

# Independence of Periodogram Ordinates at the Fourier Frequencies

Andee Kaplan & Maggie Johnson

December 16, 2013

## 1 Motivation

To explore whether it is reasonable to estimate the spectral density of a stationary time series using Bayesian methods. To do this, it is necessary to rely on the asymptotic distributional properties of periodogram ordinates. The motivation behind this work is the following model.

$$I_n(\omega_r)|f(\omega_r) \sim \text{indep Exp}(2\pi f(\omega_r)) \quad (1)$$

$$f(\omega_r) \sim \pi(\theta) \quad (2)$$

$$f(\omega_r)|I_n(\omega_r) \propto 2\pi f(\omega_r)e^{-2\pi f(\omega_r)y}\pi(\theta) \quad (3)$$

The periodogram values at  $\omega_r$  are only asymptotically distributed as independent exponentials, so to construct the Bayesian model it is essential that the asymptotic behavior in Equation 1 holds. This research uses simulations to assess two results by Lahiri, which state conditions necessary for periodogram ordinates to be asymptotically independent [4].

## 2 Background and Problem Statement

Spectral analysis, or “frequency domain analysis” is the analysis of stationary time series  $\{X_t\}$  using the decomposition  $\{X_t\}$  into sinusoidal componenets. Spectral analysis is equivalent to “time domain” analysis based on the autocovariance function, but can be more useful in applications where sinusoidal behavior is relevant. The spectral density of a mean-zero stationary process  $\{X_t\}$  is used to describe the frequency decomposition of the autocovariance function  $\gamma(\cdot)$  and the process  $\{X_t\}$  [1]. It is the function  $f(\cdot)$  defined by

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\omega} \gamma(h), \quad -\infty < \omega < \infty \quad (4)$$

where  $e^{i\omega} = \cos(\omega) + i\sin(\omega)$  and  $i = \sqrt{-1}$  [1]. The periodogram,  $I_n(\cdot)$  of  $\{X_t\}$  can be regarded as a sample estimator of  $2\pi f(\cdot)$  [1]. The periodogram of  $\{x_1, \dots, x_n\}$  at a frequency  $\omega$  is the function

$$I_n(\omega) = \frac{1}{n} \left| \sum_{t=1}^n x_t e^{-it\omega} \right|^2 \quad (5)$$

**Result 1.** *Let  $f(\omega_r) \neq 0, 1 \leq r \leq k$  where  $f$  denotes the spectral density of a stationary time series  $\{X_t\}$ . Then when  $n \rightarrow \infty$  the joint distribution of the periodogram at  $\omega_r$ ,  $I_n(\omega_r)$ , tends to that of  $k$  mutually independent random variables distributed as  $\text{Exponential}(2\pi f(\omega_r))$  for  $0 < \omega_r < \pi$  [1].*

Result 1 describes the asymptotic distribution of a periodogram at fixed frequencies. In practice, this result has often been used with the Fourier frequencies,  $\{\omega_j = 2\pi j/n : j = 1, \dots, n\}$ , in a fixed time period. [1]. However, it has been shown that the asymptotic behavior of the periodogram at the Fourier frequencies does not hold as  $n \rightarrow \infty$  by SN Lahiri [4]. In his paper, “A necessary and sufficient condition for asymptotic independence of discrete Fourier transforms under short-and long-range dependence,” Lahiri also derived two results that describe behavior required for a set of frequencies to satisfy the asymptotic independence assumption in result 1.

**Result 2.** *In the absence of data tapering,*

- (a) *the periodogram values at asymptotically distant ordinates ( $I_n(\omega_r)$ ) are asymptotically independent. Asymptotically distant ordinates  $\{\omega_{ln}\}, \{\omega_{kn}\}$  satisfy  $|n(\omega_{ln} - \omega_{kn})| \rightarrow \infty$  as  $n \rightarrow \infty$ .*
- (b) *the periodogram values at asymptotically close ordinates which are asymptotically distant from the sequence  $\{0\}$  are asymptotically independent. Asymptotically close ordinates  $\{\omega_{ln}\}, \{\omega_{kn}\}$  satisfy  $|n(\omega_{ln} - \omega_{kn})| \rightarrow 2\pi l$  for some nonzero integer  $l$  as  $n \rightarrow \infty$  [4].*

Results 1 and 2 lend to the question: “Can a subset of sufficiently spaced and distant from zero Fourier frequencies be constructed such that the periodogram ordinates at these frequencies are asymptotically independent and exponentially distributed?” We explore these two results to determine if, for large  $n$ ,

1. result 1 will not, in fact, hold at the Fourier frequencies, and
2. if (1) is true, a usable subset of Fourier frequencies can be constructed such that result 1 holds.

### 3 Models

To explore the asymptotic behavior of the periodogram at the Fourier frequencies, we used ARMA time series models, as these are the only class of models with known spectral densities. A subset

of the models considered are: a) IID Gaussian(0, 1); b) AR(1) with  $\phi = 0.5$ ; c) AR(4) with  $\phi = [0.08, 0.33, 0.1, 0.45]$ ; d) MA(1) with  $\theta = 0.7$ ; and e) ARMA(4,1) with  $\phi = [0.08, 0.33, 0.1, 0.45]$  and  $\theta = 0.7$ .

### IID Gaussian

The following result about the IID Gaussian model provides knowledge of the exact, rather than asymptotic, behavior of the periodogram at any frequency.

**Result 3.** For  $\{X_t\} \stackrel{IID}{\sim} N(0, \sigma^2)$ , the periodogram values  $\{I_n(\omega_j) : \omega_j \in \mathcal{F}_n, \omega_j \notin \{0, \pi\}\}$  are IID Exponential( $\sigma^2$ ) random variables [1].

This model was used as an initial baseline check to support the testing procedure of independent exponentially distributed ordinates. If our procedure showed failure with the IID Gaussian white noise model, then this would be an indication of an issue with the procedure, rather than the frequency spacings. The spectral density of the IID Gaussian model is  $f_X(\omega) = \sigma^2/2\pi, \omega \in [-\pi, \pi]$ .

### ARMA(p,q)

We exclusively used ARMA models as our experimental models because the spectral density of an ARMA model has a known and closed form. An ARMA model is a process  $\{X_t\}$  with the form  $\phi(B)X_t = \theta(B)Z_t$  where  $B$  is the backshift operator and  $\{Z_t\}$  is WN(0,  $\sigma^2$ ). Models of this form have spectral density

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2}, \omega \in [-\pi, \pi] [1]. \quad (6)$$

## 4 Methods

For result 1 to hold for  $\{X_t\} = \{x_1, \dots, x_n\}$  at the full Fourier frequencies, or a subset of Fourier frequencies, the joint asymptotic distribution of the periodogram ordinates  $I_n(\omega)$  must be equivalent to the product of  $n$  independent exponential distributions with means equal to  $2\pi f(\omega)$ . It is difficult to test the distribution of the full asymptotic joint distribution, so we split the problem into two parts: (1) a test of asymptotic exponential distribution at each  $\omega_j, j = 1, \dots, n$ , and (2) a method to test independence over  $\omega$ . Independence is tested pairwise across all frequencies of interest. We implement the following testing procedure and repeat  $s$  simulations of the tests to determine if the tests reject exponentiality and independence at similar rates as the Type I errors of our tests ( $\alpha$ ).

## Procedure

1. Simulate  $M$  draws from  $X_1, \dots, X_n$ , where  $\{X_t\}$  is a stationary time series from one of the five models.
2. Obtain  $M$  periodograms using the Fourier frequencies from  $(0, \pi)$ ,  $\omega_j = \frac{2\pi j}{n} : j = 1, \dots, \lfloor n/2 \rfloor$ .
3. Obtain  $\lfloor \frac{n-1}{2} \rfloor$  values of the spectral density at each Fourier frequency  $f_X(\omega_j)$ . These will be known by design.
4. Simulate  $M$  draws from the  $\lfloor \frac{n-1}{2} \rfloor$  different distributions  $\text{Exp}(2\pi f_X(\omega_r))$ .
5. Test the distribution of the  $M$  values of periodograms at each Fourier frequency separately. Store the number of failed tests at each frequency.
6. Test the pairwise independence of neighboring  $M$  values of periodograms at each Fourier frequency. Store the number of failed tests at each frequency.
7. Inspect the results from steps 5 and 6 to determine if different spacings of frequencies are needed. If so, use sparser frequencies and repeat from step 2 at the chosen frequencies.

## 4.1 Tests

We use the following nonparametric hypothesis tests to assess independence and distribution of the periodogram values.

**Spearman's Rank Correlation** Spearman's rank correlation  $\rho$ , is a measure of correlation between a bivariate random sample of size  $n$ . The calculation of this statistic corresponds to Pearson's  $r$  computed on the ranks (and average ranks in the case of ties) of the data. It is used as a test statistic to test the null hypothesis,  $H_0$  : The bivariate random sample  $X_i$  and  $Y_i$  are mutually independent, against the alternative hypothesis,  $H_1$  : Either there is a tendency for the larger values of  $X$  to be paired with the larger values of  $Y$  or there is a tendency for the smaller values of  $X$  to be paired with the larger values of  $Y$ . An asymptotic  $t$  distribution is used to calculate p-values [2].

We used a two-sided Spearman's rank correlation test at the  $\alpha = 0.05$ -level to test whether periodogram values at neighboring frequencies (with different inbetween spacings)  $\omega_j$  and  $\omega_k$  were pairwise independent.

**Kolmogorov-Smirnov Test of Distribution** The Kolmogorov-Smirnov (KS) test is a procedure that uses the maximum vertical distance between an empirical cumulative distribution function and

a named cumulative distribution function as a measure of how much the two functions resemble each other. The only assumption of this test is that the data are a random sample. The KS test uses test statistic  $T = \sup_x |F^*(x) - S(x)|$  to test the null hypothesis,  $H_0 : F(x) = F^*(x)$  for all  $x$ , where  $F^*(x)$  is the completely specified hypothesized distribution function,  $F(x)$  is the unknown distribution function of the data, and  $S(x)$  is the empirical distribution function of the data. For data of length  $n \leq 100$ , exact p-values are available, and for data of length  $n > 100$ , an asymptotic approximation is used [2]. The asymptotic distribution is called the Kolmogorov distribution, and is of the form  $P(T \leq \lambda) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 \lambda^2}$  [3].

We used a two-sided KS test at the  $\alpha = 0.05$ -level to test whether the distributions of our periodogram values at each frequency  $\omega_j$  were exponentially distributed with mean parameter  $2\pi f_X(\omega_j)$ , where  $f_X(\omega_j)$  is the spectral density of  $X_t$  evaluated at  $\omega_j$ .

## 5 Results

### 5.1 Full Fourier Decomposition

We simulated data, periodograms, and spectral densities for each of the five models discussed in section 3 using  $M = 1000$ ,  $n = 500$ ,  $s = 200$  and  $\sigma^2 = 1$ . Tests of asymptotic exponential distribution and pairwise independence for neighboring frequencies were conducted first using the entire set of Fourier frequencies,  $\{w_j = \frac{2\pi j}{n}, j = 1, \dots, 249\}$  for each of the five models.

#### 5.1.1 IID Gaussian

The IID Gaussian white noise model should have periodogram ordinates  $I_n(\omega)$  distributed as exactly independent exponentials, regardless of the length of the time series. Figure 1 shows the estimated periodogram for one simulated IID Gaussian time series, as well as the constant spectral density. The independence in the time series is seen in the randomness of the periodogram.

Tests of exponential distribution and pairwise independence of neighboring Fourier frequencies gave rejection rates relatively similar to the  $\alpha$ -level of 0.05 over 200 simulations, shown in figure 2. This suggested that the IID Gaussian model results in independent exponentially distributed periodogram ordinates, as expected, and provided some assurance that the testing procedure was performing as expected.

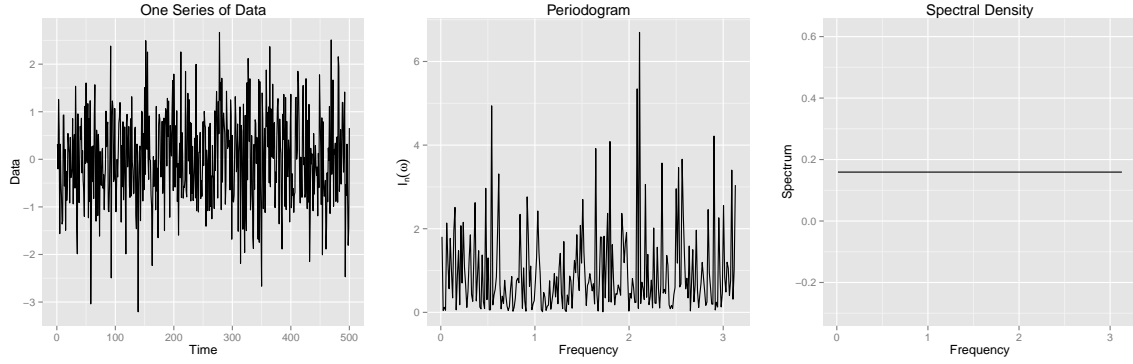


Figure 1: One draw, periodogram, and spectral density of a Gaussian IID Model.

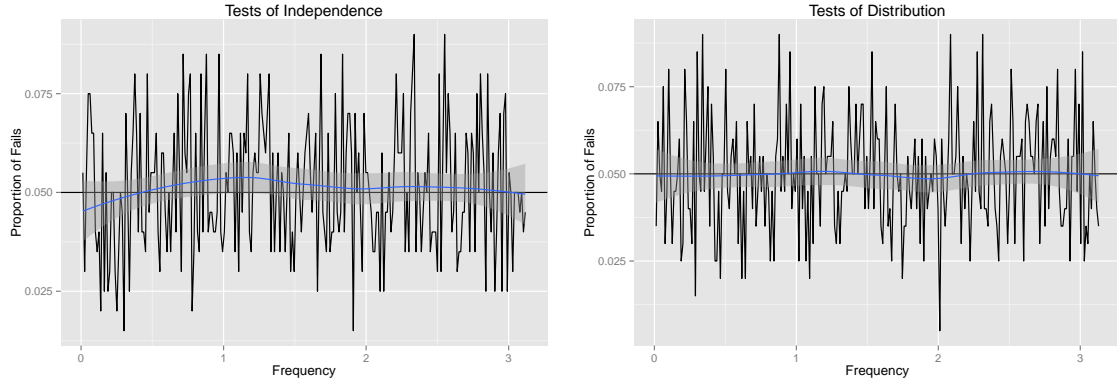


Figure 2: Tests of independence and distribution for a Gaussian IID Model.

### 5.1.2 AR(1)

The first model considered to test the asymptotic assumptions in results 1 and 2 was an AR(1) model with  $\phi = 0.5$ . For a time series  $\{X_t\}$ , this model has the form

$$X_t = \phi X_{t-1} + Z_t \quad (7)$$

where  $Z_t \sim WN(0, \sigma^2)$ . Figure 3 shows that the periodogram for this model has more structure than that of the IID Gaussian with frequencies closer to zero showing up as of higher importance than those further from zero.

The tests of exponential distributions and pairwise independence also gave rejection results consistent with the Type I error rate. Therefore, there was no indication that result 1 does not hold at the Fourier frequencies for an AR(1) model with  $\phi = 0.5$ . This was not entirely surprising as an AR(1) model has only slightly more structure than a white noise model as it represents dependence between

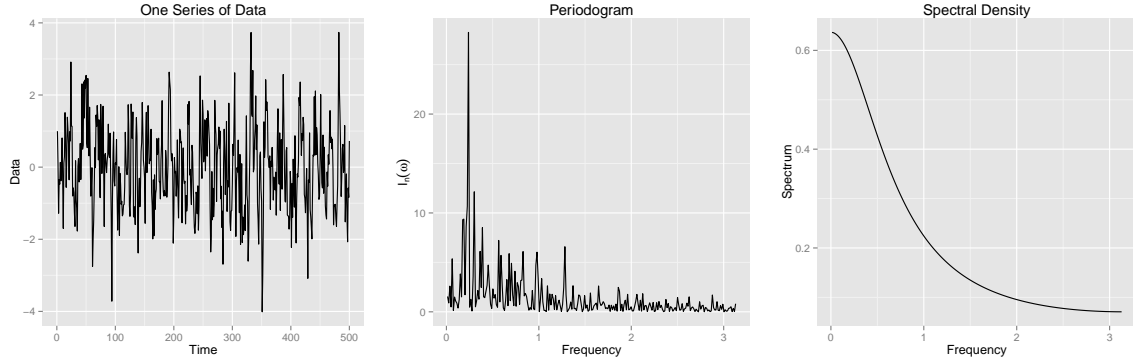


Figure 3: One draw, periodogram, and spectral density of an AR(1) model.

neighboring time points only.

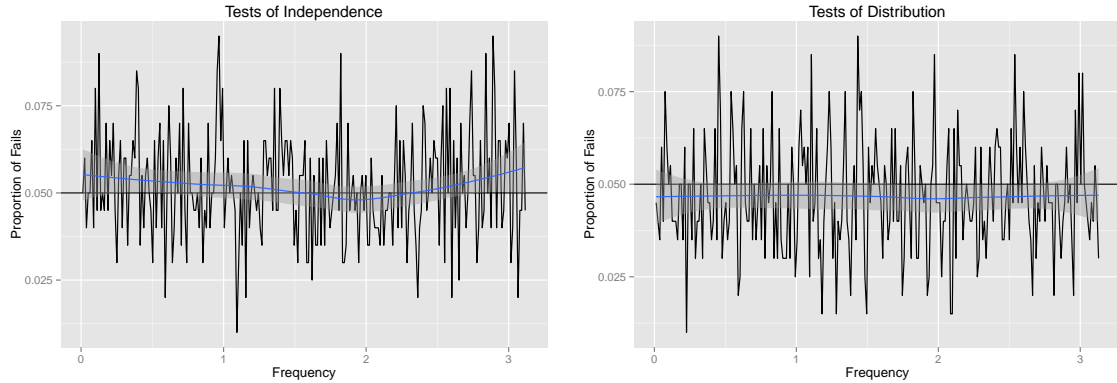


Figure 4: Tests of independence and distribution for an AR(1) model.

### 5.1.3 AR(4)

To experiment with a time series with a more long-term structure, we implemented an AR(4) model with  $\phi = [0.08, 0.33, 0.1, 0.45]$ . The estimated periodogram, for simulated time series following this model, is dominated by frequencies very close to zero, figure 5.

Unlike the IID Gaussian and AR(1) models, this AR(4) model does not have rejection rates all similar to  $\alpha$  for the independence and exponential tests at neighboring Fourier frequencies. At frequencies closer to zero, the pairwise independence tests fail at rates consistently higher than  $\alpha$  and likewise, especially for frequencies less than one, the periodogram ordinates could not be considered to be asymptotically exponentially distributed, see figure 6. Constructing a sparse subset of Fourier

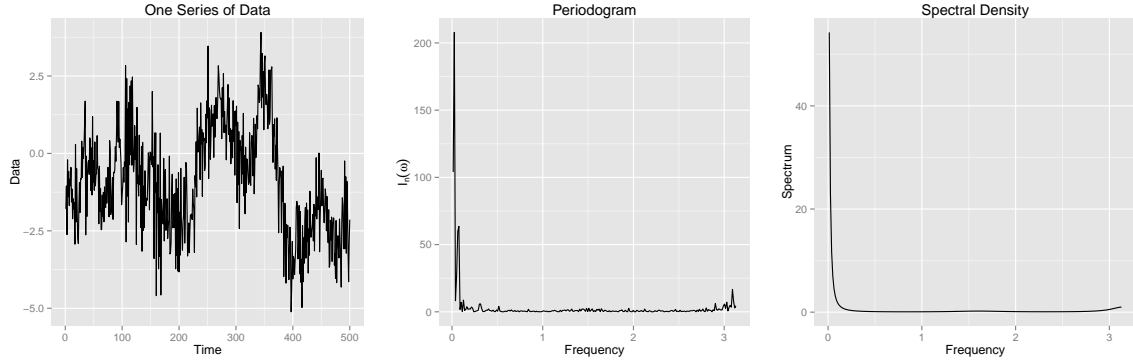


Figure 5: One draw, periodogram, and spectral density of an AR(4) model.

frequencies near zero would not result in asymptotically exponential periodogram ordinates at these frequencies. This is a shortcoming of limiting our method to only looking at differently spaced Fourier frequencies. In this paper, we explore if sparcifying frequencies near zero would satisfy asymptotic independence; see section 5.2. However future work could address the question of asymptotic distribution by spacing of non-Fourier frequencies.

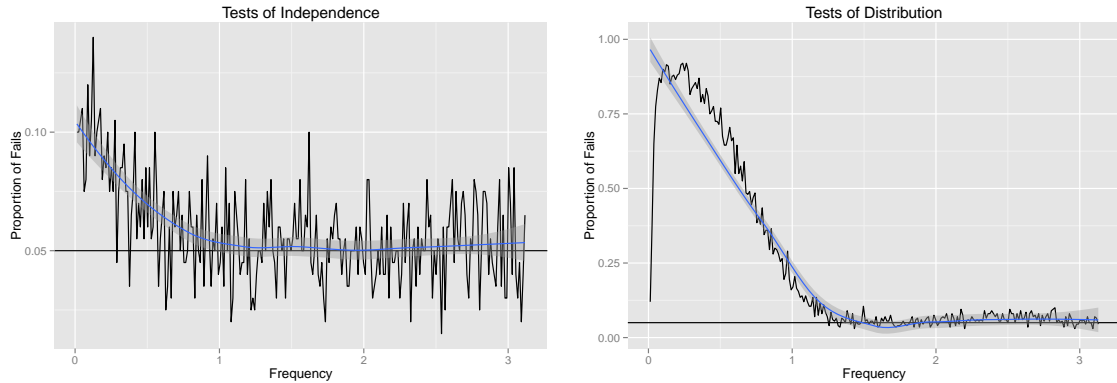


Figure 6: Tests of independence and distribution for an AR(4) model.

#### 5.1.4 MA(1)

We also considered models with moving average structure such as a MA(1) model with  $\theta = 0.7$ . A time series with this behavior has the form

$$X_t = Z_t + \theta Z_{t-1} \quad (8)$$



where  $Z_t \sim WN(0, \sigma^2)$ . Figure 7 shows that the estimated periodogram for a simulated time series of this form puts more importance on frequencies closer to zero than to  $\pi$ , as in the AR(1) model, but shows a more random structure than the AR(1).

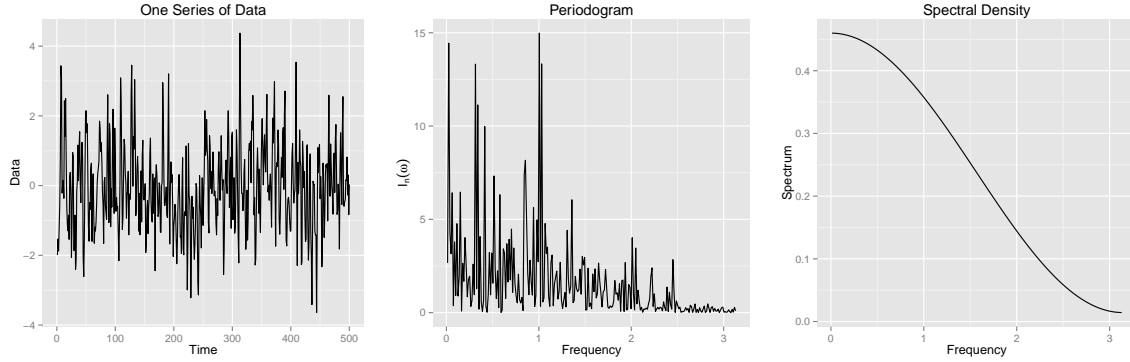


Figure 7: One draw, periodogram, and spectral density of an MA(1) model.

The tests of independence gave relatively similar rejection rates to  $\alpha$ , (figure 8) indicating no serious departure from pairwise independence at the neighboring Fourier frequencies. The Kolmogorov-Smirnov test gave reasonable rejection rates except for frequencies nearing  $\pi$ . This was the opposite of what was seen with the AR(4) model, and as sparsifying the Fourier frequencies would not result in asymptotically independent exponentially distributed periodogram ordinates at these frequencies, the Fourier frequencies nearing  $\pi$  would need to be excluded. These frequencies did not appear to be as important in the periodogram in figure 7, so it could be reasonable to proceed with a subset of frequencies of this type to use in Bayesian modeling of the spectral density.

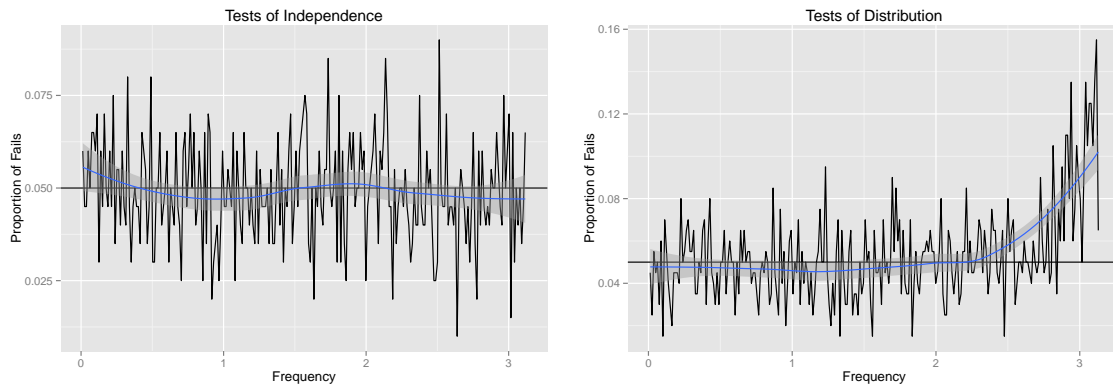


Figure 8: Tests of independence and distribution for an MA(1) model.

### 5.1.5 ARMA(4,1)

Lastly, we considered an ARMA model with both autoregressive (AR) and moving average (MA) components. A time series with this behavior has the form

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \phi_4 X_{t-4} + Z_t + \theta Z_{t-1} \quad (9)$$

where  $Z_t \sim WN(0, \sigma^2)$ . Specifically, we used the AR and MA coefficients equal to those of the AR(4) and MA(1) models discussed previously.

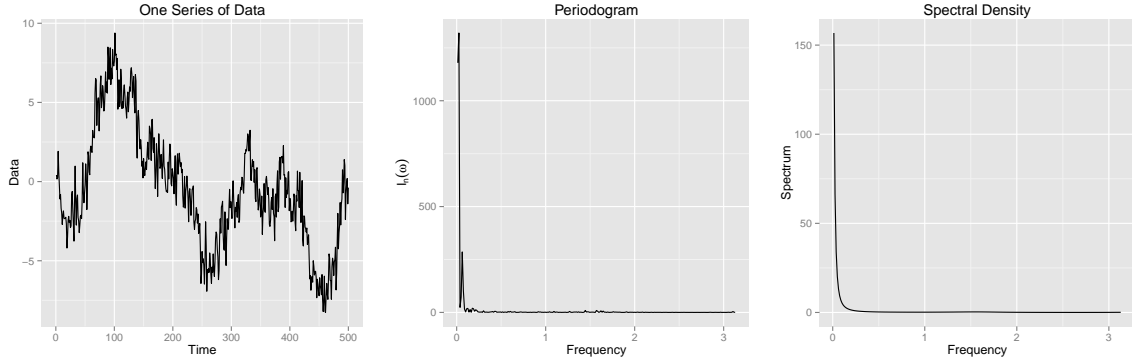


Figure 9: One draw, periodogram, and spectral density of an ARMA(4,1) model.

This ARMA model failed the pairwise Spearman rank correlation tests at frequencies close to zero, in a similar fashion as the AR(4) model. An attempt at sparcifying frequencies close to zero could be made to satisfy the independence conditions of Lahiri 2. However, the asymptotic exponential assumption is violated at the majority of the Fourier frequencies, as seen in figure 10. The combination of similar results from the AR(4) at frequencies near to zero and failures at frequencies near to  $\pi$  from the the MA(1) model result in the ARMA(4,1) satisfying asymptotic exponential distributional properties only at frequencies near  $\pi/2$ . Once again, this issue can not be resolved by our proposed method and would need to be explored at non-Fourier frequencies.

## 5.2 Sparce Frequencies for AR(4)

Lahiri's result 2 suggests that, for sparce frequencies and dense frequencies distant from zero, asymptotic independence should hold. We attempted to apply this concept to the AR(4) model to see if we could find a subset where pairwise independence would hold at all frequencies. In figure 6, asymptotic independence appeared to be satisfied at frequencies roughly larger than  $2\pi * 60/500$ . Therefore, we retained all the Fourier frequencies greater than  $2\pi * 60/500$  and focused on evenly

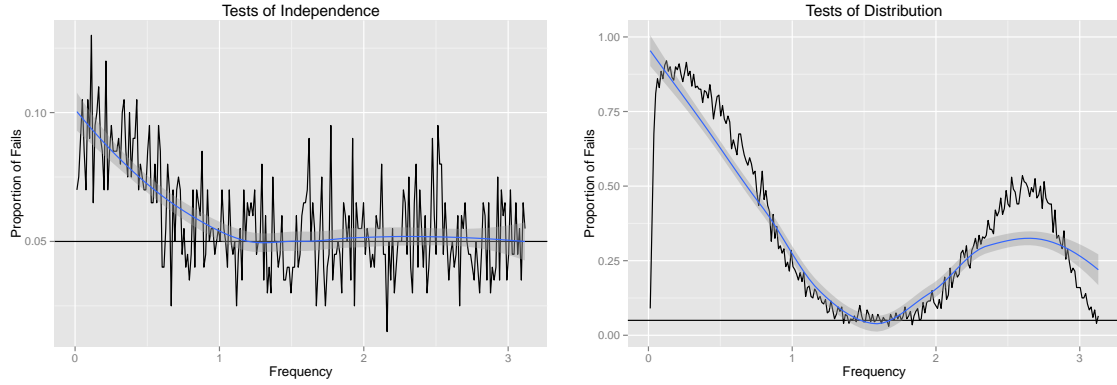


Figure 10: Tests of independence and distribution for an ARMA(4,1) model.

spaced subsets of frequencies from  $2\pi/500$  to  $2\pi * 60/500$ . We considered all possible subsets of Fourier frequencies from  $2\pi/500$  to  $2\pi * 60/500$  and tested the pairwise independence for each. Choice of starting frequency for each subset was also considered. For example, removing every other frequency would result in two possible subsets, one beginning at  $2\pi/500$ , the other at  $2\pi * 2/500$ . Examples of spacing frequencies by one (keep every other frequency) subsets and spacing frequencies by two (keep every third frequency) subsets and the rejection rates for pairwise tests of independence are shown in figures 11 and 12.

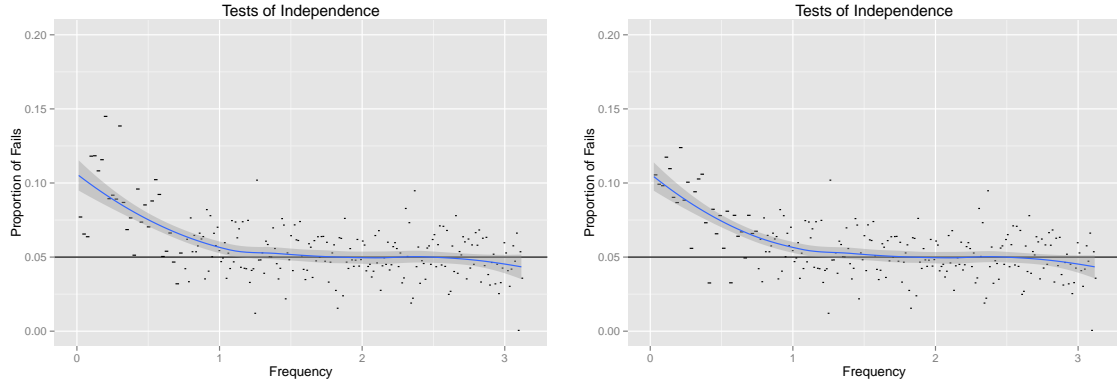


Figure 11: Tests of pairwise independence for every other Fourier frequency starting at the first and second Fourier frequencies, respectively, including the remaining full set of Fourier frequencies.

The pairwise rejection rates for the subsets shown in figures 11 and 12 decrease slightly from that in figure 6 but still show higher rejection rates than the  $\alpha$ -level of 0.05. It could be argued that for frequencies roughly greater than  $2\pi * 30/500$  in the second subset of every other spacings shown in figure 11, the asymptotic independence assumption holds. However, considering all other possible

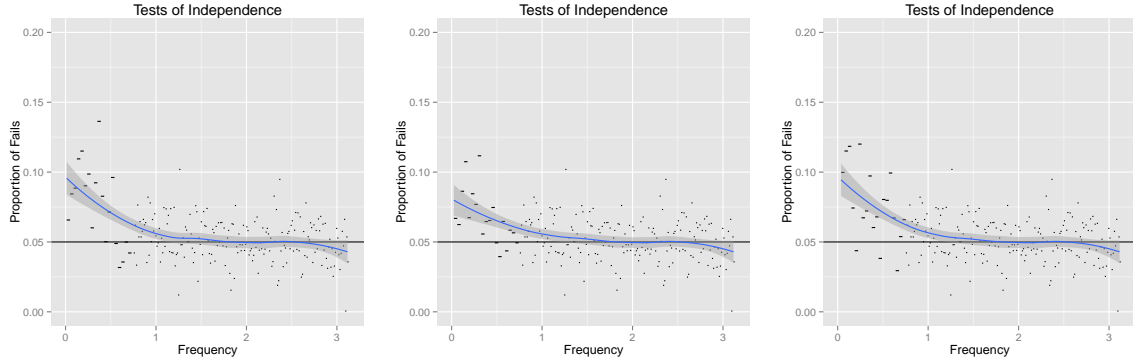


Figure 12: Tests of pairwise independence for every third Fourier frequency starting at the first, second and third Fourier frequencies, respectively, including the remaining full set of Fourier frequencies.

spacings between frequencies, we were unable to find a subset that resulted in pairwise independence at frequencies less than  $2\pi * 30/500$ . This could indicate that frequencies greater than  $2\pi * 25/500$  are sufficiently “distant from zero”, satisfying result 2. Therefore, a subset of Fourier frequencies where the asymptotic independence result would reasonably hold would be  $\{\omega_j = 2\pi j/500, j = 30, 32, 34, \dots, 60\} \cup \{\omega_j = 2\pi j/500, j = 61, 62, \dots, 249\}$ . Unfortunately, the estimated periodograms for time series of this AR(4) class are dominated by frequencies very close to 0, figure 5, so this subset would not be practical for a Bayesian analysis, or any analysis of the spectral density.

## 6 Discussion

This research focused on assessing asymptotic independence and exponential distributional assumptions of periodograms at the Fourier frequencies. In models where asymptotic pairwise independence did not hold at all of the frequencies, we explored the possibility of constructing a subset of frequencies where the asymptotic independence of result 2 was satisfied. In all of the ARMA models we considered, including those discussed, we were either unable to find a lack of pairwise independence at the full Fourier frequencies, or if independence was not satisfied, we were unable to find a suitable subset of frequencies such that result 2 held without removing a large number of important frequencies. For example, in the AR(4) example, pairwise asymptotic independence was satisfied for a subset of frequencies if those near zero were removed. This was a large number of frequencies, roughly  $\{2\pi*/500, \dots, 2\pi*30/500\}$  and happened to be the only frequencies that appeared important in the estimated periodograms. Therefore, this subset of frequencies would be relatively useless if anyone wanted to proceed with further practical analysis.

It should be noted that if a set of random variables are all pairwise independent, this does not imply they are jointly independent. In our tests of asymptotic independence, it is possible that in models in which we found an indication of pairwise asymptotic independence, a higher order dependent structure was still present. The use of pairwise independence tests also presented a multiple comparison issue which we have not accounted for here. To test all possible pairs of Fourier frequencies,  $\sum_{i=1}^k (k-1)$  pairwise tests are required where  $k = \lfloor \frac{n-1}{2} \rfloor$ . A familywise error rate correction such as Bonferroni may be too conservative in this application, so other multiple comparison corrections should be considered if pairwise independence is to be used to assess asymptotic independence of the periodogram.

In this paper we have focused largely on the assumption of asymptotic independence and the results by Lahiri 2. However, the asymptotic exponential assumption is equally as important if one were to implement a Bayesian analysis of the spectral density. If our tests of distribution failed at one or more frequencies, the subset of Fourier frequencies that could produce asymptotically independent exponential periodogram ordinates could not include these frequencies. In models such as the MA(2) and ARMA(4,1) we considered, this was a very large number of frequencies and therefore, using the Fourier frequencies for these specific models in a Bayesian analysis would be incredibly ineffective. In order to address the issue of asymptotic exponential distributions, we could examine spacings of non-Fourier frequencies near zero. However, the benefit of this option has not been explored.

Additionally, our research focused only on stationary ARMA time series due to the known closed-form of the spectral density function in these models. It should be noted that in more interesting cases, i.e. models with a seasonal structure, these models do not have known closed-form spectral densities, and thus we were unable to include them in this research. This is because without the values of the spectral density at tested frequencies, we would be unable to use the KS test for distributional fit. There are two alternatives that can be visited in future work. The first is using an unbiased and consistent estimator of the spectral density (like a window estimator [1]) as the mean of the scaled periodogram. A sensitivity analysis could be conducted to determine bandwidth. The second is to use a modification of the KS test called the Lilliefors test for the exponential distribution. This test is designed to test the null hypothesis,  $H_0$  : The random sample has the exponential distribution with unknown mean, against the alternative hypothesis,  $H_1$  : The distribution is not exponential [2]. At the time of this writing, the Lilliefors test is not implemented in R. Through the use of one of these modifications, it may be possible to apply this method to models with a more general structure.

As a final note, in our experience the spacings of frequencies depended strongly on what model the data was generated from. In practice, one will never know what process underlies the data

(thus a model fitting procedure is employed). Because of this, the spacing procedure will not be generalizable across multiple models. Furthermore, this method is impossible to employ in a live situation even with the modifications mentioned above because one will never have thousands of realizations of their data to check approximate exponential distribution and independence between frequencies of the periodograms.

## References

- [1] Peter J Brockwell and Richard A Davis. *Introduction to time series and forecasting*. Taylor & Francis US, 2002.
- [2] William J Conover. *Practical nonparametric statistics*. New York, Chichester: Wiley, 1998.
- [3] A Kolmogorov. “On the empirical determination of a distribution function”. In: *Breakthroughs in statistics*. Springer, 1992, pp. 106–113.
- [4] SN Lahiri. “A necessary and sufficient condition for asymptotic independence of discrete Fourier transforms under short-and long-range dependence”. In: *The Annals of Statistics* 31.2 (2003), pp. 613–641.