrm{E}[u^2]$ and the bound goes to zero.

Consider $u(k)$, which is denoted by u for simplicity:

$$\begin{aligned} u^2 &\leq u \cdot s \\ &\leq u \cdot A\mathbf{I}_{\{s<A\}} + u \cdot s\mathbf{I}_{\{s\geq A\}} \\ &\leq u \cdot A + s^2 \cdot \mathbf{I}_{\{s\geq A\}}, \end{aligned} \tag{6}$$

where $\mathbf{I}_{\{\cdot\}}$ is a standard indicator function.

A is chosen for given ϵ such that $\mathrm{E}[s^2\mathbf{I}_{\{s\geq A\}}] \leq \epsilon$. Thus, a big A makes ϵ small. Then,

$$\mathrm{E}[u^2] \leq \mathrm{E}[uA + s^2\mathbf{I}_{\{s\geq A\}}] \leq \mathrm{E}[u] \cdot A + \epsilon \leq (\mu - \lambda)A + \epsilon. \tag{7}$$

If $\lambda \to \mu$, $\mathrm{E}[u^2] \to \epsilon$ for any given ϵ. This means

$$\mathrm{E}[u^2] = 0 \ \text{ as } \ \lambda \to \mu, \tag{8}$$

which completes the proof.