APPROXIMATING HEAVY TRAFFIC WITH BROWNIAN MOTION

HARVEY BARNHARD

ABSTRACT. This expository paper shows how waiting times of certain queuing systems can be approximated by Brownian motion. In particular, when customers exit a queue at a slightly faster rate than they enter, the waiting time of the nth customer can be approximated by the supremum of reflected Brownian motion with negative drift. Along the way, we introduce fundamental concepts of queuing theory and Brownian motion. This paper assumes some familiarity of stochastic processes.

CONTENTS

Date: November 21, 2018.

1. Introduction

This paper discusses a Brownian Motion approximation of waiting times in stochastic systems known as queues. The paper begins by covering the fundamental concepts of queuing theory, the mathematical study of waiting lines. After waiting times and heavy traffic are introduced, we define and construct Brownian motion. The construction of Brownian motion in this paper will be less thorough than other texts on the subject. We instead emphasize the components of the construction most relevant to the final result—the waiting time of customers in a simple queue can be approximated with Brownian motion.

Throughout the paper, we use the terminology of a convenience store: *customers* populate a queue and the amount of time it takes for a customer to exit the system once at the front of the line is called a *service time*. This paper focuses on one of the simplest queuing models, and seeks to answer one simple question: *when customers are being serviced at a rate only slightly faster than customers are arriving, how long can a customer expect to wait in line before being serviced?*

2. Queuing Theory

Consider a queue in which inter-arrival times and service times are exponentially distributed. Note that a process with exponential inter-arrival times may also be regarded as a Poisson process where the number of customers arriving in a given time period follows the Poisson distribution. Let N_t denote the number of customers in line at time t. Let λ_N and μ_N be the rates at which customers arrive and are serviced given that there are N customers in line. For example, if $\lambda_N = 1$ and the unit of time is seconds, then we expect one arrival per second. If $\lambda_N = 1/2$ then we expect two arrivals per second. Denote $a(t)$ as the probability that the next arrival will occur t units of time from now, and $s(t)$ as the probability that the next service will be complete t units of time from now. The probability distributions are thus given by

$$a_N(t) = \lambda_N e^{-\lambda_N t}, \qquad s_N(t) = \mu_N e^{-\mu_N t}$$

So if $A \sim a_N(t)$ and $S \sim s_N(t)$ are random variables, then

$$\mathbb{E}\left[A\right] = \frac{1}{\lambda_N} \qquad \text{and} \qquad \mathbb{E}\left[S\right] = \frac{1}{\mu_N}$$

Figure 1 depicts such a queue. Rectangles represent customers waiting in line and the circle represents customers currently being serviced.

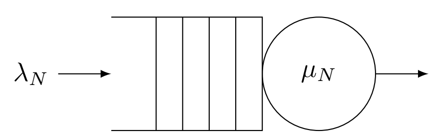

Figure 1.

We can also consider the possible number of customers in a queue as states, so that the state space consists of the non-negative integers, where λ_N and μ_N are the probabilistic rate of transition between states.

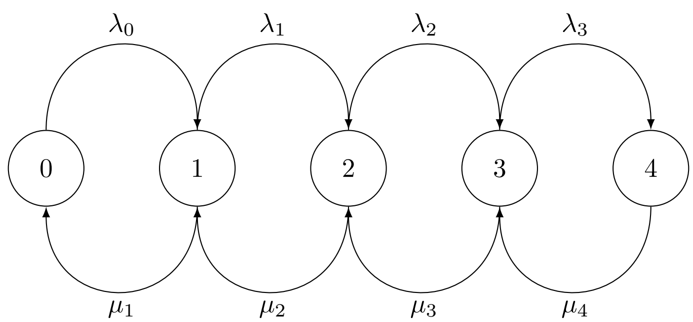

FIGURE 2.

2.1. **Continuous-Time Markov Chains.**
A system with a countable state-space and probabilistic transitions between states suggests the use of Markov chains. However, standard Markov chains of elementary probability theory operate in a discrete-time setting. In this paper, we consider exponential service and inter-arrival times. As a result, transitions between states can occur anytime in the continuous timeline of non-negative reals. We therefore aim to construct a countable-space, continuous-time version of a discrete-time Markov chain. We first introduce two properties that define continuous-time Markov chains—the Markov property and time-homogeneity.

Definition 2.1. Consider a stochastic process, $\{X_t\}$ taking values in a state space S. $\{X_t\}$ is said to exhibit the *Markov property* if for $y \in S$ and $t \geq s$

$$\mathbb{P}\{X_t = y \mid X_r, 0 \leq r \leq s\} = \mathbb{P}\{X_t = y \mid X_s\}$$

Definition 2.2. A *time-homogenous* Markov chain is a stochastic process $\{X_t\}_{t\geq 0}$ such that

$$\mathbb{P}\{X_t = y \mid X_s = x\} = \mathbb{P}\{X_{t-s} = y \mid X_0\}$$

If a process exhibits time-homogeneity, transition probabilities depend only on the state of the process, not the absolute time.

Definition 2.3. A *continuous-time Markov chain* is a stochastic process $\{X_t\}_{t\geq 0}$ taking values in a state space S and satisfying

$$\mathbb{P}\{X_{t+\Delta t} = i \mid X_t = i\} = 1 - q_{ii}\Delta t + o(\Delta t)$$
$$\mathbb{P}\{X_{t+\Delta t} = j \mid X_t = i\} = q_{ij}\Delta t + o(\Delta t)$$

where q_{ij} represents the transition rate from $i \in S$ to $j \in S$, and

$$q_{ii} \equiv \sum_{j \neq i} q_{ij}$$

The use of q_{ij} as a rate of transition between states requires further explanation. Each q_{ij} is the derivative of $p_{ij}(t)$, the probability of starting in state i and ending in state j after Δt units of time have elapsed (here we have assumed differentiability of $p_{ij}(t)$ at $\Delta t = 0$).

$$q_{ij} = \lim_{\Delta t \to 0} = \frac{\mathbb{P}\{X_{t+\Delta t} = j \mid X_t = i\}}{\Delta t}$$

The matrix $\mathbf{P}(t)$ is the *transition matrix* and its (i, j)th element is $p_{ij}(t)$:

$$\mathbf{P}(t) = \begin{bmatrix} p_{11}(t) & p_{12}(t) & \cdots \\ p_{21}(t) & p_{22}(t) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

The matrix $\mathbf{Q}$ is called the *infinitesimal generator* and its (i, j)th element is q_{ij}. Note that if zero time has elapsed, then there is zero probability of transitioning out of the original state, so $p_{ii}(0) = 1$ and $\mathbf{P}(0) = \mathbf{I}$. We can therefore write

$$\mathbf{Q} = \mathbf{P}'(0)$$

and

$$\mathbf{P}(t) = \mathbf{I} + \mathbf{Q}\Delta t$$

Example 2.5 will illustrate a deeper connection between q_{ij} and $p_{ij}(t)$, but we must first define what it means for two states to be *communicable*.

Definition 2.4. Let i and j be two states of a Markov chain. We say that i and j *communicate* if there is a positive probability that one state will lead to the other state given a certain amount of time. In notation, i and j communicate if there exists time durations $r > 0$ and $s > 0$ such that

$$\mathbb{P}\{X_{t+r} = i \mid X_t = j\}, \qquad \mathbb{P}\{X_{t+s} = j \mid X_t = i\}$$

A *communication class* is the collection of all such states that communicate with one another. The set of communication classes partition the state space into disjoint sets. If there exists only one communication class, then the Markov chain is said to be *irreducible*. In this paper, the state-space of queues comprises the non-negative integers and is irreducible.

Example 2.5. *Let $\{X_t\}$ be an irreducible, continuous-time Markov chain. Show that for each i, j and every $t > 0$*

$$\mathbb{P}\{X_t = j \mid X_0 = i\} > 0$$

Proof.
Fix i and j. Since $\{X_t\}$ is irreducible, there exists some time t such that

$$\mathbb{P}\{X_t = j | X_0 = i\} > 0$$

We want to show that this holds for all t. Consider the sequence of events that occur between i and j, and denote them as $k_1, \ldots, k_n$, where k_1 occurs immediately after i and k_n occurs immediately before j. It is clear that

$$q_{k_l, k_{l+1}} > 0 \qquad \text{for all } l$$

Or equivalently for some times $s < t$,

$$\mathbb{P}\{X_t = k_{l+1} \mid X_s = k_l\} > 0$$

Now pick any time $t > 0$. Note that there are potentially many paths from i to j, and $k_1, \ldots, k_n$ represents one such path. Therefore, there is a non-zero probability that i will go to j by this transition path. In notation,

$$\mathbb{P}\{X_t = j \mid X_0 = i\} \geq \prod_{l=1}^{n-1} \mathbb{P}\left\{X_{\frac{(l+1)t}{n}} = k_{l+1} | X_{\frac{lt}{n}} = k_l\right\} > 0$$

□

We now extend the result of the example to show how each continuous-time Markov chain induces a discrete-time Markov chain. Let $\{X_t\}_{t\geq 0}$ be an irreducible, continuous-time Markov chain. Let S be a countable state-space, and i, j be states in S. Let t_0, t_1, t_2 be distinct times with $t_0 < t_1 < t_2$. Define

$$p_{ij}(t_0, t_2) \equiv \mathbb{P}\{X_{t_2} = j \mid X_{t_0} = i\}$$

to be the probability of $\{X_t\}$ being in state j at time t_2 given that the process was in state i at time t_0. Then for each transition probability we can write

$$p_{ij}(t_0, t_2) = \sum_{j\in S} p_{ij}(t_0, t_1) p_{jk}(t_1, t_2)$$

Enumerate the sequence of states in order of occurrence:

$$\begin{aligned} n_0 &= 0 \\ n_1 &= \inf\{t : X_t \neq X_0\} \\ n_2 &= \inf\{t \geq n_1 : X_t \neq X_{n_1}\} \\ &\vdots \\ n_m &= \inf\{t \geq n_{m-1} : X_t \neq X_{n_{m-1}}\} \end{aligned}$$

then $\{X_n\} \equiv \{X_n : n \in \{n_0, n_1, \ldots\}\}$ forms a discrete-time Markov chain out of a continuous-time Markov chain. This process is known as *embedding*, and $\{X_n\}$ is called an *embedded Markov chain*. A similar procedure will be used later in Section 3.3 when we construct a random walk, a discrete-time process, out of a Brownian motion, a continuous-time process. Note that the embedding process causes us to lose information about the *holding times* of each state, the amount of time spent in a state before transitioning to another state. We therefore do not lose any further information by normalizing the time between states to some constant $c > 0$,

$$n_m - n_{m-1} = c \qquad \text{for all } \ m \geq 1$$

So after normalization, the transition-rate matrix for $\{X_n\}$ simply becomes $\mathbf{P}(c)$ as defined above.

2.2. **Birth-Death Processes.**

A *birth-death process* is a non-negative integer valued, continuous-time Markov chain. That is, a stochastic process, $\{N_t\}$, fulfilling the following two conditions

(1) $N_t \in \{0, 1, 2, 3, \ldots\}$
(2) as $s \to 0$, $(N_{t+s} - N_t) \in \{0, -1, 1\}$

The second condition states that for an infinitesimal change in time, the population of the system will either increase by one, decrease by one, or remain the same. A birth-death process is a system where the rate of exit out of and entrance into the system are known, and the collection of such processes forms a large family of stochastic processes. For instance, a Poisson process is one of the simplest birth-death processes where the "birth" rate is constant and the "death" rate is zero. In particular, a queuing process is a birth-death process where customer arrivals are "births" and service completions are "deaths".

2.3. **Queuing Models.**

Example 2.6 ($M/M/c$)**.** In this model, inter-arrival times and service times are exponentially distributed with rates λ and μ, respectively. In this queue, λ and μ are constant and independent of the current state. That is, $\lambda_N = \lambda$ and $\mu_N = \mu$ for all N. Service is received on a "first come, first served" basis, and c customers may be serviced at a time. The simplest case is when $c = 1$, where only one customer is serviced at a time. In an $M/M/c$ queue where customers are serviced as soon as they arrive, regardless of the population of the system, we set $c = \infty$. The $M/M/c$ model assumes that no customers leave the system between arrival and completion of service.

This paper focuses on the $M/M/1$ queue, but we introduce a few more models as examples of the breadth of situations that can be analyzed through the lens of queuing theory.

Example 2.7 ($G/G/c$)**.** This model is the same as the $M/M/c$ model except inter-arrival times and service times follow some general (represented by G), not necessarily exponential, distribution.

Example 2.8 ($G/G/c/K$)**.** This model is the same as the $G/G/c$ model but with an upper bound, K, on the number of customers that can occupy the system at a given time. This type of queue is called a *truncated queue*.

Example 2.9 (A/B/C/D/E)**.** The above queuing models are identified using *Kendall Notation*, a combination of letters and slashes. Each position in the notation corresponds to a certain queuing characteristic (e.g. inter-arrival time distribution), and the letter specifies the characteristic (e.g. exponentially distributed inter-arrival times). The following table covers a wide selection of queuing models that are easily be described using Kendall notation.

Characteristic	Symbol	Explanation
(A) Inter-arrival time distribution	M	Exponential
(B) Service time distribution	D	Deterministic
	E_k	Erlang type k
	G	General
(C) # of parallel servers	$1, 2, \dots, \infty$	
(D) Max. system capacity	$1, 2, \dots, \infty$	
(E) Order of service	$FCFS$	First come, first served
	$LCFS$	Last come, first served
	RSS	Random selection
	PR	Priority
	GD	General

The table above is by no means a definitive list of queuing characteristics. For instance, inter-arrival and service times may be state-dependent as introduced in the beginning of this section with rates λ_N and μ_N. Queues can also be cyclic in the sense that once a customer is serviced, the customer immediately returns to a position in line to await service once again.

2.4. **Properties of Queuing Models.**

The first two properties one may ask of a queue are the expected number of customers in line and the duration of time a customer must wait before being serviced. Let N_t be the number of customers in a $G/G/c$ queue at time t. If we let $p_n = \mathbb{P}\{N_t = n\}$, the expected population of the system is

$$\mathbb{E}[N_t] = \sum_{n=0}^{\infty} np_n$$

and the expected population of those in the queue (and not being serviced) is

$$\mathbb{E}[N_{t,q}] = \sum_{n=0}^{\infty}(n-c)p_n$$

Definition 2.10. The nth *waiting time* is the amount of time the nth customer spends waiting in line prior to entering service. We denote waiting time as the random variable $W_q^{(n)}$ and the total time a customer is in the system, including service time, as $W^{(n)}$. The nth customer is the nth overall customer to enter the system, not the customer in the nth position in line.

We now consider the waiting time of the nth customer to enter a queuing system where the initial population is zero, that is $N_0 = 0$. Let $S^{(n)}$ be the nth service time, and let $I^{(n)}$ be the nth inter-arrival time, i.e. the time between the $(n-1)$st customer and the nth customer arriving in the system. Define

$$U^{(n)} \equiv S^{(n)} - I^{(n)}$$

to be the time between the nth inter-arrival time and the nth service time.

Theorem 2.11 (Lindley's Equation)**.** *In a single-server queue where customers are serviced on a first-come first-served basis, the waiting time of the* $(n+1)$*th customer is recursively given by*

$$W_q^{(n+1)} = \max\left\{0, W_q^{(n)} + S^{(n)} - I^{(n)}\right\}$$

Proof.
Since the initial population of the system is zero, the first customer to arrive will be serviced immediately, so $W_q^{(1)} = 0$. The waiting time for the second customer will be the time it takes for the first customer to finish being serviced. If the first customer has already been serviced by the time the second customer arrives, then the waiting time for the second customer will be zero:

$$W_q^{(2)} = \max\left\{0, S^{(1)} - I^{(1)}\right\} = \max\left\{0, W_q^{(1)} + S^{(1)} - I^{(1)}\right\}$$

For $n \geq 3$ the waiting time of the nth customer is simply the waiting time of the $(n-1)$th customer and the amount of time it takes for the $(n-1)$th customer to be serviced, or zero if the $(n-1)$th customer has already been serviced by the time the nth customer enters the system. □

Lemma 2.12. *Under the same conditions as the preceding theorem,* $W_q^{(n)}$ *may be rewritten as*

$$W_q^{(n)} = \max\left\{U^{(1)} + \cdots + U^{(n-1)}, U^{(2)} + \cdots + U^{(n-1)}, \ldots, U^{(n-1)}, 0\right\}$$

Proof.
We prove the lemma by iterating over $W_q^{(n)}$.

$$\begin{aligned} W_q^{(n)} &= \max\left\{W_q^{(n-1)} + U^{(n-1)}, 0\right\} \\ &= \max\left\{\max\left\{W_q^{(n-2)} + U^{(n-2)}, 0\right\} + U^{(n-1)}, 0\right\} \\ &= \max\left\{\max\left\{W_q^{(n-2)} + U^{(n-2)} + U^{(n-1)}, U^{(n-1)}\right\}, 0\right\} \\ &= \max\left\{W_q^{(n-2)} + U^{(n-2)} + U^{(n-1)}, U^{(n-1)}, 0\right\} \end{aligned}$$

Continue this process in a recursive manner until we have

$$W_q^{(n)} = \max\left\{W_q^{(1)} + U^{(1)} + \cdots + U^{(n-1)}, U^{(2)} + \cdots + U^{(n-1)}, \ldots, U^{(n-1)}\right\},$$

The result of the lemma is now immediate since we have assumed that $W_q^{(1)} = 0$. □

The goal of this paper is to approximate $W_q^{(n)}$, the waiting time of the nth customer. From the preceding lemma, it is clear that the waiting time depends on the waiting times of all preceding customers in line. More precisely, the waiting time of the nth customer is the maximum of the partial sums of the decreasing sequence $\left\{U^{(n-i)}\right\}_{i=1}^{n}$, an i.i.d. sequence of random variables. Section 3 will show that this decreasing sequence is a random walk, and when n is sufficiently large this sequence can be approximated by Brownian motion. For such an approximation to be accurate, however, each $U^{(n-i)}$ must be sufficiently small small. This situation arises in queues that exhibit heavy traffic.

Definition 2.13. The *traffic intensity*, ρ, of a given queue is defined as the ratio of arrival times to service times. So for an $M/M/c$ queue, the traffic intensity is

$$\rho = \frac{\lambda}{c\mu}$$

where λ and μ are the arrival and service rate, respectively, of customers in the queue. For a single server $M/M/1$ queue, the traffic intensity is

$$\rho = \frac{\lambda}{\mu}$$

Remark 2.14. When ρ is close to zero, waiting times approach zero as queue length shortens and new customers are serviced immediately upon arrival. When ρ is greater than one, waiting times successively increase as the queue population "explodes" to infinity. The interesting case occurs when ρ is close to—but does not exceed—one. A system is said to be in *heavy traffic* when $\rho \in (1-\varepsilon, 1)$ and $\varepsilon > 0$ is small.

3. Brownian Motion

This section introduces the stochastic process of Brownian motion viewed as the limit of random walks. In this paper, we confine our study of Brownian motion to the one-dimensional case.

3.1. **Random Walks.**

Definition 3.1. A one-dimensional random walk is a stochastic process constructed as the sum of i.i.d random variables. That is, if $\{X_i\}$ is a sequence of i.i.d. random variables with $X_i \in \mathbb{R}$ (regarded as steps), then the sum of the first n such random variables,

$$R_n = \sum_{i=1}^{n} X_i$$

is the n-th value of a random walk. A random walk is said to be *symmetric* if $X_i \in \mathbb{R}$ is symmetrically distributed about zero. A random walk is called *simple* if $\mathbb{P}\{X_i = 1\} + \mathbb{P}\{X_i = -1\} = 1$.

Example 3.2. Figure 3 depicts a simple random walk simulated using the statistical computing software R with $\mathbb{P}\{X_i = -1\} = \mathbb{P}\{X_i = 1\} = \frac{1}{2}$, and 15 total steps. Taking $R_0 = X_0 = 0$, the values of X_i and R_i, are

$$\{X_i\}_{i=1}^{15} = \{1, -1, 1, -1, -1, -1, -1, 1, -1, -1, 1, 1, -1, 1, -1\}$$
$$\{R_i\}_{i=1}^{15} = \{0, 1, 0, 1, 0, -1, -2, -3, -2, -3, -4, -3, -2, -3, -2, -3\}$$

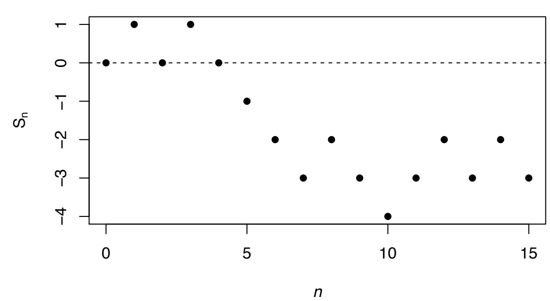

FIGURE 3.

An important property of random walks is that they exhibit *independent increments*, meaning that for any selection of positive integers, $t_1 < t_2 < \cdots < t_n$, the random variables

$$R_{t_2} - R_{t_1},\ \ldots,\ R_{t_n} - R_{t_{n-1}}$$

are independent. This property immediately follow from the fact that each increment is the sum of i.i.d random variables.

3.2. **Brownian Motion, Definition.**
We now introduce the continuous-time analog to a random walk—Brownian motion. A one-dimensional standard Brownian motion with variance parameter σ^2 is a real-valued process, $\{B_t\}_{t\geq 0}$, defined in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that has the following properties

(I) If $t_0 < t_1 < \cdots < t_n$, then $B_{t_0}, B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent.
(II) If $s, t \geq 0$, then
$$P(B_{s+t} - B_s \in A) = \int_A \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp(-x^2/2\sigma^2 t)$$
(III) With probability 1, $t \mapsto B_t$ is continuous.

The first two conditions may be concisely reworded as

(I) Independent increments
(II) Increment lengths are normally distributed with mean 0 and variance $\sigma^2 t$:
$$B_{s+t} - B_s \sim N(0, \sigma^2 t)$$

Remark 3.3.
Brownian motion is a collection of stochastic processes. When we say *a* Brownian motion or *the* Brownian motion, we are talking about a particular realization of Brownian motion in the space of continuous functions, $\mathcal{C}[0,T]$, where T is a stopping time.

The above definition establishes the necessary conditions for a stochastic process to be Brownian, but the existence of such a process requires further work. Through an application of the Kolmogorov extension theorem, existence is proved in [4]. However, such a proof is bereft of intuition for queueing theory and approximating waiting times. Instead, we take the existence of Brownian motion as given and show that Brownian motion can be constructed as a limit of random walks—the result of Donsker's Theorem. Brownian motion constructed in such a manner is called the standard Wiener process.

Before constructing Brownian motion, we must first establish what type of convergence we are discussing.

Definition 3.4. A sequence of random variables $\{X_i\}_{i=1}^{\infty}$ *converges almost surely* to a random variable X in the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (denoted $X_n \xrightarrow{a.s.} X$) if
$$\mathbb{P}\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} = 1$$

We aim to show almost sure convergence of a scaled random walk to a Brownian motion in $(\mathcal{C}[0,T], d)$, the metric space of continuous functions on the interval $[0,T]$ with d as the supremum norm:
$$d(f,g) = \sup_{t\in[0,T]} |f(t) - g(t)|$$

Note that random walks and Brownian motions are stochastic processes while d measures the distance between two continuous functions. In order to measure the distance between random walks and Brownian motion we instead consider the paths taken by these stochastic processes:
$$t \mapsto R_t \qquad \text{and} \qquad t \mapsto B_t$$

3.3. **Brownian Motion, Construction.** In this paper, the construction of Brownian motion comprises the following steps:

(1) Construct a simple random walk out of a Brownian motion path using Skorokhod embedding
(2) Linearly interpolate and scale the embedded random walk, creating a continuous function from $[0, n] \mapsto \mathbb{R}$
(3) Take the uniform limit of this scaled random walk to derive the original Brownian motion path

The paper presents an alternative formulation to the canonical construction of Brownian motion which involves defining Brownian motion on the dyadic rationals and using continuity to extend the definition to the reals. To motivate this particular construction, we will proceed out of order; we will begin with step (2), backtrack to step (1), and conclude by stringing together steps (1) through (3) in subsection 3.3.3 (Strong Approximation).

3.3.1. *Scaled Random Walks.*
Let $\{R_m\}$ be a random walk where $\mathbb{E}[X_i] = 0$, $\text{Var}\{X_i\} = \sigma^2$, and

$$R_m = \sum_{i=1}^{m} X_i.$$

Definition 3.5. A *scaled random walk* with scaling parameter n is a real valued random process, $\left\{R^{(n)}\right\}$, defined on $t \geq 0$. For t such that nt is an integer,

$$R_t^{(n)} = \frac{1}{\sqrt{n\sigma^2}} R_{nt}$$

For t such that nt is not an integer, $S^{(n)}(t)$ is defined by linear interpolation:

$$R_t^{(n)} = \frac{1}{\sqrt{n\sigma^2}} \left[R_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor) \left(R_{\lceil nt \rceil} - R_{\lfloor nt \rfloor} \right) \right]$$

This method of scaling operates on two aspects of random walks; as n increases, the time between state transitions decreases and increment lengths decrease. Let $t_0 = 0$ and $t_k = \inf_t \{kt \in \mathbb{Z}, nt > nt_{(k-1)}\}$, then as $n \to \infty$

$$t_k - t_{(k-1)} \to 0 \qquad \text{and} \qquad R_{t_k}^{(n)} - R_{t_{(k-1)}}^{(n)} \to 0$$

As a result, $R^{(n)}(t)$ is continuous with probability 1 on $t \geq 0$ for all n. The scaling factor $\frac{1}{\sqrt{n\sigma^2}}$ ensures constant variance for a given t:

$$\text{Var}\left\{\frac{1}{\sqrt{n\sigma^2}} R_{nt}\right\} = \text{Var}\left\{\frac{1}{\sqrt{n\sigma^2}} \sum_{i=1}^{nt} X_i\right\} = \left(\frac{1}{\sqrt{n\sigma^2}}\right)^2 nt\text{Var}\{X_i\} = t\frac{\text{Var}\left\{X_i\right\}}{\sigma^2} = t$$

By observing the limiting behavior of $R_t^{(n)}$ (i.e. as $n \to \infty$), we establish a construction of Brownian motion.

3.3.2. *Skorokhod Embedding.*
To construct Brownian motion out of random walks, we must first construct random walks out of Brownian motion. Structuring the proof in this matter is not without reason; there are uncountably many realizations of Brownian motion, so we want to ensure that the random walks to which we are taking the limit do, in fact, converge to the Brownian motion of interest. Such a procedure of deriving a random walk from a Brownian motion is called *Skorokhod embedding.* The procedure is given in the following pseudocode:

```
Set t_0 = 0 and n = 0.  Let T be a stopping time of Brownian motion {B_t}
so that {B_t} is defined on [0,T].  Let R_0 = 0 be the initial value
of the Skorokhod embedding.
While t_n + t ≤ T{
  Initialize t = 0.
  While |B_t − B_{t_n}| ≤ 1 {
    Continue B_t until |B_t − B_{t_n}| = 1 for the first time.
    Set t_(n+1) = inf{t : |B_t − B_{t_n}| = 1}.
    If B_{t_(n+1)} − B_{t_n} = 1, then set X_(n+1) = 1.
    Otherwise, B_{t_(n+1)} − B_{t_n} = −1, so set X_(n+1) = −1.
    Let the (n+1)th value of the Skorokhod embedding be
    R_(n+1) = R_n + X_{n+1}.
    Increment to the next step of the inner while loop:  n = n + 1.
  }
}
```

The resulting sequence $\{R_n\}_{n=1,2,\dots}$ is the Skorokhod embedding of $\{B_t\}_{t\in[0,T]}$.

Theorem 3.6. *Let B_t be a standard Brownian motion. Set $t_0 = 0$ and let t_n be the stopping time where $|B_t| = 1$ for the nth time:*

$$t_n \equiv \inf\{t \geq 0 : |B_t - B_{t_n}| = 1\}$$

If we define $R_n \equiv B_{t_n}$, then R_n is a simple random walk.

Proof.
To prove that $B_0, B_{t_1}, B_{t_2}, \dots$ is a simple random walk, we merely need to show that

(1) $\mathbb{P}\{B_{t_{n+1}} - B_{t_n} = 1\} = \mathbb{P}\{B_{t_{n+1}} - B_{t_n} = -1\} = \frac{1}{2}$ for all $n \in \mathbb{N}$
(2) $t_{n+1} - t_n$ are i.i.d. random variables

The first item follows from the symmetric property of Brownian increments, property (II) in the definition of Brownian motion: $B_{s+t} - B_s \sim N(0, \sigma^2 t)$.

The second item follows from the strong Markov property of Brownian motion, which states that if τ is a stopping time with respect to $\{B_t\}$, then $\{B_t - B_\tau\}$ is also a Brownian motion. We do not include a rigorous treatment of stopping times or a proof of the strong Markov property. □

We now state a theorem that places a probabilistic bound on the difference between a Brownian motion and its Skorokhod embedding.

Theorem 3.7. *Define* Θ *to be the maximum distance between a Brownian motion and its Skorokhod embedding:*

$$\Theta(B, R; T) = \max_{0 \le t \le T} \{|B_t - R_t|\}$$

Note that Θ *is equivalent to the distance between* $\{B_t\}$ *and* $\{R_t\}$ *in the space of continuous functions on the time interval* $[0, T]$ *equipped with the supremum norm. There exist* $c, a \in [0, \infty)$ *such that for all* $r \le n^{1/4}$ *and all integers* $n \ge 3$

$$\mathbb{P}\left\{\Theta(B, R; T) \ge r n^{1/4} \sqrt{\log n}\right\} \le c e^{-ar}$$

Proof.
It will suffice to prove the theorem for $r \ge 9c^2$ where c is the constant such that

$$\mathbb{P}\left\{\operatorname{osc}(B; \delta, T) > r\sqrt{\delta \log(1/\delta)}\right\} \le cT\delta^{(r/c)^2}$$

and

$$\operatorname{osc}(B; \delta, T) = \sup\{|B_t - B_s| : s, t \in D; s, t \in [0, T]; |s - t| \le \delta\}$$

is the *oscillation* of B_t restricted to $t \in D$, the dyadic rationals. The proof that such a c exists can be found in [3], page 68, where it is also shown that for $n \in \mathbb{N}$,

$$\Theta(B, S; n) \le 1 + \operatorname{osc}(B; 1, n) + \max\{|B_j - B_{\tau_j}| : j = 1, \ldots, n\}$$

Now suppose $9c^2 \le r \le n^{1/4}$. If $|B_n - B_{\tau_n}|$ is large, then either $|n - \tau_n|$ is large or the oscillation of B is large. Consider the three events:

(1) $\left\{\Theta(B, R; n) \ge r n^{1/4} \sqrt{\log n}\right\}$
(2) $\left\{\operatorname{osc}(B; r\sqrt{n}, 2n) \ge (r/3) n^{1/4} \sqrt{\log n}\right\}$
(3) $\{\max_{1 \le j \le n} |\tau_j - j| \ge r\sqrt{n}\}$

From above, we know that event (1) is contained in the union of events (2) and (3). As a consequence, we need only prove the result of the theorem for events (2) and (3). We first tackle event (2).

$$\begin{aligned}
\mathbb{P}&\left\{\operatorname{osc}(B; r\sqrt{n}, 2n) > (r/3) n^{(1/4)} \sqrt{\log n}\right\} \le 3\mathbb{P}\left\{\operatorname{osc}(B; r\sqrt{n}, n) > (r/3) n^{(1/4)} \sqrt{\log n}\right\} \\
&= 3\mathbb{P}\left\{\operatorname{osc}(B; r n^{-1/2}) > (r/3) n^{-(1/4)} \sqrt{\log n}\right\} \\
&\le 3\mathbb{P}\left\{\operatorname{osc}(B; r n^{-1/2}) > (\sqrt{r}/3) \sqrt{r n^{-1/2} \log\left(n^{1/2}/r\right)}\right\}
\end{aligned}$$

If $\sqrt{r}/3 \ge c$ and $r \le n^{1/4}$, we can conclude that there exist c and a such that

$$\mathbb{P}\left\{\operatorname{osc}(B; r n^{-1/2}) > (\sqrt{r}/3) \sqrt{r n^{-1/2} \log(n^{1/2})}\right\} \le c e^{-ar \log n}$$

For event (3), we refer to a proof on page 266 in the appendix of [3] to show that there exist c, a such that

$$\mathbb{P}\left\{\max_{1 \le j \le n} |\tau_j - j| > r\sqrt{n}\right\} \le c e^{-ar^2}$$

□

3.3.3. *Strong Approximation.*
With preliminaries in order, we now show that Brownian motion can be thought of as a limit of simple random walks, a construction known as the *strong approximation of Brownian motion.* Let $\{B_t\}$ be a standard Brownian motion with variance parameter 1 as defined in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\{B_t^{(n)}\}$ be the scaled Brownian motion:

$$B_t^{(n)} = \frac{1}{\sqrt{n}} B_{nt}$$

From $\{B^{(n)}\}$, we use the Skorokhod embedding to derive the simple random walk, $\{S^{(n)}\}$. And from $\{S^{(n)}\}$ we derive the scaled simple random walk, $\{R^{(n)}\}$:

$$R_t^{(n)} = \frac{1}{\sqrt{n}} S_{nt}^{(n)}$$

From Theorem 3.7 we know that there exists $c, a \in [0, \infty)$ such that for all positive integers T,

$$\mathbb{P}\left\{\max_{0 \le t \le Tn} \left| S_t^{(n)} - B_t^{(n)} \right| \ge cr(Tn)^{1/4} \sqrt{\log (Tn)} \right\} \le ce^{-ar}$$

multiplying by $\frac{1}{\sqrt{n}}$, this becomes

$$\mathbb{P}\left\{\max_{0 \le t \le T} \left| R_t^{(n)} - B_t \right| \ge crT^{1/4} n^{-1/4} \sqrt{\log (Tn)} \right\} \le ce^{-ar}$$

Letting $r = c \log n$ where c is sufficiently large for the given T, the inequality becomes

$$\mathbb{P}\left\{\max_{0 \le t \le T} \left| R_t^{(n)} - B_t \right| \ge cn^{-1/4} \log^{3/2} n \right\} \le \frac{c}{n^2}$$

To proceed, we require the use of the Borel-Cantelli lemma, the result of which is provided below.

Lemma 3.8 (Borel-Cantelli)**.**
If A_n is a sequence of events in Ω and $\sum_{i=1}^{\infty} \mathbb{P}\{A_n\} < \infty$ then

$$\mathbb{P}\{A_n \text{ } i.o.\} = 0$$

where $\{A_n \text{ } i.o.\} = \{\limsup_{n\to\infty} A_n\}$ is the event that infinitely many A_n occur.

If we set

$$A_n = \max_{0 \le t \le T} \left| R_t^{(n)} - B_t \right| \ge cn^{-1/4} \log^{3/2} n$$

we can then apply the Borel-Cantelli lemma to the above inequality to conclude that

$$\max_{0 \le t \le T} \left| R_t^{(n)} - B_t \right| \le cn^{-1/4} \log^{3/2} n$$

with probability one for all n sufficiently large. Since

$$\lim_{n\to\infty} cn^{-1/4} \log^{3/2} n = 0$$

we conclude that

$$R_t^{(n)} \xrightarrow{a.s.} B_t \quad \text{in } \mathcal{C}[0, T]$$

The following theorem condenses the results of this section up to this point.

Theorem 3.9 (Donsker's Theorem)**.** *Let $\{X_i\}$ be a sequence of i.i.d random variables with mean 0 and variance σ^2. If $R_t^{(n)}$ is a scaled random walk as defined above, then $\{R_t^{(n)} \Rightarrow B_t\}$ where B_t is a standard Brownian motion. This result is also known as the functional central limit theorem.*

3.4. **Brownian Motion, Variants.**

The preceding construction of Brownian motion will prove particularly useful in our approximation of heavy traffic waiting times. However, this process requires further modification. Recall that when traffic intensity, ρ, is slightly less than 1, customers are being serviced at a slightly faster rate then they are entering the queue. Thus, given enough time, traffic will eventually "clear out" due to faster service times. However, this particular construction of Brownian motion has increments of mean zero,

$$\mathbb{E}[B_{s+t} - B_s] = 0$$

meaning that the approximate number of customers in line would—on average—remain unchanged regardless of how long the line is observed. Such a property is called the martingale property. Instead, we want a process that exhibits the *super*martingale property:

$$\mathbb{E}[B_{s+t} - B_s] < 0$$

Definition 3.10. Let $\mathcal{F}_n$ be an increasing sequence of σ-fields ($\mathcal{F}_n$ is called a *filtration*). Let $\{X_n\}$ be a sequence of random variables with finite mean and $X_n \in \mathcal{F}_n$ for all n. Martingales, supermartingales, and submartingales are then defined as follows for all n

$$\begin{aligned}
\textbf{Martingale}: &\quad \mathbb{E}\left[X_{n+1} \mid \mathcal{F}_n\right] = X_n \\
\textbf{Supermartingale}: &\quad \mathbb{E}\left[X_{n+1} \mid \mathcal{F}_n\right] \leq X_n \\
\textbf{Submartingale}: &\quad \mathbb{E}\left[X_{n+1} \mid \mathcal{F}_n\right] \geq X_n
\end{aligned}$$

In words, a sequence that we expect to remain the same over time is a martingale, a sequence we expect to increase is a submartingale, and a sequence we expect to decrease is a supermartingale. To approximate heavy traffic, we want a Brownian motion-like process that exhibits the supermartingale property.

Definition 3.11. A Brownian motion with drift parameter α and variance parameter σ^2, $\{B_t^\alpha\}$, is defined by the following properties:

1. Independent increments
2. Increment lengths are normally distributed with mean αt and variance $\sigma^2 t$
3. the path $t \mapsto B_t^\alpha$ is continuous

Note that these properties are the same as standard Brownian motion except increment lengths now have non-zero means. Since we have already shown the existence of standard Brownian motion, the construction of Brownian motion with drift is now easy.

Lemma 3.12. *Let $\{B_t\}$ be a standard Brownian motion with variance parameter σ^2, then $B_t^\alpha = B_t + \alpha t$ is Brownian motion with drift parameter α and variance parameter σ^2. Moreover, B_t^α fulfills the supermartingale property when $\alpha < 0$.*

Proof.
Independence of increments follows directly from independence of increments for B_t and the fact that αt is a constant for a given t. To prove that B_t^α has stationary increments, we have to show that the distribution of increments $B_{s+t}^\alpha - B_s^\alpha$ depends only the time interval t, and not the absolute time. Recall from the definition of B_t that $B_{s+t} - B_t$ are normal random variables with mean 0 and variance t for $s > 0$.

$$\begin{aligned}\mathbb{E}\left[B_{s+t}^\alpha - B_s^\alpha\right] &= \mathbb{E}\left[B_{s+t}^\alpha\right] - \mathbb{E}\left[B_s^\alpha\right] \\ &= \mathbb{E}\left[B_{s+t}^\alpha + \alpha(s+t)\right] - \mathbb{E}\left[B_s + \alpha s\right] \\ &= \alpha(s+t) - \alpha s \\ &= \alpha t\end{aligned}$$

$$\begin{aligned}\mathrm{Var}\left\{B_{s+t}^\alpha - B_s^\alpha\right\} &= \mathrm{Var}\left\{B_{s+t} - B_s + \alpha t\right\} \\ &= \mathrm{Var}\left\{B_{s+t} - B_s\right\} \\ &= \sigma^2 t\end{aligned}$$

And since for fixed t, αt is just a constant, we use the property of normal random variables to conclude that

$$B_{s+t}^\alpha - B_s^\alpha \sim N(\alpha t, \sigma^2 t) \quad \text{for all} \quad s > 0$$

Continuity of B_t^α directly follows from continuity of B_t. We have shown that B_t^α constructed as the sum of standard Brownian motion and drift αt is indeed Brownian motion with drift.

□

Brownian motion with negative drift provides a better approximation of the waiting-time of customers in a queue than does standard Brownian motion. However, the unboundedness of Brownian motion suggests that waiting times could perhaps be negative. In order to circumvent this issue, we instead use *reflected Brownian motion with drift.* Reflected Brownian motion acts much the same as Brownian motion except on a given boundary off of which the Brownian motion is—unsurprisingly—reflected. In the one-dimensional case, the boundaries may be any interval on the reals that provide a lower and upper bound to the Brownian motion. For our purposes, we only use a single boundary, zero, that gives a lower bound to Brownian motion. That is, for all $t \geq 0$, we want $B_t^\alpha \geq 0$.

Definition 3.13. Let B_t^α be a standard Brownian motion with drift parameter α, and let M_t^α be the running maximum of B_t^α:

$$M_t^\alpha = \sup_{0 \leq s \leq t} B_s^\alpha$$

Then *reflected Brownian motion* with drift parameter $-\alpha$ and boundary 0 defined on $\mathbb{R}^+$ can be constructed as follows:

$$R_t^{\alpha,0} \equiv \max\{0, M_t^\alpha\} - B_t^\alpha$$

Note that $\max\{0, S_t^\alpha\} \geq 0$ and $M_t^\alpha \geq B_t^\alpha$, $B_t^{\alpha,0} \geq 0$ for all t. Most importantly, the local behavior of $B_t^{\alpha,0}$ is exactly like Brownian motion with drift parameter $-\alpha$ since $\max\{0, S_t^\alpha\}$ is simply a constant when $B_t^\alpha \neq 0$.

We have now provided sufficient modifications of Brownian motion, and we next move on—what we will ultimately use to approximate heavy traffic waiting times. The next example provides us with the distribution of the supremum of a Brownian motion with negative drift.

Example 3.14.
Let B_t be a Brownian motion with drift $\alpha < 0$ and variance parameter σ^2. Define

$$M_t = \sup_{0 \le s \le t} B_s$$

to be the running maximum, then

$$M_\infty \sim \exp\left(\frac{2|\alpha|}{\sigma^2}\right)$$

that is

$$\mathbb{P}\{W \ge w\} = e^{-\frac{-2|\alpha|}{\sigma^2}w}, \ w \ge 0$$

$$\mathbb{E}\left[M_\infty\right] = \frac{\sigma^2}{2|\alpha|}$$

Proof.
for constants $a, b > 0$ Let $T(-a, b)$ be the first time that $B - t$ hits $-a$ or b:

$$T(-a, b) = \inf\{t : B_t = -a \ \text{ or } \ B_t = b\}$$

In [7], it is proved that

$$\mathbb{P}\left\{B_{T(-a,b)=b}\right\} = \frac{\exp(2\alpha a/\sigma^2) - 1}{\exp(2\alpha a/\sigma^2) - \exp(-2\alpha b/\sigma^2)}$$

Since $\alpha < 0$, $\exp(2\alpha a/\sigma^2) \to 0$, we have

$$\lim_{a\to\infty} \mathbb{P}\{B_{T(-a,b)} = b\} = \lim_{a\to\infty} \frac{\exp(2\alpha a/\sigma^2) - 1}{\exp(2\alpha a/\sigma^2) - \exp(-2\alpha b/\sigma^2)} = \exp(2\alpha b/\sigma^2)$$

The left-hand side becomes the probability that the process will reach b somewhere along its path (i.e. that the maximum of the process exceeds b somewhere along its path). Therefore, we have

$$\mathbb{P}\{W \ge b\} = \exp(2\alpha b/\sigma^2) = \exp\{-2|\alpha|b/\sigma^2\}$$

□

This example shows us that the supremum of a Brownian motion observed "forever" (i.e. on the interval $[0, \infty)$) is exponentially distributed with rate $2|\alpha|/\sigma^2$.

4. Heavy Traffic Approximation

We now have all we need to provide a heuristic argument for the approximation of single-servers queue in heavy traffic using Brownian Motion. For a rigorous, albeit opaque, treatment of this approximation, see [8], the seminal work on this subject by Kingman.

Theorem 4.1. *The waiting time of the nth customer in an $M/M/1$ queue can be approximated by the supremum of reflected Brownian motion with negative drift and boundary at 0.*

Proof.
Recall that an $M/M/1$ queue is a process where the number of customer arrivals follows a poisson process with rate λ and service times are exponentially distributed with rate μ. From Lindley's equation in Section 2, we can write the waiting time for the nth customer as

$$W_q^{(n)} = \max\left\{U^{(1)} + \cdots + U^{(n-1)}, U^{(2)} + \cdots + U^{(n-1)}, \ldots, U^{(n-1)}, 0\right\}$$

where $U^{(n)} = S^{(n)} - I^{(n)}$ is the difference between the nth service time and the nth inter-arrival time. Define the partial sum $P_k^{(n)}$ to be zero when $k = 0$ and

$$P_k^{(n)} \equiv \sum_{i=1}^{k} U^{(n-i)} \quad \text{for} \quad k \geq 1$$

We can then rewrite the waiting time of the nth customer as

$$W_q^{(n)} = \max_{0 \leq k \leq (n-1)} P_k^{(n)}$$

The fact that $P_k^{(n)}$ is the kth value of a random walk follows immediately from $U^{(1)}, \ldots, U^{(n-1)}$ being i.i.d random variables. Let α and σ^2 be the expectation and variance of $U^{(i)}$. We calculate α and σ^2 by first noting that $S^{(i)}$ and $I^{(i)}$ are exponentially distributed with rates μ and λ. Recall that $\rho = \lambda/\mu$ is the traffic intensity as defined in the first section of this paper.

$$\alpha = \mathbb{E}\left[U^{(i)}\right] = \mathbb{E}\left[S^{(i)}\right] - \mathbb{E}\left[I^{(i)}\right] = \frac{1}{\mu} - \frac{1}{\lambda} = \frac{\rho - 1}{\lambda}$$
$$\sigma^2 = \mathrm{Var}\left\{U^{(i)}\right\} = \mathrm{Var}\left\{S^{(i)}\right\} + \mathrm{Var}\left\{I^{(i)}\right\} = \frac{1}{\mu^2} + \frac{1}{\lambda^2}$$

We next compute the expectation and variance of $P_k^{(n)}$, using the fact that $U^{(i)}$ are i.i.d. random variables.

$$\mathbb{E}\left[P_k^{(n)}\right] = \sum_{i=1}^{k} \mathbb{E}\left[U^{(n-i)}\right] = k\alpha$$
$$\mathrm{Var}\left\{P_k^{(n)}\right\} = \sum_{i=1}^{k} \mathrm{Var}\left\{U^{(n-i)}\right\} = k\sigma^2$$

For a finite n, $W_q^{(n)}$ is therefore the maximum value of a random walk with n steps.

We now show that $\{P_k^n\}$ can be approximated by Brownian motion with drift parameter α and variance parameter σ^2 when n is sufficiently large. We restate the properties of such a Brownian motion, $\{B_t\}$, below:

(1) $B_0 = 0$
(2) B_t has stationary and independent increments
(3) $B_t \sim N(\alpha t, \sigma^2 t)$
(4) $t \mapsto B_t$ is continuous with probability one

The first property is clear as $P_0^{(n)} = 0$ for all n. The second property follows from the fact that $\{U^{(i)}\}$ are i.i.d random variables. The third property is a result of the central limit theorem for i.i.d. sequences, the result of which is provided below.

Theorem 4.2. *Let* $\{X_i\}_{i=1}^{\infty}$ *be a sequence of i.i.d. random variables with* $\mathbb{E}[X_i] = \mu$ *and* $\mathrm{Var}\{X_i\} = \sigma^2$*. If* $S_n = X_1 + \cdots X_n$ *then*

$$\sqrt{n}S_n \overset{d}{\Rightarrow} N(\mu, \sigma^2)$$

In words, as $n \to \infty$*,* $\sqrt{n}S_n$ *converges in distribution to a normal random variable with mean* μ *and variance* σ^2*.*

When k is sufficiently large, the central limit theorem implies that

$$P_k^{\infty} \equiv \lim_{n\to\infty} P_k^{(n)}$$

is approximately normally distributed with mean $k\alpha$ and variance $k\sigma^2$. In practice, taking $n \to \infty$ represents a queue that has been operating continuously with a large amount of customers.

Property (4) is where we make use of the heavy traffic assumption. Since ρ is very close to one, $|\alpha|$ is small, so increments of $\{P_k^{(n)}\}$ are, on average, small. Small increments ensure that continuity approximately holds. Therefore, P_k^{∞} (approximately) fulfills properties (1)-(4) above, so the waiting time of the nth customer can be approximated as the supremum of Brownian motion with negative drift. □

Before concluding with a computation of the expected waiting time, we take note of the conditions under which the above approximation is accurate.

(1) Service and inter-arrival times are exponentially distributed with known, constant rates.
(2) $\rho = \lambda/\mu$ is very close to, but does not exceed, one.
(3) Many other customers preceded the nth customer in line, the customer whose waiting time we wish to approximate. In common usage, the result of the central limit theorem is used when $n \geq 30$.

We conclude with a computation of the expected waiting time for large n. Example 3.14 proved that the supremum of Brownian motion with negative drift is exponentially distributed, so we conclude that $W_q^{(n)}$ can be approximated by an exponential distribution with rate $2|\alpha|/\sigma^2$ when n is at least greater than 30.

$$\mathbb{E}\left[W_q^{(n)}\right] \approx \frac{\sigma^2}{2|\alpha|} = \frac{1}{2}\frac{\frac{1}{\mu^2} + \frac{1}{\lambda^2}}{\left|\frac{\rho-1}{\lambda}\right|}$$

Acknowledgements

I would like to extend my sincerest gratitude to Peter May for providing me and my peers with the opportunity to participate in the University of Chicago Math REU. My interest in pure math now heightened and my mathematical maturity developed, I consider myself extremely fortunate to have taken part in this program. I must also thank Gregory Lawler, whose lectures and texts have proved invaluable in the writing of this paper. Finally, I am forever grateful to my mentor, Yucheng Deng, for guiding me throughout the entire writing process. I attribute much of the fruit borne by my exploration of Brownian motion and queues to Yucheng's motivation and insight.

References

1. Gross, Donald and Harris, Carl M. *Fundamentals of Queueing Theory.* John Wiley and Sons, Inc. 2008.
2. Lawler, Gregory F. *An Introduction to Stochastic Processes.* Chapman and Hall. 2006.
3. Lawler, Gregory F. and Limic, Vlada. *Random Walk: A Modern Introduction* https://www.math.uchicago.edu/∼lawler/srwbook.pdf
4. Durrett, Richard. *Probability: Theory and Examples.* Cambridge University Press. 2010.
5. Billingsley, Patrick. *Convergence of Probability Measures*, John Wiley and Sons, Inc. 2009.
6. Peskir, Goran. "On Reflecting Brownian Motion with Drift", Proc. Symp. Stoch. Syst. (Osaka, 2005), ISCIE Kyoto, 2006, (1-5), Research Report No. 3, 2005, Probab. Statist. Group Manchester (8 pp).
7. Huang, Yibi. Lecture notes for STAT 253/317, "Introduction to Probability Models", University of Chicago. 2014. https://galton.uchicago.edu/∼yibi/teaching/stat317/2014/
8. Kingman, J. F. C. "On Queues in Heavy Traffic". Journal of the Royal Statistical Society. Accessed on JSTOR 2984229. 1962.

Department of Mathematics, University of Chicago
Email address: harveybarnhard@uchicago.edu