Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Strong Embeddings for Transitory Queueing Models

Prakash Chakraborty, Harsha Honnappa

To cite this article:
Prakash Chakraborty, Harsha Honnappa (2021) Strong Embeddings for Transitory Queueing Models. Mathematics of Operations Research

Published online in Articles in Advance 11 Nov 2021

. https://doi.org/10.1287/moor.2021.1158

informs
http://pubsonline.informs.org/journal/moor

MATHEMATICS OF OPERATIONS RESEARCH
Articles in Advance, pp. 1–34
ISSN 0364-765X (print), ISSN 1526-5471 (online)

Strong Embeddings for Transitory Queueing Models

Prakash Chakraborty,[a] **Harsha Honnappa**[b]

[a] Department of Statistics, Purdue University, West Lafayette, Indiana 47906; [b] School of Industrial Engineering, Purdue University, West Lafayette, Indiana 47906

Contact: chakra15@purdue.edu, https://orcid.org/0000-0003-4958-7888 (PC); honnappa@purdue.edu, https://orcid.org/0000-0002-0834-054X (HH)

Received: June 28, 2019
Accepted: November 30, 2020
Published Online in Articles in Advance: November 11, 2021

MSC2000 Subject Classification: Primary: 60F17, 60K25; secondary: 90B22
OR/MS Subject Classification: Primary: Queues; secondary: nonstationary, limit theorems

https://doi.org/10.1287/moor.2021.1158

Abstract. In this paper, we establish strong embedding theorems, in the sense of the Komlós-Major-Tusnády framework, for the performance metrics of a general class of transitory queueing models of nonstationary queueing systems. The nonstationary and non-Markovian nature of these models makes the computation of performance metrics hard. The strong embeddings yield error bounds on sample path approximations by diffusion processes in the form of functional strong approximation theorems.

Funding: The authors are supported by the National Science Foundation [Grant CMMI/1636069].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/moor.2021.1158.

Keywords: nonstationary models • transitory models • strong approximations • queueing theory

1. Introduction

In this paper, we establish strong embedding theorems, in the sense of the Komlós-Major-Tusnády (KMT) framework, for the performance metrics of a general class of *transitory queueing models* (Bet et al. [3], Honnappa et al. [22]). Transitory queueing models assume a large but finite population of customers arrive at the system over some time horizon. Examples of such systems include hospital surgery departments and clinics, subscription-based services such as video and game streaming, app-based ride-sharing/transportation, and delivery services. In each of these cases, the pool of potential customers is known to the service provider a priori, because of appointments that are handed out to patients ahead of time in the healthcare examples and subscriptions/sign-ons in the case of streaming and app-based services. Of course, not all the potential customers may turn up for service. Nonetheless, the finite pool implies that transitory models are nonstationary, both in the sense that they are *purely* transient in nature and because the model parameters can vary temporally. This makes the computation of the performance metrics rather difficult. Consequently, we seek to approximate the performance metric stochastic processes by simpler ones that capture their most vital temporal features. The strong embedding theorems in this paper yield probabilistic error bounds between the discrete-event performance metric processes and simpler diffusion process approximations in terms of the population size n. Our results will provide practitioners and engineers with a turn-key analysis yielding error bounds in terms of the population size so that diffusion approximations can be confidently used in their performance analysis, system design, and control problems.

Strong approximations were first used for studying time homogeneous queueing models in Rosenkrantz [35] (see the survey paper of Glynn [14] for a comprehensive introduction to the use of strong approximations to $G/G/1$ queueing models in heavy traffic). In general, the transient analysis of queueing models is rather complicated, and therefore, a number of approximations have been developed in appropriate scaling regimes, typically by certain types of reflected diffusion processes (Chen and Yao [6]). As queueing models can be expressed (approximately) as functionals of random walks, strong approximations are particularly useful in this application context because the driving random walks can be directly replaced by approximating Brownian motion processes. Strong approximation analysis yields rates of convergence and, consequently, rigorous justification of the heavy-traffic approximation on a sample path basis. Our results provide similar insights for a class of nonstationary queueing models under a *population acceleration* scaling framework.

We assume that the offered load to the queueing system is time-of-day dependent and displays long-range correlations. The modeling and analysis of transitory queues is, in general, quite complicated, and we operate under the simplifying assumption that the time-of-day and correlative effects are present solely in the traffic characteristics and that the service requirements of the arriving customers are independent and identically distributed. We propose a two-variable traffic model labeled RS(G, p) wherein we impute the ith arriving customer (out of n) with the random variable tuple (T_i, ζ_i), where T_i takes values in $[0,\infty)$ and ζ_i is binary, taking the values 0 or 1. Th term T_i models the (potential) arrival epoch of customer i, and $\sum_{i=1}^n \zeta_i$ is the number of customers who actually enter the queue; here RS stands for randomly scattered. We assume that the tuples are independent and identically distributed over the population and that T_1 follows a distribution G and $\mathbb{E}\zeta_1 = p$. We also assume that the service requirements are generally distributed with finite moment generating function in the neighborhood of zero and independent of the tuple. Consequently, we label this the RS$(G,p)/G/1$ queue. We make the following contributions in this paper:

1. We prove functional strong approximation theorems (FSATs) for the workload and queue length performance metric processes of the RS$(G,p)/G/1$ queue in Theorems 2 and 3 (respectively). These FSATs yield sample path error bounds between the performance metrics and nonstationary reflected Brownian Bridge processes. The nonstationary Brownian Bridge processes capture the fact that the offered load is time-of-day dependent and has long range correlations.
2. The proofs of the FSATs are consequences of nonasymptotic functional strong embedding theorems (FSETs) proved for the RS(G, p) traffic process in Proposition 8, the workload process in Proposition 10, and the queue length process in Proposition 11 that yield exponential probability bounds as a function of the population size.
3. Our proof of the nonasymptotic probabilistic bounds require Dvoretzky-Kiefer-Wolfowitz (DKW) (Dvoretzky et al. [10]) style inequalities for Brownian motion randomly time-changed by a stochastic pure jump process, proved in Proposition 3. This generalized DKW inequality may be of independent interest and useful in proving bounds for other types of models. As a consequence, we obtain improved convergence rates for the diffusion approximations than what is obtained in Mandelbaum et al. [30] for nonhomogeneous Markovian networks.

1.1. Commentary on Main Results

Our analysis leans on strong approximations for empirical processes and random walks (Csörgo and Révész [7]) but also requires substantial innovation. The FSAT in Theorem 2 is a consequence of Proposition 10, where we prove a strong embedding result for the workload process of a RS$(G,p)/G/1$ queue, under the assumption that the service times possess finite moment generating functions in a neighborhood around zero. We show that, with high probability, for a given fixed population size n the sample paths of the workload process can be approximated by those of a reflected Brownian bridge process with time dependent drift and diffusion coefficients. Indeed, we show that the convergence rate is $O(n^{1/4}\sqrt{\log n})$. Next, the FSAT to the queue length process of the RS$(G,p)/G/1$ queue in Theorem 3 follows from Proposition 11. Paralleling the result in Proposition 10, we show that the approximating process is a reflected Brownian bridge process with time inhomogeneous drift and diffusion coefficients. However, the drift and diffusion coefficients are scaled versions of those observed in Proposition 10. We note that the analysis of the queue length strong embedding theorem is significantly more involved. The proofs of these results requires a careful construction of a DKW-style inequality for a time-changed Brownian motion process, which we did not find in the literature (Proposition 3). Again, we show that the convergence rate for the queue length process is $O(n^{1/4}\sqrt{\log n})$.

1.2. Relation with Prior Transitory Analyses

The RS(G,p) model affords flexibility for modeling service systems where the pool of potential customers is known a priori. This typically includes systems where customers subscribe to the service ahead of time; for example, clinics and surgical departments in hospitals where patients are given appointment times, video and game streaming services with subscribing customers, or ridesharing and food delivery services where the pool of customers are those who have downloaded the smartphone app. In each of these cases, the service provider has knowledge of who the potential customers are, but not all customers will use the service on a given day. The randomized arrivals in the RS(G, p) model accounts for this effect, which is ignored in the $\Delta_{(i)}/G/1$ model where $\sum_{i=1}^n \zeta_i = n$ (rendering this variable redundant). The RS(G, p) model can be extended to a periodic traffic setting, as done in Glynn and Honnappa [15], and the performance metric approximations can still be used in that setting.

The bibliography on the $\Delta_{(i)}/G/1$ model now includes pointwise limit results (Louchard [28], Newell [33]), functional strong laws and central limits (Bet et al. [3], Glynn and Honnappa [15], Honnappa et al. [21, 22]), and large deviations principles (Glynn and Honnappa [15], Honnappa [19]). In the population acceleration scaling limit, the results in Honnappa et al. [21, 22] show that the limiting diffusion for the workload and queue length processes are regulated through a *directional derivative* reflection map (Mandelbaum and Ramanan [32]). This limit can be recovered by the FSATs in Theorems 2 and 3, although the result in Honnappa et al. [21, 22] holds under the weaker condition that the service requirements have two finite moments. However, extracting performance measures (such as moments of the workload/queue length) from the directional derivative reflected process is incredibly hard. Indeed, in Bet et al. [3], a different critical scaling is used to show that the queue length converges to a reflected Brownian motion with parabolic drift when the arrival epoch distribution G is exponential. On the other hand, in Glynn and Honnappa [15], a special "critical" load condition is used to prove that the workload process is approximated by a reflected Brownian motion process. These limit processes can be recovered automatically from the FSATs proved in this paper, albeit at the cost of stronger conditions on the service requirements. We note, however, that with effort it is possible to extend the FSATs to cases where only $m > 2$ moments are available.

1.3. Relation to FSATs for Nonstationary Models

There is a large and growing literature on nonstationary queueing models covering the whole range of problems that confront the modeling of nonstationary service systems. A crucial difference between this large body of work and the growing literature on transitory models is that the former implicitly assumes an infinite population of customers, whereas transitory models are exclusively finite population. We cannot possibly do justice to the large body of work on nonstationary models (see Whitt [39] for a recent review). Instead, we focus on strong approximation results that are most closely related. To the best of our knowledge, strong approximations have been proved almost exclusively for Markovian nonstationary models; note that the literature on strong approximations for stationary queueing networks is far more extensive. The most influential papers in this genre are Mandelbaum and Massey [29] and Mandelbaum et al. [30], where the important *uniform acceleration* scaling regime was introduced. In the former, strong approximations for Markov processes were leveraged to prove an FSAT (and consequently functional strong laws and central limit theorems) for an isolated time-varying Markovian single-server queue. This analysis was significantly generalized in the latter paper to include multiserver queueing networks with abandonment. In Mandelbaum and Pats [31], strong approximations were leveraged to prove functional limits for state-dependent, nonstationary Markovian queues. Čudina and Ramanan [8] and Armony et al. [2] use uniform acceleration to establish asymptotic optimality of control policies under uniform acceleration scalings. More recently, Ko and Pender [26] consider nonstationary Markovian arrival processes (MAPs) as models of the traffic and develop a bespoke Poisson representation of the MAP process. They then exploit the strong approximations in Mandelbaum et al. [30] to prove functional strong laws and central limit theorems. All of these results are premised on the availability of strong approximation results for Markov processes (Eithier and Kurtz [11, chapter 7]). However, the performance metric processes for the $\mathrm{RS}(G,p)/G/1$ queue are not Markov (although, of course, one could do state-space enlargement), and we therefore choose to develop the strong approximation results from scratch. What is also nice is the fact that we are able to leverage strong approximation results proved for stationary random walks and empirical processes to study nonstationary queueing models without making explicit Markovian assumptions. We believe the methods highlighted in this paper can be used for analyzing other nonstationary stochastic models (such as nonstationary many-server queues, networks of nonstationary queues, and even nonstationary multiclass queues).

1.4. Technical Challenges and Contrast of Analysis Against Strong Approximations for Markovian Queues

Observe that both Mandelbaum and Massey [29] and Mandelbaum et al. [30] use strong approximations of Poisson processes to obtain an approximating diffusion process to the queue length. In Mandelbaum and Massey [29], the simplistic nature of the microscopic rates for both the arrival and departure Poisson processes implies that the diffusion approximations need no further analysis. The standard strong approximations used yields the best convergence rate of $O(\log \epsilon)$ (where ϵ is the accelerating factor). On the other hand, in Mandelbaum et al. [30, theorem 2.3 and its proof], the microscopic instantaneous rates are more general and (equation 2.24 for exact statement) are assumed to satisfy an asymptotic second-order expansion on acceleration. Consequently, further analysis of the time-changed Brownian motions is warranted. The most natural trick is to use continuity of the Brownian motion to evaluate a uniform limit of the said time-changed Brownian motion as the acceleration term $\eta \to \infty$, where $\eta = 1/\epsilon$. However, this comes at the cost of an approximation error, and in Mandelbaum et al. [30], the second-order diffusion approximation has a $o(\sqrt{\eta})$ convergence rate. This highlights a crucial difference between Mandelbaum et al. [30] and our present work. The analysis in Section 4 establishing a DKW-style inequality for randomly time-changed Brownian motion provides a way of obtaining better convergence rates, without requiring the asymptotic approximations on the microscopic rates in Mandelbaum et al. [30]. Our analysis, under reasonable DKW-style assumptions, provides a $O(n^{1/4}\sqrt{\log n})$ convergence rate as $n \to \infty$, which is a significant improvement over the $o(\sqrt{n})$ rate that the analysis done in Mandelbaum et al. [30] would yield. Our arrival process has a strong approximation courtesy the strong approximation for the empirical Cumulative Distribution Function (CDF), whereas the strong approximation for random walks comes in handy for the departure process. However, finding the best possible convergence rates led us to find approximations for Brownian motions evaluated at renewal processes and the busy time process, thereby further complicating our path to obtain the desired strong approximation with tighter convergence rates. Our DKW-style inequality can be used to prove strong approximations for other time-varying queueing models (such as the composition traffic model in Whitt [38]).

The rest of the paper is organized as follows. We start with preliminaries and main results in Section 2. In Section 3, we provide a brief primer on the strong approximation methodology, particularly the coupling arguments that underly the KMT construction. We do so to make the paper self-contained and because the KMT construction is recondite and not widely understood. Next, we present the DKW-style inequality for controlling the error between the Brownian motion and a counterpart process stochastically time-changed by a jump process in Section 4. Section 5 presents strong embeddings for the RS(G, p) traffic process. The strong embeddings for the workload and queue length processes are proved in Sections 6 and 7, respectively. We end with commentary and conclusions in Section 8.

2. Preliminaries and Main Results

2.1. A Mechanistic Model of Queueing

Consider a single server, infinite buffer queue that is nonpreemptive, nonidling, and starts empty. Service follows a first-come-first-served (FCFS) schedule. Let n be the *nominal* number of customers applying for service. Customers independently sample an arrival epoch T_i, $i = 1, \ldots, n$, from a common distribution function. In addition, all customers independently sample identical Bernoulli random variables ζ_i, $i = 1, \ldots, n$. Customer i chooses to turn up at time T_i only if $\zeta_i = 1$; we call this the dropout variable. The arrival process is the cumulative number of customers that have arrived by time t. Let Bern(p) represent the Bernoulli probability distribution with parameter p.

Assumption 1. *For every $n \geq 1$, let $T_1, \ldots, T_n$ be independent and identically distributed (i.i.d.) samples from a general distribution with distribution function G. Denote G_n to be the empirical distribution function given by*

$$G_n(t) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{T_i \leq t\}}. \tag{1}$$

Let $\zeta_1, \ldots, \zeta_n$ be i.i.d. samples from Bern(p). Then the arrival process A_n is given by

$$A_n(t) = \sum_{i=1}^{nG_n(t)} \zeta_i. \tag{2}$$

Remark 1. We call A_n in (2) as the RS(G, p) traffic model. The $\Delta_{(i)}/G/1$ model introduced in Honnappa et al. [22] is a special case of Assumption 1, corresponding to $p = 1$.

Remark 2. It is possible to consider other ways of modeling a random number of arrivals. However, the dropout model considered here is a mechanistic way of describing the traffic. The model assumes each user will sample a potential time to arrive and a binary indicator that the customer will actually enter the queue at that time. See Section 8 for further discussion.

Remark 3. Observe that while the nominal number of arrivals is n, the actual number of arrivals realized is random. This traffic model provides a mechanistic description of nonstationary arrivals: because the distribution G is nonuniform (in general), the expected number of arrivals per-unit time $\mathbf{E}[A_n(t)]/t$ can be seen to equal $npG(t)/t$, by an application of Wald's identity. This can be seen as a surrogate of an arrival rate that is clearly time varying; we have not assumed that the distribution is differentiable and consequently defining the rate as the derivative of $\mathbf{E}[A_n(t)]$ is inappropriate. A crucial point to note is that this time dependency arises from microscopic behavior as opposed to a posited time dependency in the rate function. This stands in contrast with the vast majority of nonstationary models proposed in the literature where the model description starts with posited time-varying rate functions.

Sometimes it is useful to consider arrivals from a general distribution that in turn approaches the limiting distribution G as $n \to \infty$.

Assumption 2. *For every $n \geq 1$, let $T_1, \ldots, T_n$ be i.i.d. samples from a general distribution with distribution function $G^{(n)}$, which satisfies the following condition:*

$$r_n(G) := \sup_{t \in [0,\infty)} |G^{(n)}(t) - G(t)| = O\left(\frac{1}{\sqrt{n}}\right), \tag{3}$$

for some strictly increasing and Lipschitz continuous distribution function G. In addition, assume that each $G^{(n)}$ is Lipschitz continuous and the Lipschitz coefficient increases at most polynomially in n. The arrival process A_n is now defined similar to (2) but with $G_n^{(n)}$ instead of G_n.

Remark 4. For simplicity and ease of presentation, we will assume that arrivals are supported on $[0,\infty)$, that is, $G(0) = G^{(n)}(0) = 0$.

Next, let $\{V_i, i \geq 1\}$ be a sequence of independent and identically distributed nonnegative random variables. V_i represents the service requirement in time units of the i^{th} potential customer who turns up into the system. We also assume that the sequence is independent of the arrival times T_i, $i = 1, \ldots, n$ and the corresponding indicators of turning up ζ_i, $i = 1, \ldots, n$.

Assumption 3. *For every $n \geq 1$,* let *$V_1, \ldots, V_n$ be i.i.d. samples from a distribution that admits existence of a moment generating function in a neighborhood of zero. Let μ and σ denote the mean and standard deviation, respectively, of this distribution. Let*

$$W_n(t) = \sum_{i=1}^{A_n(t)} V_i \tag{4}$$

denote the cumulative offered load to the system until time t.

We assume that the server efficiency is c_n, that is, it completes c_n jobs in unit time. Let $M_n(t)$ be the "truncated" renewal process counting the number of jobs that the server can complete by time t if working continuously with efficiency c_n (notice that only n jobs arrive to the system):

$$M_n(t) := \sup\left\{0 \leq m \leq n : \sum_{i=1}^{m} V_i \leq c_n t\right\}. \tag{5}$$

2.2. Functional Strong Approximations

In this section, we list the main results proven in the sequel. Strong approximation results are usually stated in terms of versions of the random variables we wish to approximate. In our case, we require versions of the random arrival times T_i, the indicators of turning up ζ_i, and the service times V_i. In order to avoid repetition, we do not mention this crucial requirement in the following theorem statements. However, the same version suffices for each theorem below. Let us also note that is often customary in the literature to assume that the underlying probability space is rich enough to support the random variables and the approximating stochastic processes. Our first result provides a strong embedding for the arrival process. Its proof follows from the forthcoming Proposition 8.

Theorem 1. *There exists a Brownian motion $\hat{B}$, a Brownian bridge $B^{\mathrm{br},n}$ such that if H_n be defined as*

$$H_n(t) = \begin{cases} npG(t) + \sqrt{n}\left(pB^{\mathrm{br},n}_{G(t)} + \sqrt{p(1-p)}\hat{B}_{G(t)}\right), & \text{under Assum. 1,} \\ np(G(t) + r_n(G)) + \sqrt{n}\left(pB^{\mathrm{br},n}_{G(t)} + \sqrt{p(1-p)}\hat{B}_{G(t)}\right), & \text{under Assum. 2,} \end{cases}$$

then

$$\sup_{t \in [0,\infty)} |A_n(t) - H_n(t)| \overset{\text{a.s.}}{=} O(n^{1/4}\sqrt{\log n}).$$

Remark 5. It is useful to contrast Theorem 1 with the setting in Whitt [38]. In the latter, traffic is modeled through a sequence of time-changed stochastic counting processes $\{A^n(t) := (N \circ \Lambda^n)(t)\}$, where N is a stationary stochastic counting process that satisfies an FCLT and Λ^n is a posited cumulative arrival rate function that is assumed to be such that $\hat{\Lambda}_n(t) := n^{-1/2}(\Lambda^n(nt) - nt)$ satisfies $\hat{\Lambda}_n(t) \to \hat{\Lambda}(t)$ uniformly on compact sets of $[0,\infty)$ as $n \to \infty$, for some deterministic limit function $\hat{\Lambda}$. Whitt [38, theorem 3.1] shows that the scaled traffic process $\hat{A}_n(t) := n^{-1/2}(A^n(nt) - nt)$ converges to a limit $B + \hat{\Lambda}$, where B is a Brownian motion. A vital advantage of such a traffic model is that the stochasticity and the nonstationarities/time dependencies are completely separated from each other in the limit.

On the other hand, we do not see such a clean separation in H_n immediately. However, suppose that Assumption 2 holds with $G(t) = t$ on $[0,1]$, and $n^{1/2}(G^{(n)}(t) - t) \to \hat{G}(t)$ uniformly on compact sets of $[0,\infty)$ as $n \to \infty$. Then, using the fact that $B_t^{\mathrm{br},n} = D\tilde{B}_t^n - t\tilde{B}_1^n$ for a standard Brownian motion process $\tilde{B}^n$, and (Oksendal [34, theorem 8.5.2]) it follows that

$$\begin{aligned}\frac{1}{\sqrt{np}}\Big(H_n(t) - npt\Big) &= \frac{1}{\sqrt{np}}\Big(np\Big(G(t) + r_n(G)\Big) + \sqrt{n}\Big(pB^{\mathrm{br},n}_{G(t)} + \sqrt{p(1-p)}\hat{B}_{G(t)}\Big) - npt\Big)\\ &= \sqrt{np}\Big(G^{(n)}(t) - t\Big) + \sqrt{p}O(1) + \frac{1}{\sqrt{p}}\Big(pB^{\mathrm{br},n}_{G(t)} + \sqrt{p(1-p)}\hat{B}_{G(t)}\Big)\\ &\overset{D}{=} \sqrt{np}\Big(G^{(n)}(t) - t\Big) + \sqrt{p}O(1) + (\sqrt{p} + \sqrt{1-p})\int_0^t \sqrt{G'(s)}d\bar{B}_s - \sqrt{p}G(t)\bar{B}_1,\\ &= \sqrt{p}\hat{G}(t) + \sqrt{p}O(1) + (\sqrt{p} + \sqrt{1-p})\bar{B}_t - \sqrt{p}tZ.\end{aligned}$$

where $\bar{B}$ is a standard Brownian motion process. Theorem 1 immediately shows that, for the arrival process $A_n(t)$ at fixed $t \in [0,1]$, $(np)^{-1/2}(A_n(t) - npt) \Rightarrow \sqrt{p}\hat{G}(t) + (\sqrt{p} + \sqrt{(1-p)})\bar{B}_t + \bar{Z}$ as $n \to \infty$, where $\bar{Z}$ is a Gaussian random variable with mean $O(\sqrt{p})$ and standard deviation $\sqrt{p}t$. This is reminiscent of the limit in Whitt [38, theorem 3.1] and shows that our framework can recover a separation of the macroscopic time dependencies and the mesoscopic stochasticity. The setting in Whitt [38] is important because it forms the basis for a whole series of works around nonstationary queueing models (see the survey in Whitt [39]). We also note that a more rigorous weak limit analysis for a specific choice of $G^{(n)}$ is presented in Glynn and Honnappa [15].

Our next major result proves strong embeddings for the workload process. In particular, for the cumulative load to the system, we have the following result, which follows from the forthcoming Propositions 9 and 10.

Theorem 2. *Along with the Brownian motion $\hat{B}$ and Brownian bridge $B^{\mathrm{br},n}$ as considered in Theorem 1, there exists a Brownian motion B such that if R_n be defined as*

$$R_n(t) = \sqrt{n}\sigma B_{pG(t)} + \mu H_n(t)$$

then

$$\sup_{t\in[0,\infty)} |W_n(t) - R_n(t)| \overset{\text{a.s.}}{=} O(n^{1/4}\sqrt{\log n}).$$

Let ϕ be the reflection map functional given by $\phi(f)(t) := f(t) - \inf_{u\le t} f(u)$. Then the total remaining workload at time t can be expressed as $\phi(W_n - c_n \cdot \mathrm{id})(t)$, and this satisfies

$$\sup_{t\in[0,\infty)} |\phi(W_n - c_n\cdot\mathrm{id})(t) - \phi(R_n - c_n\cdot\mathrm{id})(t)| \overset{\text{a.s}}{=} O(n^{1/4}\sqrt{\log n}),$$

where $\mathrm{id} : x \longmapsto x$ is the identity map.

Finally, Theorems 1 and 2 are used to prove a strong embedding for the queue length process, Q_n, that includes both any customer in service and all waiting customers. Recall that the queue length $Q_n(t)$ at time t is the difference between the number of arrivals and the number of job completions before time t. Denoting by $D_n(t)$ the amount of time the queue stays busy until time t, the queue length can be expressed as

$$Q_n(t) = A_n(t) - M_n(D_n(t)). \tag{6}$$

Finally, the idle time process of the server is given by

$$I_n(t) := t - D_n(t). \tag{7}$$

The following theorem is a consequence of Proposition 11.

Theorem 3. *Let B, $\hat{B}$ be the Brownian motions $B^{\mathrm{br},n}$, the Brownian bridge processes as considered in Theorems 1 and 2. Let*

$$X_n(t) = H_n(t) - \frac{c_n t}{\mu} + \sqrt{n}\frac{\sigma}{\mu}B_{E_n(t)},$$

where

$$E_n(t) = \begin{cases}\dfrac{c_n t}{n\mu} + \inf\limits_{s\le t}\Big(pG(s) - \dfrac{c_n}{n}\dfrac{s}{\mu}\Big), & \textit{under Assum. } 1,\\[2ex] \dfrac{c_n t}{n\mu} + pr_n(G) + \inf\limits_{s\le t}\Big(pG(s) - \dfrac{c_n}{n}\dfrac{s}{\mu}\Big), & \textit{under Assum. } 2.\end{cases}$$

Then the queue length $Q_n(t)$ *satisfies*

$$\sup_{t\in[0,\infty)} |Q_n(t) - \phi(X_n)(t)| \overset{\text{a.s.}}{=} O(n^{1/4}\sqrt{\log n}),$$

if $c_n = O(n^p)$ *for some* $p > 0$ *and* $\liminf_n c_n > 0$. *Else we have*

$$\sup_{t\in[0,\infty)} |Q_n(t) - \phi(X_n)(t)| \overset{\text{a.s.}}{=} O(n^{1/4}\sqrt{\log c_n}).$$

Remark 6. Observe that the queue length spends more time near zero as the server efficiency becomes super polynomial in n, resulting in a greater approximation error.

Remark 7. Theorems 2 and 3 show that the scaled workload process $Z_n := \phi(W_n - c_n \cdot \text{id})/n$ and the scaled queue length process Q_n/n are both closely approximated by nonstationary reflected Brownian motion (RBM) processes on a sample path basis. These theorems also imply the results in Honnappa et al. [21, 22] and Bet et al. [3], where functional strong laws and central limit theorems were proved for the scaled processes when $p = 1$.

3. Strong Embeddings: A Primer

Let $X_1, X_2, \ldots$ be i.i.d. random variables from a distribution with mean zero and variance one. Let $S_n = \sum_{i=1}^n X_i$ denote the n^{th} partial sum. Then the classical central limit theorem states that

$$\mathbf{P}\left(\frac{S_n}{\sqrt{n}} \le y\right) \to \Phi(y) \text{ as } n \to \infty, \tag{8}$$

where Φ is the central normal CDF. Equation (8) states that the distribution of $S_n/\sqrt{n}$ approaches that of a standard normal as $n \to \infty$. A stochastic process analog of (8) was proved in Donsker [9]. Let the stochastic process $\{S_n(t); t \in [0,1]\}$ be constructed as follows for each $n \in \mathbb{N}$:

$$S_n(t) = \frac{1}{\sqrt{n}}\Big(S_{[nt]} + X_{[nt]+1} + (nt - [nt])\Big). \tag{9}$$

Then $\{S_n(t), t \in [0,1]\}$ converges in distribution to $\{B(t), t \in [0,1]\}$ as $n \to \infty$, where B is a standard Brownian motion. More precisely,

$$h(S_n) \overset{d}{\to} h(B), \tag{10}$$

for every continuous functional $h : C(0,1) \to \mathbb{R}$. Heuristically, Equations (9) and (10) imply that for n large enough $S_{[nt]} + X_{[nt+1]}(nt - [nt])$ is close in distribution to $\sqrt{n}B_t$. Using the scaling property of Brownian motion and observing that $X_{[nt+1]}$ is negligible compared with $S_{[nt]}$ (for large n), we can concur that S_k is approximately close to B_k for all $k \in \{1, \ldots, n\}$. A bound on the difference of the two was provided in Strassen [36], who showed the existence of a probability space containing versions of all associated random variables and processes such that

$$\frac{S_k - B_k}{\sqrt{n \log\log n}} \overset{\text{a.s.}}{\to} 0, \text{ as } k \to \infty. \tag{11}$$

Equation (11) can be restated in the following form:

$$\sup_{0\le t\le 1} \frac{S_n(t) - \frac{1}{\sqrt{n}}B_{nt}}{\sqrt{\log\log n}} \overset{\text{a.s.}}{\to} 0. \tag{12}$$

A close associate of the partial sums S_n are the empirical distribution functions corresponding to a sample of iid random variables. Consider for simplicity a random sample $U_1, U_2, \ldots$ of i.i.d. $U[0,1]$ random variables. The empirical CDF is then given by

$$F_n(t) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{U_i \le t\}},\ t \in [0,1].$$

Observe that the random quantities $\mathbf{1}_{\{U_i \le t\}}$ are i.i.d. with mean t and variance $t(1-t)$. After proper scaling, and considering our previous discussion, we expect the empirical process α_n given by

$$\alpha_n(t) = \sqrt{n}\Big(F_n(t) - t\Big),$$

to be close to a normal random variable with variance $t(1-t)$. We also expect a convergence result akin to (10) in the process level. Recall that the standard Brownian bridge B^{br} is a stochastic process that may be defined as

$$B_t^{\mathrm{br}} = B_t - tB_1, \ t \in [0,1],$$

for a Brownian motion B, because B^{br} is a Gaussian process and $\mathrm{Var}(B_t^{\mathrm{br}}) = t(1-t)$, B^{br} is a possible candidate for the stochastic process approximating the empirical process. Indeed this was proved to be true in a result analogous to equation (12) in Brillinger [4], who showed the existence of a probability space containing versions of all associated random variables and processes such that

$$\sup_{0\le t\le 1} |\alpha_n(t) - B_t^{\mathrm{br}}| \overset{a.s.}{=} O\left(\left(\frac{\log n}{n}\right)^{1/4} (\log n \log\log n)^{1/4}\right). \tag{13}$$

This result immediately implies the analogue to (10), that is, $\{\alpha_n(t), t \in [0,1]\}$ converges in distribution to $\{B_t^{\mathrm{br}}, t \in [0,1]\}$.

Equations (12) and (13) are insightful and provide a rate of convergence of the partial sums and the empirical processes. However, these are not the best rates of convergence one can achieve. It was shown by Komlós-Major-Tusnády in Komlós et al. [27] that when X_i is allowed to have a finite moment generating function in a neighborhood of zero:

$$\sup_{1\le k\le n} |S_k - B_k| = O(\log n). \tag{14}$$

A similar rate is enjoyed by the empirical processes of uniforms. These two results are stated in Theorems 4 and 5, along with the novel construction (also known as the Hungarian method) of X_i's and U_i's from the Brownian motion and Brownian bridge, respectively. A new and different approach in proving such embedding results has been provided in Chatterjee [5] for the simple symmetric random walk. We will use the terminology *strong embedding* for coupling an arbitrary random variable W with a Gaussian random variable Z so that $W - Z$ has exponentially decaying tails at the appropriate scale. Theorems 4 and 5 thus provide strong embeddings to the partial sums S_n and the empirical processes α_n. As alluded to in the Introduction, we will apply these results to obtain strong embeddings for the performance metrics of a $\mathrm{RS}(G,p)/G/1$ queue.

3.1. Strong Embedding of the Random Walk

We present the KMT theorem for the strong embedding of the random walk. Proof ideas and construction can be found in the online appendix.

Theorem 4. *Let F be a distribution function with mean 0 and variance 1. In addition, suppose the moment generating function corresponding to F, $R(t) = \mathbb{E}(e^{tX})$, $X \sim F$, exists in a neighborhood of zero. Then, given a Brownian motion B, and using it, one can construct a sequence of random variables $X_1, X_2, \ldots$ that are independent and identically distributed to F. Furthermore, the partial sums of X_is are strongly coupled to the Brownian motion B in the following sense. For every $n \in \mathbb{N}$ and $x > 0$,*

$$\mathbf{P}\left(\sup_{1\le k\le n}\left|\sum_{i=1}^{n} X_i - B_n\right| > C\log n + x\right) < Ke^{-\lambda x}, \tag{15}$$

where C, K, and λ are positive constants depending only on F.

3.2. Strong Embedding of the Empirical Process

We present the strong embedding result for the empirical process. Proof ideas and construction can be found in the online appendix.

Theorem 5. *There exists a probability space with independent $U[0,1]$ random variables $U_1, U_2, \ldots$ and a sequence of Brownian bridges $B_1^{\mathrm{br}}, B_2^{\mathrm{br}}, \ldots$ such that for all $n \ge 1$ and $x \in \mathbb{R}$,*

$$\mathbf{P}\left(\sup_{s\in[0,1]} \sqrt{n}\left|\alpha_n(s) - B_n^{br}(s)\right| > C\log n + x\right) < Ke^{-\lambda x} \tag{16}$$

for some constants C, K, and λ. Here the empirical process α_n is given by

$$\alpha_n(s) = \sqrt{n}(F_n(s) - s)$$

and

$$F_n(s) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{U_i \le s\}}.$$

Remark 8. The constants C, K, and λ in Theorem 5 can be chosen as $C = 100$, $K = 10$, and $\lambda = 1/50$. See Csörgo and Révész [7, theorem 4.4.1] for more details.

Remark 9. The KMT construction relies on the generation of a sample of n uniforms $U_1, \ldots, U_n$ from a Brownian bridge $B^{\mathrm{br},n}$. It can be seen from the construction that having obtained $\{U_1, \ldots, U_n\}$, one is unable to obtain another U_{n+1} such that the new set $\{U_1, \ldots, U_{n+1}\}$ satisfies (16) with the same Brownian bridge. Instead it would be necessary to redo the construction. This necessitates the need for a different Brownian bridge $B^{\mathrm{br},n}$ for every n.

4. Control of Time-Changed Brownian Motion

Our analyses in subsequent sections provide strong embedding results for several queue length characteristics to corresponding diffusion approximations. In order to achieve those results, we need a strong control over the difference between a Brownian motion evaluated at several n-level stochastic quantities and their corresponding fluid limits as n goes to infinity (e.g., the empirical distribution of arrival epochs against the true arrival distribution). In this section, we present general results on bounding the difference between Brownian motion evaluated at some stochastic jump process and its fluid limit. Proposition 3 is rather general and might be of independent interest. We start by stating an assumption on the fluid limit.

Assumption 4. *For each $n \ge 1$, let $\xi_n : [0,\infty) \longmapsto \mathbb{R}$ be a bounded Lipschitz continuous function; that is, there exists $c_{\xi_n} > 0$ such that*

$$|\xi_n(s) - \xi_n(t)| \le c_{\xi_n}|s - t|,$$

for all $s, t \in [0,\infty)$.

We also impose regularity conditions on the stochastic jump process along which our Brownian motion will be evaluated. These are collected in the following assumption.

Assumption 5. *Let $\{T_n\}_{n\ge1}$ be a sequence of nondecreasing positive numbers. Let $\Xi_n := \{\Xi_n(s); s \in [0, L_n]\}$ be a stochastic pure jump process defined on a common probability space $(\Omega, \mathcal{F}, \mathbf{P})$ for every $n \ge 1$, such that almost surely the number of its jumps in $[0, L_n]$ is bounded above by kn^m for some fixed positive constants k and m. In addition, assume that*

$$\limsup_{n} \sup_{s\in[0,L_n]} \{|\Xi_n(s)|\} < \infty.$$

Denote $D = \limsup_n \sup_s |\Xi_n(s)|$.

In order to obtain a nonasymptotic probabilistic bound on the difference $|B_{\Xi_n(s)} - B_{\xi_n(s)}|$, where ξ_n and Ξ_n are introduced in Assumptions 4 and 5, respectively, we impose further conditions on the distribution of $|\Xi_n - \xi_n|$. In particular, we require a DKW-style inequality (Dvoretzky et al. [10]) for the tail distribution of $|\Xi_n - \xi_n|$.

Assumption 6. *For every $n \ge 1$,* let ξ_n and Ξ_n be as considered in Assumptions 4 and 5. Let there be constants k_0, k_1, k_2, k_3, and $0 < \gamma < 4$ such that the following inequality holds for every $\varepsilon > 0$:

$$\mathbf{P}\left(\sup_{s\in[0,L_n]} |\Xi_n(s) - \xi_n(s)| > \varepsilon + k_0\frac{\log n}{n}\right) \le k_1 e^{-k_2 n^{\gamma}\varepsilon^2 \wedge k_3 n^{\gamma}\varepsilon}.$$

In addition, denote α_n by

$$\alpha_n := \frac{1}{\sqrt{2}}\left(\sup_{t\in[0,L_n]} |\Xi_n(t) - \xi_n(t)|\right)^{1/2}. \tag{17}$$

In Propositions 1 and 2, we will show that the Assumptions 4–6 are satisfied for the arrival process given in (2) and the truncated renewal process given in (5). In order to prove these two lemmas, we first recall a few facts on subexponential random variables.

Lemma 1. *Let $X_1, \ldots, X_n$ be i.i.d. copies of a random variable with mean μ such that there exist parameters (ν, m) satisfying*

$$\mathbf{E}[e^{\lambda(X-\mu)}] \le e^{\frac{\nu^2\lambda^2}{2}} \text{ for all } |\lambda| < \frac{1}{m}. \tag{18}$$

Then the following holds true:

(i)

$$\mathbf{P}\left(\left|\sum_{i=1}^{n} X_i - n\mu\right| \geq nt\right) \leq \begin{cases} 2e^{-\frac{nt^2}{2\nu^2}} & \textit{for } 0 \leq t \leq \frac{\nu^2}{m}, \\ 2e^{-\frac{nt}{2m}} & \textit{for } t > \frac{\nu^2}{m}. \end{cases} \tag{19}$$

(ii)

$$\mathbf{P}\left(\sup_{0\leq k\leq n}\left|\sum_{i=1}^{k} X_i - k\mu\right| \geq nt\right) \leq \begin{cases} 2e^{-\frac{nt^2}{2\nu^2}} & \textit{for } 0 \leq t \leq \frac{\nu^2}{m}, \\ 2e^{-\frac{nt}{2m}} & \textit{for } t > \frac{\nu^2}{m}. \end{cases} \tag{20}$$

Proof.

(i) The result follows from the usual considerations for subexponential random variables (Wainwright [37, section 2.1.3]). The main ingredient is a Chernoff-type approach to obtain

$$\begin{aligned}\mathbf{P}\left(\sum_{i=1}^{n} X_i - n\mu \geq nt\right) &\leq \frac{\mathbf{E}[e^{\lambda(\sum_{i=1}^{n} X_i - n\mu)}]}{e^{n\lambda t}} \\ &\leq \exp\left(-n\lambda t + \frac{n\lambda^2\nu^2}{2}\right).\end{aligned} \tag{21}$$

Optimization of the right-hand side followed by a premultiplication by 2 to obtain the two-sided tail bound yields the desired result (19).

(ii) Observe that $M_k = \sum_{i=1}^{k} X_i - k\mu$ is a martingale. In addition, $x \longmapsto e^{\lambda x}$ is a convex function. Consequently $e^{\lambda M_k}$ is submartingale. Thus, applying Doob's martingale inequality, we obtain

$$\begin{aligned}\mathbf{P}\left(\sup_{0\leq k\leq n}\sum_{i=1}^{k} X_i - k\mu \geq nt\right) &= \mathbf{P}\left(\sup_{0\leq k\leq n} e^{\lambda(\sum_{i=1}^{k} X_i - k\mu)} \geq e^{\lambda nt}\right) \\ &\leq \frac{E\left(e^{\lambda(\sum_{i=1}^{k} X_i - k\mu)}\right)}{e^{\lambda nt}},\end{aligned}$$

thus reducing our considerations to (21). The same arguments carry forward and we obtain (20). □

Assumption (18) in Lemma 1 holds for every random variable X with a finite moment generating function in a neighborhood of zero. This is a consequence of the following lemma.

Lemma 2. *Let X be a random variable with mean μ, whose moment generating function exists in a neighborhood of zero. Then we have*

$$\mathbf{E}[e^{\lambda(X-\mu)}] \leq e^{\lambda^2\nu^2/2} \quad \textit{for all } |\lambda| < \frac{1}{m}, \tag{22}$$

where $\nu = \sqrt{2\mathrm{Var}(X)}$ and m is given by the condition

$$\mathbf{E}[e^{2\lambda|X-\mu|}] < 4 \textit{ for all } |\lambda| < \frac{1}{m}.$$

Proof. Observe that the moment generating function of X satisfies

$$\mathbf{E}[e^{\lambda(X-\mu)}] \leq \mathbf{E}\left(1 + \lambda(X-\mu) + \frac{\lambda^2}{2}(X-\mu)^2 e^{\lambda|X-\mu|}\right).$$

Noticing $\mathbf{E}X = \mu$ and by Cauchy-Schwarz inequality, we have

$$\begin{aligned}\mathbf{E}[e^{\lambda(X-\mu)}] &\leq 1 + \frac{\lambda^2}{2}\sqrt{\mathbf{E}(X-\mu)^4}\sqrt{\mathbf{E}[e^{2\lambda|X-\mu|}]} \\ &\leq 1 + \lambda^2\mathrm{Var}(X)\sqrt{\frac{\mathbf{E}[e^{2\lambda|X-\mu|}]}{4}}.\end{aligned}$$

Hence, for all λ satisfying $\mathbf{E}[e^{2\lambda|X-\mu|}] < 4$, we have

$$\mathbf{E}[e^{\lambda(X-\mu)}] \le 1 + \lambda^2 \text{Var}(X).$$

However, $1 + \lambda^2\text{Var}(X) \le e^{\lambda^2 \text{Var}(X)}$ and consequently,

$$\mathbf{E}[e^{\lambda(X-\mu)}] \le e^{\lambda^2 \text{Var}(X)}.$$

This yields (22). □

Proposition 1. *Let Assumption 1 hold with G being the uniform distribution function on* $[0,1]$. *Let* $\hat{A}_n$ *denote the corresponding arrival process in* (2). *Then Assumptions 4–6 hold with*

$$\xi_n(t) = pt \;\; and \;\; \Xi_n(t) = \frac{\hat{A}_n(t)}{n}, \; for \; t \in [0,1],$$

and $L_n = 1$ *for all* n.

Proof. Note that ξ_n is bounded and Lipschitz with $c_{\xi_n} = p$. Next, notice that $\hat{A}_n$ has at most n jumps in $[0,1]$. In addition,

$$\sup_{s\in[0,1]} \left\{ \left|\frac{\hat{A}_n(s)}{n}\right|, |ps| \right\} \le 1.$$

It remains to prove a DKW type inequality for the difference $|\frac{\hat{A}_n(t)}{n} - pt|$. To that effect, observe that

$$\begin{aligned} \frac{\hat{A}_n(s)}{n} - ps &= \frac{\sqrt{p(1-p)}}{n} \sum_{i=1}^{nF_n(s)} \frac{\chi_i - p}{\sqrt{p(1-p)}} + p\Big(F_n(s) - s\Big) \\ &= \sqrt{p(1-p)} \frac{1}{n} \sum_{i=1}^{nF_n(s)} X_i + p\Big(F_n(s) - s\Big), \end{aligned}$$

where $X_i = \frac{\chi_i - p}{\sqrt{p(1-p)}}$ are i.i.d. random variables with mean zero and variance one. Consequently, we have

$$\mathbf{P}\left(\sup_{s\in[0,1]} \left| \frac{\hat{A}_n(s)}{n} - ps \right| > \varepsilon \right) \le \mathbf{P}\left(\sqrt{p(1-p)} \sup_{0\le k\le n} \frac{1}{n} \left| \sum_{i=1}^{k} X_i \right| > \frac{\varepsilon}{2} \right) + \mathbf{P}\left(p|F_n(s) - s| > \frac{\varepsilon}{2} \right). \tag{23}$$

From the standard DKW inequality for empirical distributions (Dvoretzky et al. [10]), the second term has the standard exponentially decreasing bound given by

$$\mathbf{P}\left(\sup_{s\in[0,1]} p\Big|F_n(s) - s\Big| > \frac{\varepsilon}{2} \right) \le 2e^{-n\varepsilon^2/(4p^2)}. \tag{24}$$

For the first term, observe that the X_i's have a finite moment generating function $E[e^{\lambda X_i}]$ for all λ. Hence, appealing to Lemmas 1 and 2, we obtain

$$\mathbf{P}\left(\sup_{0\le k\le n} \left| \sum_{i=1}^{k} X_i \right| > \frac{n\varepsilon}{2\sqrt{p(1-p)}} \right) \le 2e^{-\frac{n\varepsilon^2}{8p(1-p)}}. \tag{25}$$

Consequently, there exist constants k_1 and k_2 such that

$$\mathbf{P}\left(\sup_{s\in[0,1]} \left| \frac{\hat{A}_n(s)}{n} - ps \right| > \varepsilon \right) \le k_1 e^{-k_2 n\varepsilon^2}. \quad \square$$

Proposition 2. *Let Assumption 3 hold and* M_n *be given by* (5). *Then for any sequence of nondecreasing positive reals* L_n, *Assumptions 4–6 hold with*

$$\xi_n(t) = \left(\frac{c_n}{n}\frac{t}{\mu}\right) \wedge 1 \;\; and \;\; \Xi_n(t) = \frac{M_n(t)}{n}, \;\; for \; t \in [0, L_n].$$

Proof. Note that ξ_n is bounded and Lipschitz with $c_{\xi_n} = \frac{c_n}{n\mu}$. Next notice that M_n has at most n jumps in $[0, L_n]$. In addition

$$\sup_{s\in[0,L_n]} \left| \frac{M_n(s)}{n} \right| < 1.$$

It remains to prove a DKW-style inequality for the difference $|\frac{M_n(t)}{n} - (\frac{c_n}{n}\frac{t}{\mu}) \wedge 1|$. Observe that for any L_n positive,

$$\sup_{0\le t\le L_n}\left|\frac{M_n(t)}{n} - \left(\frac{c_n}{n}\frac{t}{\mu}\right)\wedge 1\right| \le \sup_{0\le t\le \frac{S_n}{c_n}+}\left|\frac{M_n(t)}{n} - \frac{c_n}{n}\frac{t}{\mu}\right| + \sup_{\frac{S_n}{c_n}\le t\le L_n}\left|1 - \left(\frac{c_n}{n}\frac{t}{\mu}\right)\wedge 1\right|$$
$$\le \sup_{0\le t\le \frac{S_n}{c_n}+}\frac{\left|N(c_n t) - \frac{c_n t}{\mu}\right|}{n} + \frac{1}{n} + \frac{\left|\frac{S_n}{n} - \mu\right|}{\mu},$$

where $N(t) = \inf\{m \ge 0 : \sum_{i=1}^m V_i > t\}$. By a change of variable, the first term on the right-hand side has a simpler representation on which we have

$$\sup_{0\le t\le L_n}\left|\frac{M_n(t)}{n} - \left(\frac{c_n}{n}\frac{t}{\mu}\right)\wedge 1\right| \le \sup_{0\le s\le \frac{S_n}{\mu}+}\frac{|\tilde{N}_s - s|}{n} + \frac{1}{n} + \frac{|\frac{S_n}{n} - \mu|}{\mu}, \tag{26}$$

where $\tilde{N}_s = \inf\left\{m \ge 0 : \sum_{i=1}^m \frac{V_i}{\mu} > t\right\}$. From Horvath [23, lemma] and observing that $\tilde{N}(\frac{S_n}{\mu}+) = n$, we notice that

$$\sup_{0\le s\le \frac{S_n}{\mu}+}|\tilde{N}(s) - s| \le \sup_{0\le s\le n}|\tilde{U}(s) - s|, \tag{27}$$

where $\tilde{U}(s) = \sum_{i=1}^{[s]}\frac{V_i}{\mu}$. In addition,

$$\sup_{0\le s\le n}|\tilde{U}(s) - s| \le \sup_{0\le k\le n}|\tilde{S}_k - k| + 1, \tag{28}$$

where $\tilde{S}_k = \sum_{i=1}^k \frac{V_i}{\mu}$. Combining Equations (26), (27), and (28), we obtain for all $\varepsilon > 0$

$$\mathbf{P}\left(\sup_{0\le t\le L_n}\left|\frac{M_n(t)}{n} - \left(\frac{c_n}{n}\frac{t}{\mu}\right)\wedge 1\right| \ge \varepsilon + \frac{2}{n}\right) \le \mathbf{P}\left(\sup_{0\le k\le n}|\tilde{S}_k - k| > \frac{\varepsilon n}{2}\right) + \mathbf{P}\left(\left|\frac{S_n}{n} - \mu\right| > \frac{\varepsilon\mu}{2}\right). \tag{29}$$

Observe that V_is have a finite moment generating function in a neighborhood of zero. Hence, appealing to Lemmas 1 and 2, we have for all $\varepsilon > 0$

$$\mathbf{P}\left(\sup_{0\le k\le n}|\tilde{S}_k - k| > \frac{n\varepsilon}{2}\right) \le 2\exp(-k_2' n\varepsilon^2 \wedge k_3' n\varepsilon),$$

for some constants k_2 and k_3. Similarly, Lemmas 1 and 2 also imply for all $\varepsilon > 0$

$$\mathbf{P}\left(\left|\frac{S_n}{n} - \mu\right| > \frac{\varepsilon\mu}{2}\right) \le 2\exp(-k_4' n\varepsilon^2 \wedge k_5' n\varepsilon).$$

Consequently, we have constants k_1, k_2, and k_3 such that for all $\varepsilon > 0$,

$$\mathbf{P}\left(\sup_{0\le t\le L_n}\left|\frac{M_n(t)}{n} - \left(\frac{c_n}{n}\frac{t}{\mu}\right)\wedge 1\right| \ge \varepsilon + \frac{2}{n}\right) \le k_1 e^{-k_2 n\varepsilon^2 \wedge k_3 n\varepsilon}. \quad \square$$

In the following lemma, we obtain an upper bound on the expected value of α_n as denoted in (17). This result will be used in Lemma 5.

Lemma 3. *Let Assumptions 4–6 hold. Then there exists a constant C' such that*

$$\mathbf{E}[\alpha_n] \le C'\frac{\sqrt{\log n}}{n^{\gamma/4}}. \tag{30}$$

Proof. Because ξ_n is bounded, without loss of generality, let us assume

$$\sup_{s\in[0,L_n]}\{|\xi_n(s)|, |\Xi_n(s)|\} \le D.$$

As a consequence notice

$$\alpha_n = \frac{1}{\sqrt{2}}\left(\sup_{t\in[0,L_n]} |\Xi_n(t) - \xi_n(t)|\right)^{1/2} \le \sqrt{D}.$$

Thus, we have

$$\begin{aligned}\mathbf{E}[\alpha_n] &= \int_0^{\sqrt{D}} \mathbf{P}[\alpha_n > t]dt = \int_0^{\sqrt{D}} \mathbf{P}\left[\sup_{t\in[0,L_n]} |\Xi_n(t) - \xi_n(t)| > 2t^2\right]dt \\ &= \frac{1}{2\sqrt{2}}\int_0^{2D} \frac{1}{\sqrt{s}}\mathbf{P}\left[\sup_{t\in[0,L_n]} |\Xi_n(t) - \xi_n(t)| > s\right]ds,\end{aligned} \tag{31}$$

where the last step is obtained by a change of variable. Now breaking the integral into parts and using Assumption 6, we obtain

$$\begin{aligned}\mathbf{E}[\alpha_n] &\le \frac{1}{2\sqrt{2}}\int_0^{k_0\frac{\log n}{n}} \frac{1}{\sqrt{s}}ds + \int_{k_0\frac{\log n}{n}}^{2D} \frac{1}{\sqrt{s}}\,\mathbf{P}\left(\sup_{t\in[0,L_n]} |\Xi_n(t) - \xi_n(t)| > s\right)ds \\ &\le \frac{1}{\sqrt{2}}\sqrt{k_0\frac{\log n}{n}} + \frac{1}{2\sqrt{2}}\int_0^{2D-k_0\frac{\log n}{n}} \frac{1}{\sqrt{s + k_0\frac{\log n}{n}}}\mathbf{P}\left(\sup_{t\in[0,L_n]} |\Xi_n(t) - \xi_n(t)| > k_0\frac{\log n}{n} + s\right)ds \\ &\le \frac{1}{\sqrt{2}}\sqrt{k_0\frac{\log n}{n}} + \frac{1}{2\sqrt{2}}\int_0^{2D-k_0\frac{\log n}{n}} \frac{1}{\sqrt{s}}k_1 e^{-k_2 n^\gamma s^2 \wedge k_3 n^\gamma s}ds \\ &\le \frac{1}{\sqrt{2}}\sqrt{k_0\frac{\log n}{n}} + \frac{1}{4\sqrt{2}n^{\gamma/4}}\int_0^\infty k_1 e^{-k_2 z \wedge k_3\sqrt{z}n^{\gamma/2}} z^{-3/4}dz \le C'\frac{\sqrt{\log n}}{n^{\gamma/4}},\end{aligned}$$

for some constant C', where the penultimate step is obtained by a change of variable ($n^\gamma s^2 \longmapsto z$) and then use of the fact that the integral $\int_0^\infty \exp(-4k_2 z \wedge k_3\sqrt{z}n^{\gamma/2})z^{-3/4}dz$ is bounded yields our desired result (30). □

Assumption 7. *Let ξ_n and Ξ_n be as considered in Assumptions 4 and 5. Let a standard Brownian motion B be defined on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Then let $\hat{f}_s$ be the Gaussian process defined on $[0, L_n]$ by*

$$\hat{f}_s = B_{\xi_n(s)} - B_{\Xi_n(s)}.$$

In addition, let $\mathbb{P}$ denote the conditional probability on $(\Omega, \mathcal{F}, \mathbf{P})$ given Ξ_n. Thus, the conditional expectation $\mathbb{E}[Z] = \mathbf{E}[Z|\Xi_n(s); s \in [0, L_n]]$. Denote $\gamma_n := \mathbf{E}[\sup_{t\in[0,L_n]}\hat{f}_t]$.

The key ingredient to find probabilistic bounds for $\hat{f}$ as alluded to previously is the Borell-Tsirelson-Ibragimov-Sudakov (Borell-TIS) inequality included later for completeness. As a first step, we find the conditional expectation of $\sup \hat{f}$ given Ξ_n. This is obtained in the following lemma.

Lemma 4. *Let Assumptions 4, 5, and 7 hold. Then there exist constants M and $\tilde{C}$ such that*

$$\mathbb{E}\left[\sup_{t\in[0,L_n]} \hat{f}_t\right] \le M\left(\int_0^{\alpha_n}\sqrt{\log\left(\frac{c_{\xi_n}L_n}{\varepsilon^2} + 1\right)}d\varepsilon + \tilde{C}\alpha_n\sqrt{\log n}\right). \tag{32}$$

Proof. The canonical metric for $\hat{f}$ in $(\hat{\Omega}, \hat{\mathcal{F}}, \mathbb{P})$ is given by

$$\hat{d}(s,t) = \left(\mathbb{E}[(\hat{f}_s - \hat{f}_t)^2]\right)^{1/2}. \tag{33}$$

Let $\hat{D}$ denote the diameter of $[0, L_n]$ with respect to the canonical metric, that is,

$$\hat{D} = \sup_{s,t\in[0,L_n]} \hat{d}(s,t).$$

Let $\hat{N}(\varepsilon)$ be the metric entropy defined by the smallest number of balls of diameter ε (with respect to the canonical metric $\hat{d}$) that cover $[0, L_n]$. Then from Adler and Taylor [1, theorem 1.3.3], there exists a universal constant M such that

$$\mathbb{E}\left[\sup_{t\in[0,L_n]} \hat{f}_t\right] \le M\int_0^{\hat{D}/2}\left(\log\hat{N}(\varepsilon)\right)^{1/2}d\varepsilon. \tag{34}$$

It can be easily shown that the canonical metric $\hat{d}$ as defined in (33) satisfies

$$\hat{d}(s,t)^2 = [|\xi_n(s) - \Xi_n(s)| + |\xi_n(t) - \Xi_n(t)| - 2(\xi_n(s) \wedge \xi_n(t) + \Xi_n(s) \wedge \Xi_n(t) - \xi_n(s) \wedge \Xi_n(t) - \xi_n(t) \wedge \Xi_n(s))].$$

If the numbers $\xi_n(s), \Xi_n(s), \xi_n(t), \Xi_n(t)$ are arranged in ascending order from the vector (d_1, d_2, d_3, d_4), then it can be shown that

$$\hat{d}(s,t) = \sqrt{(d_4 - d_3) + (d_2 - d_1)}. \tag{35}$$

This implies, for s and t such that $\Xi_n(s) = \Xi_n(t)$,

$$\hat{d}(s,t) = \sqrt{|\xi_n(s) - \xi_n(t)|} \le \sqrt{c_{\xi_n}}\sqrt{|s-t|}.$$

In addition, note that $\hat{D}/2 = \sup_{s,t\in[0,L_n]} \hat{d}(s,t)/2 \le (\sup_{t\in[0,L_n]} |\Xi_n(t) - \xi_n(t)|)^{1/2}/\sqrt{2} = \alpha_n$.

In order to obtain an upper bound to $\hat{N}(\varepsilon)$, recall as mentioned earlier, $\hat{d}(s,t) \le \sqrt{c_{\xi_n}}\sqrt{|s-t|}$ whenever $\Xi_n(s) = \Xi_n(t)$. Because Ξ_n has at most kn^m points of discontinuity, there are at most $(kn^m + 1)$ intervals where Ξ_n is constant. Let these intervals be $R_0, \ldots, R_{kn^m}$. Then $\hat{N}(\varepsilon)$ can be bounded above as follows:

$$\hat{N}(\varepsilon) \le \sum_{i=0}^{kn^m} \left\lceil c_{\xi_n} \frac{R_i}{\varepsilon^2} \right\rceil \le \sum_{i=0}^{kn^m} c_{\xi_n} \frac{R_i}{\varepsilon^2} + (kn^m + 1) = c_{\xi_n} \frac{L_n}{\varepsilon^2} + (kn^m + 1). \tag{36}$$

Thus, using (36) we get from (34):

$$\mathbb{E}\left[\sup_{t\in[0,L_n]} \hat{f}_t\right] \le M \int_0^{\alpha_n} \sqrt{\log\left(\frac{c_{\xi_n} L'_n}{\varepsilon^2} + kn^m + 1\right)} d\varepsilon. \tag{37}$$

Observe that $\log(x+y) \le \log(x+1) + \log(y)$ for $x \ge 0$ and $y \ge 1$. Consequently, we obtain

$$\begin{aligned} \int_0^{\alpha_n} \sqrt{\log\left(\frac{c_{\xi_n} L_n}{\varepsilon^2} + kn^m + 1\right)} d\varepsilon &\le \int_0^{\alpha_n} \sqrt{\log\left(\frac{c_{\xi_n} L_n}{\varepsilon^2} + 1\right) + \log(kn^m + 1)\, d\varepsilon} \\ &\le \int_0^{\alpha_n} \sqrt{\log\left(\frac{c_{\xi_n} L_n}{\varepsilon^2} + 1\right)} d\varepsilon + \alpha_n \sqrt{\log(kn^m + 1)}, \end{aligned} \tag{38}$$

where in the last step, we used the fact that $\sqrt{x+y} \le \sqrt{x} + \sqrt{y}$. From (37) and (38), we obtain that there exists a constant $\tilde{C}$ such that

$$\mathbb{E}\left[\sup_{t\in[0,L_n]} \hat{f}_t\right] \le M\left(\int_0^{\alpha_n} \sqrt{\log\left(\frac{c_{\xi_n} L_n}{\varepsilon^2} + 1\right)} d\varepsilon + \tilde{C}\alpha_n\sqrt{\log n}\right). \quad \square$$

Having obtained the conditional expectation of $\sup \hat{f}$, we next obtain the unconditional expectation of $\sup \hat{f}$. This is achieved in the following result and is a crucial ingredient in the proof of Proposition 3.

Lemma 5. *Let Assumptions 4–7 hold. Then there exists constant C such that*

$$\gamma_n = \mathbf{E}\left[\sup_{t\in[0,L_n]} \hat{f}_t\right] \le C \frac{\sqrt{\log(L_n \vee n)}}{n^{\gamma/4}}.$$

Proof. Using integration by parts and denoting $c_{\xi_n} L_n$ by L'_n, the first term in (32) yields

$$\begin{aligned} \int_0^{\alpha_n} \sqrt{\log\left(\frac{L'_n}{\varepsilon^2} + 1\right)} d\varepsilon &= \alpha_n \sqrt{\log\left(\frac{L'_n}{\alpha_n^2} + 1\right)} + \int_0^{\alpha_n} \frac{L'_n/(L'_n + \varepsilon^2)}{\sqrt{\log\left(\frac{L'_n}{\varepsilon^2} + 1\right)}} d\varepsilon \\ &= \alpha_n \sqrt{\log\left(\frac{L'_n}{\alpha_n^2} + 1\right)} + \sqrt{L'_n} \int_{\sqrt{\log\left(\frac{L'_n}{\alpha_n^2} + 1\right)}}^{\infty} \frac{1}{\sqrt{e^{t^2} - 1}} dt, \end{aligned} \tag{39}$$

where the last step is obtained by a change of variable ($\sqrt{\log(L_n'/\varepsilon^2+1)} \longmapsto t$). It is readily checked that

$$\frac{1}{\sqrt{e^{t^2}-1}} \le e^{-t^2/2}\sqrt{1+\frac{\alpha_n^2}{L_n'}} \text{ for } t \ge \sqrt{\log\left(\frac{L_n'}{\alpha_n^2}+1\right)}.$$

Consequently, we obtain from (39):

$$\int_0^{\alpha_n}\sqrt{\log\left(\frac{L_n'}{\varepsilon^2}+1\right)}d\varepsilon \le \alpha_n\sqrt{\log\left(\frac{L_n'}{\alpha_n^2}+1\right)} + \sqrt{L_n'+\alpha_n^2}\int_{\sqrt{\log(L_n'/\alpha_n^2+1)}}^{\infty} e^{-t^2/2}dt. \tag{40}$$

Now, (40) can be represented using the standard normal distribution function as follows:

$$\int_0^{\alpha_n}\sqrt{\log\left(\frac{L_n'}{\varepsilon^2}+1\right)}d\varepsilon = \alpha_n\sqrt{\log\left(\frac{L_n'}{\alpha_n^2}+1\right)} + \sqrt{2\pi(L_n'+\alpha_n^2)}\left(1-\Phi\left(\sqrt{\log\frac{L_n'}{\alpha_n^2}+1}\right)\right). \tag{41}$$

Thus, from (32) and (41), we obtain

$$\mathbb{E}[\sup_{t\in[0,L_n]}\hat{f}_t] \le M\left(\alpha_n\sqrt{\log\left(\frac{L_n'}{\alpha_n^2}+1\right)} + \sqrt{2\pi(L_n'+\alpha_n^2)}\left(1-\Phi\left(\sqrt{\log\left(\frac{L_n'}{\alpha_n^2}+1\right)}\right)\right)\right) + C\alpha_n\sqrt{\log n}. \tag{42}$$

Having completed the first step in our attempt to bound γ_n, we now proceed to obtain an upper bound to the right-hand side of (42). The expectation of the first term:

$$\begin{aligned}\mathbf{E}[\alpha_n\sqrt{\log(L_n'/\alpha_n^2+1)}] &= \int_0^{\infty}\mathbf{P}\left(\alpha_n\sqrt{\log\left(\frac{L_n'}{\alpha_n^2}+1\right)} \ge x\right)dx\\ &= \int_0^{s_g}\mathbf{P}\left(\alpha_n \ge g^{-1}(x)\right)dx,\end{aligned}$$

where g is the function $\tilde{g}(y) = y\sqrt{\log(L_n'/y^2+1)}$ restricted to the domain $[0, s_g]$ and s_g is the point of global maxima of $\tilde{g}$, thereby making g invertible. In addition, it is readily checked by comparing values of $\tilde{g}'$ that $s_g < \sqrt{L_n'/(e-1)}$. Following these steps, we thus obtain

$$\begin{aligned}\mathbf{E}[\alpha_n\sqrt{\log(L_n'/\alpha_n^2+1)}] &= \int_0^{s_g}\mathbf{P}(\alpha_n \ge t)g'(t)dt = \int_0^{s_g}\mathbf{P}(\alpha_n \ge t)\frac{\log\left(\frac{L_n'}{t^2}+1\right) - \frac{L_n'}{L_n'+t^2}}{\sqrt{\log\left(\frac{L_n'}{t^2}+1\right)}}\\ &\le \int_0^{k_0\frac{\log n}{n}}\sqrt{\log\left(\frac{L_n'}{t^2}+1\right)}dt + \int_{k_0\frac{\log n}{n}}^{s_g}\mathbf{P}(\alpha_n \ge t)\sqrt{\log\left(\frac{L_n'}{t^2}+1\right)}dt.\end{aligned} \tag{43}$$

The first term on the right-hand side in (43) can be bounded above using (41) as follows:

$$\begin{aligned}&\int_0^{k_0\frac{\log n}{n}}\sqrt{\log\left(\frac{L_n'}{t^2}+1\right)}dt\\ &\le k_0\frac{\log n}{n}\sqrt{\log\left(\frac{n^2L_n'}{k^2(\log n)^2}+1\right)} + \sqrt{2\pi\left(L_n'+k^2\frac{(\log n)^2}{n^2}\right)}\left(1-\Phi\left(\sqrt{\log\left(\frac{n^2L_n'}{k^2(\log n)^2}+1\right)}\right)\right).\end{aligned}$$

It is readily checked by using the standard upper bound for normal tail probability that there exists a constant C' such that the right-hand side is bounded above by $C_1'n^{-\gamma/4}\sqrt{\log(L_n' \vee n)}$, and thus we have

$$\int_0^{k_0\frac{\log n}{n}}\sqrt{\log\left(\frac{L_n'}{t^2}+1\right)}dt \le C_1'\frac{\sqrt{\log(L_n'\vee n)}}{n^{\gamma/4}}. \tag{44}$$

In order to bound the second term in (43), let us perform a change of variable manipulation, namely replace $n^\gamma t^4$ by e^{-z}. We now obtain from Assumption 6:

$$\int_{k_0\frac{\log n}{n}}^{s_g} \mathbf{P}(\alpha_n \geq t)\sqrt{\log\left(\frac{L'_n}{t^2}+1\right)}dt = \int_0^{s_g-k_0\frac{\log n}{n}} k_1 e^{-k_2 4n^\gamma t^4 \wedge 2k_3 n^\gamma t^2}\sqrt{\log\left(\frac{L'_n}{(t+k_0\frac{\log n}{n})^2}+1\right)}dt$$
$$\leq \int_0^{s_g-k_0\frac{\log n}{n}} k_1 e^{-k_2 4n^\gamma t^4 \wedge 2k_3 n^\gamma t^2}\sqrt{\log\left(\frac{L'_n}{t^2}+1\right)}dt$$
$$\leq \frac{k_1}{4n^{\gamma/4}}\int_{-\infty}^{\infty} \exp\left(-4k_2 e^{-z} \wedge 2k_3 n^{\gamma/2} e^{-z/2} - \frac{z}{4}\right)\sqrt{\log\left(n^{\gamma/2}L'_n e^{z/2}+1\right)}dz$$
$$\leq \frac{k_1}{4n^{\gamma/4}}\int_{-\infty}^{\infty} \exp\left(-4k_2 e^{-z} \wedge 2k_3 n^{\gamma/2} e^{-z/2} - \frac{z}{4}\right)\sqrt{\frac{\gamma}{2}\log n + \log\left(L'_n e^{z/2}+1\right)}dz.$$

Consequently, we have

$$\int_{k_0\frac{\log n}{n}}^{s_g} \mathbf{P}(\alpha_n \geq t)\sqrt{\log\left(\frac{L'_n}{t^2}+1\right)}dt$$
$$\leq \frac{k_1\sqrt{\log(L'_n \vee n)}}{4n^{\gamma/4}}\int_{-\infty}^{\infty} \exp\left(-4k_2 e^{-z} \wedge 2k_3 n^{\gamma/2} e^{-z/2} - \frac{z}{4}\right)\sqrt{\frac{\gamma}{2} + \frac{\log(L'_n e^{z/2}+1)}{\log(L'_n \vee n)}}dz. \tag{45}$$

It is readily checked that the integral on the right is finite, and thus we have

$$\int_{k_0\frac{\log n}{n}}^{s_g} \mathbf{P}(\alpha_n \geq t)\sqrt{\log\left(\frac{L'_n}{t^2}+1\right)}dt \leq C'_2\frac{\sqrt{\log(L'_n \vee n)}}{n^{\gamma/4}}, \tag{46}$$

for some generic constant C'_2. Using (44) and (46) in (43), we now obtain

$$\mathbf{E}[\alpha_n\sqrt{\log(L_n/\alpha_n^2+1)}] \leq C'_3\left(\frac{\sqrt{\log(L'_n \vee n)}}{n^{\gamma/4}}\right), \tag{47}$$

where $C'_3 = C'_1 + C'_2$. For the second term in (42), we use the bound on the normal tail probability, namely, $1-\Phi(t) \leq e^{-t^2/2}/t\sqrt{2\pi}$. Thus, we have

$$\sqrt{2\pi(L'_n+\alpha_n^2)}\left(1-\Phi\left(\sqrt{\log\left(\frac{L'_n}{\alpha_n^2}+1\right)}\right)\right) \leq \frac{\alpha_n}{\sqrt{\log(L'_n+\alpha_n^2)}}.$$

The right-hand side can be bounded above by $(\log(L'_n+\alpha_n^2))^{-1/2} \leq (\log T'_1)^{-1/2}$ and the bound for $\mathbf{E}[\alpha_n]$ achieved in Lemma 3. Consequently, combining (30), (42), and (47), we have thus obtained

$$\mathbf{E}\left[\sup_{t\in[0,L_n]} \hat{f}_t\right] \leq C\left(\frac{\sqrt{\log\left((c_{\xi_n}L_n) \vee n\right)}}{n^{\gamma/4}}\right)$$

for some constant C. □

We finally arrive at the main result of this section; in Proposition 3, we state a general nonasymptotic probabilistic bound on the difference between a time-changed Brownian motion evaluated on a stochastic jump process and its fluid limit.

Proposition 3. *Let Assumptions 4–7 hold. Then there exist a constant C such that for all $n \geq 1$ and $x > 0$:*

$$\mathbf{P}\left(\sup_{s\in[0,L_n]} |B_{\Xi_n(s)} - B_{\xi_n(s)}| > C\frac{\sqrt{\log\left((c_{\xi_n}L_n) \vee n\right)}}{n^{1/4}} + x\right) \leq 2e^{-\frac{x^2}{2v_n^2}}, \tag{48}$$

where $v_n^2 = \sup_{s\in[0,L_n]}\mathbf{E}[|\Xi_n(s) - \xi_n(s)|]$.

Proof. Observe that we have

$$\begin{aligned}\mathbb{E}\left[\hat{f}_t^{\,2}\right] &= \mathbb{E}[B^2_{\xi_n(t)} + B^2_{\Xi_n(t)} - 2B_{\xi_n(t)}B_{\Xi_n(t)}]\\ &= [\xi_n(t) + \Xi_n(t) - 2\xi_n(t)\wedge\Xi_n(t)]\\ &= [|\Xi_n(t) - \xi_n(t)|]. \qquad (49)\end{aligned}$$

This implies

$$\sup_{t\in[0,L_n]} \mathbf{E}\left[\hat{f}_t^2\right] = \sup_{t\in[0,L_n]} \mathbf{E}[|\Xi_n(t) - \xi_n(t)|] = v_n^2. \qquad (50)$$

Recall $\gamma_n := \mathbf{E}[\sup_{t\in[0,1]}\hat{f}_t]$. Then by the Borell-TIS inequality (Adler and Taylor [1, theorem 2.1.1]), we have

$$\mathbf{P}\left(\sup_{t\in[0,L_n]}|\hat{f}_t| > x + \gamma_n\right) \le 2\exp\left(\frac{-x^2}{2v_n^2}\right). \qquad (51)$$

Now invoking Lemma 5, we obtain our desired result (48). □

Finally, we will require the following proposition for proving strong embeddings under Assumption 2. Consider the following regularity condition.

Assumption 8. *For every $n \ge 1$, let $\xi^{(n)}, \xi : [0,T] \longmapsto \mathbb{R}$ satisfy*

$$\sup_{s\in[0,T]}\left|\xi^{(n)}(s) - \xi(s)\right| = O\left(\frac{1}{\sqrt{n}}\right). \qquad (52)$$

In addition, let $\xi^{(n)}$ and ξ both be Lipschitz continuous with the Lipschitz coefficient of $\xi^{(n)}$ growing at most polynomially in n.

Proposition 4. *Let Assumption 8 hold and $B^n, n \ge 1$ be any sequence of Brownian motions defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Then there exists constants C, K, and λ such that for all $n \ge 1$ and $x > 0$:*

$$\mathbf{P}\left(\sup_{s\in[0,T]}|B^n_{\xi^{(n)}(s)} - B^n_{\xi(s)}| > C\frac{\sqrt{\log n}}{n^{1/4}} + x\right) < Ke^{-\lambda x^2\sqrt{n}}.$$

Proof. The key ingredient of the proof is again the Borel-TIS inequality. First, let us reuse the same notations as before; let $\hat{f}_s$ be the Gaussian process defined on $[0,T]$ by

$$\hat{f}_s = B^n_{\xi^{(n)}(s)} - B^n_{\xi(s)}.$$

As before, the canonical metric for $\hat{f}$ in $(\Omega, \mathcal{F}, \mathbf{P})$ is given by (33) and let $\hat{D}$ denote the diameter of $[0,T]$ with respect to the canonical metric, that is,

$$\hat{D} = \sup_{s,t\in[0,T]} \hat{d}(s,t).$$

Let $\hat{N}(\varepsilon)$ be the metric entropy defined by the smallest number of balls of diameter ε (with respect to the canonical metric $\hat{d}$) that covers $[0,T]$. Then from Adler and Taylor [1, theorem 1.3.3], there exists a universal constant M such that

$$\mathbf{E}\left[\sup_{t\in[0,T]}\hat{f}_t\right] \le M\int_0^{\hat{D}/2}\left(\log\hat{N}(\varepsilon)\right)^{1/2}d\varepsilon. \qquad (53)$$

It can be easily shown that the canonical metric $\hat{d}$ as defined in (33) satisfies

$$\hat{d}(s,t)^2 = [|\xi^{(n)}(s) - \xi(s)| + |\xi^{(n)}(t) - \xi(t)| - 2(\xi^{(n)}(s)\wedge\xi(t) + \xi^{(n)}(s)\wedge\xi(t) - \xi^{(n)}(s)\wedge\xi(t) - \xi^{(n)}(t)\wedge\xi(s))].$$

If the numbers $\xi(s), \xi^{(n)}(s), \xi(t), \xi^{(n)}(t)$ are arranged in ascending order form the vector (d_1, d_2, d_3, d_4), then it can be shown that

$$\hat{d}(s,t) = \sqrt{(d_4 - d_3) + (d_2 - d_1)}. \qquad (54)$$

This implies that we have

$$\hat{d}(s,t) \le \sqrt{|\xi^{(n)}(t) - \xi(t)| + |\xi^{(n)}(s) - \xi(s)|}.$$

Consequently, from Assumption 8, we have

$$\hat{D} = O\left(\frac{1}{n^{1/4}}\right). \tag{55}$$

In addition, (54) also implies

$$\hat{d}(s,t) \le \sqrt{|\xi^{(n)}(t) - \xi^{(n)}(s)| + |\xi(t) - \xi(s)|}.$$

Using the Lipschitz continuity of $\xi^{(n)}$ and ξ, we obtain that

$$\sup_{s,t\in[0,T]} \frac{\hat{d}(s,t)}{\sqrt{|s-t|}} = l_n, \tag{56}$$

where l_n grows polynomially in n. This implies $\hat{d}(s,t) \le l_n\sqrt{|s-t|}$ for all $s,t \in [0,T]$. Thus, $\{s \in [0,T] : |s - x_0| \le \varepsilon^2/l_n^2\}$ is contained in the ε−ball around x_0. The length of this ball is thus at least $2\varepsilon^2/l_n^2$. Hence, the number of ε−balls that cover $[0,T]$ is at most $Tl_n^2/2\varepsilon^2$. This leads to an upper bound for $\hat{N}(\varepsilon)$, namely

$$\hat{N}(\varepsilon) \le \frac{Tl_n^2}{2\varepsilon^2}. \tag{57}$$

Thus, using (57), we get from (53),

$$\mathbf{E}\left[\sup_{t\in[0,T]} \hat{f}_t\right] \le M\int_0^{\hat{D}/2} \sqrt{\log\left(\frac{Tl_n^2}{\varepsilon^2}\right)}d\varepsilon. \tag{58}$$

It is readily checked using integration by parts that

$$\mathbf{E}\left[\sup_{t\in[0,T]} \hat{f}_t\right] \le M\left(\frac{\hat{D}}{2}\sqrt{\log\left(\frac{4Tl_n^2}{\hat{D}^2}\right)} + \int_0^{\hat{D}/2} \frac{1}{\sqrt{\log\frac{Tl_n^2}{\varepsilon^2}}}d\varepsilon\right).$$

A change of variable $\sqrt{\log(Tl_n^2/\varepsilon^2)} \longmapsto t$ in the integral on the right-hand side yields

$$\mathbf{E}\left[\sup_{t\in[0,T]} \hat{f}_t\right] \le M\left(\frac{\hat{D}}{2}\sqrt{\log\left(\frac{4Tl_n^2}{\hat{D}^2}\right)} + \sqrt{2\pi T}l_n\left(1 - \Phi\left(\sqrt{2\log\left(\frac{2l_n\sqrt{T}}{\hat{D}}\right)}\right)\right)\right).$$

The standard upper bound to the normal tail probability now gives

$$\mathbf{E}\left[\sup_{t\in[0,T]} \hat{f}_t\right] \le M\left(\frac{\hat{D}}{2}\sqrt{\log\left(\frac{4Tl_n^2}{\hat{D}^2}\right)} + \frac{\hat{D}}{2}\frac{1}{\sqrt{2\log\frac{2l_n\sqrt{T}}{\hat{D}}}}\right). \tag{59}$$

Observe that because $l_n \ge 1$ and $\hat{D} \le 2$, we have

$$\frac{1}{\sqrt{2\log\frac{2l_n\sqrt{T}}{\hat{D}}}} \le \frac{1}{\sqrt{\log T}}. \tag{60}$$

In addition, because $x\log x \ge -1$ for all $x > 0$,

$$\frac{\hat{D}}{2}\sqrt{\log\left(\frac{4Tl_n^2}{\hat{D}^2}\right)} = \frac{1}{2}\sqrt{\hat{D}^2\log(4Tl_n^2) - \hat{D}^2\log\hat{D}^2} \le \frac{1}{2}\sqrt{\hat{D}^2\log(4Tl_n^2) + 1}. \tag{61}$$

Using Inequalities (60) and (61) in the right-hand side of (59), we get

$$\mathbf{E}\left[\sup_{t\in[0,T]} \hat{f}_t\right] \le M\left(\frac{1}{2}\sqrt{\hat{D}^2\log(4Tl_n^2) + 1} + \frac{\hat{D}}{2}\frac{1}{\sqrt{\log T}}\right). \tag{62}$$

Finally, using (55) and (56) in (62), we obtain

$$\mathbf{E}\left[\sup_{t\in[0,T]} \hat{f}_t\right] = O\left(\frac{\sqrt{\log n}}{n^{1/4}}\right).$$

Observe that

$$\sup_{s\in[0,T]} \mathbf{E}[|f_t^2|] = \sup_{s\in[0,T]} |\xi^{(n)}(s) - \xi(s)| = O\left(\frac{1}{\sqrt{n}}\right).$$

Then by the Borell-TIS inequality (Adler and Taylor [1, theorem 2.1.1]), there exists constants C, K, and λ such that

$$\mathbf{P}\left(\sup_{s\in[0,1]}\left|B^n_{\xi^{(n)}(s)} - B^n_{\xi(s)}\right| > C\frac{\sqrt{\log n}}{n^{1/4}} + x\right) < Ke^{-\lambda x^2\sqrt{n}}. \quad \square$$

5. A Strong Embedding for the Arrival Process

In this section, we derive a strong embedding for the arrival process. The following proposition is an extension of Theorem 4 when the length of the random walk is provided by a time-varying, not necessarily determinstic, function.

Proposition 5. *Let $X_1, \ldots, X_n$ be i.i.d. samples from a distribution that admits existence of a moment generating function in a neighborhood of zero. Let μ and σ denote the mean and standard deviation, respectively, of this distrbution. Let $J_n : [0,\infty) \longmapsto \{1,\ldots,n\}$ be any process. Then there exists a standard Brownian motion B and a version of $X_1,\ldots,X_n$, along with constants C_1, K_1, and λ_1 such that for all $x > 0$, we have*

$$\mathbf{P}\left[\sup_{t\in[0,\infty)}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{J_n(t)}(X_i - \mu) - \sigma B_{\frac{J_n(t)}{n}}\right| > C_1\frac{\log n}{\sqrt{n}} + x\right] < K_1 e^{-\lambda_1 x\sqrt{n}}.$$

Proof. From Theorem 4, we have that there exists a standard Brownian motion $\tilde{B}$, a version of $X_1,\ldots,X_n$, along with constants C, K, and λ (depending on the distribution of V) such that for all $x > 0$, we have

$$\mathbf{P}\left[\sup_{0\le k\le n}\left|\sum_{i=1}^{k}\left(\frac{X_i-\mu}{\sigma}\right) - \tilde{B}_k\right| > C\log n + x\right] < Ke^{-\lambda x}, \tag{63}$$

where $\sum_{i=1}^{k}$ is defined to be the null sum for $k = 0$. Because A_n takes values in $\{1,\ldots,n\}$, we may replace the supremum in the left-hand side of (63) by a supremum over k taking values in $\{A_n(t), t \in [0,\infty)\}$. Consider another version B of the standard Brownian motion $\tilde{B}$ such that

$$\sqrt{n}B_{k/n} \stackrel{d}{=} \tilde{B}_k.$$

Then from (63) there exist constants C_1, K_1, and λ_1 such that for all $x > 0$ the desired strong-embedding holds

$$\mathbf{P}\left[\sup_{t\in[0,\infty)}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{J_n(t)}(X_i - \mu) - \sigma B_{\frac{A_n(t)}{n}}\right| > C_1\frac{\log n}{\sqrt{n}} + x\right] < K_1 e^{-\lambda_1 x\sqrt{n}}. \quad \square$$

The following proposition is a consequence of Theorem 5 and holds for any general distribution as opposed to the uniform distributional assumption made in Theorem 5.

Proposition 6. *Let Assumption 1 hold with $p = 1$, that is, let us consider the $\Delta_{(i)}/G/1$ model as explained in Remark 1. Then for every $n \ge 1$, there exists a Brownian bridge $\{B^{\mathrm{br},n}; t \in [0,1]\}$ and a version of $T_1,\ldots,T_n$ along with constants C_2, K_2, and λ_2 such that for all $x > 0$, we have*

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\sqrt{n}\,(G_n(t) - G(t)) - B^{\mathrm{br},n}_{G(t)}\right| > C_2\frac{\log n}{\sqrt{n}} + x\right) < K_2 e^{-\lambda_2 x\sqrt{n}}. \tag{64}$$

Proof. We will first consider the random variables $\{G(T_i) : i = 1,\ldots,n\}$. Observe that the $G(T_i)$s are independent and identically distributed as $U[0,1]$ random variables. Consider the corresponding empirical distribution function F_n given by

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{G(T_i)\le t\}}. \tag{65}$$

Then by a little modification of (15), there exist a Brownian bridge $B^{\mathrm{br},n}$ (observe by Remark 9, the Brownian bridge under consideration depends on n) and constants C_2, K_2, and λ_2 such that

$$\mathbf{P}\left(\sup_{t\in[0,1]}|\sqrt{n}(F_n(t)-t)-B_t^{\mathrm{br},n}|>C_2\frac{\log n}{\sqrt{n}}+x\right)<K_2e^{-\lambda_2 x\sqrt{n}}. \tag{66}$$

Let the inverse distribution function G^{-1} be defined as

$$G^{-1}(t):=\sup\{x\in\mathbb{R}:G(x)\le t\}. \tag{67}$$

Observe that, because of our definition of G^{-1}, we have

$$G(x)\le t \;\text{ iff }\; x\le G^{-1}(t). \tag{68}$$

Applying relation (68) in (65), and using (1), we obtain

$$F_n(t)=\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{T_i\le G^{-1}(t)\}}=G_n(G^{-1}(t)). \tag{69}$$

Inserting (69) in (66) yields

$$\mathbf{P}\left(\sup_{t\in[0,1]}|\sqrt{n}(G_n(G^{-1}(t))-t)-B_t^{\mathrm{br},n}|>C_2\frac{\log n}{\sqrt{n}}+x\right)<K_2e^{-\lambda_2 x\sqrt{n}}. \tag{70}$$

In addition, observe that for any $s_1<s_2$ such that $G(s_1)=G(s_2)$, we have for all $i=1,\dots,n$:

$$\mathbf{P}[T_i\in(s_1,s_2]]=0.$$

This implies that, although $G^{-1}(G(s))\ge s$, we still have

$$\mathbf{1}_{\{T_i\le G^{-1}(G(s))\}}=\mathbf{1}_{\{T_i\le s\}}\;\text{ almost surely (a.s.)}$$

Consequently, we obtain

$$G_n(G^{-1}(G(s)))=\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{T_i\le G^{-1}(G(s))\}}=\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{T_i\le s\}}=G_n(s)\text{ a.s.} \tag{71}$$

We use this property in (70). Notice that

$$\{G(s):s\in[0,\infty)\}\subset[0,1].$$

Thus, we have the following inequality between the suprema of the same function over these two sets:

$$\sup_{t\in[0,1]}|\sqrt{n}(G_n(G^{-1}(t))-t)-B_t^{\mathrm{br},n}|\ge\sup_{s\in[0,\infty)}\left|\sqrt{n}(G_n(G^{-1}(G(s)))-G(s))-B_{G(s)}^{\mathrm{br},n}\right|. \tag{72}$$

Inserting (71) in (72), we thus get

$$\sup_{t\in[0,1]}|\sqrt{n}(G_n(G^{-1}(t))-t)-B_t^{\mathrm{br},n}|\ge\sup_{s\in[0,\infty)}\left|\sqrt{n}(G_n(s)-G(s))-B_{G(s)}^{\mathrm{br},n}\right|\text{ a.s.} \tag{73}$$

Looking at the complement probability in the left-hand side of (70), we obtain as a result of (73):

$$\begin{aligned}1-K_2e^{-\lambda_2 x\sqrt{n}}&<\mathbf{P}\left(\sup_{t\in[0,1]}|\sqrt{n}(G_n(G^{-1}(t))-t)-B_t^{\mathrm{br},n}|\le C_2\frac{\log n}{\sqrt{n}}+x\right)\\&\le\mathbf{P}\left(\sup_{s\in[0,\infty)}|\sqrt{n}(G_n(s)-G(s))-B_{G(s)}^{\mathrm{br},n}|\le C_2\frac{\log n}{\sqrt{n}}+x\right).\end{aligned}$$

This yields our desired result (64). □

Remark 10. *Observe that the constants C_2, K_2, and λ_2* in (66) do not depend on G and that the same constants satisfy (64). Thus, owing to Remark 8, we have that $C_2=100$, $K_2=10$, and $\lambda_2=1/50$ satisfy (64).

We now adapt the statement of Proposition 6 under Assumption 2.

Corollary 1. *Let Assumption 2 hold. Then for every $n \geq 1$, there exists a Brownian bridge $\{\tilde{B}^{br,n}; t \in [0,1]\}$ and a version of $T_1, \ldots, T_n$ such that for all $x > 0$, the same constants C_2, K_2, and λ_2 as in Proposition 6 satisfy*

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\sqrt{n}\left(G_n^{(n)}(t) - G^{(n)}(t) - \tilde{B}^{br,n}_{G^{(n)}(t)}\right)\right| > C_2\frac{\log n}{\sqrt{n}} + x\right) < K_2 e^{-\lambda_2 x\sqrt{n}}.$$

Proof. Observe from Remark 10 for every $k \geq 1$, there exists a Brownian bridge $B^{br,k,n}$ such that for all $x > 0$, the same constants C_2, K_2, and λ_2 as in Proposition 6 satisfy

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\sqrt{n}\left(G_n^{(k)}(t) - G^{(k)}(t)\right) - B^{br,k,n}_{G^{(k)}(t)}\right| > C_2\frac{\log n}{\sqrt{n}} + x\right) < K_2 e^{-\lambda_2 x\sqrt{n}}.$$

In particular, for $k = n$ and writing $B^{br,n,n}$ as $\tilde{B}^{br,n}_{G^{(n)}(t)}$, we have

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\sqrt{n}\left(G_n^{(n)}(t) - G^{(n)}(t)\right) - \tilde{B}^{br,n}_{G^{(n)}(t)}\right| > C_2\frac{\log n}{\sqrt{n}} + x\right) < K_2 e^{-\lambda_2 x\sqrt{n}}. \quad \square$$

In the sequel, we will require control over Brownian motion evaluated at the fluid-scaled arrival process A_n/n and the corresponding fluid limit. This is achieved for $U[0,1]$ distributed time epochs in the following proposition.

Proposition 7. *Let $T_1, \ldots, T_n$ be i.i.d. samples from the $U[0,1]$ distribution. Let $\hat{A}_n$ be the arrival process with dropouts given by*

$$\hat{A}_n(t) = \sum_{i=1}^{nF_n(t)} \zeta_i,$$

where F_n is the empirical distribution function corresponding to the sample $T_1, \ldots T_n$, and ζ_i are i.i.d. Ber(p). *Let B be a Brownian motion. Then, there exist constants C_3, K_3, and λ_3 such that for all $n \geq 1$ and $x > 0$:*

$$\mathbf{P}\left(\sup_{s\in[0,1]}\left|B_{\frac{\hat{A}_n(s)}{n}} - B_{ps}\right| > C_3\frac{\sqrt{\log n}}{n^{1/4}} + x\right) \leq K_3 e^{-\lambda_3 x^2\sqrt{n}}.$$

Proof. The proof follows from the DKW-type inequality established for the Brownian motion in Proposition 3, the conditions for which are satisfied in Proposition 1. Consequently, we obtain there exists constants C_3, K_3, and λ_3 such that

$$\mathbf{P}\left(\sup_{s\in[0,1]}\left|B_{\frac{\hat{A}_n(s)}{n}} - B_{ps}\right| > C_3\frac{\sqrt{\log n}}{n^{1/4}} + x\right) \leq 2e^{-\frac{x^2}{\sigma_n^2}}, \tag{74}$$

where

$$\sigma_n^2 = \sup_{s\in[0,1]} \mathbf{E}\left(\left|\frac{\hat{A}_n(s)}{n} - ps\right|\right). \tag{75}$$

In order to bound σ_n^2, we first apply the Cauchy-Schwarz inequality to get

$$\mathbf{E}\left(\left|\frac{\hat{A}_n(s)}{n} - ps\right|\right) \leq \sqrt{\mathbf{E}\left(\frac{\hat{A}_n(s)}{n} - ps\right)^2}. \tag{76}$$

Observe $\hat{A}_n(s)$ has the Bin(n, ps) distribution, which implies

$$\mathbf{E}\left(\frac{\hat{A}_n(s)}{n} - ps\right)^2 = \frac{\text{Var}(\hat{A}_n(s))}{n^2} = \frac{ps(1-ps)}{n}. \tag{77}$$

Combining (76) and (77), we obtain from (75):

$$\sigma_n^2 = \sup_{s\in[0,1]} \mathbf{E}\left(\left|\frac{\hat{A}_n(s)}{n} - ps\right|\right) \leq \sup_{s\in[0,1]}\sqrt{\frac{ps(1-ps)}{n}} = \frac{1}{2\sqrt{n}}. \tag{78}$$

Using Inequality (78) in (74), we now have our desired strong embedding result:

$$\mathbf{P}\left(\sup_{s\in[0,1]}\left|B_{\frac{\hat{A}_n(s)}{n}} - B_{ps}\right| > C_3\frac{\sqrt{\log n}}{n^{1/4}} + x\right) \le 2e^{-2x^2\sqrt{n}},$$

which holds for all $n \ge 1$ and $x > 0$. □

We now extend our result in Proposition 7 to generally distributed time epochs. This is the subject of the following corollary.

Corollary 2. *Let Assumption 1 holds. Let B be a Brownian motion. Then for every* $n \ge 1$ and $x > 0$, *the same constants* C_3, K_3, *and* λ_3 *as in Proposition 7 satisfy*

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|B_{\frac{A_n(t)}{n}} - B_{pG^{(n)}(t)}\right| > C_3\frac{\sqrt{\log n}}{n^{1/4}} + x\right) \le K_3 e^{-\lambda_3 x^2\sqrt{n}}.$$

Proof. The proof of this result is similar to the reasonings we adopted in Proposition 6. Consequently, let us again consider the random variables $\{G(T_i) : i = 1, \ldots, n\}$. Observe that the $G(T_i)$s are independent and identically distributed as $U[0,1]$ random variables. Consider the corresponding distribution function F_n given by

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{G(T_i)\le t\}}.$$

Let $\hat{A}_n(t) = \sum_{i=1}^{nF_n(t)}\zeta_i$. In addition, recall the definition of $G^{-1}(t)$ in (67). Using (69), we obtain

$$\sup_{s\in[0,1]}\left|B_{ps} - B_{\frac{\hat{A}_n(s)}{n}}\right| = \sup_{s\in[0,1]}\left|B_{ps} - B_{\frac{A_n(G^{-1}(s))}{n}}\right|. \tag{79}$$

In a spirit similar to what is used to obtain (73), we have from the analogue to (71):

$$\sup_{s\in[0,1]}\left|B_{ps} - B_{\frac{A_n(G^{-1}(s))}{n}}\right| \ge \sup_{s\in[0,\infty)}\left|B_{\frac{A_n(s)}{n}} - B_{pG(s)}\right| \text{ a.s.} \tag{80}$$

Our desired result now follows from Proposition 7. Observe Proposition 7 guarantees existence of constants C_3, K_3, and λ_3 such that, for all $n \ge 1$ and $x > 0$, we have

$$\mathbf{P}\left(\sup_{s\in[0,1]}\left|B_{\frac{\hat{A}_n(s)}{n}} - B_{ps}\right| > C_3\frac{\sqrt{\log n}}{n^{1/4}} + x\right) \le K_3 e^{-\lambda_3 x^2\sqrt{n}}. \tag{81}$$

We now complete our proof by looking at the complement probability in (81) and using (79), we have

$$\begin{aligned} 1 - K_3 e^{-\lambda_3 x^2\sqrt{n}} &\le \mathbf{P}\left(\sup_{s\in[0,1]}\left|B_{\frac{\hat{A}_n(s)}{n}} - B_{ps}\right| \le C_3\frac{\sqrt{\log n}}{n^{1/4}} + x\right) \\ &= \mathbf{P}\left(\sup_{s\in[0,1]}\left|B_{\frac{A_n(G^{-1}(s))}{n}} - B_{ps}\right| \le C_3\frac{\sqrt{\log n}}{n^{1/4}} + x\right). \end{aligned} \tag{82}$$

Now using (80) in (82), we obtain

$$1 - K_3 e^{-\lambda_3 x^2\sqrt{n}} \le \mathbf{P}\left(\sup_{s\in[0,\infty)}\left|B_{\frac{A_n(s)}{n}} - B_{pG(s)}\right| \le C_3\frac{\sqrt{\log n}}{n^{1/4}}\right), \tag{83}$$

which yields our desired result. □

Remark 11. *Observe that the constants* C_3, K_3, *and* λ_3 *in Corollary 2 do not depend on the particular distribution G.*

Corollaries 1 and 2 provide approximations in terms of $G^{(n)}$. *In order to further simplify our approximation processes, we require the following lemma.*

Lemma 6. *Let Assumption 2 hold. Then for any* $q > 0$ *and any sequence of Brownian motions* B^n, *there exist constants* C_4, K_4, *and* λ_4 *such that for all* $n \ge 1$ *and* $x > 0$:

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|B^n_{qG^{(n)}(t)} - B^n_{qG(t)}\right| > C_4\frac{\sqrt{\log n}}{n^{1/4}} + x\right) < K_4 e^{-\lambda_4 x^2\sqrt{n}}.$$

Proof. We apply Proposition 4. Take $T = 1$, $\xi^{(n)}(s) = qG^{(n)}(G^{-1}(s))$, $\xi(s) = qs$, and observe that continuity of G implies

$$\sup_{s\in[0,1]} |\xi^{(n)}(s) - \xi(s)| = \sup_{s\in[0,1]} q|G^{(n)}(G^{-1}(s)) - s| = \sup_{s\in[0,\infty)} q|G^{(n)}(G^{-1}(G(s))) - G(s)|.$$

Notice $G(G^{-1}(G(s))) = G(s)$ for all s. Thus, we have

$$\sup_{s\in[0,1]} |\xi^{(n)}(s) - \xi(s)| = \sup_{s\in[0,\infty)} |G^{(n)}(G^{-1}(G(s))) - G(G^{-1}(G(s)))| \le \sup_{s\in[0,\infty)} |G^{(n)}(s) - G(s)|.$$

From (3), we now obtain that

$$\sup_{s\in[0,1]} |\xi^{(n)}(s) - \xi(s)| = O\left(\frac{1}{\sqrt{n}}\right).$$

The Lipschitz continuity of ξ is obvious, whereas Lipschitz continuity of $\xi^{(n)}$ follows from that of $G^{(n)}$ and G^{-1}. In addition, from a similar property for $G^{(n)}$, the Lipschitz coefficient of $\xi^{(n)}$ grows at most polynomially in n. Thus, we have constants C_4, K_4, and λ_4 such that

$$\mathbf{P}\left(\sup_{t\in[0,1]} |B_{qG^{(n)}(G^{-1}(s))} - B_{qs}| > C_4 \frac{\sqrt{\log n}}{n^{1/4}} + x\right) < K_4 e^{-\lambda_4 x^2 \sqrt{n}}. \tag{84}$$

Because G is strictly increasing in $[0,\infty)$, we have $G^{-1}(G(s)) = s$ for $s \in [0,\infty)$. Consequently, we have

$$\sup_{s\in[0,1]} |B_{qG^{(n)}(G^{-1}(s))} - B_{qs}| = \sup_{s\in[0,\infty)} |B_{qG^{(n)}(G^{-1}(G(s)))} - B_{qG(s)}| = \sup_{s\in[0,\infty)} |B_{qG^{(n)}(s)} - B_{qG(s)}|.$$

This provides our desired result from (84). □

We can now extend Corollary 2 courtesy Lemma 6.

Corollary 3. *Suppose Assumption 2 holds and B is a Brownian motion. Then there exist constants C_5, K_5, and λ_5 such that for all $n \ge 1$ and $x > 0$, we have*

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|B_{\frac{A_n(t)}{n}} - B_{pG(t)}\right| > C_5 \frac{\sqrt{\log n}}{n^{1/4}} + x\right) < K_5 e^{-\lambda_5 x^2 \sqrt{n}}.$$

Proof. From Corollary 2 and observing Remark 11, we have

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|B_{\frac{A_n(t)}{n}} - B_{pG^{(n)}(t)}\right| > C_3 \frac{\sqrt{\log n}}{n^{1/4}} + x\right) < K_3 e^{-\lambda_3 x^2 \sqrt{n}}.$$

From Lemma 6 with $q = p$, we have constants C_4', K_4' and λ_4' such that

$$\mathbf{P}\left(\sup_{t\in 0,\infty)} \left|B_{pG^{(n)}(t)} - B_{pG(t)}\right| > C_4' \frac{\sqrt{\log n}}{n^{1/4}} + x\right) < K_4' e^{-\lambda_4' x^2 \sqrt{n}}.$$

Combining these, we have our desired result. □

Corollary 1 can be extended using the following result that again is a consequence of Lemma 6.

Corollary 4. *Let Assumption 2 holds. Then for any sequence of Brownian bridges $\tilde{B}^{\mathrm{br},n}$, there exist constants C_6, K_6, and λ_6 such that for all $n \ge 1$ and $x > 0$, we have*

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\tilde{B}^{\mathrm{br},n}_{G^{(n)}(t)} - \tilde{B}^{\mathrm{br},n}_{G(t)}\right| > C_6 \frac{\sqrt{\log n}}{n^{1/4}} + x\right) < K_6 e^{-\lambda_6 x^2 \sqrt{n}}.$$

Proof. Using the fact that the Brownian bridge $\tilde{B}^{\mathrm{br},n}$ can be represented as

$$\tilde{B}^{\mathrm{br},n}_t \stackrel{D}{=} B^n_t - tB^n_1,$$

for a Brownian motion B^n, we have that

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} |\tilde{B}^{\mathrm{br},n}_{G^{(n)}(t)} - \tilde{B}^{\mathrm{br},n}_{G(t)}| > 2z\right) \le \mathbf{P}\left(\sup_{t\in[0,\infty)} |B^n_{G^{(n)}(t)} - B^n_{G(t)}| > z\right) + \mathbf{P}\left(\sup_{t\in[0,\infty)} |G^{(n)(t)} - G(t)||B^n_1| > z\right) \quad \forall z > 0. \tag{85}$$

From Lemma 6, with $q = 1$, there exist constants C_4'', K_4'' and λ_4'' such that for $z = (C_4''\sqrt{\log n}/n^{1/4} + x)$, we have

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} |B^n_{G^{(n)}(t)} - B^n_{G(t)}| > z\right) < K_4'' e^{-\lambda_4'' x^2 \sqrt{n}}. \tag{86}$$

Recall notation $r_n(G)$ introduced in (3). Because $z \geq x$, we have

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} |G^{(n)}(t) - G(t)|\ |B_1^n| > z\right) = \mathbf{P}\left(|B_1^n| > \frac{z}{r_n(G)}\right) \leq \mathbf{P}\left(|B_1^n| > \frac{x}{r_n(G)}\right).$$

Because $r_n(G) = O(\frac{1}{\sqrt{n}})$ and $\mathbf{P}(|B_1^n| > u) \leq 2e^{-u^2/2}$, we now have

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} |G^{(n)}(t) - G(t)|\ |B_1^n| > z\right) \leq 2e^{-\frac{x^2}{2r_n(G)^2}} \leq 2e^{-\lambda_4'' x^2 n}, \tag{87}$$

for some constant λ_4''. Using (86) and (87) in (85), we obtain that there exist constants C_6, K_6, and λ_6, such that

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\tilde{B}^{\mathrm{br},n}_{G^{(n)}(t)} - \tilde{B}^{\mathrm{br},n}_{G(t)}\right| > C_6 \frac{\sqrt{\log n}}{n^{1/4}} + x\right) < K_6 e^{-\lambda_6 x^2 \sqrt{n}}. \quad \square$$

We now arrive at our main result for this section, namely a strong embedding for the arrival process A_n.

Proposition 8. *Let Assumption 1 or 2 hold. Then there exists a Brownian motion $\hat{B}$, a Brownian bridge $B^{\mathrm{br},n}$ such that if $\hat{H}_n$ be defined as*

$$\hat{H}_n(t) = \begin{cases} \sqrt{n}pG(t) + pB^{\mathrm{br},n}_{G(t)} + \sqrt{p(1-p)}\hat{B}_{G(t)}, & \text{under Assum. 1,} \\ \sqrt{n}p(G(t) + r_n(G)) + pB^{\mathrm{br},n}_{G(t)} + \sqrt{p(1-p)}\hat{B}_{G(t)}, & \text{under Assum. 2,} \end{cases}$$

then there exists a version of $T_1, \ldots, T_n$, a version of $\zeta_1, \ldots, \zeta_n$, along with constants C_7, K_7, and λ_7 such that for all $n \geq 1$ and $x > 0$:

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\frac{A_n(t)}{\sqrt{n}} - \hat{H}_n(t)\right| > C_7 \frac{\sqrt{\log n}}{n^{1/4}} + x\right) < K_7 e^{-\lambda_7 x^2 \sqrt{n}}.$$

Proof. Step 1. Assumption 1: From Proposition 5, there exists a Brownian motion $\hat{B}$ along with constants $\hat{C}_1$, $\hat{K}_1$, and $\hat{\lambda}_1$ such that for all $n \geq 1$ and $x > 0$, we have

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\frac{1}{\sqrt{n}} \sum_{i=1}^{nG_n(t)} \frac{(\zeta_i - p)}{\sqrt{p(1-p)}} - \hat{B}_{G_n(t)}\right| > \hat{C}_1 \frac{\log n}{\sqrt{n}} + x\right) < \hat{K}_1 e^{-\hat{\lambda}_1 x \sqrt{n}}. \tag{88}$$

From Proposition 6, there exists a Brownian bridge $B^{\mathrm{br},n}$ along with constants $\hat{C}_2$, $\hat{K}_2$, and $\hat{\lambda}_2$ such that for all $n \geq 1$ and $x > 0$,we have

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\sqrt{n}\Big(G_n(t) - G(t)\Big) - B^{\mathrm{br},n}_{G(t)}\right| > \hat{C}_2 \frac{\log n}{\sqrt{n}} + x\right) < \hat{K}_2 e^{-\hat{\lambda}_2 x \sqrt{n}}. \tag{89}$$

Observe that the arrival process A_n given by (2) may be decomposed as follows:

$$\frac{A_n(t)}{\sqrt{n}} = \sqrt{p(1-p)}A_{1,n}(t) + pA_{2,n}(t) + \sqrt{p(1-p)}A_{3,n}(t) + \hat{H}_n(t), \tag{90}$$

where

$$A_{1,n}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{nG_n(t)} \frac{(\zeta_i - p)}{\sqrt{p(1-p)}} - \hat{B}_{G_n(t)}$$
$$A_{2,n}(t) = \sqrt{n}(G_n(t) - G(t)) - B^{\mathrm{br},n}_{G(t)},$$

and

$$A_{3,n}(t) = \hat{B}_{G_n(t)} - \hat{B}_{G(t)}.$$

From Corollary 2, there exist constants $\hat{C}_3$, $\hat{K}_3$ and $\hat{\lambda}_3$ such that for all $n \geq 1$ and $x > 0$:

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} |A_{3,n}(t)| > \hat{C}_3 \frac{\sqrt{\log n}}{n^{1/4}} + x\right) < \hat{K}_3 e^{-\hat{\lambda}_3 x^2 \sqrt{n}}. \tag{91}$$

Using the bounds (88), (89), and (91) in Decomposition (90), we now obtain existence of constants C_7, K_7, and λ_7 such that for all $n \geq 1$ and $x > 0$:

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\frac{A_n(t)}{\sqrt{n}} - \hat{H}_n(t)\right| > C_7 \frac{\sqrt{\log n}}{n^{1/4}} + x\right) < K_7 e^{-\lambda_7 x^2 \sqrt{n}}.$$

Step 2. Assumption 2: From Proposition 5, there exists a Brownian motion $\hat{B}$ along with constants $\hat{C}_1$, $\hat{K}_1$, and $\hat{\lambda}_1$ such that for all $n \geq 1$ and $x > 0$, we have

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\frac{1}{\sqrt{n}} \sum_{i=1}^{nG_n^{(n)}(t)} \frac{(\zeta_i - p)}{\sqrt{p(1-p)}} - \hat{B}_{G_n^{(n)}(t)}\right| > \hat{C}_1 \frac{\log n}{\sqrt{n}} + x\right) < \hat{K}_1 e^{-\hat{\lambda}_1 x \sqrt{n}}. \tag{92}$$

From Corollary 1, there exists a Brownian bridge $B^{\mathrm{br},n}$ along with constants $\hat{C}_2$, $\hat{K}_2$ and $\hat{\lambda}_2$ such that for all $n \geq 1$ and $x > 0$, we have

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\sqrt{n}(G_n^{(n)}(t) - G^{(n)}(t)) - B^{\mathrm{br},n}_{G^{(n)}(t)}\right| > \hat{C}_2 \frac{\log n}{\sqrt{n}} + x\right) < \hat{K}_2 e^{-\hat{\lambda}_2 x \sqrt{n}}. \tag{93}$$

Observe that the arrival process A_n given by (2) may be decomposed as follows:

$$\frac{A_n(t)}{\sqrt{n}} = \sqrt{p(1-p)} A_{1,n}(t) + p A_{2,n}(t) + \sqrt{p(1-p)} A_{3,n}(t) + p A_{4,n}(t) + p A_{5,n}(t) + \left(\hat{H}_n(t) - \sqrt{n} p r_n(G)\right), \tag{94}$$

where

$$\begin{aligned}
A_{1,n}(t) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{nG_n^{(n)}(t)} \frac{(\zeta_i - p)}{\sqrt{p(1-p)}} - \hat{B}_{G_n^{(n)}(t)},\\
A_{2,n}(t) &= \sqrt{n}\left(G_n^{(n)}(t) - G^{(n)}(t)\right) - B^{\mathrm{br},n}_{G^{(n)}(t)},\\
A_{3,n}(t) &= \left(\hat{B}_{G_n^{(n)}(t)} - \hat{B}_{G(t)}\right),\\
A_{4,n}(t) &= B^{\mathrm{br},n}_{G^{(n)}(t)} - B^{\mathrm{br},n}_{G(t)},
\end{aligned}$$

and

$$A_{5,n}(t) = \sqrt{n}(G^{(n)}(t) - G(t)).$$

From Corollary 2 and Lemma 6, there exist constants $\hat{C}_3$, $\hat{K}_3$ and $\hat{\lambda}_3$ such that for all $n \geq 1$ and $x > 0$:

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} |A_{3,n}(t)| > \hat{C}_3 \frac{\sqrt{\log n}}{n^{1/4}} + x\right) < \hat{K}_3 e^{-\hat{\lambda}_3 x^2 \sqrt{n}}. \tag{95}$$

Next from Corollary 4, there exist constants $\hat{C}_4$, $\hat{K}_4$ and $\hat{\lambda}_4$ such that for all $n \geq 1$ and $x > 0$, we have

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} |A_{4,n}(t)| > \hat{C}_4 \frac{\sqrt{\log n}}{n^{1/4}} + x\right) < \hat{K}_4 e^{-\hat{\lambda}_4 x^2 \sqrt{n}}. \tag{96}$$

Ultimately note that according to Assumption 2, we have

$$\sup_{t\in[0,\infty)} |A_{5,n}(t)| = \sup_{t\in[0,\infty)} \sqrt{n}(G^{(n)}(t) - G(t)) = \sqrt{n} r_n(G) < \infty. \tag{97}$$

Using the Bounds (92), (93), (95), (96), and (97) in Decomposition (94), we now obtain existence of constants C_7, K_7, and λ_7 such that for all $n \geq 1$ and $x > 0$:

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\frac{A_n(t)}{\sqrt{n}} - \hat{H}_n(t)\right| > C_7 \frac{\sqrt{\log n}}{n^{1/4}} + x\right) < K_7 e^{-\lambda_7 x^2 \sqrt{n}}. \quad \square$$

6. A Strong Embedding for the Workload Process

In this section, we derive a strong embedding for the workload process, as well as the total remaining workload.

Proposition 9. *Let Assumptions 1 or 2 and 3 hold. Then there exist Brownian motions B and $\hat{B}$ and a Brownian bridge B^{br} such that if $\hat{R}_n$ be defined as*

$$\hat{R}_n(t) = \sigma B_{pG(t)} + \mu \hat{H}_n(t),$$

where $\hat{H}_n$ has been defined in Proposition 8, then there exists a version of $T_1, \ldots, T_n$, a version of $V_1, \ldots, V_n$, a version of $\zeta_i, \ldots, \zeta_n$ along with constants C_8, K_8, and λ_8 such that for all $n \geq 1$ and $x > 0$:

$$\mathbf{P}\left(\sup_{t \in [0,\infty)} \left| \frac{W_n(t)}{\sqrt{n}} - \hat{R}_n(t) \right| > C_8 \frac{\sqrt{\log n}}{n^{1/4}} + x \right) \leq K_8 e^{-\lambda_8 x^2 \sqrt{n}}.$$

Proof. From Proposition 5, there exists a Brownian motion B along with constants $\hat{C}_1$, $\hat{K}_1$, and $\hat{\lambda}_1$ such that for all $n \geq 1$ and $x > 0$, we have

$$\mathbf{P}\left(\sup_{t \in [0,\infty)} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{A_n(t)} \frac{(V_i - \mu)}{\sigma} - B_{\frac{A_n(t)}{n}} \right| > \hat{C}_1 \frac{\log n}{\sqrt{n}} + x \right) < \hat{K}_1 e^{-\hat{\lambda}_1 x \sqrt{n}}. \tag{98}$$

Observe that the workload process W_n given by (4) may be decomposed as follows:

$$\frac{W_n(t)}{\sqrt{n}} = \sigma W_{1,n}(t) + \mu W_{2,n}(t) + \sigma W_{3,n}(t) + \hat{R}_n(t), \tag{99}$$

where

$$W_{1,n}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{A_n(t)} \frac{(V_i - \mu)}{\sigma} - B_{\frac{A_n(t)}{n}}$$
$$W_{2,n}(t) = \frac{A_n(t)}{\sqrt{n}} - \hat{H}_n(t),$$

and

$$W_{3,n}(t) = B_{\frac{A_n(t)}{n}} - B_{pG(t)}.$$

From Proposition 8, there exist constants $\hat{C}_1$, $\hat{K}_1$ and $\hat{\lambda}_1$ such that for all $n \geq 1$ and $x > 0$:

$$\mathbf{P}\left(\sup_{t \in [0,\infty)} |W_{2,n}(t)| > \hat{C}_2 \frac{\sqrt{\log n}}{n^{1/4}} + x \right) < \hat{K}_2 e^{-\hat{\lambda}_2 x^2 \sqrt{n}}. \tag{100}$$

From Corollary 2, there exists constants $\hat{C}_3$, $\hat{K}_3$ and $\hat{\lambda}_3$ such that for all $n \geq 1$ and $x > 0$:

$$\mathbf{P}\left(\sup_{t \in [0,\infty)} |W_{3,n}(t)| > \hat{C}_3 \frac{\sqrt{\log n}}{n^{1/4}} + x \right) < \hat{K}_3 e^{-\hat{\lambda}_3 x^2 \sqrt{n}}. \tag{101}$$

Using the Bounds (98), (100), and (101) in Decomposition (99), we now obtain existence of constants C_8, K_8, and λ_8 such that for all $n \geq 1$ and $x > 0$:

$$\mathbf{P}\left(\sup_{t \in [0,\infty)} \left| \frac{W_n(t)}{\sqrt{n}} - \hat{R}_n(t) \right| > C_8 \frac{\sqrt{\log n}}{n^{1/4}} + x \right) \leq K_8 e^{-\lambda_8 x^2 \sqrt{n}}. \quad \square$$

The following result helps in extending strong embedding of processes to strong embedding of their Skorohod reflections defined in the sequel.

Lemma 7. *Let f and g be real-valued functions defined on $[0,\infty)$. Assume f and g satisfy the following property:*

$$\sup_{t \in [0,\infty)} |f(t) - g(t)| < \delta. \tag{102}$$

Then we have the following:

$$\sup_{t\in[0,\infty)}\left|\inf_{0\le u\le t} f(u)-\inf_{0\le u\le t} g(u)\right|\le\delta. \tag{103}$$

Proof. Observe that (103) is immediate if the following holds for every $\varepsilon>0$:

$$\sup_{t\in[0,\infty)}\left|\inf_{0\le u\le t} f(u)-\inf_{0\le u\le t} g(u)\right|<\delta+\varepsilon. \tag{104}$$

Let us prove (104) through contradiction. First, assume the contrary, namely, there exists $t_0\in[0,\infty)$ such that

$$\left|\inf_{0\le u\le t_0} f(u)-\inf_{0\le u\le t_0} g(u)\right|\ge\delta+\varepsilon.$$

Consequently, assume without loss of generality:

$$\inf_{0\le u\le t_0} f(u)<\inf_{0\le u\le t_0} g(u)-(\delta+\varepsilon). \tag{105}$$

Observe there exist points $t_1^\varepsilon, t_2^\varepsilon$ such that

$$f(t_1^\varepsilon)<\inf_{0\le u\le t_0} f(u)+\varepsilon, \text{and } g(t_2^\varepsilon)<\inf_{0\le u\le t_0} g(u)+\varepsilon. \tag{106}$$

From (102) and (106), we obtain

$$g(t_1^\varepsilon)-\delta<f(t_1^\varepsilon)<\inf_{0\le u\le t_0} f(u)+\varepsilon. \tag{107}$$

Finally, from (105) and (107), we obtain

$$g(t_1^\varepsilon)<\inf_{0\le u\le t_0} g(u),$$

which contradicts the definition of infimum and is not true. Hence our Assumption (105) is wrong, and we must have

$$\inf_{0\le u\le t_0} f(u)\ge\inf_{0\le u\le t_0} g(u)-(\delta+\varepsilon). \tag{108}$$

Interchanging f and g in (108) allows us to conclude (104), and hence (103) as well. □

We now arrive at a strong embedding of the total remaining workload.

Proposition 10. *Let ϕ be the reflection map functional given by*

$$\phi(f)(t)=f(t)-\inf_{u\le t} f(u).$$

Then, under the same assumptions and notations as in Proposition 9, there exist constants C_9, K_9, and λ_9 such that for all $n\ge 1$ and $x>0$:

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\frac{1}{\sqrt{n}}\phi(W_n-c_n\cdot \mathrm{id})(t)-\phi\left(\hat{R}_n-\frac{c_n}{\sqrt{n}}\cdot\mathrm{id}\right)(t)\right|>C_9\frac{\sqrt{\log n}}{n^{1/4}}+x\right)<K_9e^{-\lambda_9x^2\sqrt{n}},$$

where $\mathrm{id}: x\longrightarrow x$ *denotes the identity function, and c_n is a positive constant denoting the server efficiency rate.*

Proof. Observe from Proposition 9, we have

$$1-K_8e^{-\lambda_8x^2\sqrt{n}}<\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\frac{W_n(t)}{\sqrt{n}}-\hat{R}_n(t)\right|<C_8\frac{\sqrt{\log n}}{n^{1/4}}+x\right). \tag{109}$$

From Lemma 7, we have

$$\begin{aligned}
&\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\frac{W_n(t)}{\sqrt{n}}-\hat{R}_n(t)\right|<C_8\frac{\sqrt{\log n}}{n^{1/4}}+x\right)\\
&=\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\frac{W_n(t)-c_nt}{\sqrt{n}}-\left(\hat{R}_n(t)-\frac{c_nt}{\sqrt{n}}\right)\right|<C_8\frac{\sqrt{\log n}}{n^{1/4}}+x\right)\\
&\le\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\inf_{0\le u\le t}\frac{W_n(u)-c_nu}{\sqrt{n}}-\inf_{0\le u\le t}\left(\hat{R}_n(u)-\frac{c_nu}{\sqrt{n}}\right)\right|\le C_8\frac{\sqrt{\log n}}{n^{1/4}}+x\right). \qquad (110)
\end{aligned}$$

Combining Equations (109) and (110), and recalling the definition of ϕ, we obtain

$$1-K_9e^{-\lambda_9x^2\sqrt{n}}<\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\frac{1}{\sqrt{n}}\phi(W_n-c_n\cdot\mathrm{id})(t)-\phi\left(\hat{R}_n-\frac{c_n}{\sqrt{n}}\cdot\mathrm{id}\right)(t)\right|\le C_9\frac{\sqrt{\log n}}{n^{1/4}}+x\right),$$

for some constants C_9, K_9, and λ_9. □

7. A Strong Embedding for the Queue Length Process

In this section, we obtain a strong approximation to the queue length process. Control of the truncated renewal process M_n (recall relation 5) would lead to a strong approximation of the queue length.

Lemma 8. *Let Assumption 3 holds and* $Z_n(t)$ be given by

$$Z_n(t)=c_n\left(\frac{t}{\mu}-\frac{M_n(t)}{c_n}\right).$$

Then there exist constants C_{10}*,* K_{10}*, and* λ_{10} *such that for all* $n\ge 1$ *and* $x>0$:

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\frac{Z_n(t)}{\sqrt{n}}-\frac{\sigma}{\mu}B_{\frac{M_n(t)}{n}}\right|>C_{10}\frac{\log n}{\sqrt{n}}+x\right)<K_{10}e^{-\lambda_{10}x\sqrt{n}}. \qquad (111)$$

Proof. Notice $Z_n(t)$ can be further decomposed as

$$Z_n(t)=\sum_{i=1}^{M_n(t)}\frac{(V_i-\mu)}{\mu}-\frac{1}{\mu}\left(\sum_{i=1}^{M_n(t)}V_i-c_nt\right). \qquad (112)$$

Henceforth, the two terms in (112) will be approximated. Using the definition of $M_n(t)$, the second term in (112) can be bounded as follows:

$$\frac{1}{\mu}\left|\sum_{i=1}^{M_n(t)}V_i-c_nt\right|\le\frac{V_{(M_n(t)+1)\wedge n}}{\mu}.$$

Hence for any constant $C>0$, we obtain

$$\begin{aligned}
\mathbf{P}\left(\sup_{t\in[0,\infty)}\frac{1}{\sqrt{n}\mu}\left|\sum_{i=1}^{M_n(t)}V_i-c_nt\right|>C\frac{\log n}{\sqrt{n}}+x\right)&\le\mathbf{P}\left(\frac{V_i}{\mu}>C\log n+x\sqrt{n},\ \text{for all } i=1,\dots,n\right)\\
&=\left(\mathbf{P}\left(V_1>C\mu\log n+\mu x\sqrt{n}\right)\right)^n.
\end{aligned}$$

Using Chebyshev's inequality, we obtain

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\frac{1}{\sqrt{n}\mu}\left|\sum_{i=1}^{M_n(t)}V_i-c_nt\right|>C\frac{\log n}{\sqrt{n}}+x\right)\le\left(\frac{\mathbf{E}\left[e^{\delta\frac{V_1}{n}}\right)}{e^{C\delta\mu\frac{\log n}{n}+\delta\mu\frac{x}{\sqrt{n}}}}\right)^n\frac{\mathbf{E}[e^{\delta V_1}]}{e^{C\delta\mu\log n+\delta\mu x\sqrt{n}}}\le Ke^{-\lambda x\sqrt{n}}, \qquad (113)$$

for some constants K and λ, where the last step is obtained by choosing δ sufficiently small.

In order to approximate the first term in (112), observe that from Proposition 5, there exist constants C_1, K_1, and λ_1 such that

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{M_n(t)}(V_i-\mu)-\sigma B_{\frac{M_n(t)}{n}}\right|>C_1\frac{\log n}{\sqrt{n}}+x\right)<K_1e^{-\lambda_1 x\sqrt{n}},$$

which implies

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{M_n(t)}\frac{(V_i-\mu)}{\mu}-\frac{\sigma}{\mu}B_{\frac{M_n(t)}{n}}\right|>\frac{C_1}{\mu}\frac{\log n}{\sqrt{n}}+x\right)<K_1e^{-\lambda_1\mu x\sqrt{n}}. \tag{114}$$

Using (113) and (114), our desired Inequality (111) is obtained. □

We now approximate the Brownian motion evaluated at M_n appearing in Lemma 8.

Lemma 9. *Let Assumption 3 holds. Then there exist constants C_{11}, K_{11}, and λ_{11}. such that for all $n\geq 1$ and $x>0$:*

$$\mathbf{P}\left(\sup_{0\leq t\leq L_n}\left|B_{M_n(t)/n}-B_{(\frac{c_n t}{n\,\mu})\wedge 1}\right|>C_{11}\frac{\sqrt{\log\left(\frac{c_n}{n}\frac{L_n}{\mu}\vee n\right)}}{n^{1/4}}+x\right)<K_{11}e^{-\lambda_{11}x^2\sqrt{n}}.$$

Proof. It suffices to check the conditions in Proposition 3, which are satisfied by Proposition 2. Combining Equations (26), (27), and (28), we obtain

$$\mathbf{E}\left[\sup_{s\in[0,L_n]}\left|\frac{M_n(t)}{n}-\frac{c_n}{n}\frac{t}{\mu}\wedge 1\right|\right]\leq\frac{\sup_{0\leq k\leq n}|\tilde{S}_k-k|}{n}+\frac{2}{n}+\frac{\mathbf{E}|S_n-n\mu|}{n\mu}.$$

Using the fact that $\mathbf{E}\left(\sup_{0\leq k\leq n}|\tilde{S}_k-k|\right)\leq C\mathbf{E}|\tilde{S}_n-n|$, we obtain

$$\mathbf{E}\left[\sup_{s\in[0,L_n]}\left|\frac{M_n(t)}{n}-\frac{c_n}{n}\frac{t}{\mu}\wedge 1\right|\right]\leq\frac{C}{\sqrt{n}}.$$

This yields our desired result. □

Remark 12. Notice from definition, $M_n(t)$ equals n for all $t>S_n/c_n$. Hence, control of $B_{\frac{M_n(t)}{n}}$ for $t\in[0,\infty)$ reduces to a control of $B_{\frac{M_n(t)}{n}}$ for $t\in[0,S_n/c_n]$. However Lemma 9 leads a control of $B_{\frac{M_n(t)}{n}}$ over $t\in[0,L_n]$ for a predetermined and fixed sequence L_n. Hence, we need S_n/c_n to be in an interval $[0,L_n]$ with exponentially high probability (i.e., the complement event has exponentially decreasing probability). This is achieved in the following lemma.

Lemma 10. *Let $S_n=V_1+\ldots+V_n$. Then for every $\eta>0$, there exists $\delta>0$ such that*

$$\mathbf{P}\left(\frac{S_n}{c_n}\geq L_n\right)\leq\exp(-(\delta c_n L_n-n\eta)).$$

Proof.

$$\mathbf{P}\left(\frac{S_n}{c_n}\geq L_n\right)\leq\frac{\mathbf{E}e^{tS_n}}{e^{tc_nL_n}}\leq\frac{(\mathbf{E}[e^{tV_1}])^n}{e^{tc_nL_n}}\leq e^{-(\delta c_nL_n-n\eta)},$$

for some δ small enough such that $\mathbf{E}e^{\delta V_1}<e^{\eta}$. □

From Lemma 9 and Remark 12, we obtain the following result, which controls $B_{\frac{M_n(t)}{n}}$ for all t positive.

Lemma 11. *Let Assumption 3 holds. Then for all $n\geq 1$ and $x>0$:*

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|B_{M_n(t)/n}-B_{(\frac{c_n t}{n\,\mu})\wedge 1}\right|>C_{11}\frac{\sqrt{\log\left(\frac{c_nL_n}{n\mu}\vee n\right)}}{n^{1/4}}+x\left|\frac{S_n}{c_n}\leq L_n\right.\right)<K_{11}e^{-\lambda_{11}x^2\sqrt{n}}.$$

Finally, we arrive at the main result for this section, namely a strong embedding for the queue length Q_n.

Proposition 11. *Let Assumptions 1 or 2 and 3 hold. Let $\tilde{E}_n$ be given by*

$$\tilde{E}_n(t) = \begin{cases} pG(t) - \frac{c_n}{n}\frac{t}{\mu}, & \text{under Assum. 1,} \\ p\Big(G(t) + r_n(G)\Big) - \frac{c_n}{n}\frac{t}{\mu}, & \text{under Assum. 2.} \end{cases} \tag{115}$$

Recall $\hat{H}_n$ defined in Proposition 8 and define the process $\hat{Y}_n$ as follows:

$$\hat{Y}_n(t) = \left(\hat{H}_n(t) - \frac{c_n t}{\sqrt{n}\mu}\right) + \frac{\sigma}{\mu} B_{\left(\frac{c_n t}{n\mu} + \inf_{s\le t} \tilde{E}_n(s)\right)}.$$

Then there exists a version of $T_1, \ldots, T_n$, a version of $V_1, \ldots, V_n$, a version of $\zeta_i, \ldots, \zeta_n$ along with constants C_{12}, K_{12}, λ_{12} and ζ independent of n, such that for all $n \ge 1$ and $x > 0$:

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\frac{Q_n(t)}{\sqrt{n}} - \phi(\hat{Y}_n)(t)\right| > C_{12}\frac{\sqrt{\log n}}{n^{1/4}} + x\right) < K_{12} e^{-\lambda_{12}\sqrt{n}x^2 \wedge x} + e^{-n\zeta},$$

if $c_n = O(n^m)$ for some $m > 0$ and $\liminf_n c_n > 0$, else

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\frac{Q_n(t)}{\sqrt{n}} - \phi(\hat{Y}_n)(t)\right| > C_{12}\frac{\sqrt{\log c_n}}{n^{1/4}} + x\right) < K_{12} e^{-\lambda_{12}\sqrt{n}x^2 \wedge x} + e^{-n\zeta},$$

Proof. Observe that from (6) the diffusion-scaled queue length $Q_n/\sqrt{n}$ can be further decomposed as

$$\frac{Q_n(t)}{\sqrt{n}} = Y_n(t) + \frac{c_n}{\sqrt{n}\mu} I_n(t),$$

where the idle time process I_n has been defined in (7), Y_n is given by

$$Y_n(t) := \left(\frac{A_n(t)}{\sqrt{n}} - \hat{H}_n(t)\right) + \frac{c_n}{\sqrt{n}}\left(\frac{D_n(t)}{\mu} - \frac{M_n(D_n(t))}{c_n}\right) + \left(\hat{H}_n(t) - \frac{c_n}{\sqrt{n}}\frac{t}{\mu}\right), \tag{116}$$

and $\hat{H}_n$ has been defined in Proposition 8. By the Skorohod reflection theorem, we have that

$$\frac{c_n}{\sqrt{n}\mu} I_n(t) = -\inf_{s\le t} Y_n(s), \tag{117}$$

and the busy time process is given by

$$D_n(t) = t + \frac{\sqrt{n}\mu}{c_n}\inf_{s\le t} Y_n(s).$$

The diffusion-scaled queue length process is now given by the Skorohod reflection of Y_n:

$$\frac{Q_n(t)}{\sqrt{n}} = \phi(Y_n)(t). \tag{118}$$

Thus, a strong embedding of $Q_n/\sqrt{n}$ would follow from a strong embedding of Y_n. Notice that we already have a strong embedding of the arrival process courtesy of Proposition 8, namely there exist constants C_7, K_7, and λ_7, such that for all $n \ge 1$ and $x > 0$:

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\frac{A_n(t)}{\sqrt{n}} - \hat{H}_n(t)\right| > C_7\frac{\log n}{\sqrt{n}} + x\right) < K_7 e^{-\lambda_7 x^2 \sqrt{n}}. \tag{119}$$

Thus, to complete the strong embedding of Y_n, we need to approximate $\tilde{Z}_n$ given by

$$\tilde{Z}_n(t) := c_n\left(\frac{D_n(t)}{\mu} - \frac{M_n(D_n(t))}{c_n}\right).$$

Observe that the busy time process $D_n(t)$ is nondecreasing and takes values in $[0, S_n/c_n]$, where $S_n = \sum_{i=1}^n V_i$. Consequently from Lemma 8, there exist constants $\hat{C}_1$, $\hat{K}_1$, and $\hat{\lambda}_1$ such that for all $n \ge 1$ and $x > 0$:

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\frac{\tilde{Z}_n(t)}{\sqrt{n}} - \frac{\sigma}{\mu} B_{\frac{M_n(D_n(t))}{n}}\right| > \hat{C}_1\frac{\log n}{\sqrt{n}} + x\right) < \hat{K}_1 e^{-\hat{\lambda}_1 x\sqrt{n}}. \tag{120}$$

Let

$$\tilde{Y}_n(t) = \left(\hat{H}_n(t) - \frac{c_n t}{\sqrt{n}\mu}\right) + \frac{\sigma}{\mu} B_{\frac{M_n(D_n(t))}{n}}. \tag{121}$$

Using (119) and (120) in (116), we thus have constants $\hat{C}_2$, $\hat{K}_2$, and λ_2 such that for all $n \geq 1$ and $x > 0$:

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} |Y_n(t) - \tilde{Y}_n(t)| > \hat{C}_2 \frac{\log n}{\sqrt{n}} + x\right) < \hat{K}_2 e^{-\hat{\lambda}_2 x\sqrt{n}}. \tag{122}$$

From the converse of Lemma 7, there exist constants $\hat{C}_3$, $\hat{K}_3$, and $\hat{\lambda}_3$ such that for all $n \geq 1$ and $x > 0$:

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\inf_{s\leq t} Y_n(s) - \inf_{s\leq t} \tilde{Y}_n(s)\right| > \hat{C}_3 \frac{\log n}{\sqrt{n}} + x\right) < \hat{K}_3 e^{-\hat{\lambda}_3 x\sqrt{n}}.$$

Recalling Expression (7) and using (117), we have for every $n \geq 1$ and $x > 0$:

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\frac{c_n}{\sqrt{n}\mu}(t - D_n(t)) + \inf_{s\leq t} \tilde{Y}_n(s)\right| > \hat{C}_3 \frac{\log n}{\sqrt{n}} + x\right) < \hat{K}_3 e^{-\hat{\lambda}_3 x\sqrt{n}}.$$

Consequently, for every $n \geq 1$ and $x > 0$:

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\frac{c_n}{n\mu} D_n(t) - \left(\frac{c_n}{n}\frac{t}{\mu} + \frac{1}{\sqrt{n}} \inf_{s\leq t} \tilde{Y}_n(s)\right)\right| > \hat{C}_3 \frac{\log n}{n} + x\right) < \hat{K}_3 e^{-\hat{\lambda}_3 x n}. \tag{123}$$

Recall $\tilde{E}_n$ given by (115). From the expression of $\tilde{Y}_n$ in (121) and recalling $\hat{H}_n$ from Proposition 8, we have

$$\begin{aligned}
&\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\frac{\tilde{Y}_n(t)}{\sqrt{n}} - \tilde{E}_n(t)\right| > \varepsilon\right) \\
&\leq \mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\frac{pB^{\text{br},n}_{G(t)}}{\sqrt{n}}\right| + \sup_{t\in[0,\infty)} \left|\frac{\sqrt{p(1-p)}\hat{B}_{G(t)}}{\sqrt{n}}\right| + \sup_{t\in[0,\infty)} \left|\frac{\sigma}{\mu}\frac{B_{M_n(D_n(t))/n}}{\sqrt{n}}\right| > \varepsilon\right) \\
&\leq \mathbf{P}\left(\sup_{t\in[0,\infty)} |B^{\text{br},n}_{G(t)}| > \frac{\sqrt{n}\varepsilon}{3p}\right) + \mathbf{P}\left(\sup_{t\in[0,\infty)} |\hat{B}_{G(t)}| > \frac{\sqrt{n}\varepsilon}{3\sqrt{p(1-p)}}\right) + \mathbf{P}\left(\sup_{t\in[0,\infty)} |B_{M_n(D_n(t))/n}| > \frac{\sqrt{n}\mu}{\sigma}\frac{\varepsilon}{3}\right).
\end{aligned} \tag{124}$$

Observe that both $G(t)$ and $M_n(D_n(t))/n$ are less than one. Using the tail probability for the supremum of the standard Brownian bridge on $[0,1]$, we have

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|B^{\text{br},n}_{G(t)}\right| > \frac{\sqrt{n}\varepsilon}{3p}\right) \leq 2\exp\left(\frac{2n\varepsilon^2}{9p^2}\right). \tag{125}$$

Using the tail probability for the supremum of the Brownian motion on $[0,1]$, we have

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} |\hat{B}_{G(t)}| > \frac{\sqrt{n}\varepsilon}{3\sqrt{p(1-p)}}\right) \leq 4\int_{\sqrt{n}\varepsilon/3\sqrt{p(1-p)}}^{\infty} \frac{e^{-s^2/2}}{\sqrt{2\pi}}, \tag{126}$$

and

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} |B_{M_n(D_n(t))/n}| > \frac{\sqrt{n}\mu}{\sigma}\frac{\varepsilon}{2}\right) \leq 4\int_{\sqrt{n}\varepsilon\mu/2\sigma}^{\infty} \frac{e^{-s^2/2}}{\sqrt{2\pi}}. \tag{127}$$

Using (125), (126), and (127) in (124), we have that there exist constants k_1 and k_2 such that

$$\mathbf{P}\left(\sup_{t\in[0,\infty)} \left|\frac{\tilde{Y}_n(t)}{\sqrt{n}} - \tilde{E}_n(t)\right| > \varepsilon\right) \leq k_1 \exp(-k_2 n\varepsilon^2).$$

From the converse of Lemma 7, there exist constants $\hat{k}_1$ and $\hat{k}_2$, such that

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\inf_{s\le t}\frac{\tilde{Y}_n(s)}{\sqrt{n}}-\inf_{s\le t}\tilde{E}_n(s)\right|>\varepsilon\right)\le \hat{k}_1\exp(-\hat{k}_2 n\varepsilon^2). \tag{128}$$

Now, using (123) and (128), we have

$$\begin{aligned}
&\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\frac{c_n}{n}\frac{D_n(t)}{\mu}-\left(\frac{c_n t}{n\mu}+\inf_{s\le t}\tilde{E}_n(s)\right)\right|>\hat{C}_3\frac{\log n}{n}+2\varepsilon\right)\\
&\le \mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\frac{c_n}{n}\frac{D_n(t)}{\mu}-\left(\frac{c_n}{n}\frac{t}{\mu}+\frac{1}{\sqrt{n}}\inf_{s\le t}\tilde{Y}_n(s)\right)\right|>\hat{C}_3\frac{\log n}{n}+\varepsilon\right)\\
&+\ \mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\frac{1}{\sqrt{n}}\inf_{s\le t}\tilde{Y}_n(s)-\inf_{s\le t}\tilde{E}_n(s)\right|>\varepsilon\right)\le \hat{K}_3e^{-\hat{\lambda}_3 n\varepsilon}+\hat{k}_1\exp(-\hat{k}_2 n\varepsilon^2).
\end{aligned} \tag{129}$$

Because $|x\wedge 1-y\wedge 1|\le|x-y|$ we have from (129) constants $\hat{k}_3$, $\hat{k}_4$, and $\hat{k}_5$ such that

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\left(\frac{c_n}{n}\frac{D_n(t)}{\mu}\right)\wedge 1-\left(\frac{c_n t}{n\mu}+\inf_{s\le t}\tilde{E}_n(s)\right)\wedge 1\right|>\hat{C}_3\frac{\log n}{n}+\varepsilon\right)\le \hat{k}_3e^{-\hat{k}_4 n\varepsilon^2\wedge\hat{k}_5 n\varepsilon}. \tag{130}$$

From Proposition 2, we have constants $\hat{k}_6$ and $\hat{k}_7$ such that

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\frac{M_n(D_n(t))}{n}-\left(\frac{c_n}{n}\frac{D_n(t)}{\mu}\right)\wedge 1\right|>\frac{2}{n}+\varepsilon\ \middle|\ \frac{S_n}{c_n}\le L_n\right)\le \hat{k}_6e^{-\hat{k}_7 n\varepsilon^2}. \tag{131}$$

Combining (130) and (131), we have constants $\tilde{k}_0$, $\tilde{k}_1$, $\tilde{k}_2$, and $\tilde{k}_3$ such that

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\frac{M_n(D_n(t))}{n}-\left(\frac{c_n t}{n\mu}+\inf_{s\le t}\tilde{E}_n(s)\right)\wedge 1\right|>\tilde{k}_0\frac{\log n}{n}+\varepsilon\ \middle|\ \frac{S_n}{c_n}\le L_n\right)\le \hat{k}_1e^{-\hat{k}_2 n\varepsilon^2\wedge\hat{k}_3 n\varepsilon}.$$

Now, from Proposition 3, we obtain constants $\hat{C}_4$, $\hat{K}_4$, and $\hat{\lambda}_4$ such that

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|B_{\frac{M_n(D_n(t))}{n}}-B_{(\frac{c_n t}{n\mu}+\inf_{s\le t}\tilde{E}_n(s))}\right|>\hat{C}_4\frac{\sqrt{\log(\frac{c_nL_n}{n\mu}\vee n)}}{n^{1/4}}+x\middle|\frac{S_n}{c_n}\le L_n\right)\le \hat{K}_4e^{-\hat{\lambda}_4 x^2\sqrt{n}}. \tag{132}$$

Let $\hat{Y}_n$ be given by

$$\hat{Y}_n(t)=\left(\hat{H}_n(t)-\frac{c_n t}{\sqrt{n}\mu}\right)+\frac{\sigma}{\mu}B_{(\frac{c_n t}{n\mu}+\inf_{s\le t}\tilde{E}_n(s))}.$$

We now obtain from (121) and (132) existence of constants $\hat{C}_5$, $\hat{K}_5$, and $\hat{\lambda}_5$ such that

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}|\tilde{Y}_n(t)-\hat{Y}_n(t)|>\hat{C}_5\frac{\sqrt{\log(\frac{c_nL_n}{n\mu}\vee n)}}{n^{1/4}}+x\middle|\frac{S_n}{c_n}<L_n\right)\le \hat{K}_5e^{-\hat{\lambda}_5 x^2\sqrt{n}}.$$

Hence, from (122), we have constants $\hat{C}_6$, $\hat{K}_6$, and $\hat{\lambda}_6$ such that

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}|Y_n(t)-\hat{Y}_n(t)|>\hat{C}_6\frac{\sqrt{\log(\frac{c_nL_n}{n\mu}\vee n)}}{n^{1/4}}+x\middle|\frac{S_n}{c_n}\le L_n\right)\le \hat{K}_6e^{-\hat{\lambda}_6\sqrt{n}x\wedge x^2}.$$

Recalling (118), we now obtain that there exist constants $\hat{C}_7$, $\hat{K}_7$, and $\hat{\lambda}_7$ such that

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\frac{Q_n(t)}{\sqrt{n}}-\phi(\hat{Y}_n)(t)\right|>\hat{C}_7\frac{\sqrt{\log(\frac{c_nL_n}{n\mu}\vee n)}}{n^{1/4}}+x\middle|\frac{S_n}{c_n}<L_n\right)\le \hat{K}_7e^{-\hat{\lambda}_7\sqrt{n}x\wedge x^2}.$$

Observe that for any two sets A and B, we have that $P(A)\le P(A|B)+P(B^c)$. Thus, using Lemma 10, we have

$$\mathbf{P}\left(\sup_{t\in[0,\infty)}\left|\frac{Q_n(t)}{\sqrt{n}}-\phi(\hat{Y}_n)(t)\right|>\hat{C}_7\frac{\sqrt{\log(\frac{c_nL_n}{n\mu}\vee n)}}{n^{1/4}}+x\right)\le \hat{K}_7e^{-\hat{\lambda}_7\sqrt{n}x\wedge x^2}+e^{-(\delta c_nL_n-n\eta)}.$$

Finally choosing $L_n=n$ yields the desired result. □

8. Commentary and Conclusions

From a philosophy of science perspective, one can consider the bulk of the nonstationary queueing model literature as *phenomenological* (Frigg and Hartmann [12]) in nature, that is, accurately reflecting empirical evidence but not necessarily first principles. For instance, as noted in Remark 5 (see Whitt [39] as well), a widely used nonstationary traffic model uses a composition construction, where a cumulative intensity function that captures the time-varying effects is posited. However, these models are not necessarily a first principles explanation of how customers choose to arrive at a service system. This distinction between phenomenological and mechanistic modeling is not crucial from a performance analysis/prediction perspective, but it can be important from a system design or optimization and control perspective. For instance, in Armony et al. [2], Green and Savin [16], Hassin and Mendel [18], and Kim et al. [25], the problem of designing an optimal appointment schedule to a single server queueing system is studied. In this instance, the standard composition nonstationary traffic models (Whitt [39]) are not appropriate. In Armony et al. [2], in particular, the authors use a transitory model and use its corresponding diffusion limit to solve for an asymptotically optimal schedule. Models of rational arrival behavior have also received significant interest in the literature (Glazer and Hassin [13], Hassin [17], Honnappa and Jain [20], Jain et al. [24]). However, standard traffic models are, in general, awkward to use for computing Nash equilibrium strategies, which are more naturally modeled through a finite pool model as done in Honnappa and Jain [20] and Jain et al. [24]. The $\mathrm{RS}(G,p)/G/1$ model provides a mechanistic, flexible description of queueing behavior and can be used in a broad range of optimization/control and game theoretic models.

The $\mathrm{RS}(G,p)/G/1$ model generalizes the $\Delta_{(i)}/G/1$ model that has been studied in the literature. However, computing performance metrics for the discrete event $\mathrm{RS}(G,p)/G/1$ queueing model is quite difficult, because of the complicated time dependencies in the model. The strong embeddings (and FSATs) proved in this paper provide error bounds from tractable diffusion approximations. Furthermore, these results can be specialized to yield prior diffusion limits obtained via weak convergence in Honnappa et al. [21, 22] and Bet et al. [3]. We anticipate that our FSAT results will be immensely useful for optimization and control problems involving queueing systems.

There are several avenues for further exploration. First, our most general conditions on the traffic model in Assumption 2 allows the arrival epoch distribution to depend on the population size, allowing for the possibility that an increase in the population will change customer behavior because there is an increase in demand for services. We currently assume that the dropout probability is stationary. A more general model would allow for time-dependent/nonstationary dropout probabilities. It seems possible to extend the current FSETs and FSATs to this setting. More complicated is establishing analogous results for a multiserver queue. In Honnappa et al. [22], diffusion limits were established for a fixed multiserver queue in the large population asymptotic limit, relying on the fact that in the large sample limit, the regulator process is identical to the single server case. In our current paper, however, the FSETs (which are for finite n) are much harder to prove, because we can no longer use the asymptotic simplification. This issue is compounded when the servers are not identical, and we will investigate these results in future papers. A further avenue for investigation is how to prove FSETs in a scaling regime that is analogous to the many-server heavy-traffic (MSHT) scaling. In this case, we anticipate that the diffusion approximation should be some type of a nonstationary Halfin-Whitt diffusion process, but we have not been able to prove the FSET, and it appears we might require some new mathematical innovations to achieve this result.

References

[1] Adler RJ, Taylor JE (2009) *Random Fields and Geometry* (Springer Science & Business Media, Berlin).
[2] Armony M, Atar R, Honnappa H (2019) Asymptotically optimal appointment schedules. *Math. Oper. Res.* 44(4):1345–1380.
[3] Bet G, van der Hofstad R, van Leeuwaarden JS (2019) Heavy-traffic analysis through uniform acceleration of queues with diminishing populations. *Math. Oper. Res.* 44(3):821–864.
[4] Brillinger DR (1969) An asymptotic representation of the sample distribution function. *Bull. Amer. Math. Soc.* (New Series) 75(3):545–547.
[5] Chatterjee S (2012) A new approach to strong embeddings. *Probability Theory Related Fields* 152(1-2):231–264.
[6] Chen H, Yao DD (2013) *Fundamentals of Queueing Networks: Performance, Asymptotics, and Optimization*, vol. 46 (Springer Science & Business Media, Berlin).
[7] Csörgo M, Révész P (2014) *Strong Approximations in Probability and Statistics* (Academic Press, New York).
[8] Čudina M, Ramanan K (2011) Asymptotically optimal controls for time-inhomogeneous networks. *SIAM J. Control Optim.* 49(2):611–645.
[9] Donsker MD (1951) An invariance principle for certain probability limit theorems. *Memoirs Amer. Math. Soc.* 6:12.
[10] Dvoretzky A, Kiefer J, Wolfowitz J, et al (1956) Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* 27(3):642–669.
[11] Ethier SN, Kurtz TG (2009) *Markov Processes: Characterization and Convergence*, vol. 282 (John Wiley & Sons, Hoboken, NJ).
[12] Frigg R, Hartmann S (2018) Models in science. Accessed 25 September, 2021, https://plato.stanford.edu/archives/sum2018/entries/models-science.
[13] Glazer A, Hassin R (1983) m/1: On the equilibrium distribution of customer arrivals. *Eur. J. Oper. Res.* 13(2):146–150.

[14] Glynn PW (1998) *Strong Approximations in Queueing Theory. Asymptotic Methods in Probability and Statistics* (Elsevier, New York).
[15] Glynn PW, Honnappa H (2017) On Gaussian limits and large deviations for queues fed by high intensity randomly scattered traffic. Preprint, submitted August, https://arxiv.org/abs/1708.05584.
[16] Green LV, Savin S (2008) Reducing delays for medical appointments: A queueing approach. *Oper. Res.* 56(6):1526–1538.
[17] Hassin R (2016) *Rational Queueing* (Chapman and Hall/CRC).
[18] Hassin R, Mendel S (2008) Scheduling arrivals to queues: A single-server model with no-shows. *Management Sci.* 54(3):565–572.
[19] Honnappa H (2017) Rare events of transitory queues. *J. Appl. Probabilities* 54(3):943–962.
[20] Honnappa H, Jain R (2015) Strategic arrivals into queueing networks: The network concert queueing game. *Oper. Res.* 63(1):247–259.
[21] Honnappa H, Jain R, Ward AR (2014) On transitory queueing. Preprint, submitted XX, https://arxiv.org/abs/1412.2321.
[22] Honnappa H, Jain R, Ward AR (2015) A queueing model with independent arrivals, and its fluid and diffusion limits. *Queueing Systems* 80(1-2):71–103.
[23] Horvath L (1984) Strong approximation of renewal processes. *Stochastic Processing Appl.* 18(1):127–138.
[24] Jain R, Juneja S, Shimkin N (2011) The concert queueing game: To wait or to be late. *Discrete Event Dynamic Systems* 21(1):103–138.
[25] Kim SH, Whitt W, Cha WC (2018) A data-driven model of an appointment-generated arrival process at an outpatient clinic. *INFORMS J. Comput.* 30(1):181–199.
[26] Ko YM, Pender J (2018) Strong approximations for time-varying infinite-server queues with non-renewal arrival and service processes. *Stochastic Models* 34(2):186–206.
[27] Komlós J, Major P, Tusnády G (1975) An approximation of partial sums of independent rv'-s, and the sample df. i. *Z. Wahrscheinlichkeitstheor. Verwandte Geb.* 32(1-2):111–131.
[28] Louchard G (1994) Large finite population queueing systems. The single-server model. *Stochastic Processing Appl.* 53(1):117–145.
[29] Mandelbaum A, Massey WA (1995) Strong approximations for time-dependent queues. *Math. Oper. Res.* 20(1):33–64.
[30] Mandelbaum A, Pats G (1995) State-dependent queues: Approximations and applications. *Stochastic Networks* 71:239–282.
[31] Mandelbaum A, Ramanan K (2010) Directional derivatives of oblique reflection maps. *Math. Oper. Res.* 35(3):527–558.
[32] Mandelbaum A, Massey WA, Reiman MI (1998) Strong approximations for Markovian service networks. *Queueing Systems* 30(1–2):149–201.
[33] Newell C (2013) *Applications of Queueing Theory*, vol. 4 (Springer Science & Business Media, Berlin).
[34] Oksendal B (2013) *Stochastic Differential Equations: An Introduction with Applications* (Springer Science & Business Media, Berlin).
[35] Rosenkrantz WA (1980) On the accuracy of kingman's heavy traffic approximation in the theory of queues. *Probability Theory Related Fields* 51(1):115–121.
[36] Strassen V (1964) An invariance principle for the law of the iterated logarithm. *Z. Wahrscheinlichkeitstheor. Verwandte Geb.* 3(3): 211–226.
[37] Wainwright MJ (2019) *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, vol. 48 (Cambridge University Press, Cambridge, UK).
[38] Whitt W (2016) Heavy-traffic limits for a single-server queue leading up to a critical point. *Oper. Res. Lett.* 44(6):796–800.
[39] Whitt W (2018) Time-varying queues. *Queueing Models Service Management* 1(2):79–164.