Resenhas IME-USP 2000, Vol. 4, No. 4, 423 - 431.

# Examining an Irregularly Sampled Time Series for Whiteness

David R. Brillinger

**Abstract:** Suppose there is a stationary time series $\{Y(t)\}$ whose values at scattered times, $\{Y(\tau_1),\ Y(\tau_2),\ \ldots\}$, are available. Suppose it is of interest whether the series itself is white noise. The empirical Fourier transform is proposed to address this question. In the white noise case, the expected value of the mod-squared (i.e. of the periodogram) will be constant. Under regularity conditions the Fourier transform itself will be asymptotically complex normal and confidence limits for the power spectrum may be constructed using the exponential distribution. In practice $Y$ might be the residuals of a time series model fit.

**Key words:** Fourier inference, goodness of fit, periodogram, stationary increments, stationary point process, stationary time series, white noise.

## 1. Introduction

In a variety of empirical studies the assessment of a fitted model is carried out by an examination of the residuals. In particular this is often the case when working with time series, see Box and Jenkins (1970). A researcher may wonder if the residuals of a time series model fit are white, that is correspond to a sequence of independent identically distributed observations. Being white often means a model fitting well. In practice the series may have been recorded at irregularly spaced time points $\{\tau_1,\ \tau_2,\ \ldots\}$ and so the residuals are themselves at irregularly placed points. Missing values is one source of irregular spacings, e.g. Jones (1971), Brillinger (1972), Dunsmuir and Robinson (1981). So too is the situation where points are randomly deleted, see Bloomfield (1970) and Thrall (1980). On the other hand recording at equi-distant time points may be inconvenient or simply impossible.

In this work the specific desire is to develop a statistic useful for assessing the whiteness of a process measured at irregularly spaced time points having in mind the alternative of a nonconstant power spectrum. One reason for the research is that under an assumption of independence, the (approximate) distributions of derived quantities computed further along in a study may be simpler.

One basic statistic in the analyses proposed is the empirical Fourier transform.

$$d_X^T(\lambda) \;=\; \sum_{1}^{N(T)} Y(\tau_j) exp\{-i\lambda\tau_j\}, \qquad -\infty \;<\; \lambda \;<\; \infty \tag{1}$$

where $N(T)$ is the number of time points available in the interval $(0,T)$. (The reason for using the subscript $X$ here will be apparent later.)

The basic quantity proposed for examination of the hypothesis of whiteness is the periodogram

$$I_{XX}^T(\lambda) = \frac{1}{2\pi T}|d_X^T(\lambda)|^2 \tag{2}$$

The periodogram will be seen to be asymptotically exponential under stationary mixing assumptions and seen to have expected value constant in $\lambda$ under an assumption of constant power spectrum.

In Section 2 the three types of process of concern in this paper are discussed. In Section 3 linear and quadratic expressions in the data are discussed. Section 4 discussed central limit theorems for pertinent Fourier transforms. Section 5 presents an example of the procedure applied to the residuals of a fit to some river flow data collected near Manaus. In Section 6 extensions are mentioned and there is discussion in Section 7.

## 2. Point processes, time series and stationary increment processes

### *Point processes*

In this work the time series values considered are those recorded at times $\tau_1, \tau_2, ..., \tau_{N(T)}$. The developments of the paper are asymptotic and the $\tau$-values will be viewed as a stretch of a doubly infinite sequence $\{\tau_j\}$. The sequence will be assumed in turn to be a realization of a stationary stochastic point process or of a deterministic sequence satisfying generalized harmonic analysis type conditions.

For example $N(t) = \#\{\tau_j \leq t\}$ may be a stationary point process with spectral representation

$$N(t) = \int \frac{exp\{i\lambda t\} - 1}{i\lambda t} dZ_N(\lambda)$$

with the rate and power spectrum parameters given by $EdN(t) = p_N dt$ and

$$cov\{dZ_N(\lambda), dZ_N(\mu)\} = \delta(\lambda - \mu) f_{NN}(\lambda) d\lambda d\mu$$

respectively. Particular cases are the homogeneous Poisson, the renewal process and the jittered regular process ($\tau_j = jh + \epsilon_j$, for some $h > 0$, the $\epsilon_j$ being i.i.d.).

In the deterministic case the $\tau_j$ maybe given by $\tau_j = G^{-1}(j/T)$ for some c.d.f. $G$. In Brillinger (1973), they are assumed to satisfy

*Assumption II.* $\tau_1, \tau_2, ...,$ is an increasing sequence of positive numbers with the properties,

$$|N(s) - N(t)| \leq A + B|s - t|$$

$-\infty < s, t < \infty$ for some finite $A, B$. Also the limits

$$\lim_{T\to\infty} N(T)/T = p_N$$

and

$$\lim_{T\to\infty} T^{-1} \int_0^T [N(t+u) - N(t)]dN(t) \ = \ M_{NN}(u) \ = \ up_N^2 + C_{NN}(u)$$

exist for almost all $u$.

A power spectrum may be defined here by

$$f_{NN}(\lambda) \ = \ p_N/2\pi + \int exp\{-i\lambda t\}dC_{NN}(u)/2\pi$$

*Time series*

It is common to assume that time series data are a stretch $Y(t),\ t = 0, ..., T-1$ of a stationary time series. With an assumption of mixing, one can develop the large sample properties of computed statistics.

In the work here there will be a latent time series $\{Y(t)\}$ with $-\infty < t < \infty$ or $t = 0, \pm 1, \pm 2, ....$ There will be a basic time interval, $(0, T)$, with the observation times $\tau_1, ..., \tau_{N(T)}$ falling in it. The values available for analysis are the $\tau_j$ and the $Y(\tau_j)$.

Suppose that the series $Y$ is stationary with spectral density $f_{YY}(\cdot)$. Then $Y$ has a Cramér representation

$$Y(t) \ = \ \int e^{it\lambda} dZ_Y(\lambda)$$

where $Z_Y$ has uncorrelated increments and

$$cov\{dZ_Y(\lambda), dZ_Y(\mu)\} \ = \ \delta(\lambda - \mu) f_{YY}(\lambda) d\lambda d\mu$$

*Stationary increment processes*

A process with stationary increments is a convenient mathematical concept for the present work. It contains both point processes and time series as particular cases. The definition is: a process $\{X(t), -\infty < t < \infty\}$ whose joint distributions of differences $X(t_1+u) - X(t_1), ..., X(t_k+u) - X(t_k)$ do not depend on $u$ for all $K = 1, 2, ....$ These processes have spectral representations

$$X(t) \ = \ \int \frac{exp\{i\lambda t\} - 1}{i\lambda t} dZ_X(\lambda) \tag{3}$$

with

$$cov\{dZ_X(\lambda), dZ_X(\mu)\} \ = \ \delta(\lambda - \mu) f_{XX}(\lambda) d\lambda d\mu$$

The above point and time series are particular cases, e.g. , Kolmogorov (1940) and Brillinger (1972).

The case of a sampled series, $Y$, corresponds to $dX = YdN$. This is the reason for using the subscript $X$ in expression (1), in particular

$$d_X^T(\lambda) = \int_0^T exp\{-i\lambda t\}Y(t)dN(t)$$

## 3. First- and second-order moments

Now expression (1) may be written

$$d_X^T(\lambda) = \int \Psi^T(\lambda - \alpha)dZ_Y(\alpha)$$

with

$$\Psi^T(\lambda) = \sum_{j=1}^{N(T)} exp\{-i\lambda\tau_j\}.$$

Supposing $Y$ has mean 0 and that the $\tau_j$ are fixed then $Ed_X^T(\lambda) = 0$ and

$$E|d_X^T(\lambda)|^2 = var\{d_X^T(\lambda)\} = \int |\Psi^T(\lambda - \alpha)|^2 f_{YY}(\alpha)d\alpha. \tag{4}$$

One sees that the expected value smooths the spectrum $f_{YY}$, i.e. convolves it with the function $|\Psi^T|^2$. The notable thing in the present context is that if $f_{YY}$ is constant in $\lambda$, then so is $E|d_Y^T|^2$ .

Consider next the case of $\{\tau_j\}$ a stationary point process, independent of the zero mean stationary time series $Y$. The process $dX = YdN$ has stationary increments and power spectrum

$$f_{XX}(\lambda) = p_N^2 f_{YY}(\lambda) + \int f_{NN}(\lambda - \alpha)f_{YY}(\alpha)d\alpha \tag{5}$$

and

$$var\{d_X^T(\lambda)\} = \int |\Psi^T(\lambda - \alpha)^2 f_{XX}(\alpha)d\alpha.$$

If $Y$ is zero mean white noise, i.e. $f_{YY}(\cdot)$ is constant, then $E|d_X^T|^2 = var\{d_X^T\}$ is also constant in $\lambda$.

## 4. Central limit theorems

A variety of central limit theorems have been developed for empirical Fourier transforms. In particular, Brillinger (1973) provides a central limit theorem for statistics of the form

$$\sum_{1}^{N(T)} Y(\tau_j)$$

for the cases of $\{\tau_j\}$: i) a stationary mixing point process and ii) a deterministic sequence satisfying Assumption II above. The same arguments and assumptions may be applied here to find that $d_X^T(\lambda)$ is asymptotically complex normal with mean 0 and variance $2\pi T f_{XX}(\lambda)$ with $f_{XX}(\lambda)$ given by (5).

Specifically one has

*Theorem 1.* Let $Y(t), -\infty < t < \infty$ be a zero mean stationary time series and let $N(t)$ an independent stationary point process. Suppose that the stationary increment process $X$ of $dX = YdN$ satisfies the mixing Assumption 2.2 of Brillinger (1972), then the empirical Fourier transform (1) is asymptotically normal.

and

*Theorem 2.* Let $Y(t), -\infty < t < \infty$ be a zero mean stationary time series satisfying Assumption I of Brillinger (1973) and let $N(t)$ be a deterministic step function satisfying Assumption II above, then the empirical Fourier transform (1) is asymptotically normal.

The asymptotic distribution, $N^C(0, 2\pi T f_{XX}(\lambda))$, of the empirical FT in these two cases leads to the periodogram (2) being asymptotically exponential with parameter $f_{XX}(\lambda)$, see Brillinger (1975). Large sample marginal confidence intervals may then be set directly using the exponential distribution as in Brillinger (1975).

## 5. A Brazilian example

Close to the confuence of the Rio Negro, near Manaus, the River Solimões divides into two channels, the Careiro and the Amazon. In collaborative work with Professor H. O'Reilly Sternberg, Brillinger, Preisler, Ager and Kie (2000), the relationship of the flow rates of the Solimões and the downstream Careiro is being studied. The concern is that if the flow rates change, as well they might, this would lead to substantial difficulties for the population of the region.

In collecting the data because of the vagaries of ship availability the measurements were made at irregular times. Figure 1 shows time series plots of the cumulative counts of the measurement times. The slope of the curve gives the rate at which data were being collected. It would be constant if the data had been collected regularly. One notes some periods, eg. around 1995, with little data collection taking place.

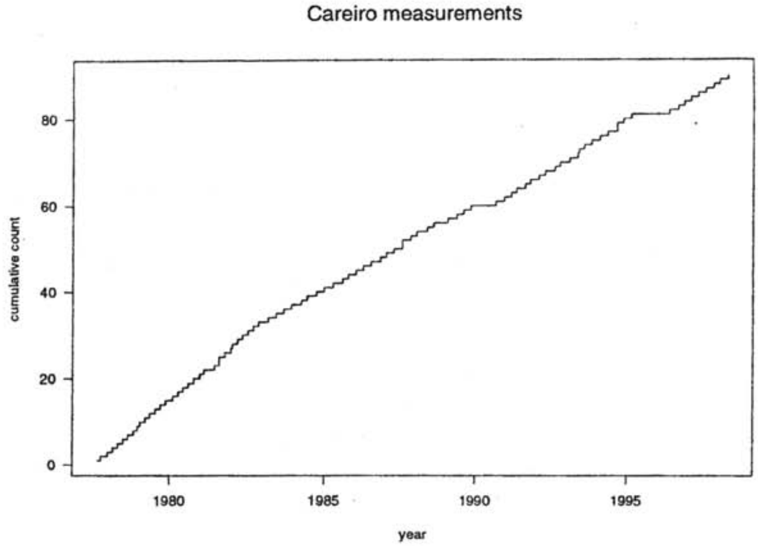


Figure 1: Cumulative count of measurements time.

As part of the analyses of Brillinger, Preisler, Ager and Kie (2000) the Careiro values were predicted assuming a model of trend plus annual cosinusoid and residuals of the fit computed. A question was whether the error series could be approximated as white in further stages of the analysis. Figure 2 is a time series plot of the residual series. One notes gaps when no measurments were being made. The horizontal line is at the level 0 and the values are seen to be fluctuating about that level.

Figure 3 provides the periodogram statistic (2). Approximate 95 % confidence intervals have been added to the figure. It was a pleasant surprise to note that 4.87 % of the values fell outside of those limits. In consequence an assumption of whiteness in the future work did not appear inappropriate here.

## 6. Extensions

The research work in the paper Brillinger, Preisler, Ager and Kie (2000) stimulated the development of variants of the above results to apply to the case of processes with stationary increments. In the work of Brillinger, Preisler, Ager

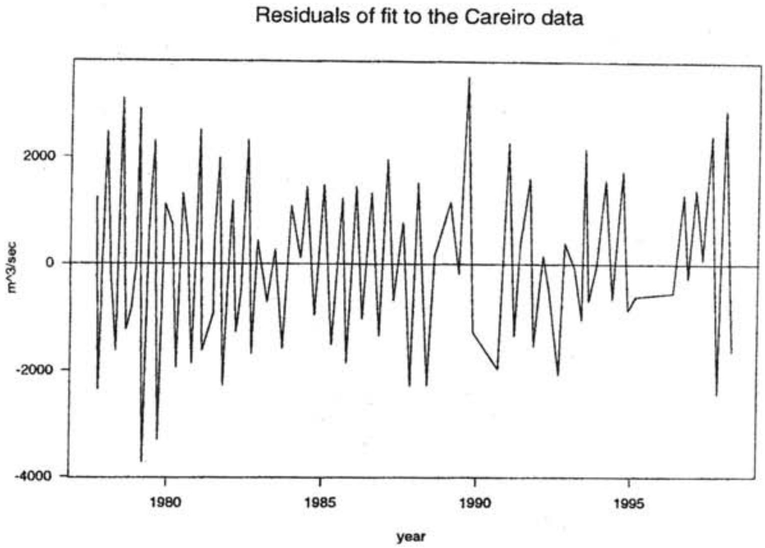


Figure 2: Residuals of fit to Careiro values via trend plus annual cosinusoid function.

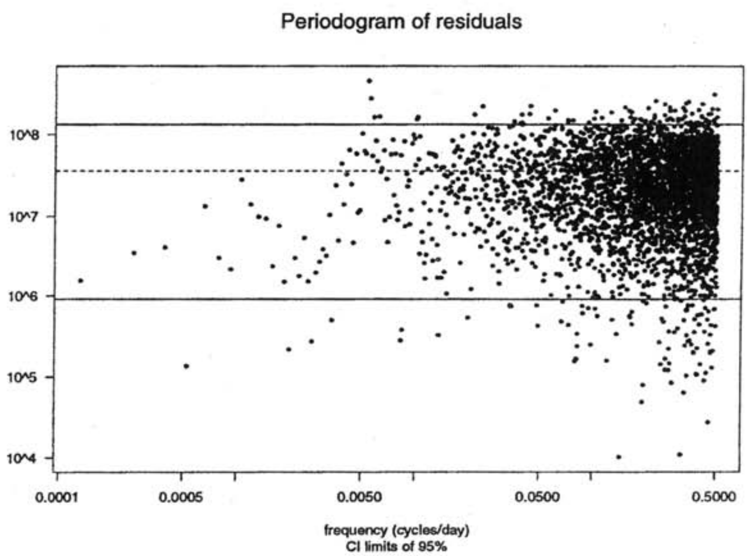


Figure 3: The periodogram statistic (2).

and Kie (2000) paths of elk movement were modelled as realizations of bivariate diffusion proceses. Sampled locations along the paths were available. Expressions for the drift and variance terms were assumed and it was desired to assess the model, particularly whether the residuals were approximately white. The times of measurement were unequally spaced.

Suppose then that $X$ is a process with stationary increments and that the values $(\tau_j, X(\tau_j))$ are available. Extending the preceding arguments a statistic to consider is:

$$d_X^T(\lambda) \;=\; \sum_{1}^{N(T)-1} [X(\tau_{j+1}) - X(\tau_j)] exp\{-i\lambda\tau_j\}.$$

Using the spectral representation (3) one quickly obtains an expression for $d_X^T(\lambda)$ and thereby sees that the expected periodogram is constant in $\lambda$ when the spectrum is.

Another situation where the work of this paper is pertinent occurs when one has a spatial point process with points of two types and one seeks to predict the values at points at one type from values at the other type points. This happens in seismology when one has acceleration values and Modified Mercalli intensity values. It is of interest whether the residuals are approximately white as once again this will simplify later analyses.

## 7. Discussion

Reasons as to why the Fourier approach was adopted include that the approximate sampling properties are clear and that plausible alternatives to whiteness are often simply described in the frequency domain.

A warning is necessary: apparent constancy of the periodogram may not imply whiteness. It may well be that all the lag values values of the autocovariance function do not occur amongst the $\tau_j - \tau_k$ even as $T$ grows to $\infty$.

In some cases it may be appropriate to use the tapered variant

$$\sum_j h^T(\tau_j) Y(\tau_j) exp\{-i\lambda\tau_j\}$$

of (1), eg. when the spectrum falls of rapidly.

## Acknowledgement

This research was supported by the NSF Grants DMS-9704739, DMS-9971309 and INT-9600251.

## References

BLOONFIELD, P. (1970) Spectral analysis with randomly missing observations, *J. Royal Statistical Soc. B* **32** , 369-380.

G. E. P. BOX AND G. M. JENKINS (1970) Time Series Analysis; Forecasting and Control Holden-Day, San Francisco.

BRILLINGER, D. R. (1972) The spectral analysis of stationary interval functions. *Proc. Sixth Berkeley Symp. Math. Stat. Prob.* **1**, 483-513.

BRILLINGER, D. R. (1973) Estimation of the mean of a stationary time series by sampling. *J. Applied Probability* **10**, 419-431.

BRILLINGER. D. R. (1975) Time Series: Data Analysis and Theory. Holt Rienhart, New York.

BRILLINGER, D. R. AND STERNBERG, H. O'REILLY (1999) Análise de séries temporais bidimensionais em que os componentes são amostrados em instantes diferentes. Talk at ESTE 8, Nova Friburgo.

BRILLINGER, D. R. , PREISLER,H. K. , AGER, A. A. AND KIE, J. G. (2000) Modelling the movements of free ranging animals. Submitted.

JONES, R. H. (1971) Spectrum estimation with missing observations. *Ann. Inst. Statist. Math.* **23**, 387-398.

KOLMOGOROV, A. N. (1940) Curves in Hilbert space invariant with regard to a one parameter group of motions. *Dokl. Akad. Nauk. SSSR* **26** 6-9.

DUNSMUIR, W. AND ROBINSON, P. (1981) Estimation of time series models in the presence of missing data. *J. Amer. Statist. Assoc.* **76**, 560-568.

THRALL, A. D. (1980) A spectral analysis of a time series in which probabilities of observation are periodic. *Time Series* (Ed. O. D. Anderson), 357-371. North-Holland.

**David R. Brillinger**
Department of Statistics
University of California
Berkeley, CA 94720
brill@stat.berkeley.edu
**USA**