Fourier Trajectory Analysis For System Discrimination

Lucy E. Morgan,

Lancaster University Management School, Lancaster University
Lancaster, LA1 4YR, UK

Russell R. Barton

Department of Supply Chain and Information Systems, The Pennsylvania State University
University Park, PA 16802, USA

ABSTRACT

With few exceptions, simulation output analysis has focused on static characterizations, to determine a property of the steady-state distribution of a performance metric such as a mean, a quantile, or the distribution itself. Analyses often seek to overcome difficulties induced by autocorrelation of the output stream. But sample paths generated by stochastic simulation exhibit dynamic behavior that is characteristic of system structure and associated distributions. In this technical report, we investigate these dynamic characteristics, as captured by the Fourier transform of a dynamic simulation trajectory. We find that Fourier coefficient magnitudes can have greater discriminatory power than the usual test statistics, and with simpler analysis resulting from the statistical independence of coefficient estimates at different frequencies. Theoretical and Empirical results are provided.

1 Introduction

Until recently the norm in stochastic simulation has been to perform experiments that produce long-run performance measures averaged over replications such as the expected number in a system. Although useful these measures ignore the dynamic behavior within each run of a simulation. For example averaging over the expected number in a system does not provide information on potential peaks and troughs in the numbers through time (Fishman and Kiviat, 1967). Dynamic simulation behavior can be better observed by considering simulation trajectories. We exploit the dynamic behavior of simulation trajectories using weighted Fourier magnitudes. Our illustrations indicate that superior discriminatory power can be gained by utilizing dynamic system behavior when compared with a long-run performance measure based batch means statistic.

In this paper we examine properties of the Fourier transform applied to the number-in-system dynamic trajectory output of a stochastic simulation, and present a test for the discrimination of systems using the Fourier coefficient magnitudes. The test presented is predominantly useful for identifying a difference between dynamic systems, rather than estimating a property of the steady-state distribution characterizing some performance metric. It is therefore natural to assume that this type of test would be useful for optimization. Its value also obtains for model validation as it allows for deeper comparisons of model vs. system trajectories. The surge in interest for digital twin technologies also prompts the need for methods that enable the detection of the divergence of two systems. The test presented in this paper is a step towards this.

The use of Fourier methodology in simulation is not new, but our use of simulation trajectories for discrimination is novel. Until now only exploratory analysis has been performed for using dynamic system

behavior for system discrimination. In this paper we formalise these early findings and provide new statistics with higher discriminatory power. In Section 2 we discuss where this contribution sits within other uses of Fourier analysis in the simulation literature. In Section 3 we introduce the Fourier representation of a simulation trajectory and statistical consideration of the properties of Fourier magnitudes are presented and evaluated in the context of an M/M/1 queueing model. In Section 4 a test for discrimination is introduced and then evaluated, and in Section 5 we conclude the paper with some comments on future extensions.

2 Background

Output analysis for discrete event simulation has largely focused on static characterizations such as expected value, variance, and quantiles, of output streams such as number-in-system or waiting time. The technique of batch means collected over a single run is designed to eliminate the impact of dynamic behavior through choice of sufficiently long and/or separated batches.

The dynamic behavior of simulation output affects the statistical properties of such estimates. It has been most commonly characterized by the autocovariance of the output stream. For a covariance stationary process, the variance of the sample mean of a sequence of n consecutive observations in a single output stream can be characterized as the average of all n^2 variance/autocovariance terms. Fourier and wavelet modeling of the autocorrelation function (spectral density) has been used to characterize transient behavior (Morse, 1955; Eick et al., 1993; Massey, 2002) and the steady-state mean and variance (Fishman and Kiviat, 1967; Fishman, 1971; Heidelberger and Welch, 1983; Lada et al., 2007).

Autocovariances are difficult to estimate and characterize: the usual estimates are dependent across lags. Further, conditions on the autocovariance are required for the existence of finite frequency domain characterizations of steady-state behavior via the Fourier transform of the estimated autocorrelation (Parzen, 1957). Consistent frequency domain characterizations must be obtained by means other than increasing the time interval T, since for a general class of stochastic functions, the variance of the estimated Fourier coefficient tends to a constant as the size of the interval (and thus the discrete number of samples available) tends to infinity (Bartlett, 1950; Parzen, 1961). Estimates for different frequencies are statistically independent, and this motivates our use of averages of large numbers of Fourier transform coefficient magnitudes to gain discriminatory power.

Fourier methods have also been used to study renewal intensities and other transition probabilities for queues and other Markov chains (Feller, 1966). In some cases time series models have been used to construct confidence intervals and test statistics for static properties such as the steady-state mean of of a performance metric. Steiger et al. (2005) fit an AR(1) model to appropriately chosen batch means to calculate confidence interval based on an autocorrelation adjusted variance.

Nelson (2016) suggested the need for the field of simulation analytics to place greater emphasis on the statistical analysis of the entire simulation data series, characterizing behavior over time, rather than solely on steady-state characterizations. Work in this area has focused on dynamically varying forecasts of system behavior, such as expected value of waiting time conditional on number-in-system or time of day, or indicator variables for expected system behavior, given current state, which is time-dependent and thus dynamic (Lin et al., 2019; Laidler et al., 2020). These statistics characterize behavior at an instant that changes dynamically over time: a *dynamic statistic* of *static behavior*. Our approach has been to characterize the dynamic behavior of a simulation output stream over an interval of time: a *static statistic* of *dynamic behavior*. Dong and Whitt (2015) provide an interesting alternative: modeling the dynamic behavior of number-in-system over time by a fitted state-dependent birth-death process.

Dynamic output streams from discrete event simulation models have been characterized through Fourier analysis - for sensitivity analysis and gradient estimation using sinusoidally varied simulation inputs (Schruben and Cogliano, 1987; Morrice, 1995; Jacobson et al., 1991; Sargent and Som, 1992). In that research, simulation input parameters were deliberately varied periodically to determine sensitivity of simulation output to the input parameters. By varying different input parameters sinusoidally at (carefully chosen) different frequencies, sensitivities for multiple parameters could be determined from only two simulation runs.

The index used for the discrete Fourier calculations in that research was typically an entity index. Input parameters were varied with the entity index, and discrete-time output statistics were analyzed based on the index of the associated entities. Jacobson et al. (1988) did examine using simulation clock time for the driving

frequencies, but still analyzed output using an entity index. This makes interpretation problematic, since the entity indices are spaced unevenly in time. Although it is possible to construct the Fourier transform for such data (Scargle, 1982), popular software implementations of the FFT expect evenly spaced signal data. Hazra et al. (1997) used fixed-time increments, binning arrival and departure events to characterize the frequency response of an M/G/1 queue. Eick et al. (1993) and Green et al. (1991) examined queue trajectories with sinusoidal variation of arrival rates as well, characterizing the impact on the usual steady-state performance measures as well as the mean behavior conditioned on time.

We do not seek sensitivity information, and do not sinusoidally vary parameters for that purpose, rather we seek to characterize the dynamic behavior of a system by its Fourier transform signature, for use in discriminant analysis, optimization, validation and process monitoring. In this work we focus on the number in system, a continuous time quantity, and sample the values at high frequency at equally spaced intervals.

A prior computational study using Fourier analysis for system discrimination can be found in Wu and Barton (2016). In that work, averages of sets of high-frequency Fourier transform coefficient magnitudes for number-in-system dynamic data were shown to be more effective in discriminating between high- and low-intensity queues. The following sections provide theoretical explanation for those findings, and, based on this more formal characterization, provide new statistics with greater discriminating power.

3 Fourier Representation for Simulation Output

We consider a dynamic simulation output as some function over time, $f(t)$. Continuous-time simulation outputs have a natural representation. For example, the number-in-system, say $L(t)$, is piecewise constant, taking generally nonnegative integer values, which may change at each event time. Discrete-time statistics such as time in system are indirect functions of time: the time in system for an entity departing at time t is based on delays occurring at times before t. It may be possible to construct continuous-time representations for discrete-time data from statistical models as in Lin et al. (2019) or Pedrielli and Barton (2019).

A function can be represented in both time and frequency domains, its representation on each domain is linked by the Fourier transform. The Fourier transform for a function $f(t)$ is

$$c_f(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt, \tag{1}$$

where i represents $\sqrt{-1}$. For any frequency, ω, $c(\omega)$ gives the projection of the (possibly complex-valued) function onto the complex sinusoid $\cos(\omega t) - i\sin(\omega t)$ where frequency, ω, is typically measured in radians. Here $c(\omega)$ represents the projection coefficient. A shift in time of t_0 changes the real and complex components of $c(\omega)$: $c_{t0}(\omega) = c(\omega)e^{i\omega t_0}$ but does not change the magnitude since $|e^{i\omega t_0}| = 1$.

Although the queue length function and other continuous-time statistics of simulation output are, as their name indicates, valued continuously in time (though not continuous functions of time), we represent the trajectory as a discrete time series with equal spacings. A small example portion is shown in Figure 1. In this case, Fourier characterization is via the the Fast Fourier transform (FFT) implementation of the discrete Fourier transform (DFT). The DFT characterizes a discrete set of frequencies, ω_k, with corresponding complex coefficients c_k where ω_k denotes k cycles/$T = 2\pi k$ radians/T.

Epstein (2005) shows that the discrete Fourier transform coefficient $\{c_k\}$ can be considered a Riemann sum approximation for the Fourier transform coefficient $c(\omega_k)$ for frequency k cycles/T. Theorem 4.1 of Epstein (2005) gives the following bound on error for DFT coefficients from their FT counterparts.

Theorem 1. *Consider a piecewise continuous function $f(t), t \in [-T/2, T/2]$ with N jumps in $[-T/2, T/2]$ of size $\alpha_q t_q + \beta_q$ for $q \in \{1, \ldots, N\}$. A discrete Fourier transform of $n = 2l+1$ samples of $f(t)$ evenly spaced over $[-T/2, T/2]$ approximates the Fourier transform coefficients $c(\omega), \omega \in [-\pi, \pi]$ with error bounded by*

$$|c_k - c(\omega_k)| \leq 6mc_f\left(\frac{1}{2\pi l}\right) + \frac{C\sum_{q=1}^{N}(|\alpha_q| + |\beta_q|)}{n}, \forall |k| \leq l,$$

where $mc_f(\frac{1}{2\pi l})$ is the modulus of continuity of the continuous portion of the trajectory for t spacings at most $\frac{1}{2\pi l}$, and where $C \approx 10$.

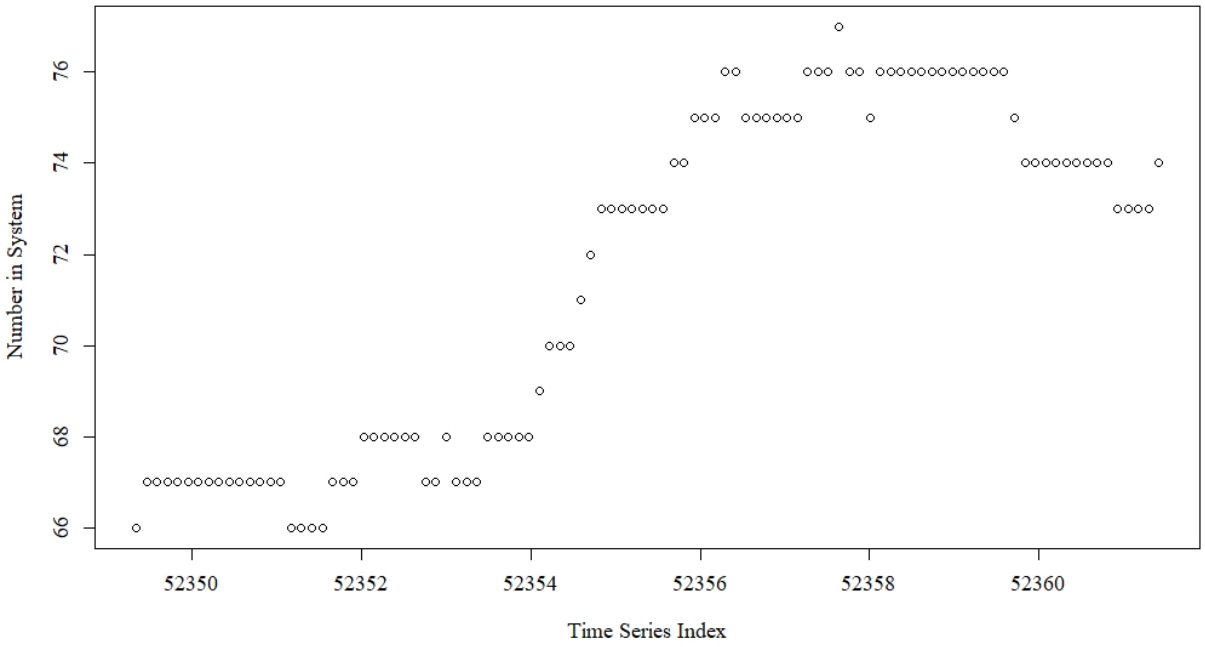

Figure 1: Discretization of number-in-system continuous time data for an M/M/1 queue, $\rho = .95$.

In the development below we derive the distribution of Fourier coefficient magnitudes for a 'busy' segment of a number-in-system trajectory, i.e., a time interval where the number-in-system never becomes zero. For such trajectory data, the bound simplifies.

Corollary 1. *Consider a busy period number-in-system trajectory $f(t), t \in [-T/2, T/2]$ with N_1 jumps of +1 and N_2 jumps of -1 and $N_3 = N_1 + N_2$. A discrete Fourier transform with $n = 2l + 1$ samples evenly spaced over $[-T/2, T/2]$ approximates the Fourier transform coefficients $c(\omega), \omega \in [-\pi, \pi]$ with error bounded by*

$$|c_k - c(\omega_k)| \leq \frac{CN_3}{n}, \forall |k| \leq l,$$

where $C \approx 10$.

This follows from i) the continuous portion of the trajectory is piecewise constant, making $mc_f(\frac{1}{2\pi l}) = 0$, and ii) $|\alpha_q| = 0$ and $|\beta_q| = 1$. The implications of the corollary are that, to use the FFT approximation up to a frequency ω_{kmax}, we need to make l sufficiently large for the error $CN_3/(2l + 1)$ to be small relative to $|c_{kmax}|$.

As is typical for Fourier analysis applied to time series, we analyze a time series over a fixed period. The connection between the Fourier transform for this finite series and the Fourier series for its infinite periodic extension is direct if the length for the series analyzed by DFT, and effectively the period of the periodic extension, T, exceeds the length of non-zero entries in the segment that is represented (Stankovic et al., 2014). In some of the analysis below, we estimate the Fourier transform for aperiodic signals by the usual approach of using the limiting periodic calculation as $T \to \infty$ (Stankovic et al., 2014).

3.1 Characteristics of the Discrete Fourier Transform

Consider a simulation trajectory over a finite time window $f(t), a \leq t \leq b$. In Fourier analysis this is often referred to as a *signal*. The width of the signal window $b - a$ is often denoted T, and often it is convenient to rescale to a 'standard' time window $a = -\pi, b = \pi$ and so $T = 2\pi$. The discrete Fourier transform assumes

that the interval is divided into n segments, with the segment endpoints having values $t_j, j = 0, \ldots, n$. In constructing the Fourier transform, the assumption is that the signal is periodic over the window, so that $f(t_n) = f(t_0)$. In computing the discrete Fourier transform, the final sample point $f(t_n)$ is not included, since the value is assumed to be a return to the first value.

In this setting, frequencies are usually thought of in terms of cycles/T, or radians/T. So one cycle per time period T is 2π radians/T for the standard time window.

The discrete Fourier transform (DFT) is represented by a discrete set of frequencies, ω_k and corresponding complex coefficients c_k such that:

$$f(t) \approx p(t) = c_0 + c_1 e^{it} + c_2 e^{2it} + \cdots + c_{n-1} e^{(n-1)it} = \sum_{k=0}^{n-1} c_k e^{ikt}. \tag{2}$$

Each (complex-valued) Fourier coefficient c_k in the DFT represents the magnitude and phase of frequency content at $\omega_k = k$ cycles/$T = 2\pi k$ radians/T. It is found by averaging (over sampled values indexed by j) the product $f(t_j)e^{-ikt_j}$:

$$c_k = \frac{1}{n} \sum_{j=0}^{n-1} f(t_j) e^{-ikt_j}. \tag{3}$$

In fact, the positive representation of the signal (from 0 to 2π) leads to high frequencies in the top half of k in equation (2) and so a better representation is to center the time scale about zero time. Alternatively, have the index k in equation (2) run from $-l$ to l (if $n-1$ is even), where $c_{-k} = c_{n-k}$ from equation (3).

Several properties of the Fourier transform make it attractive for characterizing system trajectories. These can be found in any text on Fourier analysis, e.g. Stankovic et al. (2014). While these properties hold for both Fourier transforms and discrete Fourier transforms, the third property is stated in terms of the discrete Fourier transform to simplify notation.

Property 3.1. Unchanged magnitude for time-shifts: If $c_f(\omega)$ represents the (complex) Fourier coefficient for $f(t)$, then the corresponding Fourier coefficient for $g(t) \equiv f(t - t_0)$ is given by $c_g(\omega) = c_f(\omega)e^{-it_0\omega}$. Since $|e^{-it_0\omega}| = 1, |c_g(\omega)| = |c_f(\omega)| \; \forall \; t_0, \omega$.

Property 3.2. Linearity: $c_{af_1+bf_2}(\omega) = ac_{f_1}(\omega) + bc_{f_2}(\omega) \; \forall \; a, b, \omega$. (Note that a and b may be complex.)

Property 3.3. Orthogonality: Let $\langle \cdot, \cdot \rangle$ denote the complex inner product (this reduces to the usual inner product for real vectors), e_k be the vector $[e^{-ikt_0}, \ldots, e^{-ikt_{n-1}}]$ in equation 3. Then for any k, k', $\langle e_k, e_{k'} \rangle = n$ if $k = k'$, $= 0$ otherwise. Note that in these terms, equation 3 can be given as $c_k = \langle f(\cdot), e_k \rangle$, where $f(\cdot)$ is the vector $[f(t_0), \ldots, f(t_{n-1})]$. So each Fourier coefficient is a linear combination of signal values, and the e_k vectors are orthogonal.

The time-shift property is an advantage for steady-state trajectories, and in contrast to wavelet transforms. It implies that the location of a particular pattern in the trajectory does not matter; the signature in terms of Fourier coefficient magnitudes remains the same. The linearity property means that we can characterize the coefficients of the Fourier transform of a composite trajectory as the sum of the coefficients of its component parts. This is important in the analysis that follows. Finally, the orthogonal basis implies that, when applied to signals with random components, the Fourier coefficients are statistically independent. This is in contrast to autocorrelations estimated from a stochastic signal, which are dependent. Together, these properties provide an intuitive interpretation of the frequency content of the trajectory signal: each magnitude $|c(\omega)|$ provides the sinusoidal frequency content in the trajectory signal at the corresponding frequency ω.

The discretization of the signal makes the DFT an approximation to the Fourier transform for an underlying continuous-time signal. For the continuous formulation substitute $c(\omega)$ for c_k in in equation (2) and (3) and replace the sums with integrals over ω and t, respectively. The bound for FFT error is quite good up to the Nyquist frequency ($n/2$ cycles/T) for continuous functions, and up to a smaller level $<< n/6$ for piecewise continuous functions (Epstein, 2005), keeping in mind the magnitude requirement for n imposed by Corollary 1. Further, because the DFT and Fourier coefficients are linear functions of the data, the Fourier or DFT coefficients for a sum of two signals are just the sum of the individual Fourier coefficients.

Another effective transform family for time series signal data is the wavelet transform (Chui, 1997). Because our focus is on the discrimination between steady-state signal trajectories, the usual advantages

of combined time- and frequency-location capability of wavelet transforms do not obtain. In this setting, wavelets have been shown to be comparable or inferior to Fourier transforms in signal characterization (Houtveen and Molenaar, 2001; Koo et al., 2011).

3.2 Statistical characteristics for noisy data

We will assume that a simulation output trajectory is composed of a dynamic signal that characterizes the process that is modeled, with Fourier coefficients d_k, plus additive random noise, say with Fourier coefficients e_k. There is no assumption of independence at this level. Because of the sum-of-signals property, each estimated Fourier coefficient can be thought of as a sum of the process coefficient plus a noise coefficient. When this noise is independent identically distributed across elements of the signal vector with variance σ^2, the central limit theorem guarantees that the Fourier coefficients will be independent and approximately normally distributed (Freedman and Lane (1980)) with, for any value of k:

$$\mathrm{E}(c_k) = \mathrm{E}(d_k) + \mathrm{E}(e_k) \text{ and } \mathrm{Var}(c_k) = \mathrm{Var}(e_k) = \sigma^2/n. \tag{4}$$

The mean and variance for a particular c'_k of interest then depends on e'_k, whose mean and variance can be estimated from relatively high-frequency Fourier coefficients, $k > k_{max}$, where k_{max} indicates the highest frequency with meaningful dynamic signal content, and one would expect higher frequencies to have roughly similar magnitudes and variance.

3.3 Statistical characteristics for number in system data

The number-in-system, $L(t)$, for a queueing model is a continuous time statistic and varies stochastically, but it does not follow an additive Gaussian model. Especially for situations where values typically fall in single digits, the discrete nature of $L(t)$ immediately violates the Gaussian noise assumption. Further, the natural characterization of randomness is with respect to time rather than amplitude.

The following result characterises the the distribution of coefficient magnitudes for the Fourier transform of a number in system trajectory for an M/M/1 queueing system. Supporting properties and Lemmas are provided in Appendix A.

Theorem 2. *Consider an M/M/1 queue with arrival rate λ and service rate μ during a busy interval. If in total N_3 events are observed within this interval, N_1 arrival and N_2 departure events, then the squared Fourier magnitude at frequency ω, $|c_{M/M/1_{busy}}(\omega)|^2$, has expectation*

$$N_3/\omega^2 \tag{5}$$

with a random zero-expectation perturbation

$$\frac{1}{\omega^2} \sum_{j,k=1, j\neq k}^{N_3} e^{i\omega(t_j - t_k)} \tag{6}$$

Proof. Conditioning on a long non-empty period T, the number-in-system trajectory, $L(t)$, can be viewed as a composition of single-unit step functions that occur at random times. By the superposition property of Poisson processes and Lemma 5, the random events operate as a Poisson process with rate $\lambda + \mu$.

Over a fixed time period T the number of positive jumps has a Poisson distribution with mean λT and the number of negative jumps follows a Poisson distribution with mean μT. Thus using the additive property of the Fourier transform, the Fourier coefficient at frequency ω is:

$$c_{M/M/1_{busy}}(\omega) = \frac{1}{i\omega}\left(\sum_{j=1}^{N_1} e^{-i\omega t_j} - \sum_{j=1}^{N_2} e^{-i\omega t_j}\right), \tag{7}$$

where $N_1 \sim \text{Poisson}(\lambda T)$ and $N_2 \sim \text{Poisson}(\mu T)$, and $N_3 = N_1 + N_2$.

As each event time t_l for $l = 1, 2, \ldots, N_3$ is independent and generated by a Poisson process the components of the sum in Equation 7 have independent, uniformly distributed orientations, θ_l, on the unit circle.

They can therefore be combined into a single summation with N_3 elements. Using Lemma 4 the expectation of the squared magnitude of this sum equals N_3, and thus the squared Fourier magnitude at frequency ω, $|c_{M/M/1_{busy}}(\omega)|^2$, has expectation

$$N_3/\omega^2$$

with a random zero-expectation perturbation

$$\frac{1}{\omega^2} \sum_{j,k=1,j\neq k}^{N_3} e^{i\omega(t_j - t_k)}.$$

□

The quantity in (6) has mean zero, and the elements are independent. This does not invalidate the assumption that computed Fourier coefficients will be approximately Gaussian. If T is large relative to $1/\lambda$ and $1/\mu$, then N_3 will tend to be large, and the sum in (6) might be approximated by a normal distribution. Further, different values of ω rotate these random vectors and their spacings around the unit circle, which will result in different characteristics for the unit vectors in the sum. Except for the very lowest adjacent frequencies, the result should be substantial independence of the randomness at one frequency to that at another. Thus with Poisson N_3, the quantity would follow a type of compound Poisson distribution, for large N_3.

Note that the expected value of the vector magnitude will be smaller than the rms value. For large N the approximate density is $2re^{-r^2/N}$, and $\mathrm{E}[|c_{M/M/1_{busy}}(\omega)|]$ can be approximated by $\sqrt{\frac{\pi}{4}N_3}/\omega \approx \sqrt{.785N_3}/\omega$, see (Rayleigh (Strutt), 1905).

3.4 Empirical Tests with the M/M/1 Queue

It is possible to check these characteristics empirically. For this purpose, we computed FFTs for signal data from an M/M/1 queue over a long busy period, T_{busy} time units, where T_{busy} was set large relative to $1/\lambda$ and $1/\mu$. The run length of each simulation was made much longer, in order to have a good chance of a busy period greater than or equal to the required length. The longest busy period from each replication was accepted if the length met the T_{busy} requirement. The middle interval of length T_{busy} was used from each identified busy period, so that all would have the same length, and so the period would not enter into the model calculations. We repeated this computation multiple times, and used the magnitudes and squared magnitudes of the Fourier coefficients from the replications in the following way.

For a representative set of indices $\{k\}$ we found the mean squared magnitude/s at frequency $\omega_k = 2\pi k$ corresponding to the FFT coefficient c_k by averaging the magnitudes and squared magnitudes over the replications. We checked these against the expectation for model (7): $\sqrt{.785\,\mathrm{E}(N_3)}/\omega$ and $\mathrm{E}(N_3)/\omega_k^2$, respectively, where $\mathrm{E}(N_3) = T(\lambda + \mu)$. To reduce bias, we used $\widehat{\mathrm{E}(N_3)}$ in place of $\mathrm{E}(N_3)$, where $\widehat{\mathrm{E}(N_3)}$ was the observed N_3 averaged over the replications. We examined utilizations of $\rho = .75$ and $\rho = .95$. For the higher utilization it was important to use a longer busy period, here $T_{busy} = 4000$ was used. When using shorter periods, the Fourier coefficients for $\rho = .95$ exhibited bias, likely due to the high bias from a relatively short observational period. Parameter estimation for M/M/1 queues is notoriously difficult at high utilizations; for estimating queue length to within 5% with $\rho = .95$ requires a run length of more than 500,000 mean service times. See the derivations in Whitt (1989a) or Whitt (1989b). For the lower utilization of $\rho = .75$, busy periods of such length are unlikely; but, the statistical requirements are less. A busy period of $T_{busy} = 200$ provided a good empirical fit to the model predictions for the lower utilization.

Figure 2 shows the average (across 100 replications) of Fourier coefficient magnitudes, $|c_k|$, and squared magnitudes, $|c_k|^2$, for both $\rho = .75$ and $\rho = .95$ over the representative set of indices $\{k = 1 : 500\}$. The model fits appear to be very good. A more discerning set of plots (of model values vs. actual values) are shown in Figure 3. For these plots the expectation is a scatter of points around the actual value = model value, i.e. a linear relationship. Again, the fits look quite good, but in both figures, heterosckedasticity of variance of the fitted Fourier magnitudes is apparent.

Next we examine the distribution of (6) by computing the differences $|c_{kr}|^2 - \mathrm{E}(N_{3r}/\omega_k^2)$, where r denotes a specific replication. Since for each replication, r, the sum in (6) will generally have a different number of elements, the differences across replications are typically samples from slightly different approximately

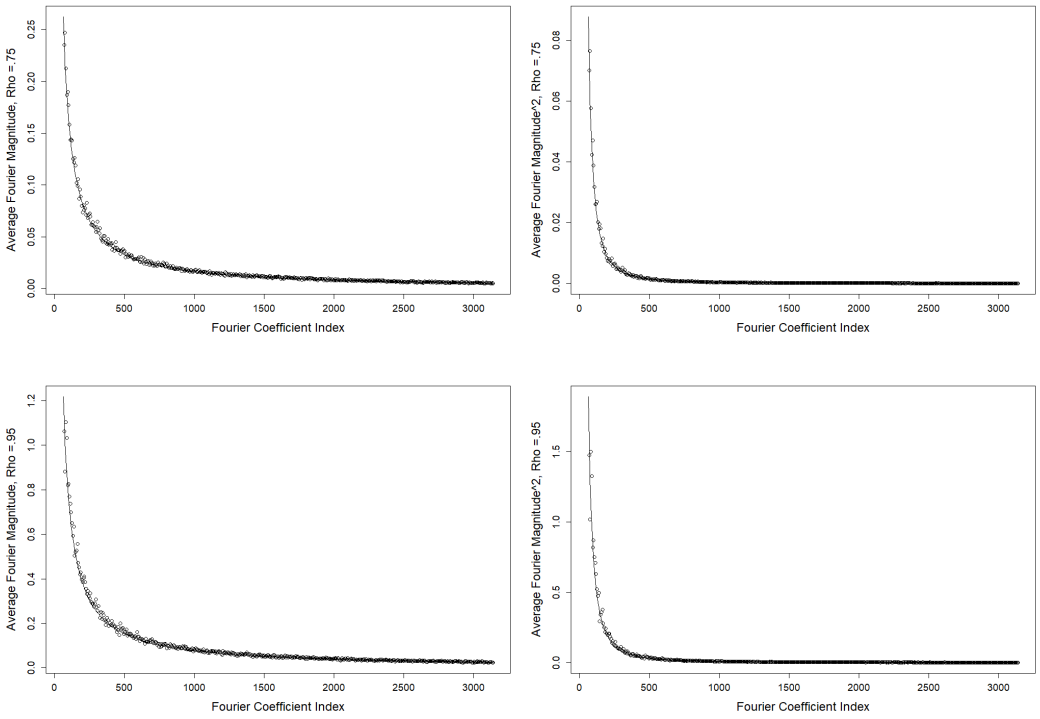

Figure 2: Model (7) and actual coefficient magnitudes and squared magnitudes, $\rho = .75$ and $\rho = .95$.

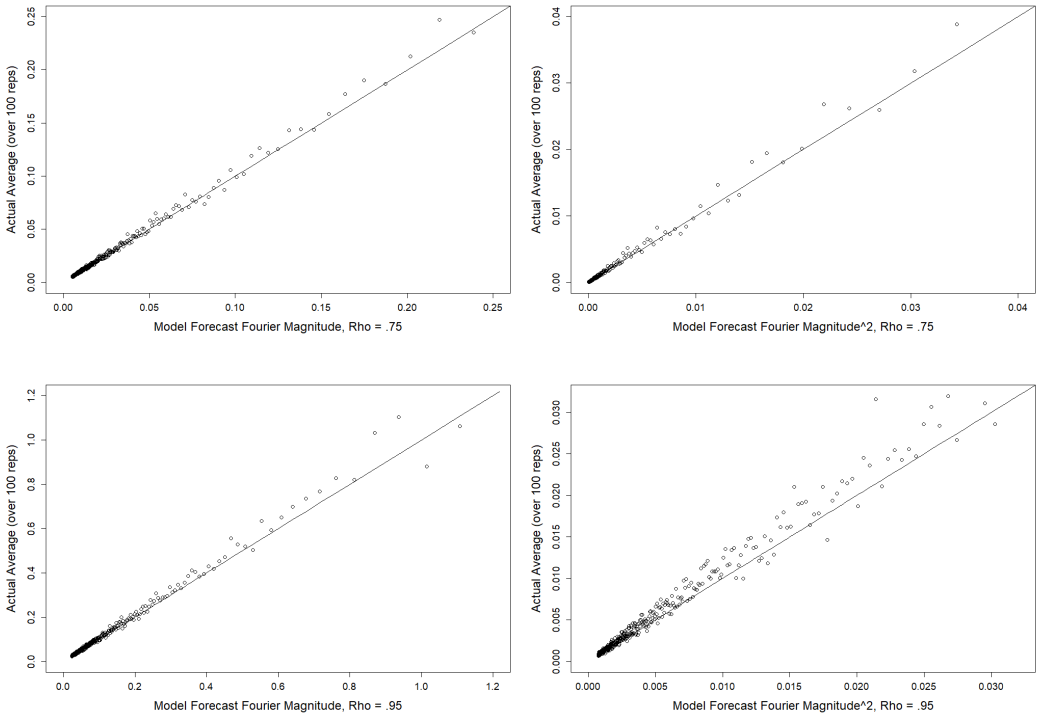

Figure 3: Actual versus Model (7) coefficient magnitudes and squared magnitudes, $\rho = .75$ and $\rho = .95$.

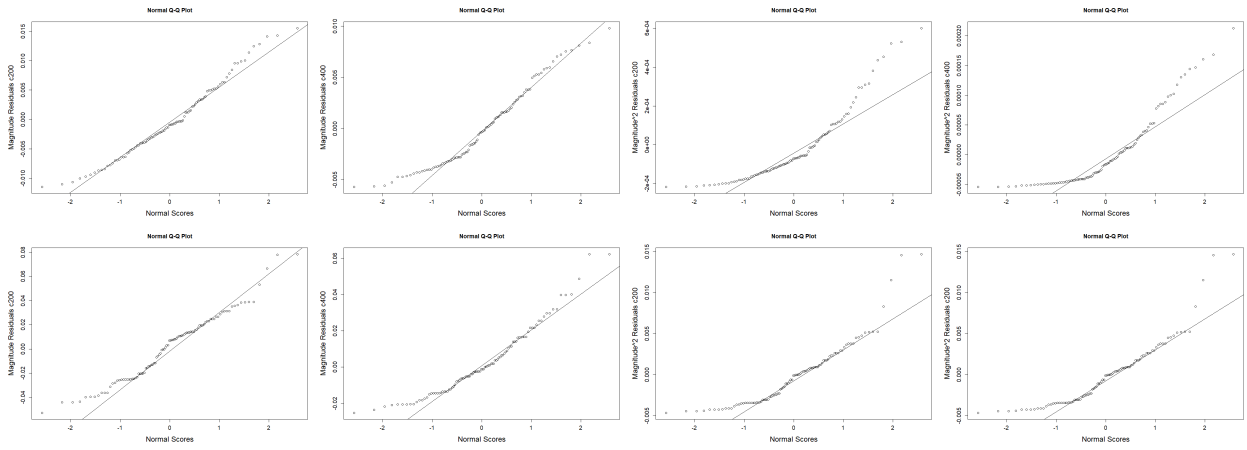

Figure 4: Normal probability plots for (actual - model) coefficient magnitudes and squared magnitudes for c_{200} and c_{400}; $\rho = .75$ (upper) and $\rho = .95$ (lower).

Gaussian distributions. An adjustment could be made to take into account the differing values of N_{3r}, but the standard deviation of the Poisson distribution is only 5.3% of the mean count for $\rho = .75$ and $T_{busy} = 200$, and only 1.1% of the mean count for $\rho = .95$ and $T_{busy} = 4000$, so this adjustment seems unnecessary.

Figure 4 shows normal probability plots for the error between model prediction and observed values at two Fourier magnitudes, c_{200} and c_{400}, over $G = 100$ replications. The residuals for the squared magnitudes look significantly non-normal, which suggests two things: i) the Gaussian assumption is more appropriate for variation in Fourier coefficient magnitudes, rather than squared magnitudes, and ii) for more Gaussian squared magnitude variations, the busy interval length, T_{busy}, for the less congested system must be increased from 200.

To investigate the assumption of independence of the Fourier magnitudes, we examine the correlation of Fourier coefficient magnitudes across two selected sets of frequencies: $k = 200 : 220$ and $k = 400 : 420$. The correlation plots in Figure 5 show generally low empirical correlations (narrow pie slices), supporting the expected independence assumption.

A visual examination of the normal probability plots in Figure 4 show that the standard deviation of variation, (6), for magnitude c_{200} at $\rho = .75$ is approximately 0.0063, and 0.0045 for magnitude c_{400}. This provides some illustration of the heteroskedasticity across magnitudes. More formally, however, we checked the assumption that the standard deviation of the estimated Fourier magnitudes scale as $1/\omega$. Figure 6 shows plots of the standard deviation of the c_k estimates across replications for $k = 1 : 500$, scaled by $\omega_k = 2\pi k$. It is clear that this scaling effectively corrects for the heteroskedasticity. In Section 4 we exploit this scaling to formulate a test statistic within a hypothesis test for the discrimination of two M/M/1 systems.

3.5 Fourier Magnitudes During General Intervals

In the last section all empirical results were presented using busy periods of the trajectory. In reality although long busy periods do occur within trajectories they are not a realistic view of general behavior. It is more difficult to characterize the behavior of Fourier coefficient magnitudes for general intervals, where the number-in-system frequently remains at zero for periods of time.

We bound the Fourier coefficient magnitudes for number-in-system trajectories for M/M/1 and M/G/1 systems for general periods by examining what happens as the utilization goes to zero. We assume $\mu = 1$ remains fixed (it is always possible to choose the time scale so that the mean is 1), and $\lambda \to 0$. Recall, for the M/M/1 system, equation (7), applies to both busy and general time intervals. The difference for general intervals is i) the distribution of the number of arrivals and service events in [0,T] and ii) dependence of times of arrivals and departures for individual entities. We will explore this by considering renewal intervals. Note that our definition of busy period does not begin and end with zero (or any other fixed number) in the system, but is chosen as a (small) segment of a general busy period. It is not equivalent to a renewal

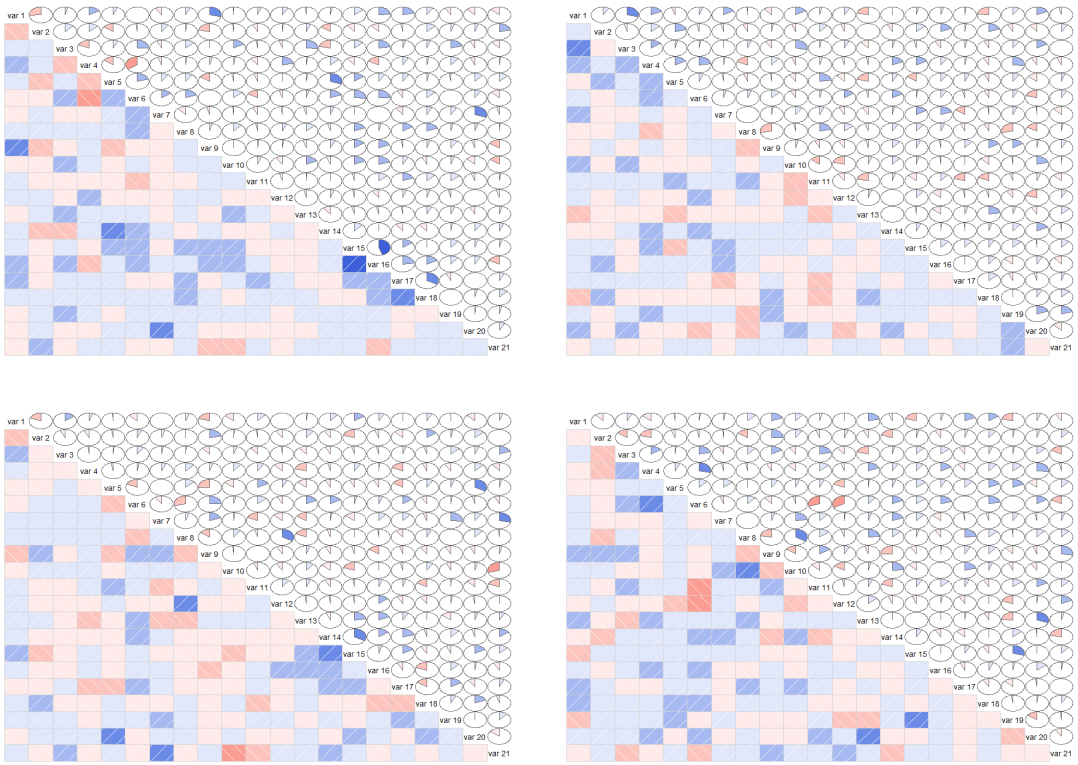

Figure 5: Selected correlations of Fourier coefficient magnitudes, $k = 200 : 220$ (left), $k = 400 : 420$ (right), $\rho = .75$ (upper) and $\rho = .95$ (lower).

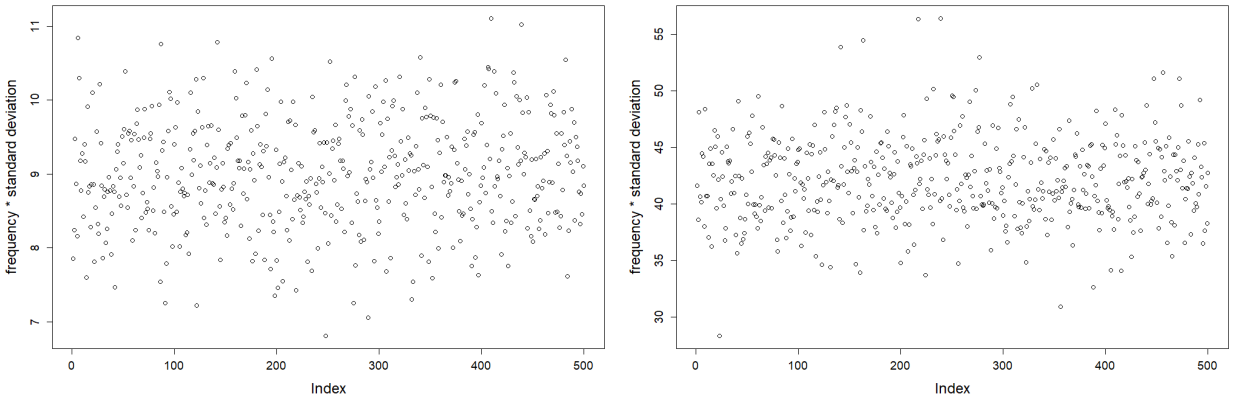

Figure 6: Plots of $\omega_k = 2\pi k$ times the across-replicate sample standard deviation of the magnitude of c_k, $k = 1 : 500$, for $\rho = .75$ (left) and $\rho = .95$ (right).

interval. Recall

$$c_{M/M/1_{busy}}(\omega) = \frac{1}{i\omega}\left(\sum_{j=1}^{N_1} e^{-i\omega t_j} - \sum_{j=1}^{N_2} e^{-i\omega t_j}\right),$$

and

$$\mathrm{E}(|c_{M/M/1_{busy}}(\omega)|^2) = \frac{1}{\omega^2}\mathrm{E}(N_3 \equiv N_1 + N_2).$$

Consider a renewal interval of length T where the starting and ending states are an empty and idle system. During the renewal interval, suppose that there are N arrivals and N service completions, $N_3 \equiv 2N$. This period might be extended to include multiple renewal periods without change in what follows. Rewriting equation (7) with a single index j, and a_j indicating the arrival time of the j^{th} entity and s_j the time of its departure from the system gives

$$c_{M/M/1_T}(\omega) = \frac{1}{i\omega}\left(\sum_{j=1}^{N} e^{-i\omega a_j} - e^{-i\omega s_j}\right). \tag{8}$$

The paired structure in equation (8) allows calculations that bound the magnitude of $c_{M/M/1_T}(\omega)$, as presented in the following theorem and corollary.

Theorem 3 (Bound on Expected Value of Squared Fourier Coefficient Magnitude, Renewal Interval, M/M/1)**.** *For number-in-system trajectory data from an M/M/1 queue with $\mu = 1$, a composition of renewal intervals of combined length $T \geq T_0 \equiv \frac{-k\pi \ln(\varepsilon_1)}{\varepsilon_2}$ can be chosen to give, for any frequency $\omega = \omega_k = 2\pi k/T$, a bound on the expected squared coefficient magnitude of:*

$$\mathrm{E}(|c_{M/M/1_T}(\omega)|^2) \leq \frac{1}{\omega^2} 2\mathrm{E}(N)\left((1-\varepsilon_1)\varepsilon_2^2 + \varepsilon_1\right) < \frac{1}{\omega^2}\mathrm{E}(N_3 \equiv 2N))$$

As the renewal period T gets longer, $\mathrm{E}(N)$ gets larger, but the fraction multiplier gets arbitrarily small. We see that a similar bound holds for the M/G/1 setting, see the proof of 2 in the Appendix B.

Corollary 2. *For number-in-system trajectory data from an M/G/1 queue with $\mu = 1$, as $\lambda \to 0$, a composition of renewal intervals of combined length $T \geq T_0 \equiv \frac{k\pi}{\varepsilon_1\varepsilon_2}$ can be chosen to give, for any frequency $\omega = \omega_k = 2\pi k/T$, a bound on the expected squared coefficient magnitude of:*

$$\mathrm{E}(|c_{M/M/1_T}(\omega_k)|^2) \leq \frac{1}{\omega^2} 2\mathrm{E}(N)\left((1-\varepsilon_1)\varepsilon_2^2 + \varepsilon_1\right) < \frac{1}{\omega^2}\mathrm{E}(N_3 \equiv 2N)$$

Proofs of Theorem 3 and Corollary 2 can be found in the appendix. Note that both the theorem and corollary hold for $\lambda > 0$ although the T_0 requirement depends on the quantile of the time-in-system distribution rather than the service time distribution. Also, the effect of decreasing squared magnitudes is more pronounced for small k (e.g, $k = 1$), since $T_0 \propto k$ for both M/M/1 and M/G/1 cases.

To explore the behavior in computational experiments, it is more convenient to consider general time intervals, with fixed T rather than renewal intervals where T varies. We computed Fourier coefficient magnitude by frequency for a general segment of number-in-system data from a simulated M/M/1 queue with $\rho = .75$. To remove warm up bias, the observed interval was initialised with the number-in-system equal to the mean (4), and was set to be of length $T = 200$ as for the busy period case. Figure 7 shows, over 100 replications, the average magnitude of Fourier coefficients versus the model prediction using $N_3' = N_1 + \rho N_2$. The left plot presents low frequencies $k = 1:500$. The actual magnitudes are well below the model prediction, especially at the lowest frequencies. The right plot presents higher frequencies, $k = 5000:5500$, showing much lower model discrepancy, although the model still has some bias. Residuals for the model were similar to results for the busy period analysis, but at lower frequencies they deviated from normality somewhat, with a shortened lower tail. As for the busy period analyses, and as expected, Fourier magnitudes did not exhibit correlation across frequencies.

Figure 8 shows plots of standard deviation (across 100 replications) of the Fourier magnitudes multiplied by ω. The left plot shows that the smaller than expected means at low frequencies exhibited in Figure 7 are accompanied by small standard deviations, at least at very low frequencies. This condition becomes more evident with larger T (here only 200). The right plot shows that this effect is not apparent at higher frequencies.

These findings have the following implications. Statistical analysis based on ω-weighted means of higher frequency values is justified given the i.i.d. and approximately Gaussian behavior at higher frequencies. Including low frequencies in a sample average would make characterization of the test statistic more difficult. In Section 4 we introduce a hypothesis test for the discrimination of two M/M/1 systems. This test will be evaluated over a number of groups of high and low frequencies to evaluate our findings in relation to the analysis presented in this section. Following the discussion above we expect discriminatory power to be higher when the magnitudes of high frequency components are included in our test statistic.

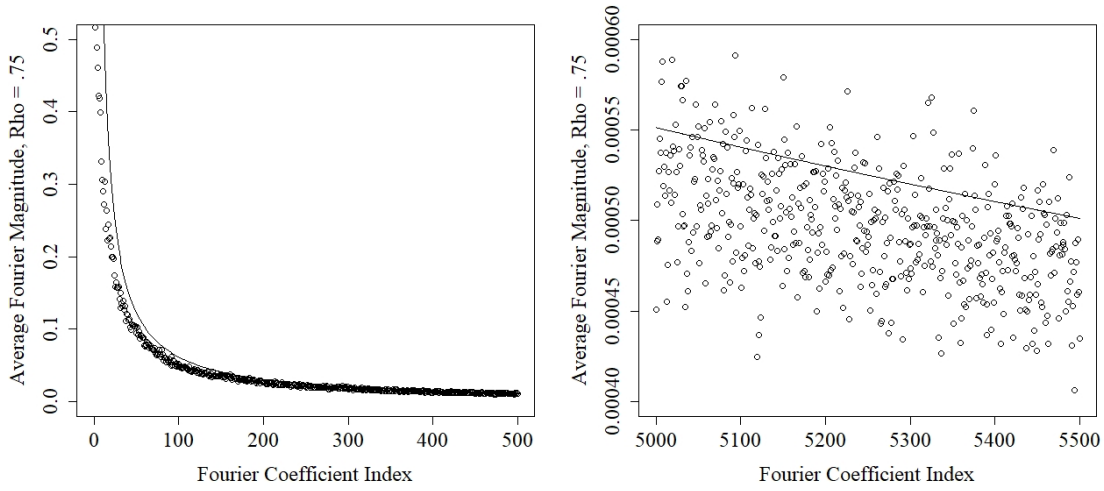

Figure 7: Model (7) with $N_3' = N_1 + \rho N_2$ and actual average coefficient magnitudes for a general segment, $\rho = .75$, $k = 1:500$ and $k = 5000:5500$.

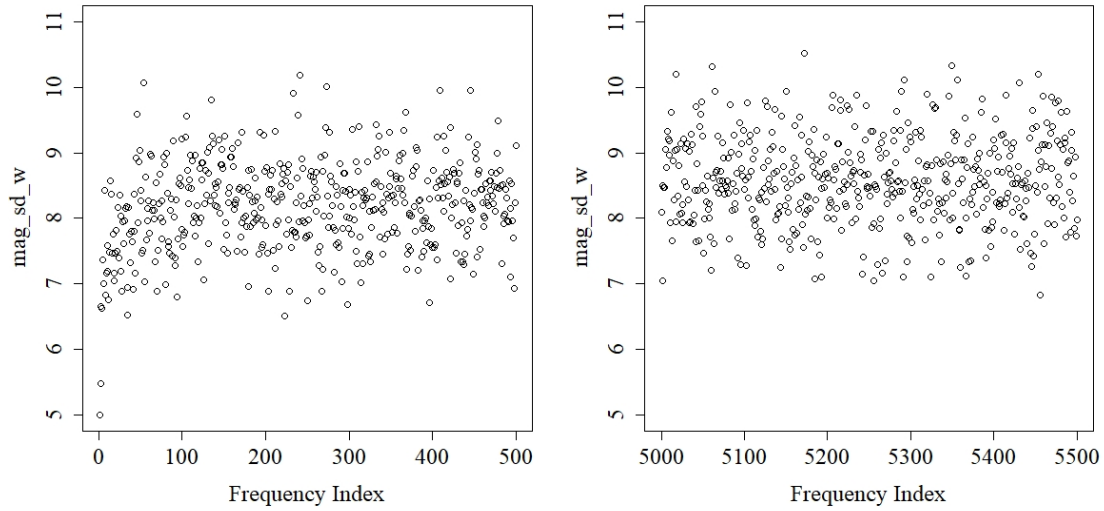

Figure 8: Plots of $\omega_k = 2\pi k$ times the across-replicate sample standard deviation of the magnitude of c_k, $k = 1:500$ and 5000:5500 for $\rho = .75$ and a general segment.

4 Fourier Analysis for Discrimination

In this section we start by introducing a test for the discrimination of systems using the Fourier magnitudes of their sample trajectories. We then illustrate the use of this test for two queueing systems under multiple scenarios. Firstly for discrimination given sample trajectories from two M/M/1 queueing models and second for discrimination given sample trajectories from two M/H$_2$/1 queueing models. Here H_2 denotes a hyper-exponential distribution, in this case referring to hyper-exponentially distributed service times. The hyper-exponential distribution has two rate parameters μ_1 and μ_2, used with probability p and $(1-p)$ respectively, for the generation of service times within the simulation.

4.1 A Test for Discrimination

We construct a family of t-tests for discrimination of two systems (system I and system II) using batch averaged ω-weighted Fourier coefficient magnitudes: $waFm$. The tests differ in which frequencies are included in the weighted average. This family of tests is based on the assumption that if two trajectories were from the same underlying process then their Fourier coefficient magnitudes over the same time increments would follow the same distribution.

We assume that trajectories from both systems are observed over a general interval of total length T, where T is some finite window. We then split this interval into B intervals of equal length, which we shall refer to as batches. The trajectories within each batch are then converted into time series with a common minimum increment. Note that the minimum increment of the time series is chosen such that n, the total number of segments, is a power of 2, and the probability of a multiple events occurring within an interval is small (< 0.05). This gives a time series within each batch that is observed over the same time increments. The implication of this is that when we calculate the Fourier coefficient magnitudes within each batch they correspond to the same set of frequencies.

The requirement for batching comes from (4); the variability of the expected Fourier magnitude depends on a sum of N_3' components, and thus the variability in the number of events that may occur, N_3', must be accounted for when estimating the variance of the expected Fourier magnitudes. To estimate this variability, within batch j, we take the mean of m Fourier magnitudes weighted by the frequency at which they were estimated, $\overline{|c|}_j$ for $j = 1, 2, \ldots, B$

$$\overline{|c|}_j = \frac{1}{m} \sum_{k=i}^{i+m+1} \omega_k |c_{kj}|, \tag{9}$$

and from this calculate a grand sample mean and variance across batches

$$\overline{\overline{|c|}} = \frac{1}{B} \sum_{j=1}^{B} \overline{|c|}_j \quad \text{and} \quad s^2 = \frac{1}{(B-1)} \sum_{j=1}^{B} (\overline{\overline{|c|}} - \overline{|c|}_j)^2.$$

Note that frequencies are limited by the Nyquist rate, so $i + m - 1 \leq n/2$. We consider the hypothesis H_0: There is no difference in the mean of the Fourier coefficient magnitudes between system I and system II against the alternative H_1: There is a difference in the mean of the Fourier coefficient magnitudes between the systems, and propose a t-test with unequal variances assumed (Welch's t-test) to test this hypothesis. This gives test statistic

$$Z = \frac{\overline{\overline{|c|}}_I - \overline{\overline{|c|}}_{II}}{\sqrt{\frac{s_I^2 + s_{II}^2}{B}}}.$$

It is assumed that $Z \sim t_\nu$ where $\nu = ((s_I^2 + s_{II}^2)/B)^2 / \sqrt{(s_I^4 + s_{II}^4)/(B^2(B-1))}$ denotes the degrees of freedom of the test. In Section 4.2 we look at the performance of the $waFm$ test statistic over a number of different frequency ranges.

4.2 Illustrating Test Performance

In this section we compare the batch averaged $waFm$ tests to a traditional batch means based test and a test for discrimination using inter-arrival times by illustrating their power for discrimination.

Let us denote the traditional batch means test statistic by Z', this test statistic considers whether there is a significant difference in the expected number in the system across the two systems. Note that this test does not use dynamic trajectory information but instead compares two systems in terms of an average long-run performance. In general the batch means approach is a method that can be used to estimate the variance of the sample mean of a performance measure when only a single trajectory of the system has been observed, (Nelson, 2013). As in the batch averaged magnitudes approach, the batch means approach splits a trajectory of length T into B equal length batches. The choice of the number of batches, B, given T is a trade off between the number of batches B and the size of the resulting batches. Schmeiser (1982) showed that $10 \leq B \leq 30$ is appropriate when the original trajectory length T is large enough to precisely estimate the true expected number in the system. In this paper we are guided by the analysis of Law and Carson

(1979) for selecting T large enough to formulate the batch means test statistics for both the M/M/1 and $M/H_2/1$ systems. In the following experiments we use B from the best performing batch means test statistic as our choice for B in the batch averaged ω-weighted magnitudes test.

We denote the inter-arrival means test statistic by Z''. Unlike the batch means based test statistics already introduced, for this test we use only the inter-arrival times from both systems observed over the entire trajectory of length T. For a general interval our model for N_3' amounts to $N_3' = N_1 + \rho N_2$ where $\rho = \lambda/\mu$. Taking the expectation of N_3' we find $\mathrm{E}(N_3') = \lambda T + \rho\mu T = 2\lambda T$, which suggests that the arrival rate, λ, drives the discriminatory power of the batch average magnitude based test. For this reason we aimed to compare Fourier magnitude tests for discrimination with a test considering a difference in the mean inter-arrival times alone. Note that the sum of inter-arrival times is a sufficient statistic for the arrival rate λ. If it were true that λ alone was the driver of the discriminatory power of our test we would expect the power of the ω-weighted test statistic to be bounded above by the power of this test. In testing for a difference in the mean inter-arrival times we use Welch's statistic. Note that this assumes that both samples of inter-arrival times are normally distributed which is known to be false; in both systems studied the inter-arrival times are exponentially distributed. As an alternative the F-test for testing for a difference in the mean of exponentially distributed data could be used here, or a Poisson test bases on total arrivals in T, but in practice the p-values of the tests were found to be comparable. It is important to remember that for many systems, recording inter-arrival times and/or the number of arrivals in a given time window T can be more difficult that recording number-in-system, so this test will not be possible in some settings.

We start our illustration of test performance by comparing the type one error rate in the discrimination of two equivalent M/M/1 systems. Both systems were set to have arrival rate $\lambda = 0.75$ and service rate $\mu = 1$. We illustrate how the magnitude based test for discrimination performs over the following frequency ranges $k = \{1:50, 100:150, 1:500, 500:1000, 1:150000, 5000:150000\}$ compared to two batch means tests with $B = 20$ and 40 batches and the test for a difference in mean inter-arrival times. To facilitate a comparison we perform $G = 100$ macro replications of each test using common random numbers and record the absolute value of the test statistics in the form of box plots in Figure 9. In addition to the $waFm$ test introduced in Section 4.1, the batch means tests and the inter-arrival mean test, we also display $G = 100$ test statistics at each frequency range for a batch average magnitude based test where the Fourier magnitudes are not weighted by frequency, ω, as in (9). We denote these test statistics by $\tilde{Z}$. A similar statistic was proposed in the preliminary work of Wu and Barton (2016). As previously mentioned, including low frequencies in the calculation of this unweighted sample average makes the characterization of the test statistic, $\tilde{Z}$, more complex. The statistic is a mixture distribution, with one sample from the distribution at each frequency.

All figures in this section are organized so that box plots of test statistics Z and $\tilde{Z}$ at each set of frequencies are recorded side-by-side. Test statistics of the form Z are represented by green box plots and test statistics $\tilde{Z}$ orange box plots. The box plots for the traditional batch means test statistics, Z', are blue and the inter-arrival means test statistics, Z'', purple. In each box plot vertical dashed grey lines are used to visually separate the tests conducted at different frequencies and the alternative test statistics. To facilitate easier comparison between the batch means and Fourier magnitude test statistics a horizontal dashed grey line has been added to each plot running through the largest median batch mean test statistic.

In Figure 9 we see that most of the test statistics for each test range between 0 and 2, as expected, and appear to have comparable type I error rate, empirical .95-quantiles were all approximately 2.

We will now consider the performance of the tests for discrimination when the two systems under study are not equivalent. We start by considering the discrimination of two M/M/1 queueing models. System I with arrival rate $\lambda = 0.8$ and system II with arrival rate $\lambda = 0.6$. Both systems are set to have service rate $\mu = 1$.

We complete this experiment for trajectories of length $T = 32960$ and $T = 88960$. Law and Carson (1979) report these as values of average sample size (average number to have been served) for achieving low and high average relative precision of batch means estimates for an M/M/1 system with traffic intensity $\rho = 0.8$, and batch size $B = 40$. Note that in systems with service rate of $\mu = 1$ we expect the average sample size to be similar to the length of the observation interval T. Of the traditional batch means test statistics considered, the one with highest discriminatory power was achieved at $B = 40$ in all experiments, so we use $B = 40$ for the $waFm$ test in all experiments. In Table 1 we report the observed trajectory length, T, and the length of each batch, T/B, for each experiment alongside; n, the total number of increments in each batch; T/Bn the increment length; the highest expected inter-event rate from the two systems under

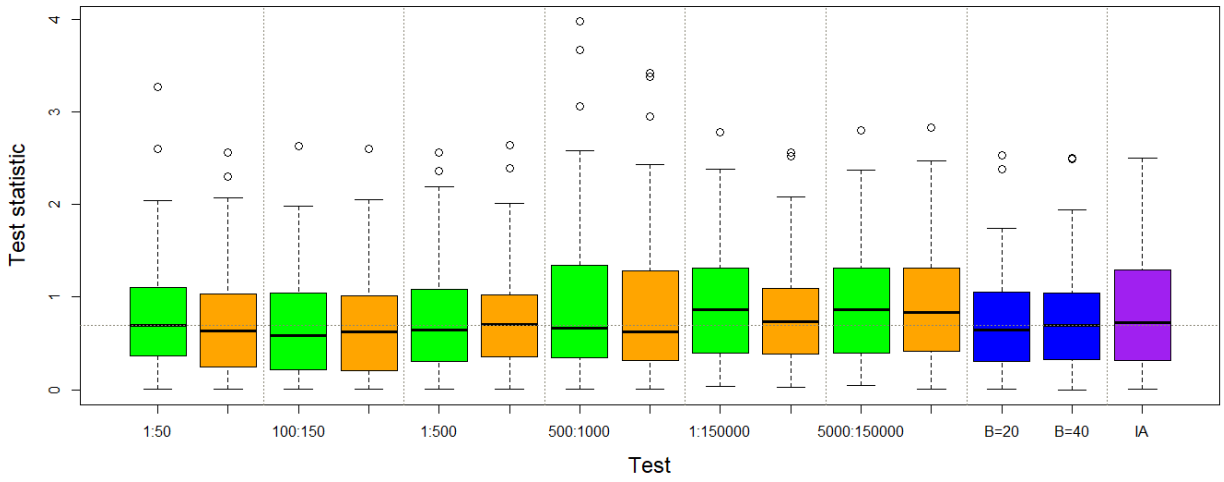

Figure 9: Boxplots of test statistics Z (green), $\tilde{Z}$ (orange), Z' (blue) and Z'' (purple) over $G = 100$ observed trajectories from an M/M/1 queue observed over the period [0, 88960].

consideration, $\lambda + \mu$, and the probability that an inter-event time, X, less than the increment length could occur, $\mathrm{F}(T/Bn) = P(X \leq T/Bn)$. In all experiments the increment length was set so that the probability of an inter-event time smaller than it occurring was less than 5%, i.e. a small percentile of the inter-event distribution, and so that the same range of high frequency magnitudes could be included in the test statistics across all tests.

Figures 10 and 11 show the test statistic values over G =100 replications. Higher test statistic values indicate greater discriminatory power. We see clearly that the Fourier magnitude based test statistics are superior to the traditional batch means statistics as higher frequencies are included. In fact in both experiments the magnitude based statistics (green and orange) have similar or superior discriminatory power than the traditional batch means statistics (blue). Comparing the the ω-weighted and unweighted magnitude based test statistics we see that when low frequencies are not included in the test statistic the power of the two tests is comparable, and when they are included the *omega*-weighting leads to higher power. This is because the weights, ω_k, become increasingly similar at high frequencies. Recall that the ω weighting effectively corrects for heteroskedasticity, known to improve the efficiency of test statistics. As expected the test for a difference in the mean inter-arrival times across the two systems performs very well in both cases.

Observe that there is a drop in power of all tests as the observation interval is shortened from $T = 88960$ in Figure 11 to $T = 32960$ in Figure 10. On closer inspection the median of the ω-weighted Fourier based test statistic drops by 55% in comparison to 60% for the inter-arrival means statistic and 65% for the batch means statistic. This is an indication that the Fourier based test may be robust even when the run length of the simulation is short.

As a second experiment we illustrate the discriminatory power of our test when comparing two $\mathrm{M/H_2/1}$

Table 1: A summary of the increment settings of the time series for each experiment and the probability of an inter-event time less than the increment length.

System	T	T/B	n	T/B	$\lambda + \mu$	$F(T/Bn) = P(X \leq T/Bn)$
M/M/1	32960	824	2^{18}	0.003	1.8	0.005
M/M/1	88960	2224	2^{18}	0.008	1.8	0.014
$\mathrm{M/H_2/1}$	203280	5082	2^{18}	0.019	1.8	0.034
M/M/1	88960	2224	2^{18}	0.008	1.75	0.014
$\mathrm{M/H_2/1}$	10000	250	2^{18}	0.001	1.8	0.002

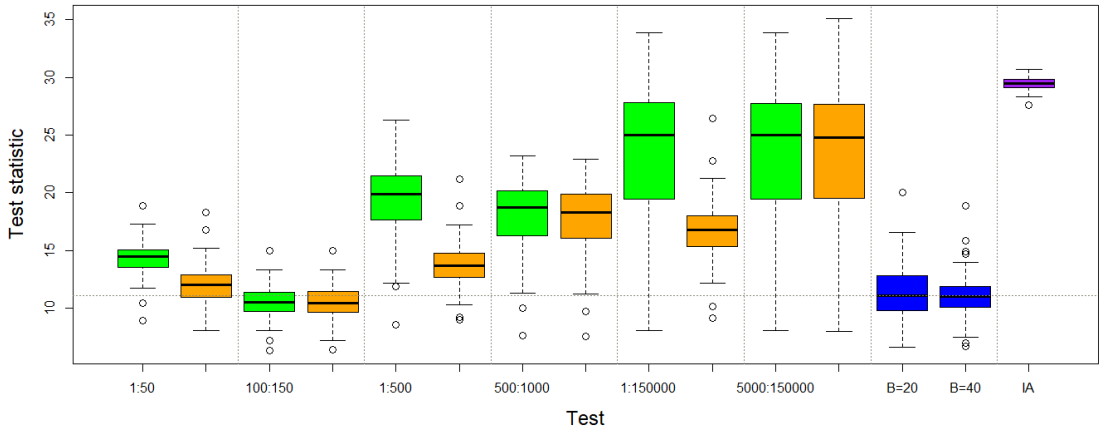

Figure 10: Boxplots of test statistics Z (green), $\tilde{Z}$ (orange), Z' (blue) and Z'' (purple) over $G = 100$ observed trajectories from an M/M/1 queue observed over the period $[0,\ 32960]$.

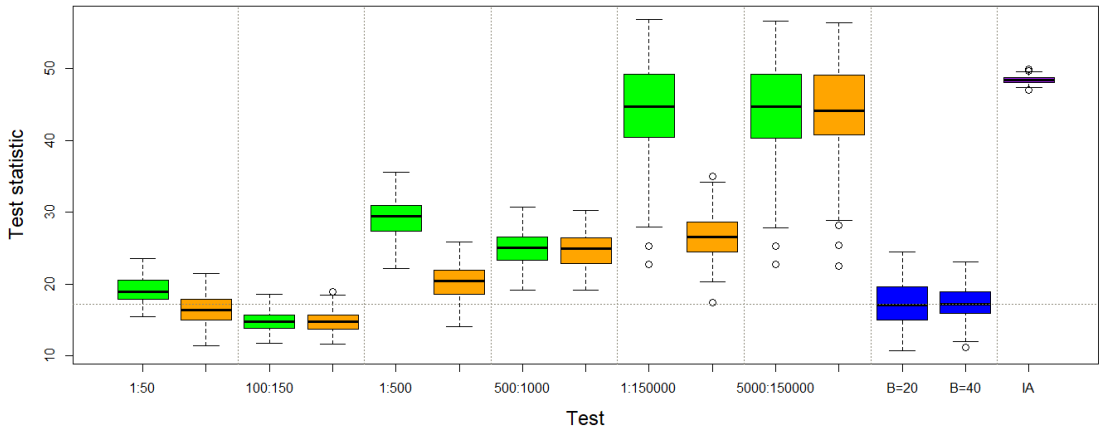

Figure 11: Boxplots of test statistics Z (green), $\tilde{Z}$ (orange), Z' (blue) and Z'' (purple) over $G = 100$ observed trajectories from an M/M/1 queue observed over the period $[0,\ 88960]$.

queueing models. The construction of the $waFm$ test statistic Z was based on the statistical characterization of the M/M/1 queueing model; in running the test for the discrimination of two M/H_2/1 systems we seek to evaluate how robust the test is to a change in distributional assumptions. The M/H_2/1 queueing model is a single sever queueing model with hyperexponentially distributed service times. We consider system I with arrival rate $\lambda = 0.8$ and system II with arrival rate $\lambda = 0.6$. The service distribution for both systems is set to be a hyperexponential distribution with parameters $\mu_1 = 0.65$, $\mu_2 = 2.25$ and $p = 0.5$. This parameterization gives an average service rate of $\mu = .991$ and a coefficient of variation of $cv = 1.989$. These parameters were chosen to match the experimental parameters in Law and Carson (1979) closely. A simulation run length of $T = 203280$ was chosen to approximate the average sample size suggested in Law and Carson (1979) to attain high average relative precision, and divides evenly into batches when $B = 40$. To evaluate our tests on the M/H_2/1 system we executed the same tests performed for the M/M/1 system i.e. using the same frequency ranges for discrimination and the same number of batches, B, in the batch means statistics. Figure 12 displays box plots of the test statistics from $G = 100$ replications of the tests for discrimination.

Interestingly we see a very similar pattern of behavior in Figure 12 as we did in Figures 10 and 11 despite

the service distribution departing from exponential. In fact all tests appear to have achieved higher power. Note that the ω-weighted test statistic retains highest power when high frequencies are included in the test statistic.

One application of $waFm$ discrimination tests would be change detection for simulated or real output. Nedényi (2018) provides details of a procedure for online change detection of a stochastic process. For change detection, much shorter run lengths would be necessary. To test the effect of the observed trajectory length, T, on the power of the tests we repeat the same tests for the $M/H_2/1$ models, but reduce the trajectory length to $T = 10,000$, a reduction of the interval to approximately 5% of its original length. As this is a large reduction in the observed interval we now report test statistics from the traditional batch means method with the number of batches set to $B = 10$ in addition to $B = 20$ and 40. Note that $B = 10$ is the minimum number recommended by Schmeiser (1982). In Figure 13 box plots of the test statistics from $G = 100$ macro replications of the tests are presented.

In Figure 13 it is clear that, across all frequency ranges, the batch averaged Fourier magnitude test have greater power than the traditional batch means test statistic. The mean inter-arrivals test clearly performs

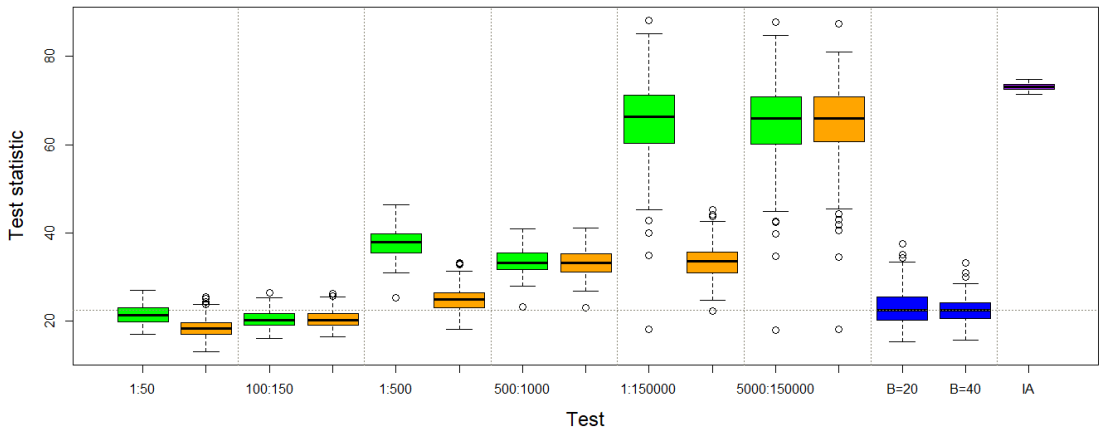

Figure 12: Boxplots of test statistics Z (green), $\tilde{Z}$ (orange), Z' (blue) and Z'' (purple) over $G = 100$ observed trajectories from an $M/H_2/1$ queue observed over the period $[0,\ 203280]$.

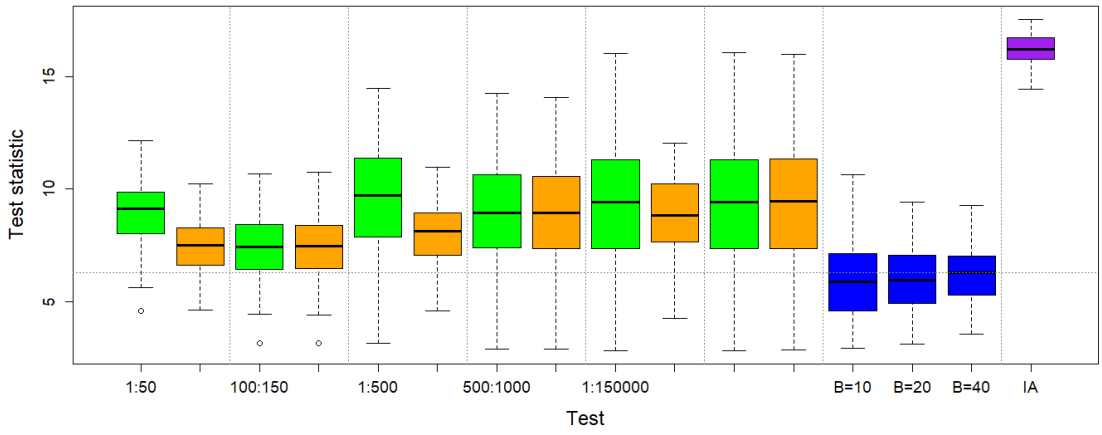

Figure 13: Boxplots of test statistics Z (green), $\tilde{Z}$ (orange), Z' (blue) and Z'' (purple) over $G = 100$ observed trajectories from an $M/H_2/1$ queue observed over the period $[0,\ 10000]$.

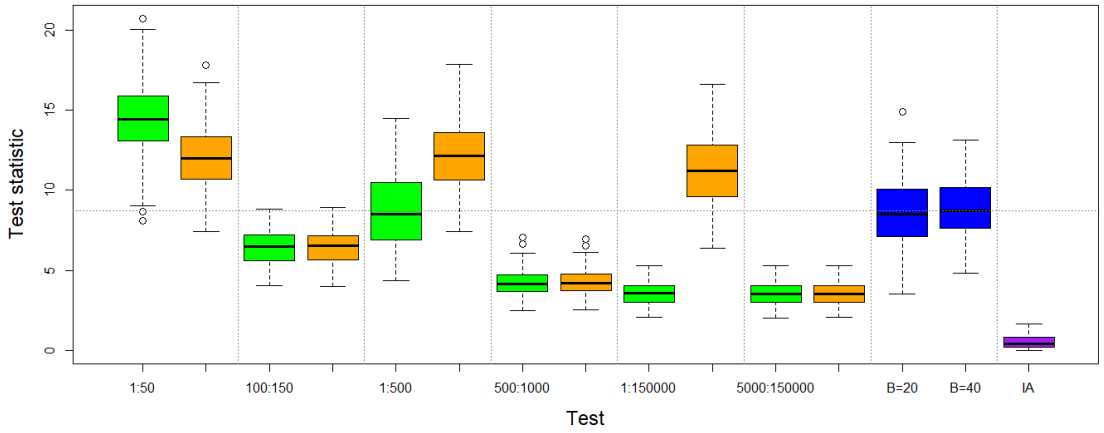

Figure 14: Boxplots of test statistics Z (green), $\tilde{Z}$ (orange), Z' (blue) and Z'' (purple) over $G = 100$ observed trajectories of an M/M/1 queue observed over the period $[0,\ 88960]$.

better than any of the batch means statistics. The outcome of this experiment is an indicator that the $waFm$ test may retain discriminatory power even over short intervals which would make change detection viable.

We end this illustration by considering two M/M/1 systems with equal arrival rate, $\lambda = 0.75$, but unequal service rate. System I was set to have service rate $\mu = 1$ and system II $\mu = 0.8$. Our interest in this experiment is in evaluating whether the discriminatory power of Fourier magnitude based tests is driven by differences in the arrival rate λ alone. Figure 14 shows the box plots of the test statistics for all tests in this scenario. Note that when the focus is on detection of changes in arrival rate, one might focus on that statistic directly. But in a number of circumstances, such as packet networks and service queues in uncontrolled settings, queue counts are more directly available than arrival counts or inter-arrival times.

From Figure 14 we see that the power of Fourier coefficient magnitude tests is reduced when the arrival rates of the two systems are equal. But, when low frequencies are included in the range of frequencies for the $waFm$ test statistic, the test appears to retain reasonable discriminatory power. This power is higher than the power of the traditional batch means tests when only the lowest frequencies are included. Intuitively this makes sense as the lowest frequency magnitude provides an estimate of the mean number-in-system which we would expect to be different given systems with different service rates, μ. It is also likely that and the lowest frequencies are picking up on the highest auto-correlation.

As expected, we see that the test for a difference in inter-arrival time means has no discriminatory power in this scenario. Although the test performs well in all other scenarios, when the arrival rate of two systems is equal it has no discriminatory power. Also recall that this test is not possible if inter-arrival times cannot be recorded.

In summary this illustration provides evidence that test statistics exploiting ω-weighted magnitudes ($waFm$) perform better than the commonly used traditional batch means test for discrimination. This suggests a gain in power by exploiting our knowledge of the statistical characterisation of the dynamic behavior of queueing models. We also compared $waFM$ to a test exploiting inter-arrival times and found it could perform well relative to this optimal test when only the arrival rates differed, and better when the arrival rate of the two systems under study was equal.

5 Summary and Further Work

In this paper we presented a hypothesis test for the discrimination of simulation trajectories using ω-weighted average Fourier magnitudes ($waFm$). This test was shown to be superior to a traditional batch means test statistic, and to the unweighted average Fourier magnitudes statistic in Wu and Barton (2016).

The $waFm$ test was constructed using characteristics of M/M/1 and M/G/1 queueing systems. This

allowed us to carefully evaluate and present both the test statistic and the assumptions underlying it. Computational investigations showed that the power for discrimination was high for both M/M/1 and two $M/H_2/1$ systems. The *waFm*-type statistic can be used in an optimization setting when dynamic behavior is an issue, and for validation of dynamic systems, as it allows for deeper comparisons of trajectory behavior. Current developments in simulation, including the uptake of digital twin technology prompts the need for methods that enable the detection of the divergence of two systems. We believe the test presented in this paper is a step towards the development of such methodology.

A consideration of more general queueing models where the inputs are a superposition of renewal processes (so only Poisson in the limit) would be a potential area to extend this paper. It may also be of interest to consider more complex systems i.e. Jackson networks. We leave these considerations for future work.

Thinking more broadly, another area that would benefit from the *waFm* statistic is the field of statistical process control (SPC). SPC is used in management to react to changes in a dynamic system. It considers whether there has been a change in the parameters driving a single trajectory, and may be used reactively to make decisions, for example to add or remove servers as the utilisation of a system fluctuates. Although there are obvious parallels to what we have presented in this paper the SPC setting makes our problem much more complex. SPC is driven by making decisions in operational time which means considering whether there has been a change in the system within a small time window. In Section 4.2 we illustrated that our test performs well for high frequencies in relatively short time intervals. The ω-weighted test retained high discriminatory power compared to the batch means tests, but further consideration is needed to see if this extends to window lengths of practical use in the SPC context. Also, in this setting, wavelet methods may hold an advantage. Another issue with SPC is the potential for lag effects in the trajectory. A change in a system parameter may take a while to effect the shape of the trajectory which could delay discrimination. While a numerical exploration of the unweighted average statistic in Wu and Barton (2016) showed promise, we leave further consideration of the SPC context to future work.

Acknowledgements

We thank Steve Chick, Barry Nelson and Lee Schruben for helpful reviews of a preliminary version of this paper. The authors acknowledge financial and computing support from Penn State's Smeal College of Business, the Durham University Business School, the STOR-i Centre for Doctoral Training (CDT) at Lancaster University, and the The High End Computing facility at Lancaster University.

References

Bartlett, M. S. (1950). Periodogram Analysis and Continuous Spectra. *Biometrika*, 37(1/2):1–16. Publisher: [Oxford University Press, Biometrika Trust].

Borwein, J. M., Nuyens, D., Straub, A., and Wan, J. (2011). Some arithmetic properties of short random walk integrals. *The Ramanujan Journal*, 26(1):109.

Chui, C. K. (1997). *Wavelets: A Mathematical Tool for Signal Analysis*. Mathematical Modeling and Computation. Society for Industrial and Applied Mathematics, Philadelphia.

Dong, J. and Whitt, W. (2015). Stochastic grey-box modeling of queueing systems: fitting birth-and-death processes to data. *Queueing Systems: Theory and Applications*, 79(3-4):391–426.

Eick, S. G., Massey, W. A., and Whitt, W. (1993). Mt/G/ Queues with Sinusoidal Arrival Rates. *Management Science*, 39(2):241–252. Publisher: INFORMS.

Epstein, C. (2005). How well does the finite Fourier transform approximate the Fourier transform? *Communications on Pure and Applied Mathematics*, 58:1421–1435.

Feller, W. (1966). On the Fourier Representation for Markov Chains and the Strong Ratio Theorem. *Journal of Mathematics and Mechanics*, 15(2):273–283. Publisher: Indiana University Mathematics Department.

Fishman, G. S. (1971). Estimating Sample Size in Computing Simulation Experiments. *Management Science*, 18(1):21–38. Publisher: INFORMS.

Fishman, G. S. and Kiviat, P. J. (1967). The Analysis of Simulation-Generated Time Series. *Management Science*, 13(7):525–557.

Freedman, D. and Lane, D. (1980). The Empirical Distribution of Fourier Coefficients. *The Annals of Statistics*, 8(6):1244–1251.

Green, L., Kolesar, P., and Svoronos, A. (1991). Some Effects of Nonstationarity on Multiserver Markovian Queueing Systems. *Operations Research*, 39(3):502–511.

Hazra, M. M., Morrice, D. J., and Park, S. K. (1997). A simulation clock-based solution to the frequency domain experiment indexing problem. *IIE Transactions*, 29:769–782.

Heidelberger, P. and Welch, P. D. (1983). Simulation Run Length Control in the Presence of an Initial Transient. *Operations Research*, 31(6):1109–1144. Publisher: INFORMS.

Heyman, D. P. and Sobel, M. J. (2004). *Stochastic Models in Operations Research: Stochastic Processes and Operating Characteristics*, volume 1. Courier Corporation. Google-Books-ID: IcVlwPS0qCwC.

Houtveen, J. H. and Molenaar, P. C. M. (2001). Comparison between the Fourier and Wavelet methods of spectral analysis applied to stationary and nonstationary heart period data. *Psychophysiology*, 38(5):729–735. Publisher: Cambridge University Press.

Jacobson, S. H., Buss, A. H., and Schruben, L. W. (1991). Driving Frequency Selection for Frequency Domain Simulation Experiments. *Operations Research*.

Jacobson, S. H., Morrice, D., and Schruben, L. W. (1988). The global simulation clock as the frequency domain experiment index. In *Proceedings of the 1988 Winter Simulation Conference*, pages 558–563, Piscataway, NJ, USA. IEEE.

Koo, I., Zhang, X., and Kim, S. (2011). Wavelet- and Fourier-Transform-Based Spectrum Similarity Approaches to Compound Identification in Gas Chromatography/Mass Spectrometry. *Analytical Chemistry*, 83(14):5631–5638. Publisher: American Chemical Society.

Lada, E. K., Wilson, J. R., Steiger, N. M., and Joines, J. A. (2007). Performance of a Wavelet-Based Spectral Procedure for Steady-State Simulation Analysis. *INFORMS Journal on Computing; Linthicum*, 19(2):150–160. Num Pages: 11 Place: Linthicum, United States, Linthicum Publisher: Institute for Operations Research and the Management Sciences.

Laidler, G., Morgan, L. E., Nelson, B. L., and Pavlidis, N. G. (2020). Metric learning for simulation analytics. In Bae, K.-H., Feng, B., Kim, S., Lazarova-Molnar, S., Zheng, Z., Roeder, T., and Thiesing, R., editors, *Proceedings of the 2020 Winter Simulation Conference*, Piscataway, NJ, USA. IEEE.

Law, A. M. and Carson, J. S. (1979). A sequential procedure for determining the length of a steady-state simulation. *Operations Research*, 27(5):1011–1025.

Lin, Y., Nelson, B. L., and Pei, L. J. (2019). Virtual Statistics in Simulation via k Nearest Neighbors. *INFORMS Journal on Computing*.

Massey, W. A. (2002). The Analysis of Queues with Time-Varying Rates for Telecommunication Models. *Telecommunication Systems*, 21(2):173–204.

Morrice, D. J. (1995). A comparison of frequency domain methodology and conventional factor screening methods. *Operations Research Letters*, 17(4):165–174.

Morse, P. M. (1955). Stochastic Properties of Waiting Lines. *Journal of the Operations Research Society of America*, 3(3):255–261.

Nedényi, F. K. (2018). An online change detection test for parametric discrete-time stochastic processes. *Sequential Analysis*, 37(2):246–267.

Nelson, B. (2013). *Foundations and Methods of Stochastic Simulation: A First Course*. Springer Science & Business Media.

Nelson, B. L. (2016). Some tactical problems in digital simulation for the next 10 years. *Journal of Simulation*, 10(1):2–11. Publisher: Taylor & Francis _eprint: https://doi.org/10.1057/jos.2015.22.

Parzen, E. (1957). On Consistent Estimates of the Spectrum of a Stationary Time Series. *Annals of Mathematical Statistics*, 28(2):329–348. Publisher: Institute of Mathematical Statistics.

Parzen, E. (1961). Mathematical Considerations in the Estimation of Spectra. *Technometrics*, 3(2):167–190. Publisher: [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality].

Pedrielli, G. and Barton, R. R. (2019). Metamodel-Based Quantile Estimation for Hedging Control of Manufacturing Systems. In *Proceedings of the 2019 Winter Simulation Conference*, pages 452–463, Piscataway, NJ, USA. IEEE. ISSN: 1558-4305.

Rayleigh (Strutt), J. W. (1905). The problem of the random walk. *Nature*, page 318.

Sargent, R. G. and Som, T. K. (1992). Current Issues in Frequency Domain Experimentation. *Management Science*, 38(5):667–687. Publisher: INFORMS.

Scargle, J. D. (1982). Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. *The Astrphhysical Journal*, 263:835–853.

Schmeiser, B. (1982). Batch size effects in the analysis of simulation output. *Operations Research*, 30(3):556–568.

Schruben, L. W. and Cogliano, V. J. (1987). An experimental procedure for simulation response surface model identification. *Communications of the ACM*, 30(8):716–730.

Stankovic, L., Dakovi, M., and Thayaparan, T. (2014). *Time-Frequency Signal Analysis with Applications*. Artech House. Google-Books-ID: b8N23zsn6H8C.

Steiger, N. M., Lada, E. K., Wilson, J. R., Joines, J. A., Alexopoulos, C., and Goldsman, D. (2005). ASAP3: a batch means procedure for steady-state simulation analysis. *ACM Transactions on Modeling and Computer Simulation*, 15(1):39–73.

Weisstein, E. W. (2020). Random Walk–2-Dimensional.

Whitt, W. (1989a). Planning Queueing Simulations. *Management Science*, 35(11):1341–1366. Publisher: INFORMS.

Whitt, W. (1989b). Simulation run length planning. In *Proceedings of the 21st conference on Winter simulation*, WSC '89, pages 106–112, Washington, D.C., USA. Association for Computing Machinery.

Wu, X. and Barton, R. R. (2016). Fourier trajectory analysis for identifying system congestion. In *Proceedings of the 2016 Winter Simulation Conference*, WSC '16, pages 401–412, Piscataway, NJ, USA. IEEE.

A Results required for Theorem 2

To characterize the distribution of the coefficient magnitudes for the Fourier transform of a number in system trajectory, we need the following properties and lemmas.

Property A.1. Positive unit step: The Fourier transform for a positive unit step at time t_0 is $c_{+step}(\omega) = \frac{1}{i\omega}e^{-i\omega t_0}$.

Property A.2. Negative unit step: The Fourier transform for a positive unit step at time t_0 is $c_{-step}(\omega) = \frac{-1}{i\omega}e^{-i\omega t_0}$.

These results follow from the transform for $f(t) = \text{sign}(t)$ in Example 1.9 in Stankovic et al. (2014) and the Fourier transform time-shift and linearity properties given above. Note that $e^{-i\omega t_0}$ can be thought of as a unit vector in the complex plane at angle ωt_0, and that $-e^{-i\omega t_0}$ is a unit vector at angle $(\omega t_0 + \pi) \mod 2\pi$. Note that if t_0 is generated by a Poisson process, the resulting angle is uniformly distributed between 0 and 2π. We will need to know the distribution of the magnitudes of a sum of such random unit vectors to characterize the distribution of coefficient magnitudes for the Fourier transform of a number-in-system trajectory.

Lemma 4. *Consider a sum of N unit vectors in the complex plane, each with orientation θ_j on $[0, 2\pi)$. The position z resulting from the sum is*

$$z = \sum_{j=1}^{N} e^{i\theta_j}$$

which has magnitude squared of

$$|z|^2 = N + \sum_{\substack{j,k=1 \\ j\neq k}}^{N} e^{i(\theta_j - \theta_k)}.$$

If the orientations θ_j are independent, uniformly distributed on $[0, 2\pi)$, then

$$\text{E}(|z|^2) = N \text{ and } |z|_{rms} = \sqrt{N}. \tag{10}$$

and $\text{E}(|z|^2)$ has random zero-expectation perturbation

$$\sum_{j,k=1, j\neq k}^{N_3} e^{i\omega(t_j - t_k)} \tag{11}$$

Proof. The lemma and proof come directly from Weisstein (2020). □

The expected value of the vector magnitude is smaller than the rms value. For large N the approximate density is $2re^{-r^2/N}$ with mean value $\sqrt{\frac{\pi}{4}N} \approx \sqrt{.785N}$ and standard deviation $\sqrt{(1 - \frac{\pi^2}{16})N} \approx \sqrt{.383N}$ (Rayleigh (Strutt), 1905).

Lemma 5. *Consider an M/M/1 queue during a busy interval. During this period arrivals and departures from the queue operate as independent Poisson processes with rates λ and μ respectively.*

Proof. This follows directly from the continuous time Markov chain representation of the number-in-system queue state. The transition rate for any state $i \neq 0$ is $q_i = \lambda + \mu$ with the probability of unit increase = $\lambda/(\lambda + \mu)$ and the probability of unit decrease = $\mu/(\lambda + \mu)$, independent of the previous transitions. □

B Proof of Theorem 3 and Corollary 2

Before the proofs, we reformulate parts of (8), and examine related bounds. Using

$$e^{-i\omega x} = \cos x - i \sin x$$

makes the sum in equation (8), after rearranging terms,

$$\sum_{j=1}^{N} -(\cos \omega s_j - \cos \omega a_j) + i(\sin \omega s_j - \sin \omega a_j). \tag{12}$$

Further, since

$$\cos x - \cos y = -2\sin\left(\frac{x+y}{2}\right)\sin\left(\frac{x-y}{2}\right) \quad \text{and} \quad \sin x - \sin y = 2\cos\left(\frac{x+y}{2}\right)\sin\left(\frac{x-y}{2}\right)$$

the paired summand in (12) can be written

$$\begin{aligned} & 2\sin\left(\frac{\omega(a_j+s_j)}{2}\right)\sin\left(\frac{\omega(s_j-a_j)}{2}\right) \\ & +2i\cos\left(\frac{\omega(a_j+s_j)}{2}\right)\sin\left(\frac{\omega(s_j-a_j)}{2}\right) \\ & = 2\left(\sin\left(\frac{\omega(a_j+s_j)}{2}\right)+i\cos\left(\frac{\omega(a_j+s_j)}{2}\right)\right)\sin\left(\frac{\omega(s_j-a_j)}{2}\right). \end{aligned} \tag{13}$$

For a general renewal interval, consider the magnitude of the complex quantity in (13), say $|W_j|$, where the dependence on ω is suppressed to simplify notation. Then,

$$|W_j| = 2\left|\sin\left(\frac{\omega(s_j-a_j)}{2}\right)\right|. \tag{14}$$

Consider a frequency that is a multiple of the period length, $\omega = \omega_k = 2k\pi/T$. Note that $s_j - a_j$ is the time in system and can be represented as a random variable S, which for an M/M/1 system has distribution (Heyman and Sobel, 2004)

$$F_S(t) = 1 - \rho + \rho(1 - e^{-(\mu-\lambda)t}). \tag{15}$$

Then equation (14) becomes

$$|W_j(\lambda,\mu)| = 2\left|\sin\left(\frac{k\pi S}{T}\right)\right|, \tag{16}$$

where now we include explicitly the dependence on λ and μ. In the busy interval analysis, the members of the pair $\{e^{-i\omega a_j}, e^{-i\omega s_j}\}$ are randomly located on the unit circle, and the expected magnitude of the two-step random walk for any pair is $|W_j(\lambda,\mu)| \approx 1.273$ (Borwein et al., 2011). We will compute the expected length of the two-step random walk corresponding to $|W_j(\lambda,\mu)|$ in the limiting case $\lambda = \lambda_n \equiv \lambda_0/n \to 0$ as $n \to \infty$ with $\mu = 1$. First, denote the random variable S when $\lambda = \lambda_n$ by S_n, and the limiting random variable as $n \to \infty$ by $\tilde{S}$. Then by (15), $\tilde{S}$ has an exponential distribution with mean 1, and since the function in equation (16) is continuous, we have the following.

Lemma 6 (Pair Magnitude Distribution as $\lambda \to 0$)**.** *Let $W_j(\lambda,\mu)$ represent the quantity in (13), generated from the number-in-system trajectory of an M/M/1 queue with service mean $1/\mu = 1$, over a renewal period of length T, and $\tilde{S}$ be a random variable from an exponential distribution with mean 1. Then*

$$\lim_{\substack{n\to\infty \\ \lambda_n\to 0}} |W_j(\lambda_n, 1)| = 2\left|\sin\left(\frac{k\pi\tilde{S}}{T}\right)\right|.$$

Proof. This result follows directly from the continuity of the magnitude function, the definition of λ_n, and the continuous mapping theorem. □

Now we can bound the magnitude of this random variable as T grows.

Lemma 7 (Bound on Expected Pair Magnitude as $\lambda \to 0$)**.** *Given the result in Lemma (6), for any chosen values $0 < \varepsilon_1, \varepsilon_2 < 1$, there is a $T_0 \equiv \frac{-k\pi\ln(\varepsilon_1)}{\varepsilon_2}$ such that any composition of renewal intervals of combined length $T \geq T_0$ gives*

$$\lim_{\lambda\to 0} \mathrm{E}|W_j| \leq 2(1-\varepsilon_1)\varepsilon_2 + 2\varepsilon_1. \tag{17}$$

Proof. Let $X(\tilde{S})$ be a function of the nonnegative random variable $\tilde{S}$ defined by:

$$X(\tilde{S}) = \left|\sin\left(\frac{k\pi\tilde{S}}{T}\right)\right|.$$

and define a discrete random variable $Y(\tilde{S})$, taking on only two values, that is stochastically greater than $X(\tilde{S})$:

$$\begin{aligned} Y(\tilde{S}) &= \frac{k\pi s_{\varepsilon_1}}{T}, 0 \le \tilde{S} \le s_{\varepsilon_1}, \\ &= 1, \text{otherwise.} \end{aligned} \tag{18}$$

Since $\tilde{S} \ge 0$, $\sin\left(\frac{k\pi\tilde{S}}{T}\right) \le \left(\frac{k\pi\tilde{S}}{T}\right)$. Further, $|X(\tilde{S})| \le 1$, and combined with (18) implies $X(\tilde{s}) \le Y(\tilde{s})$ for any realization $\tilde{s}$ of $\tilde{S}$. Then

$$\mathrm{E}X(\tilde{S}) \le \mathrm{E}Y(\tilde{S}). \tag{19}$$

Choose s_{ε_1} such that $\mathrm{Prob}(\tilde{S} \le s_{\varepsilon_1}) \ge 1 - \varepsilon_1$; that is, $s_{\varepsilon_1} = q_{\tilde{S},1-\varepsilon_1}$, the $(1-\varepsilon_1)$-quantile of $\tilde{S}$. Now choose T_0 such that

$$\frac{k\pi q_{\tilde{S},1-\varepsilon_1}}{T_0} \le \varepsilon_2. \tag{20}$$

Then choose a value $T \ge T_0$ that is the combined length of a composition of one or more renewal intervals, giving

$$\begin{aligned} EY(\tilde{S}) &= (1-\varepsilon_1)\frac{k\pi q_{\tilde{S},1-\varepsilon_1}}{T} + \varepsilon_1 \\ &\le (1-\varepsilon_1)\varepsilon_2 + (\varepsilon_1)(1). \end{aligned} \tag{21}$$

Note that as $\lambda \to 0, \tilde{S}$ has an exponential distribution with mean 1. Then $q_{\tilde{S},1-\varepsilon_1} = -\ln(\varepsilon_1)$ and

$$T_0 \ge \frac{-k\pi\ln(\varepsilon_1)}{\varepsilon_2}. \tag{22}$$

The bound holds with $T \ge T_0$ specified by (22), again by the Continuous Mapping Theorem. □

So as T grows and $\lambda \to 0$, the paired steps on the unit complex circle summed in (8) each have expected magnitude that goes to zero, in contrast with the busy-period independent value of 1.273. To bound the sum in (8), consider the squared magnitude of $c_{M/M/1_T}(\omega)$. We are now ready to prove Theorem 3.

Proof. Beginning with the representation in equation (8),

$$\begin{aligned} \mathrm{E}(|c_{M/M/1_T}(\omega_k)|^2) &= \mathrm{E}\left(\left|\frac{1}{\omega_k}\left(\sum_{j=1}^{N} e^{-i\omega_k a_j} - e^{-i\omega_k s_j}\right)\right|^2\right) \\ &= \frac{1}{\omega_k^2}\mathrm{E}\left(\left|\sum_{j=1}^{N} e^{-i\omega_k a_j} - e^{-i\omega_k s_j}\right|^2\right) \\ &\le \frac{1}{\omega_k^2}\left(\mathrm{E}\sum_{j=1}^{N}\left|e^{-i\omega_k a_j} - e^{-i\omega_k s_j}\right|^2\right) \\ &= \frac{1}{\omega_k^2}\mathrm{E}(N)\mathrm{E}\left(\left|e^{-i\omega_k a_j} - e^{-i\omega_k s_j}\right|^2\right) \\ &\le \frac{1}{\omega_k^2} 2\mathrm{E}(N)\left((1-\varepsilon_1)\varepsilon_2^2 + \varepsilon_1(1^2)\right). \end{aligned}$$

The first inequality follows from the bound on the magnitude of the sum of vectors (in the complex plane), the next equality from the composition rule for a random number N of i.i.d. random variables $X = \mathrm{E}(N)\mathrm{E}(X)$. The second inequality follows from Lemma 6 and Lemma 7 for any chosen values $0 < \varepsilon_1, \varepsilon_2 < 1$. □

The proof of Corollary 2 follows.

Proof. The result differs from that of Theorem 3 only in the determination of T_0 from (22). Bound $q_{\tilde{S},1-\varepsilon_1}$ using the Markov bound:

$$\text{Prob}(\tilde{S} \geq s'_{\varepsilon_1}) \leq \frac{\text{E}(\tilde{S})}{s'_{\varepsilon_1}} = \frac{1}{s'_{\varepsilon_1}} = \varepsilon_1$$

and so replacing $s_{\varepsilon_1} \equiv q_{\tilde{S},1-\varepsilon_1} = -\ln(\varepsilon_1)$ by $s'_{\varepsilon_1} = \frac{1}{\varepsilon_1}$ gives the bound on T_0. □