Independence of Periodogram Ordinates at the Fourier Frequencies

Andee Kaplan & Maggie Johnson

December 14, 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.

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

$$f(\omega_r) \sim \pi(\theta) \tag{2}$$

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

The periodogram values at ω_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 $\{X_t\}$ by decomposing $\{X_t\}$ into sinusoidal componenets. Spectral analysis is equivalent to "time domain" analysis based on the autocovariance function, but can be more useful in certain situations. 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 $\{X_t\}$. It is the function $f(\cdot)$ defined by

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

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

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

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 $n \to \infty$ the joint distribution of the periodogram at ω_r, $I_n(\omega_r)$, tends to that of k mutually indepependent random variables distributed as 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 in place of fixed frequencies reference this $\{2\pi j/n : j = 1, ..., n\}$. However, it has been shown that the asymptotic behavior of the periodogram at the Fourier frequencies does not hold as $n \to \infty$ (dig into reference to flesh this out). Lahiri [4] derived two results that describe behavior required for a set of frequencies to satisfy Result 1

Result 2. *In the absense 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 $|n(\omega_{ln} - \omega_{kn})| \to \infty$ as $n \to \infty$.*

(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 $|n(\omega_{ln} - \omega_{kn})| \to 2\pi l$ for some nonzero integer l as $n \to \infty$ [4].*

Result 2 lends to the idea that a subset of sufficiently spaced and and distant from 0 `Fourier` frequencies can be constructed such that the periodogram ordinates at these frequencies should be asymptotically independent. We explore these two results to determine if, for large n,

1. Result 1 will not hold at the Fourier frequencies, and

2. if (1) is true, a subset of Fourier frequencies that are sufficiently spaced and far from 0 can be constructed such that Result 1 holds.

3 Models

To explore the asymptotic behavior of the periodogram at the Fourier frequencies, we used several time series models with known spectral densities. The included models are *a)* IID Gaussian(0, 1); *b)* AR(1) with $\phi = 0.5$; *c)* AR(4) with $\boldsymbol{\phi} = [0.08, 0.33, 0.1, 0.45]$; *d)* MA(1) with $\theta = 0.7$; *e)* MA(2) with $\boldsymbol{\theta} = [0.7, 0.3]$; *f)* ARMA(4,1) with $\boldsymbol{\phi} = [0.08, 0.33, 0.1, 0.45]$ and $\theta = 0.7$; and *g)* ARMA(4,2) with $\boldsymbol{\phi} = [0.08, 0.33, 0.1, 0.45]$ and $\boldsymbol{\theta} = [0.7, 0.3]$.

IID Gaussian

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

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

This model was used as an initial baseline check to support our 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,σ^2). Models of this form have spectral density

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

Without the exact knowledge of the spectral density function, we would not be able to test the distributions of the periodogram at each fequencies. mention moving window?

4 Methods

For Result1 to hold for a known spectral density at the full Fourier frequencies, or a subset of Fourier frequencies $\boldsymbol{\omega}$, 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 in to two parts: one, a test of asymptotic exponential distribution at each ω_i, and two, a method to test independence over ω. 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 indepence at similar rates as the Type-I errors of our tests (α).

Procedure

1. Simulate M draws from $X_1, \ldots, X_n$ where $\{X_t\}$ is a stationary time series from one of three models.

2. Obtain M periodograms using the fourier frequencies from $(0, \pi)$, $\omega_j = \frac{2\pi j}{n} : j = 1, \ldots, \lfloor n/2 \rfloor$.

3. Obtain $\lfloor \frac{n-1}{2} \rfloor$ values of the spectral density at each fourier frequency $f(\omega_j)$. These will be known by design.

4. Simulate M draws from the $\lfloor \frac{n-1}{2} \rfloor$ different distributions $\text{Exp}(2\pi f(\omega_r))$

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 each 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 sparcer frequencies and repeat from step 2 at the chosen frequencies rather than fourier frequencies.

4.1 Tests

Spearman's Rank Correlation Spearman's rank correlation ρ, 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) ω_j and ω_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 $n \leq 100$, an 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 ω_j were exponentially distributed with mean parameter $2\pi f(\omega_j)$, where $f(\omega_j)$ is the spectral density evaluated at ω_j.

5   Results

5.1   Tests for 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 were conducted first using the entire set of Fourier frequencies, $\{w_j = \frac{2\pi j}{n}, j = 1, ..., 249\}$ 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 clearly seen in the randomness of the periodogram.
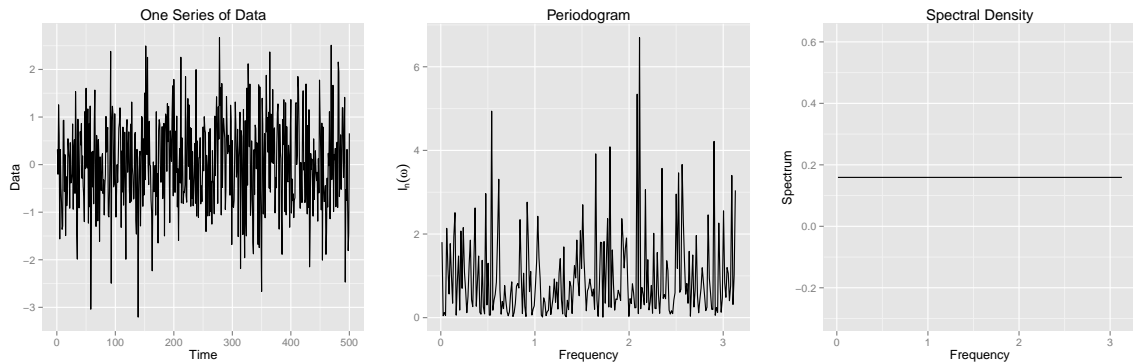

Figure 1: One draw, periodogram, and spectral density of a Gaussian IID Model.

Tests of exponential distribution and pairwise independence of neighboring Fourier frequencies gave rejection rates relatively similar to the α-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.
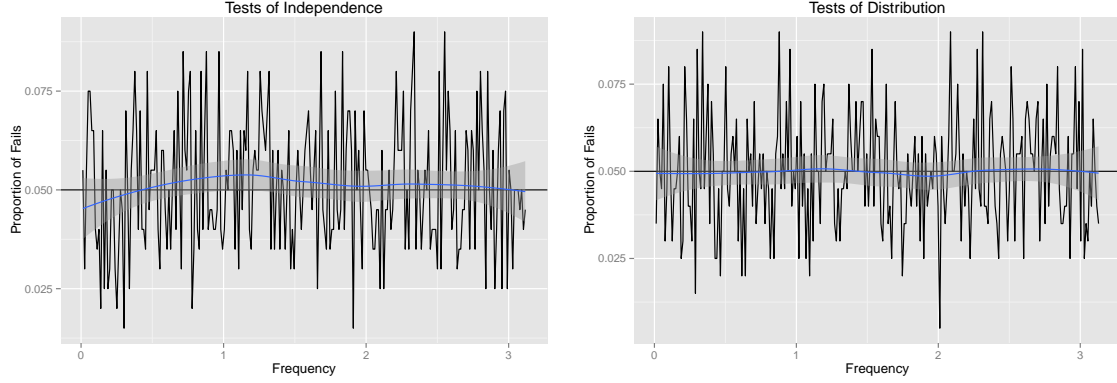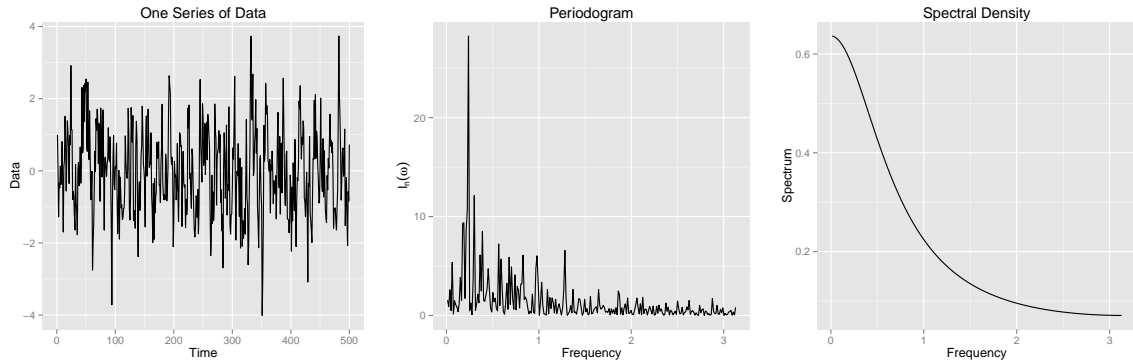

Figure 2: Tests of independence and distribution for a Gaussian IID Model.

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 $\phi = 0.05$. For a timeseries $\{X_t\}$, this model has the form

$$x_t = \phi x_{t-1} + \epsilon_t \tag{7}$$

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

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

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 neighboring time points only.
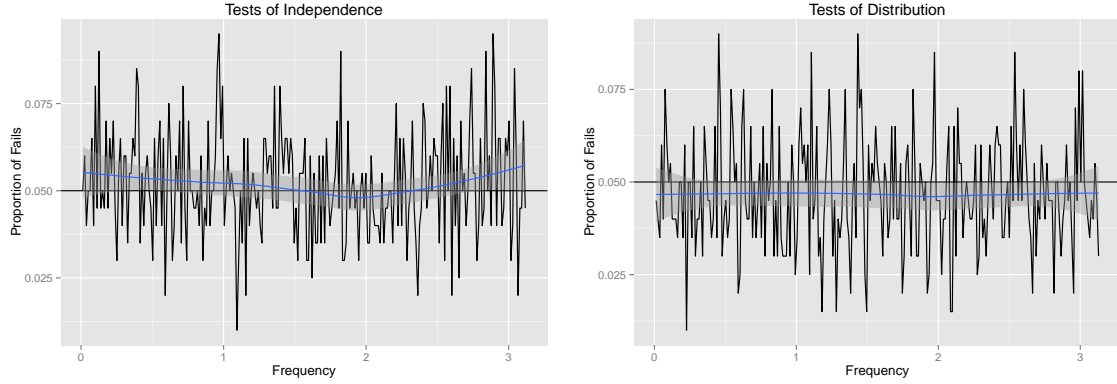

Figure 4: Tests of independence and distribution for an AR(1) model.

5.1.3   AR(4)

To experiment with a timeseries with a slightly 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 timeseries following this model, is dominated by frequencies very close to 0, Figure 5.
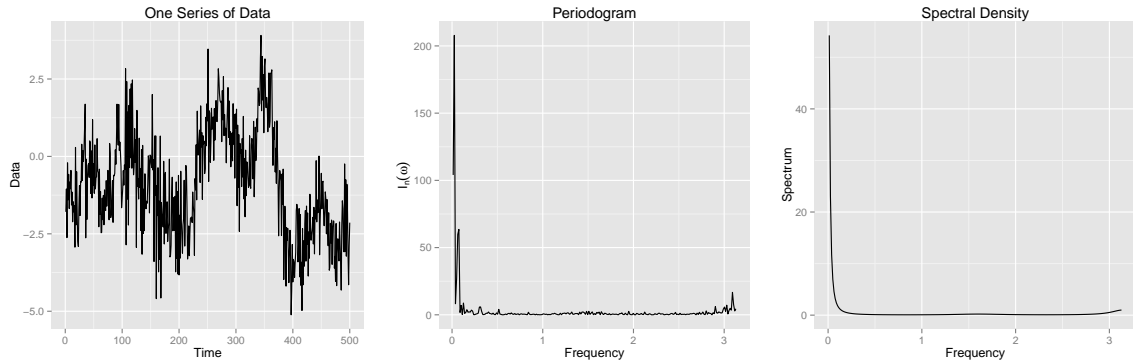

Figure 5: One draw, periodogram, and spectral density of an AR(4) model.

7

Unlike the IID Gaussian and AR(1) models, this AR(4) model does not have rejection rates all close to α for the independence and exponential tests at neighboring Fourier frequencies. At frequencies closer to 0, the pairwise independence tests fail at rates consistently higher than α and likewise, especially for frequencies less than 1, the periodogram ordinates should not be considered to be asymptotically exponentially distributed, see Figure 6. Sparcifying the Fourier frequencies closer to 0 will not change that the periodograms at these frequencies are asymptotically exponential. However, we explore if sparcifying the frequencies closer to 0 will satisfy Result 2 in Section 5.2.
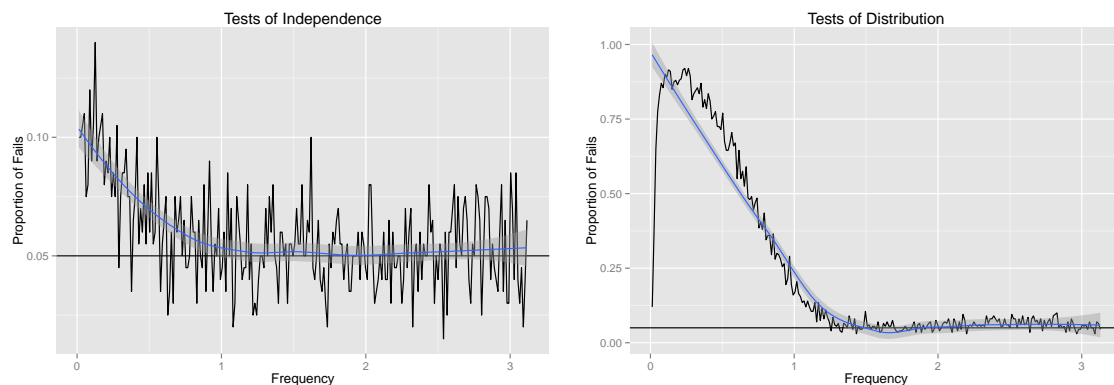

Figure 6: Tests of independence and distribution for an AR(4) model.

5.1.4 MA(1)

We also experimented with a MA(1) model with $\theta = 0.3$ where a timeseries with this behavior has the form

$$x_t = \epsilon_t - \theta\epsilon_{t-1} \tag{8}$$

where $\epsilon_t \sim WN(0, \sigma^2)$. Figure 7 shows that the estimated periodogram for a simulated timeseries of this form puts more importance on frequencies closer to 0 than to π, as in the AR(1) model, but shows a more random structure than the AR(1).
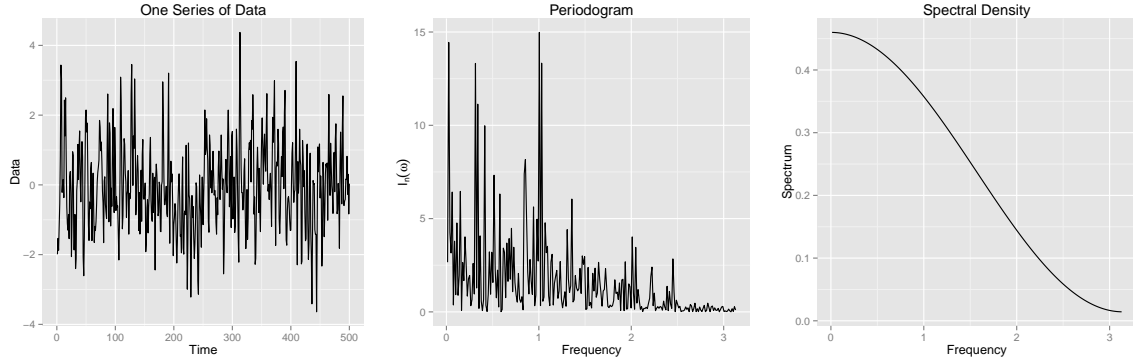
Figure 7: One draw, periodogram, and spectral density of an MA(1) model.

The tests of independence seem to give relatively similar rejection rates to α, (Figure 8) indicating no serious departure from pairwise independence at the neighboring Fourier frequencies. The Kolmogorov-Smirnov test gives reasonable rejection rates except for frequencies nearing π. This is the opposite of what was seen with the AR(4) model, and as sparcifying the Fourier frequencies would not change whether the periodogram ordinates at these frequencies are exponentially distributed. This model would only result in asymptotically independent exponentially distributed periodogram ordinates if the Fourier frequencies nearing π were excluded. As these frequencies do not appear to be as important in the periodogram in Figure 7, it could be reasonable to proceed with a subset of frequencies to use in Bayesian modeling of the spectral density.
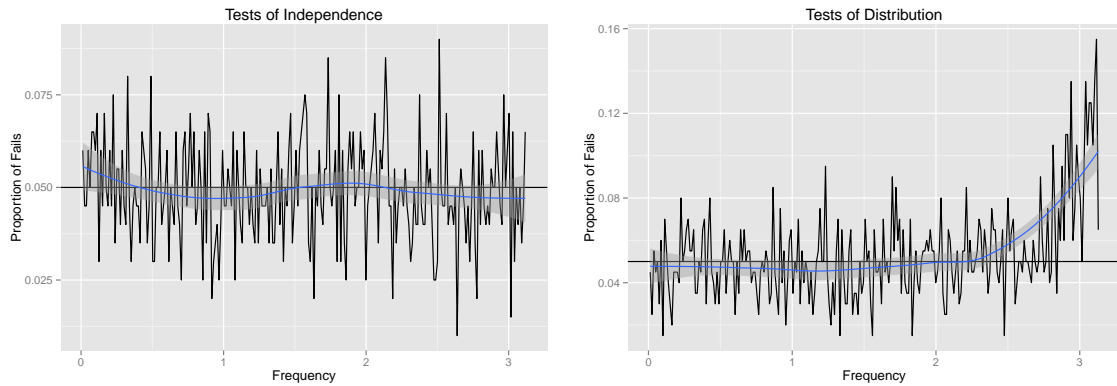

Figure 8: Tests of independence and distribution for an MA(1) model.

5.1.5 ARMA(4,1)

Lastly, we considered an ARMA model with both autoregressive and moving average components. Specifically, the AR and MA coefficients were equal to that of the AR(4) and MA(1) models discussed previously.
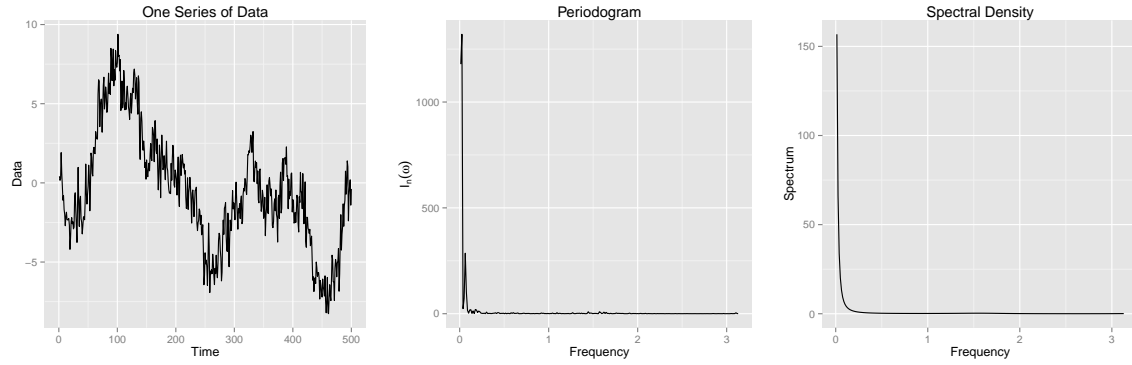

Figure 9: One draw, periodogram, and spectral density of an ARMA(4,1) model.

This ARMA model fails the pairwise independence tests at frequencies close to 0, similarily to the AR(4) model. An attempt at sparcifying frequencies close to 0 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, Figure 10. Similar results from the AR(4) at frequencies near to 0, and failures at frequences near to π from the the MA(1) model appear to result in asymptotic exponential distributional properties only at frequencies surround $\pi/2$.
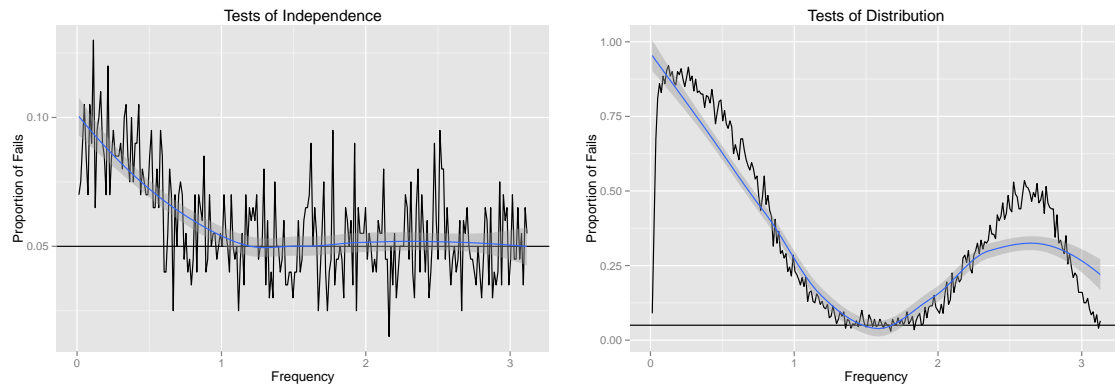

Figure 10: Tests of independence and distribution for an ARMA(4,1) model.

5.2 Sparcify Frequencies for AR(4)

Lahiri's result 2 suggests that, for sparce frequencies and frequencies distant from 0, asymptotic dependence should hold. We attempted to apply this to the AR(4) model to see if we could find a subset where pairwise independence would hold at all frequencies. We looked at all possible evenly spaced subsets of frequencies from $2\pi/500$ to $2\pi 60/500$, For example, removing every other frequency would result in two possible subsets, one beginning at $2\pi/500$, the other at $2\pi/250$. Examples of every other frequency and every third frequency subsets and the rejection rates for pairwise tests of independence are shown in Figures 11 and 12.
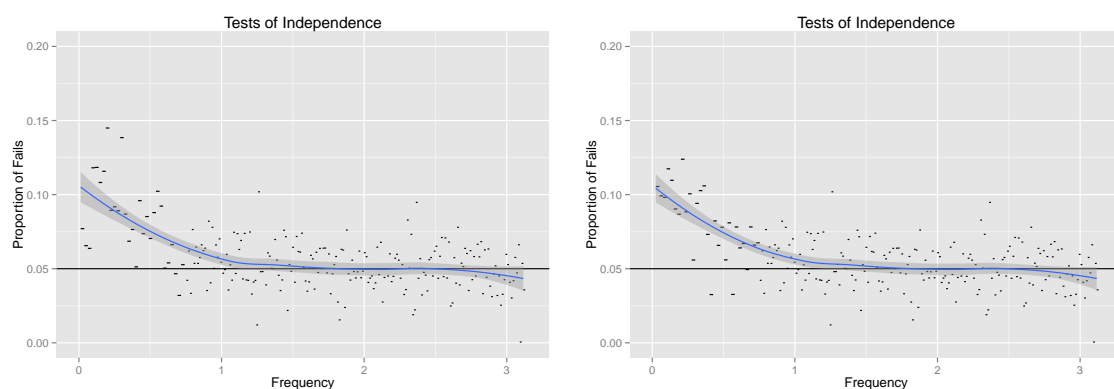

Figure 11: Tests of pairwise independence for every other Fourier frequency starting at the first and second Fourier frequencies, respectively.
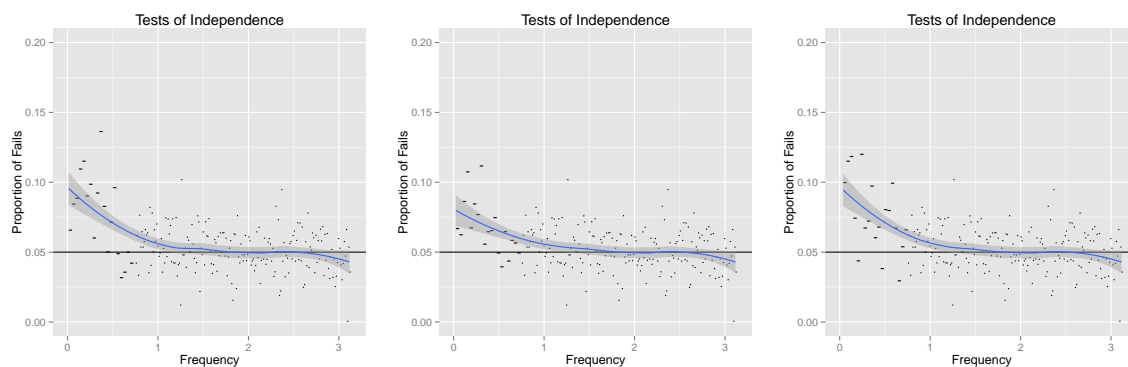

Figure 12: Tests of pairwise independence for every third Fourier frequency starting at the first, second and third Fourier frequencies, respectively.

Ultimately, considering all subsets of frequencies from $2\pi/500$ to $2\pi 60/500$, the pairwise independence result will not hold for frequencies less than $2\pi 25/500$, see Appendix ?. This could indicate that frequencies greater than $2\pi 25/500$ are sufficiently "distant from 0," satisfying Result 2. more

6   Discussion

Issues

1. Multiple comparison

2. ARMA(4,2), MA(1), MA(2)

3. Exponential Fails

4. Usability

   - ARMA useless, Models w/more structure
   - spacing depends on model
   - can never really be used in practice

5. Computation time

6. Pairwise independence $\neq$ Joint independence

References

[1]   Peter J Brockwell and Richard A Davis. *Introduction to time series and forecasting.* Taylor & Francis US, 2002.

[2]   William Jay Conover. "Practical nonparametric statistics". In: (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.