Poisson Maximum Likelihood Spectral Inference

Steven Kay*
Dept. of Electrical, Computer, and Biomedical Engineering
University of Rhode Island
Kingston, RI 02881
401-874-5804 (voice) 401-782-6422 (fax)
kay@ele.uri.edu

Keywords: signal processing, spectral analysis, pattern classification

November 16, 2017

Abstract

The standard method of spectral inference assumes a Gaussian model for the data. A less known but alternative spectral representation can be based on a nonhomogeneous Poisson process in frequency. This leads to a new likelihood function that can be used for spectral inference. In particular, the very important problems of spectral estimation and spectral classification can be approached with this new likelihood function. If an exponential model is assumed, then the parameter estimation reduces to a simple convex optimization for the spectral estimation problem. For the classification problem with known spectra the classification performance based on the Poisson likelihood function is shown by simulation to outperform the Gaussian classifier in terms of robustness. Finally, a perfect analogy between the Poisson likelihood measure and the Kullback-Leibler measure for probability density functions is established and discussed.

1 Introduction

The spectral representation for a wide sense stationary (WSS) random process is of utmost importance in modern signal processing theory and practice [1]. It forms the basis for the the science (and sometimes "art") of spectral estimation [7]. This representation relies on summing together a set of sinusoids of *fixed frequencies* with random amplitudes and random phases. Using a complex

*This work was partially supported by the Air Force of Office of Scientific Research and administered by Syracuse University under grant AFOSR-FA9550-09-0064.

representation, one can view the complex amplitude in the context of a random spectral measure, which is uncorrelated over disjoint frequency sets. Although not well known, one could also model the frequencies as random point events distributed according to a nonhomogeneous Poisson process. This leads to a different spectral representation, and consequently to a different likelihood function for spectral inference. A similar representation had been proposed earlier that uses *two* independent Poisson point processes [6]. Its intent was to allow Gaussian random processes to be easily simulated on a computer. The representation presented herein is simpler in that it uses only a single Poisson process and hence is more intuitive. It should also be mentioned that our representation is more flexible in that not only can it be used to generate a random process with a given power spectral density (PSD) but also with *an arbitrary first-order probability density function (PDF)*, if the PDF is infinitely divisible [11].

Once this spectral representation is adopted, one can derive its likelihood function [18]. With some modifications necessary to account for the fact that the random point frequencies are not actually observed in most practical situations, a realizable approximation to the likelihood function is proposed. One situation for which the frequencies are available is in the modeling of aural neural responses, in which the frequencies correspond to neural impulses along the auditory nerve [17]. We will term the approximate likelihood function to be derived as the *realizable Poisson likelihood function* (RPLF).

Using the RPLF we propose two new approaches to the important problem of spectral inference. The first is a model-based spectral estimate. As will be shown, the RBLF naturally suggests that for mathematical tractability and statistical efficiency the spectrum be modeled as the exponential power spectral density (EPSD), first proposed by Bloomfield [2]. The use of the RPLF leads to a simple convex maximization to estimate the model parameters that is easily carried out numerically. In contrast, Bloomfield's use of the asymptotic Gaussian likelihood function leads to an intractable set of maximum likelihood equations to solve. This occurs even though Bloomfield uses the well-known Whittle asymptotic Gaussian approximation. The second application of the RPLF is in discerning between random processes with different PSDs. It is shown via computer simulation that the use of the Gaussian likelihood function is inferior to the RPLF in terms of robustness. In particular, when the data set is heaving perturbed by noise, the classification performance of the Gaussian approach degrades rapidly, but not so for the RPLF.

Lastly, it is shown that maximization of the RPLF is equivalent to minimizing the KL measure between PDFs. Thus, there is a perfect analogy between PSDs and PDFs, allowing one to use all the theoretical machinery developed for PDF estimation to be applied to spectral inference. This analogy underlies the use of the "exponential PDF family" as a model for the PSD, i.e, Bloomfield's spectral estimator. It is actually a remarkable correspondence, obvious in hindsight, since both PSDs and PDFs are nonnegative functions, with the only difference being a normalization so that the PDF integrates to one. For a PSD the total power can be anything but in practice since the

overall power is usually unknown anyway, it makes little difference if we assume an overall power of one. If not, we can always scale the estimated PSD by an estimate of the overall power, i.e., the variance.

The content of the paper is as follows. In Section 2 we introduce the Poisson spectral representation with its properties summarized in Section 3. Section 4 indicates how our use of a similar representation is actually a special case. The derivation of the Poisson likelihood function and its realizable approximation is discussed in Section 5, while Sections 6 and 7 give examples of its usefullness. A key relationship with the Kullback-Leibler distance is described in Section 7. Finally, Section 8 summarizes the results and gives extensions.

2 The Poisson Spectral Representation

The background for this section can be found in [12]. For a discrete-time wide sense stationary (WSS) random process it is well known that the process has the *spectral representation* [1]

$$X[n] = \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp(j2\pi f n) X(df) \tag{1}$$

where $X(df)$ is the complex *spectral measure* and can be viewed as a complex random variable that gives the amplitude and phase of the sinusoid whose frequency is at f. The random variable has a zero mean and is uncorrelated for different frequency sinusoids. In our Poisson representation it will be necessary to assume that not only are the sinusoidal complex amplitudes uncorrelated but that they are also independent. Of course if the random process were Gaussian, the representation of (1) would also imply independence. The power at frequency f is given by $E[|X(df)|^2]$ and hence the PSD is just $P_X(f) = E[|X(df)|^2]/df$. In the proposed Poisson Spectral Representation (PSR) we simplify the discussion by considering only a *real* random process and thus, use the real representation

$$X[n] = \frac{1}{\sqrt{\lambda_0/2}} \sum_{k=1}^{N} A_k \cos(2\pi F_k n + \Phi_k) \qquad -\infty < n < \infty \tag{2}$$

where $\{A_1, A_2, \ldots, A_N\}$ are independent and identically distributed (IID) positive amplitude random variables, $\{\Phi_1, \Phi_2, \ldots, \Phi_N\}$ are IID phase random variables uniformly distributed on $[0, 2\pi)$, and with the amplitudes independent of the phases. The number of sinusoids N is a Poisson random variable with mean λ_0, and the frequencies $\{F_1, F_2, \ldots, F_N\}$ are the *point events in frequency* of a *nonhomogeneous Poisson random process* on the interval $0 \le f \le 1/2$. The Poisson random process, denoted by Π, is independent of the amplitudes and phases. In contrast to the usual spectral representation of (1) where the frequencies are fixed, and usually chosen in the limit as *uniformly spaced* in frequency, in the PSR the frequencies are randomly distributed throughout the frequency interval as a nonhomogeneous Poisson process. This representation then gives rise to a given PSD via a nonuniform distribution of sinusoidal frequency components.

The representation of (2) in which we sum a function evaluated at the points of a Poisson process and for which the function also depends upon the outcomes of other random variables is called a *marked Poisson process*. It can alternatively be written as

$$X[n] = \sum_{k=1}^{N} g_n(F_k, (A_k, \Phi_k)) \tag{3}$$

where

$$g_n(F, (A, \Phi)) = \frac{1}{\sqrt{\lambda_0/2}} A \cos(2\pi F n + \Phi).$$

In this case, the random variable associated with the kth frequency event of the Poisson process is (A_k, Φ_k) and is called the "mark". It is independent of the marks at the other frequencies and also of the other frequency events. As such, it produces a multidimensional Poisson process denoted by Π^* with events as depicted in Figure 1.

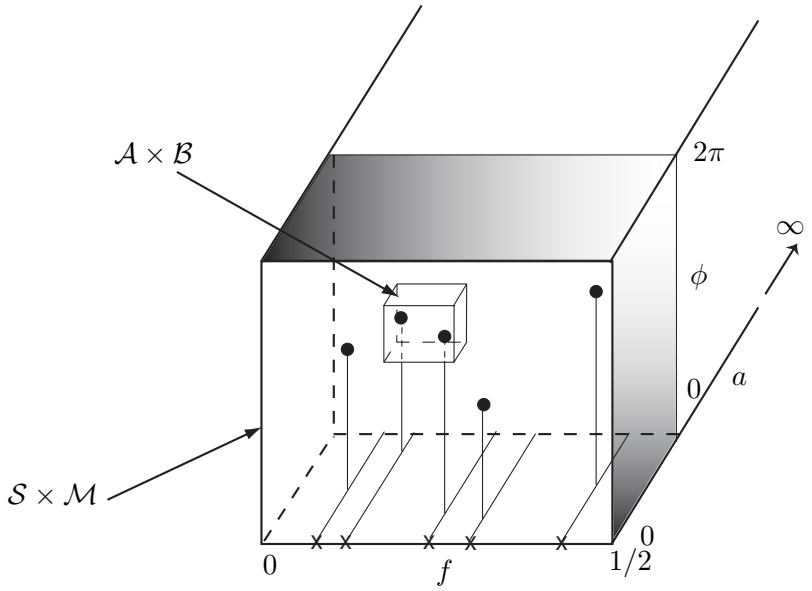

Figure 1: Illustration of an outcome of a marked Poisson process Π^* (shown as solid dots) with event F (shown as x's) and mark (A, Φ). Note that the number of events in $\mathcal{A} \times \mathcal{B}$ is two for the pictured realization and the average number of events is the mean measure $\mu(\mathcal{A} \times \mathcal{B})$.

The space in which the marked events occur will be denoted by $\mathcal{S} \times \mathcal{M}$. Here $\mathcal{S}$ denotes the frequency interval $0 \leq f \leq 1/2$, where the frequency resides and $\mathcal{M}$ denotes the marked space $[0, \infty) \times [0, 2\pi)$, where the amplitude and phase reside. The number of events in a subset V of this space is given by $N(V)$, which is a Poisson random variable with mean *measure* $E[N(V)] = \mu(V)$. Also, the number of events in disjoint subsets are independent of each order, according to the Poisson assumption. Since a projection of a marked Poisson process is also a Poisson process, the

frequencies form a nonhomogenous Poisson process with mean measure denoted by

$$\mu([f_1, f_2] \times \mathcal{M}) = E[N([f_1, f_2] \times \mathcal{M})] = \int_{f_1}^{f_2} \lambda(f) df$$

where we have assumed that all the measures are absolutely continuous. We interpret this assumption as saying that the Poisson process in frequency is nonhomogeneous and has an *intensity* of $\lambda(f)$. Thus the frequency realization is such that there are more events in frequency intervals where $\lambda(f)$ is large, i.e, where the "arrival rate" or intensity is large. Overall, it can be shown that the mean measure for the Poisson process Π^* for an arbitrary volume $\mathcal{A} \times \mathcal{B}$, where $\mathcal{A} \in \mathcal{S}$ and $\mathcal{B} \in \mathcal{M}$ is

$$\mu(\mathcal{A} \times \mathcal{B}) = \int_{\mathcal{A} \times B} \lambda(f) p_A(a) p_\Phi(\phi) d\phi \, da \, df \tag{4}$$

where $p_A(a)$ and $p_\Phi(\phi)$ are the PDFs of the amplitude and phase, respectively. We normalize the intensity by letting

$$\lambda(f) = \lambda_0 p(f)$$

where $\int_0^{1/2} p(f) df = 1$ so that $p(f)$ can be interpreted as a PDF in frequency. With this normalization the expected number of events events in $\mathcal{S} \times \mathcal{M}$ is

$$\mu(\mathcal{S} \times \mathcal{M}) = \int_0^{1/2} \int_0^{\infty} \int_0^{2\pi} p_\Phi(\phi) p_A(a) \lambda_0 p(f) d\phi \, da \, df = \lambda_0.$$

With these assumptions the PSR can be written as

$$X[n] = \int_{\mathcal{S} \times \mathcal{M}} \frac{a}{\sqrt{\lambda_0/2}} \cos(2\pi f n + \phi) N(df \times (da, d\phi)) \tag{5}$$

where $N(\mathcal{A} \times \mathcal{B})$ is the number of events occuring in the "rectangle" $\mathcal{A} \times \mathcal{B}$. This is similar to (1) except for the mark (A, Φ) and the property that the random counting measure, i.e., random variable N, is Poisson and is independent, and not just uncorrelated, in nonoverlapping sets in $\mathcal{S} \times \mathcal{M}$. We can alternatively think of the random counting measure as

$$N(df \times (da, d\phi)) = \sum_{k=1}^{N(\mathcal{S} \times \mathcal{M})} \delta(f - f_k, a - a_k, \phi - \phi_k) df \, da \, d\phi$$

where δ is a three-dimensional Dirac delta function. Inserting this into (5) will produce (2). Note that if $N(\mathcal{S} \times \mathcal{M}) = 0$ then we define $X[n]$ as zero, although this will be a low probability occurrence, especially for large λ_0. The assumption of large λ_0 is desirable in that the $X[n]$ process can be shown to be ergodic in the autocorrelation sequence only as $\lambda_0 \to \infty$, and is necessary for a practical representation. We next state the properties of (2) or equivalently (5) with the derivations to be found in the Appendices.

3 Properties of the Representation

The $X[n]$ random process as defined by (2) or equivalently by (5) with its accompanying assumptions can be shown to possess the following properties:

1. The process is zero mean, i.e., $E[X[n]] = 0$ for $-\infty < n < \infty$. This is due to the assumption that the phase is uniformly distributed.

2. The process possesses an autocorrelation sequence, which together with the first property, shows that it is WSS.

3. The PSD is given by

$$P_X(f) = \frac{E[A^2]}{2} p(|f|) \qquad -1/2 \leq f \leq 1/2. \tag{6}$$

Thus, the PSD is specified by choosing the intensity of the nonhomogeneous Poisson process in frequency since $\lambda(f) = \lambda_0 p(f)$ on the interval $0 \leq f \leq 1/2$. The total power is seen to be $E[X^2[n]] = \int_{-\frac{1}{2}}^{\frac{1}{2}} P_X(f)df = E[A^2]$, and is independent of λ_0 (the reason for the chosen normalization in (2) of $\sqrt{\lambda_0/2}$).

4. The process is ergodic in the mean. This is due to the assumption of an absolutely continuous spectral measure since this implies an absence of a delta function in the PSD at $f = 0$ [1].

5. The process is ergodic in the autocorrelation sequence as $\lambda_0 \to \infty$. This is reasonable in that as the intensity of the Poisson process in frequency increases, more frequency events occur and hence a more accurate temporal modeling of the process is obtained.

6. As $\lambda_0 \to \infty$, the first-order PDF of $X[n]$ becomes Gaussian with zero mean and variance $E[A^2]$, and this holds independently of the PDF of A (as long as the PDF does not depend on the outcome of N, as we have assumed. See [11] for the utility of allowing this dependence). Furthermore, as $\lambda_0 \to \infty$, the process becomes a Gaussian random process in a similar fashion to the well known convergence property of a filtered Poisson process as the intensity becomes large [16].

4 Rapprochement with Previous Results

The previous representation used in [11] was

$$X[n] = \frac{1}{\sqrt{M/2}} \sum_{i=1}^{M} A_i \cos(2\pi F_i n + \Phi_i)$$

where the random variables (A_i, Φ_i) were IID with the amplitudes and phases independent of each other, and the phases were uniformly distributed. It was assumed that M is a given *constant*.

These are nearly the same assumptions as for (2). The difference lies with the assumption on the frequency random variables. We have assumed these were IID with PDF $p_F(f)$ on the interval $[0, 1/2]$ and independent of the amplitude and phase random variables. In the PSR we consider the frequencies as random events distributed according to a nonhomogeneous Poisson random process with intensity $\lambda(f) = \lambda_0 p(f)$. However, if in the PSR model we *fix* the number of events N as the constant M, or equivalently condition on the number of frequency events in $[0, 1/2]$, then the PSR reduces to our previous model. This is a well known result that a nonhomogeneous Poisson process with intensity $\lambda(f)$, conditioned on the number of events, has the same distribution as the *order statistics* of M IID random variables with the PDF [18]

$$p_F(f) = \frac{\lambda(f)}{\int_0^{1/2} \lambda(f) df} \qquad 0 \le f \le 1/2.$$

But for the PSR $\lambda(f) = \lambda_0 p(f)$ so that

$$p_F(f) = \frac{\lambda_0 p(f)}{\int_0^{1/2} \lambda_0 p(f) df} = p(f)$$

since $\int_0^{1/2} p(f) df = 1$. Hence, if we condition on the number of frequency events in $[0, 1/2]$, then $X[n]$ has the properties listed in [11].

5 The Realizable Poisson Likelihood Function

Since the PSD is related to the intensity in the PSR as

$$\lambda(f) = \lambda_0 p(f) = \frac{2\lambda_0 P_X(f)}{E[A^2]} \qquad 0 \le f \le 1/2 \tag{7}$$

we can use the intensity for spectral inference. To simplify the discussion we let $E[A^2] = 1$, since as already stated we are only interested in the normalized PSD. As a result, from (6) we have that $\int_{-\frac{1}{2}}^{\frac{1}{2}} P_X(f) df = 1$. In summary, we have that $\lambda(f) = 2\lambda_0 P_X(f)$ for $0 \le f \le 1/2$, where $\int_0^{1/2} \lambda(f) df = \lambda_0$. Since the marks of the Poisson process are independent of the frequency events and since $P_X(f)$ only depends on the intensity, we can base any decision on just the intensity realization, i.e., the observed number of frequency events per unit cycles/sample. It can be shown that the part of the log-likelihood that depends on the intensity is given by [18]

$$l = -\int_0^{1/2} \lambda(f) df + \int_0^{1/2} \ln \lambda(f) N(df) \tag{8}$$

Next we assume that the frequency events are not observable but only $x[n]$ is observed. Some applications for which the frequency events can be observed are in neural auditory coding [17]. For the unobservable frequency case we proceed by noting that

$$E[N(df)] = \lambda(f) df = 2\lambda_0 P_X(f) df \approx 2\lambda_0 \bar{I}(f) df$$

where $\bar{I}(f)$ is the *normalized* periodogram, which is given by

$$\bar{I}(f) = \frac{\frac{1}{M}\left|\sum_{m=0}^{M-1} x[m]\exp(-j2\pi fm)\right|^2}{\int_{-\frac{1}{2}}^{\frac{1}{2}} I(f)df}$$

and $I(f)$ is the unnormalized periodogram

$$I(f) = \frac{1}{M}\left|\sum_{m=0}^{M-1} x[m]\exp(-j2\pi fm)\right|^2.$$

Note that $\int_{-\frac{1}{2}}^{\frac{1}{2}} \bar{I}(f)df = 1$, in accordance with $\int_{-\frac{1}{2}}^{\frac{1}{2}} P_X(f)df = 1$ and $\bar{I}(f)$ being an estimate of $P_X(f)$. The data set $x[m]$ for $m = 0, 1, \ldots, M-1$ is assumed to have been observed. We now have

$$\begin{aligned} l' &\approx -\int_0^{1/2} \lambda(f)df + \int_0^{1/2} \ln(\lambda(f))2\lambda_0\bar{I}(f)df \\ &= -\lambda_0 + 2\lambda_0\int_0^{1/2} \bar{I}(f)\ln(2\lambda_0 P_X(f))df \\ &= -\lambda_0 + 2\lambda_0\int_0^{1/2} \bar{I}(f)\ln(2\lambda_0)df + 2\lambda_0\int_0^{1/2} \bar{I}(f)\ln P_X(f)df \\ &= -\lambda_0 + \lambda_0\ln(2\lambda_0) + 2\lambda_0\int_0^{1/2} \bar{I}(f)\ln P_X(f)df. \end{aligned}$$

Finally, ignoring the terms that do not depend on the PSD and the arbitrary scaling via λ_0, we have

$$l_R = \int_{-\frac{1}{2}}^{\frac{1}{2}} \bar{I}(f)\ln P_X(f)df \tag{9}$$

as our RPLF. We have now included the negative frequency components of the PSD for convenience.

References

[1] Brockwell, P.J., Davis, R.A., *Time Series: Theory and Methods*, pg. 334, Springer-Verlag, New York, 1987.

[2] Bloomfield, P, "An Exponential Model for the Spectrum of a Scalar Time Series", *Biometrika*, 1973.

[3] Burg, J., "Maximum Entropy Spectral Analysis", PhD dissertation, Stanford Univ., 1975.

[4] Carbone, C., S. Kay, "A Novel Normalization Algorithm Based on the Three-Dimensional Minimum Variance Spectral Estimator", *IEEE Trans. on Aerospace and Electronics*, Jan., 2012.

[5] Doob, J.L., *Stochastic Processes*, J. Wiley, New York, 1953.

[6] Grigoriu, M. , "A Spectral Representation Based Model for Monte Carlo Simulation", *Probabilistic Engineering Mechanics*, pp. 365–370, 2000.

[7] Kay, S., *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[8] Kay, S., *Fundamentals of Statistical Signal Processing: Estimation*, Prentice-Hall, Englewood Cliffs, NJ, 1993.

[9] Kay, S., *Fundamentals of Statistical Signal Processing: Detection*, Prentice-Hall, Englewood Cliffs, NJ, 1998.

[10] Kay, S., "A New Approach to Fourier Synthesis with Application to Neural Encoding and Speech Classification", *IEEE Signal Processing Letters*, Julyu 2010.

[11] Kay, S., "Representation and Generation of Non-Gaussian Wide-Sense Stationary Random Processes with Arbitrary PSDs and a Class of PDFs", *IEEE Trans. on Signal Processing*, pp. 3448 - 3458, July, 2010.

[12] Kingman, J.F.C., *Poisson Processes*, Clarendon Press, Oxford, England, 1993.

[13] Kullback, S., *Information Theory and Statistics*, Dover Pub., New York, 1968.

[14] Mardia, K.V., P. Jupp, *Directional Statistics* J. Wiley, NY, 1999.

[15] Oppenheim, A. V., R.W. Schafer, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1975.

[16] Parzen, E., *Stochastic Processes*, Holden-Day, San Francisco, 1962.

[17] Sachs, M.B., E.D. Young, "Effect of Nonlinearities on Speech Encoding in the Auditory Nerve",*Journal Acoustical Society of America*, pp. 858–875, Sept. 1980.

[18] Snyder, D.L., *Random Point Processes*, J. Wiley, New York, 1975.

A Derivation of Mean and Covariance

Using the standard notation of Kingman [12], we let the Poisson process be denoted by Π and the marked Poisson process by Π^*. The event is denoted by the vector in R^3 as $\mathbf{x}$. Then we wish to determine the mean and covariance of

$$Z_m = \sum_{\mathbf{X} \in \Pi^*} g_m(\mathbf{X}) \tag{10}$$

It is shown in Appendix C that the first four moments are given as

$$E[Z_m] = \int_{\mathcal{S} \times \mathcal{M}} g_m(\mathbf{x}) \mu(d\mathbf{x}) \tag{11}$$

$$E[Z_m Z_n] = \int_{\mathcal{S} \times \mathcal{M}} g_m(\mathbf{x}) g_n(\mathbf{x}) \mu(d\mathbf{x}) \tag{12}$$

By letting $\mathbf{x} = [f\, a\, \phi]^T$ and

$$g_m(\mathbf{x}) = \frac{a}{\sqrt{\lambda_0/2}} \cos(2\pi f m + \phi)$$

and also

$$\mu(d\mathbf{x}) = \lambda_0 p(f) p_A(a) p_\Phi(\phi) d\phi\, da\, df$$

we can find the moments of $X[n]$.

The mean is derived first by using (11).

$$E[X[n]] = \int_0^{1/2} \int_0^\infty \int_0^{2\pi} \frac{a}{\sqrt{\lambda_0/2}} \cos(2\pi f n + \phi) \lambda_0 p(f) p_A(a) p_\Phi(\phi) d\phi\, da\, df = 0$$

due to the integration over ϕ.

Next the autocorrelation sequence and the PSD are found.

$$\begin{aligned} E[X[m]X[n]] &= \int_0^{1/2} \int_0^\infty \int_0^{2\pi} \frac{a^2}{\lambda_0/2} \cos(2\pi f m + \phi) \cos(2\pi f n + \phi) \lambda_0 p(f) p_A(a) p_\Phi(\phi) d\phi\, da\, df \\ &= 2E[A^2] \int_0^{1/2} \int_0^{2\pi} \frac{1}{2} \cos[2\pi f(m-n)] + \frac{1}{2} \cos[2\pi f(m+n) + 2\phi] \frac{1}{2\pi} d\phi\, p(f) df \\ &= 2E[A^2] \int_0^{1/2} \frac{1}{2} \cos[2\pi f(m-n)] p(f) df \\ &= r_X[m-n] \end{aligned}$$

so that the autocorrelation sequence is

$$r_X[k] = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{E[A^2] p(|f|)}{2} \cos(2\pi f k) df$$

and the PSD is seen to be

$$P_X(f) = \frac{E[A^2] p(|f|)}{2} \qquad -1/2 \le f \le 1/2.$$

B Derivation of Ergodicity of Sample Autocorrelation

From [5] there are two conditions that are necessary and sufficient for ergodicity. They are

1. The fourth moment $E[X[n_0]X[k+n_0]X[j+n_0]X[j+k+n_0]]$ should not depend on n_0, which is a form of stationarity for this moment.

2. If the sample autocorrelation is given by

$$\hat{r}_X[k] = \frac{1}{M+1}\sum_{j=0}^{M+1} X[j]X[j+k]$$

then we require

$$\lim_{M\to\infty}\frac{1}{M+1}\sum_{j=0}^{M+1} E[X[0]]E[X[k]]E[X[j]]E[X[j+k]] = r_X^2[k]$$

and is equivalent to requiring the variance of $\hat{r}_X[k]$ to go to zero as the data record length M goes to ∞. We first verify the stationarity of the fourth-order moment.

To do so we let

$$g_n(\mathbf{x}) = X[n] = \frac{a}{\sqrt{\lambda_0/2}}\cos(2\pi f n + \phi)$$

and use the fourth-order moment results derived in Appendix C. We have

$$\begin{aligned}
E[X[n_0]X[k+n_0]X[j+n_0]X[j+k+n_0]] &= \int_{\mathcal{S}\times\mathcal{M}} g_{n_0}(\mathbf{x})g_{k+n_0}(\mathbf{x})g_{j+n_0}(\mathbf{x})g_{j+k+n_0}(\mathbf{x})\mu(d\mathbf{x}) \\
&+ \int_{\mathcal{S}\times\mathcal{M}} g_{n_0}(\mathbf{x})g_{k+n_0}(\mathbf{x})\mu(d\mathbf{x})\int_{\mathcal{S}\times\mathcal{M}} g_{j+n_0}(\mathbf{x})g_{j+k+n_0}(\mathbf{x})\mu(d\mathbf{x}) \\
&+ \int_{\mathcal{S}\times\mathcal{M}} g_{n_0}(\mathbf{x})g_{j+n_0}(\mathbf{x})\mu(d\mathbf{x})\int_{\mathcal{S}\times\mathcal{M}} g_{k+n_0}(\mathbf{x})g_{j+k+n_0}(\mathbf{x})\mu(d\mathbf{x}) \\
&+ \int_{\mathcal{S}\times\mathcal{M}} g_{n_0}(\mathbf{x})g_{j+k+n_0}(\mathbf{x})\mu(d\mathbf{x})\int_{\mathcal{S}\times\mathcal{M}} g_{k+n_0}(\mathbf{x})g_{j+n_0}(\mathbf{x})\mu(d\mathbf{x})
\end{aligned}$$

and from Appendix A

$$\int_{\mathcal{S}\times\mathcal{M}} g_m(\mathbf{x})g_n(\mathbf{x})\mu(\mathbf{x}) = r_X[m-n].$$

Thus, the last three terms are

$$r_X^2[k] + r_X^2[j] + r_X[j+k]r_X[j-k]$$

and clearly do not depend on n_0. Considering the first term, which we denote by I, we have

$$\begin{aligned}
I &= \int_0^{1/2}\int_0^{\infty}\int_0^{2\pi} \frac{1}{(\lambda_0/2)^2}a^4\cos[2\pi f n_0 + \phi]\cos[2\pi f(k+n_0)+\phi] \\
&\cdot\ \cos[2\pi f(j+n_0)+\phi]\cos[2\pi f(j+k+n_0)+\phi]\lambda_0 p_\Phi(\phi)p_A(a)p(f)d\phi\,da\,df \\
&= \frac{4E[A^4]}{\lambda_0}\int_0^{1/2}\int_0^{2\pi}[\cos[2\pi f n_0+\phi]\cos[2\pi f(k+n_0)+\phi] \\
&\cdot\ \cos[2\pi f(j+n_0)+\phi]\cos[2\pi f(j+k+n_0)+\phi]]\,\frac{1}{2\pi}p(f)d\phi\,df
\end{aligned}$$

To evaluate the integral over ϕ we let $z_i = \exp(j\theta_i)$, $i = 1, 2, 3, 4$ with

$$\begin{aligned}
\theta_1 &= 2\pi f n_0 + \phi \\
\theta_2 &= 2\pi f(k + n_0) + \phi \\
\theta_3 &= 2\pi f(j + n_0) + \phi \\
\theta_4 &= 2\pi f(j + k + n_0) + \phi
\end{aligned}$$

so that the fourth-order product of cosines in brackets becomes

$$\frac{1}{16}\prod_{i=1}^{4}(z_i + z_i^*).$$

When multiplied out, only the product terms that have two unconjugated z_i's and two conjugated z_i's so that the term does not depend on ϕ will produce a nonzero contribution to the integral. It can be shown that this results in the terms

$$\frac{1}{8}\left[\cos(\theta_1 + \theta_2 - \theta_3 - \theta_4) + \cos(\theta_1 - \theta_2 + \theta_3 - \theta_4) + \cos(\theta_1 - \theta_2 - \theta_3 + \theta_4)\right]$$

which is

$$\frac{1}{8}\left[\cos(-2\pi fj - 2\pi fj) + \cos(-2\pi fk - 2\pi fk) + \cos(-2\pi fk + 2\pi fk)\right] = \frac{1}{8}\left[\cos(4\pi fj) + \cos(4\pi fk) + 1\right].$$

At this point we see that the fourth-order moment does not depend on n_0 and hence the first condition for ergodicity is satisfied. Continuing on we compute the fourth-order moment $E[X[0]X[k]X[j][j+k]]$, which is just the previous expression with $n_0 = 0$.

We now continue the evaluation of I.

$$\begin{aligned}
I &= \frac{4E[A^4]}{\lambda_0}\int_0^{1/2}\int_0^{2\pi}\left[\cos[2\pi f n_0 + \phi]\cos[2\pi f(k + n_0) + \phi]\right. \\
&\cdot \left.\cos[2\pi f(j + n_0) + \phi]\cos[2\pi f(j + k + n_0) + \phi]\right]\frac{1}{2\pi}p(f)d\phi df \\
&= \frac{E[A^4]}{2\lambda_0}\int_0^{1/2}\int_0^{2\pi}\left[\cos(4\pi fj) + \cos(4\pi fk) + 1\right]\frac{1}{2\pi}p(f)d\phi df \\
&= \frac{E[A^4]}{2\lambda_0}\int_0^{1/2}\left[\cos(4\pi fj) + \cos(4\pi fk) + 1\right]p(f)df \\
&= \frac{E[A^4]}{2\lambda_0 E[A^2]}\int_{-\frac{1}{2}}^{\frac{1}{2}}\left[\cos(2\pi fj) + \cos(4\pi fk) + 1\right]\frac{E[A^2]p(|f|)}{2}df \\
&= \frac{E[A^4]}{2\lambda_0 E[A^2]}[r_X[2j] + r_X[2k] + r_X[0]]
\end{aligned}$$

so that

$$E[X[0]X[k]X[j]X[j+k]] = \frac{E[A^4]}{2\lambda_0 E[A^2]}[r_X[2j] + r_X[2k] + r_X[0]] + r_X^2[k] + r_X^2[j] + r_X[j+k]r_X[j-k].$$

Thus,

$$\begin{aligned}\lim_{M\to\infty}\frac{1}{M+1}\sum_{j=0}^{M}E[X[0]X[k]X[j]X[j+k]] &= \lim_{M\to\infty}\frac{1}{M+1}\sum_{j=0}^{M}\frac{E[A^4]}{2\lambda_0 E[A^2]}(r_X[2j]+r_X[2k]+r_X[0]) \\ &+ \lim_{M\to\infty}\frac{1}{M+1}\sum_{j=0}^{M}(r_X^2[j]+r_X[j+k]r_X[j-k])+r_X^2[k]\end{aligned}$$

and assuming that $\sum_{j=0}^{M}|r_X[j]|<\infty$ and $\sum_{j=0}^{M}|r_X^2[j]+r_X[j+k]r_X[j-k]|<\infty$, which will be true for an absolutely continuous spectral measure, we see that

$$\lim_{M\to\infty}\frac{1}{M+1}\sum_{j=0}^{M}E[X[0]X[k]X[j]X[j+k]]=\frac{E[A^4]}{2\lambda_0 E[A^2]}(r_X[2k]+r_X[0])+r_X^2[k]$$

which will only approach $r_X^2[k]$ as $\lambda_0\to\infty$.

C Derivation of Joint Characteristic Function

The principal approach to determining properties of a Poisson process is the characteristic function and Campbell's theorem [12]. The general fourth-order moments necessary do not appear in the literature and so this appendix fills that gap. In the process we will also derive the lower-order moments, some of which are in [12], as well as many other references. We use a general procedure to allow the application to any Poisson process.

The joint characteristic function of $\mathbf{Z}=[Z_1\,Z_2\ldots Z_p]^T$ as given by (10) can be shown to be

$$\psi_{\mathbf{z}}(\boldsymbol{\omega})=E[\exp(j\boldsymbol{\omega}^T\mathbf{Z})]=\exp\left[\int_{\mathcal{S}\times\mathcal{M}}\left(\exp[j\boldsymbol{\omega}^T\mathbf{g}(\mathbf{x})]-1\right)\mu(d\mathbf{x})\right]$$

where $\boldsymbol{\omega}=[\omega_1\,\omega_2\ldots\omega_p]^T$, $\mathbf{g}(\mathbf{x})=[g_1(\mathbf{x})\,g_2(\mathbf{x})\ldots g_p(\mathbf{x})]^T$, and $\mu(\mathcal{A})$ is the mean measure of the set $\mathcal{A}$. It is asumed that the integral exists, which is assured if $\mu(\mathcal{S}\times\mathcal{M})<\infty$. It can be shown by Campbell's theorem that

$$E[g_i(\mathbf{X})]=\int_{\mathcal{S}\times\mathcal{M}}g_i(\mathbf{x})\mu(d\mathbf{x})$$

and assuming this equals zero, we have that

$$\begin{aligned}\psi_{\mathbf{z}}(\boldsymbol{\omega}) &= \exp\left[\int_{\mathcal{S}\times\mathcal{M}}\left(\exp[j\boldsymbol{\omega}^T\mathbf{g}(\mathbf{x})]-j\boldsymbol{\omega}^T\mathbf{g}(\mathbf{x})-1\right)\mu(d\mathbf{x})\right] \\ &= \exp\left[\int_{\mathcal{S}\times\mathcal{M}}\sum_{k=2}^{\infty}\frac{(j\boldsymbol{\omega}^T\mathbf{g}(\mathbf{x}))^k}{k!}\mu(d\mathbf{x})\right] \\ &= \exp\left[\sum_{k=2}^{\infty}\int_{\mathcal{S}\times\mathcal{M}}\frac{(j\boldsymbol{\omega}^T\mathbf{g}(\mathbf{x}))^k}{k!}\mu(d\mathbf{x})\right]\end{aligned}$$

with the last step justified via the Beppo-Levi theorem and the assumption that

$$\sum_{k=2}^{\infty}\int_{\mathcal{S}\times\mathcal{M}}\left|\frac{(j\boldsymbol{\omega}^T\mathbf{g}(\mathbf{x}))^k}{k!}\right|\mu(d\mathbf{x})<\infty.$$

Next to differentiate the characteristic function it is convenient to let

$$G(\boldsymbol{\omega}, \nu, h) = \sum_{k=\nu}^{\infty} \int_{\mathcal{S}\times\mathcal{M}} \frac{\left(\sum_{i=1}^{p} j\omega_i g_i(\mathbf{x})\right)^k}{k!} h(\mathbf{x})\mu(d\mathbf{x})$$

where $\nu \geq 0$ so that we have

$$\psi_{\mathbf{z}}(\boldsymbol{\omega}) = \exp(G(\boldsymbol{\omega}, 2, e))$$

and $e(\mathbf{x}) = 1$. Note that

$$\frac{\partial G(\boldsymbol{\omega}, \nu, e)}{\partial \omega_m} = \begin{cases} G(\boldsymbol{\omega}, \nu - 1, jg_m) & \nu \geq 1 \\ G(\boldsymbol{\omega}, 0, jg_m) & \nu < 1 \end{cases}$$

so that if the second argument of G is less than zero, it should be set to zero. Similarly, we have

$$\begin{aligned} \frac{\partial^2 G}{\partial\omega_m \partial\omega_n} &= G(\boldsymbol{\omega}, \nu - 2, j^2 g_m g_n) \\ \frac{\partial^3 G}{\partial\omega_m \partial\omega_n \partial\omega_r} &= G(\boldsymbol{\omega}, \nu - 3, j^3 g_m g_n g_r) \\ = \frac{\partial^4 G}{\partial\omega_m \partial\omega_n \partial\omega_r \partial\omega_s} &= G(\boldsymbol{\omega}, \nu - 4, j^4 g_m g_n g_r g_s). \end{aligned}$$

Also we make use of the relationship

$$G(\boldsymbol{\omega}, \nu, h)|_{\boldsymbol{\omega}=\mathbf{0}} = \begin{cases} \int_{\mathcal{S}\times\mathcal{M}} h(\mathbf{x})\mu(d\mathbf{x}) & \nu = 0 \\ 0 & \nu \geq 1 \end{cases}$$

As a result we obtain the moments as follows. They are

$$\begin{aligned} E[Z_m] &= \frac{1}{j} \left.\frac{\partial \psi_{\mathbf{z}}}{\partial \omega_m}\right|_{\boldsymbol{\omega}=\mathbf{0}} \\ &= \frac{1}{j} \psi_{\mathbf{z}}(\mathbf{0}) G(\mathbf{0}, 1, jg_m) = 0. \end{aligned}$$

$$E[Z_m Z_n] = \frac{1}{j^2} \left.\frac{\partial^2 \psi_{\mathbf{z}}}{\partial\omega_m \partial\omega_n}\right|_{\boldsymbol{\omega}=\mathbf{0}}$$

and

$$\begin{aligned} \frac{\partial^2 \psi_{\mathbf{z}}}{\partial\omega_m \partial\omega_n} &= \frac{\partial}{\partial\omega_m} \left[\psi_{\mathbf{z}}(\boldsymbol{\omega}) G(\boldsymbol{\omega}, 1, jg_m)\right] \\ &= \psi_{\mathbf{z}}(\boldsymbol{\omega}) G(\boldsymbol{\omega}, 0, j^2 g_m g_n) + \psi_{\mathbf{z}}(\boldsymbol{\omega}) G(\boldsymbol{\omega}, 1, jg_n) G(\boldsymbol{\omega}, 1, jg_m). \end{aligned}$$

Evaluating this at $\boldsymbol{\omega} = \mathbf{0}$ produces $G(\mathbf{0}, 0, j^2 g_m g_n)$ or finally

$$E[Z_m Z_n] = \int_{\mathcal{S}\times\mathcal{M}} g_m(\mathbf{x}) g_n(\mathbf{x}) \mu(d\mathbf{x}).$$

The first and second moment are just Campbell's theorem. Next

$$E[Z_m Z_n Z_r] = \frac{1}{j^3} \left.\frac{\partial^3 \psi_{\mathbf{z}}}{\partial\omega_m \partial\omega_n \partial\omega_r}\right|_{\boldsymbol{\omega}=\mathbf{0}}$$

and

$$\begin{aligned}\frac{\partial^3\psi_{\mathbf{z}}}{\partial\omega_m\partial\omega_n\partial\omega_r} &= \frac{\partial}{\partial\omega_r}\left[\psi_{\mathbf{z}}(\boldsymbol{\omega})G(\boldsymbol{\omega},0,j^2g_mg_n)+\psi_{\mathbf{z}}(\boldsymbol{\omega})G(\boldsymbol{\omega},1,jg_n)G(\boldsymbol{\omega},1,jg_m)\right]\\ &= \psi_{\mathbf{z}}(\boldsymbol{\omega})G(\boldsymbol{\omega},0,j^3g_mg_ng_r)+\psi_{\mathbf{z}}(\boldsymbol{\omega})G(\boldsymbol{\omega},1,jg_r)G(\boldsymbol{\omega},0,j^2g_mg_n)\\ &+ \psi_{\mathbf{z}}(\boldsymbol{\omega})\left[G(\boldsymbol{\omega},0,j^2g_ng_r)G(\boldsymbol{\omega},1,jg_m)+G(\boldsymbol{\omega},1,jg_n)G(\boldsymbol{\omega},0,j^2g_mg_r)\right]\\ &+ \psi_{\mathbf{z}}(\boldsymbol{\omega})G(\boldsymbol{\omega},1,jg_r)G(\boldsymbol{\omega},1,jg_n)G(\boldsymbol{\omega},1,jg_m).\end{aligned}$$

Finally, we have

$$E[Z_mZ_nZ_r]=\int_{\mathcal{S}\times\mathcal{M}}g_m(\mathbf{x})g_n(\mathbf{x})g_r(\mathbf{x})\mu(d\mathbf{x}).$$

The fourth-order moment is found similarly as

$$E[Z_mZ_nZ_rZ_s]=\frac{1}{j^4}\left.\frac{\partial^4\psi_{\mathbf{z}}}{\partial\omega_m\partial\omega_n\partial\omega_r\partial\omega_s}\right|_{\boldsymbol{\omega}=\mathbf{0}}.$$

Note that only the third derivative terms above that have a factor of $G(\cdot,0,\cdot)$ after being differentiated will be nonzero when $\boldsymbol{\omega}=\mathbf{0}$. This produces the fourth-order derivative evaluated at $\boldsymbol{\omega}=\mathbf{0}$ of

$$\begin{aligned}&\psi_{\mathbf{z}}(\boldsymbol{\omega})G(\boldsymbol{\omega},0,j^4g_mg_ng_rg_s)+\psi_{\mathbf{z}}(\boldsymbol{\omega})G(\boldsymbol{\omega},0,j^2g_rg_s)G(\boldsymbol{\omega},0,j^2g_mg_n)\\ +\;&\psi_{\mathbf{z}}(\boldsymbol{\omega})G(\boldsymbol{\omega},0,j^2g_ng_r)G(\boldsymbol{\omega},0,j^2g_mg_s)+\psi_{\mathbf{z}}(\boldsymbol{\omega})G(\boldsymbol{\omega},0,j^2g_ng_s)G(\boldsymbol{\omega},0,j^2g_mg_r)\end{aligned}$$

and finally we have

$$\begin{aligned}E[Z_mZ_nZ_rZ_s] &= \int_{\mathcal{S}\times\mathcal{M}}g_m(\mathbf{x})g_n(\mathbf{x})g_r(\mathbf{x})g_s(\mathbf{x})\mu(d\mathbf{x})+\int_{\mathcal{S}\times\mathcal{M}}g_r(\mathbf{x})g_s(\mathbf{x})\mu(d\mathbf{x})\int_{\mathcal{S}\times\mathcal{M}}g_m(\mathbf{x})g_n(\mathbf{x})\mu(d\mathbf{x})\\ &+ \int_{\mathcal{S}\times\mathcal{M}}g_n(\mathbf{x})g_r(\mathbf{x})\mu(d\mathbf{x})\int_{\mathcal{S}\times\mathcal{M}}g_m(\mathbf{x})g_s(\mathbf{x})\mu(d\mathbf{x})\\ &+ \int_{\mathcal{S}\times\mathcal{M}}g_n(\mathbf{x})g_s(\mathbf{x})\mu(d\mathbf{x})\int_{\mathcal{S}\times\mathcal{M}}g_m(\mathbf{x})g_r(\mathbf{x})\mu(d\mathbf{x}).\end{aligned}$$

D Derivation of Convergence to Gaussian Random Process

Consider an arbitrary number of samples K at arbitrary times $\{n_1,n_2,\ldots,n_K\}$. The characteristic function of $\mathbf{Z}=[X[n_1]\,X[n_2]\ldots X[n_K]]^T$ was shown in Appendix C to be given by

$$\psi_{\mathbf{z}}(\boldsymbol{\omega})=\exp\left[\sum_{k=2}^{\infty}\int_{\mathcal{S}\times\mathcal{M}}\frac{(j\boldsymbol{\omega}^T\mathbf{g}(\mathbf{x}))^k}{k!}\mu(d\mathbf{x})\right]$$

where $\mathbf{x}=[f\;a\;\phi]^T$ and

$$x[n_i]=g_i(\mathbf{x})=\frac{1}{\sqrt{\lambda_0/2}}a\cos(2\pi fn_i+\phi).$$

Thus, we have

$$\begin{aligned}
\ln \psi_{\mathbf{z}}(\boldsymbol{\omega}) &= \sum_{k=2}^{\infty} \int_0^{1/2} \int_0^{\infty} \int_0^{2\pi} \frac{1}{k!} \left(j \sum_{i=1}^{K} \omega_i \frac{1}{\sqrt{\lambda_0/2}} a \cos(2\pi f n_i + \phi) \right)^k \lambda_0 p(f) p_A(a) p_\Phi(\phi) d\phi \, da \, df \\
&= \sum_{k=2}^{\infty} \int_0^{1/2} \int_0^{\infty} \int_0^{2\pi} \frac{1}{\lambda_0^{k/2-1} k!} \left(j\sqrt{2} \sum_{i=1}^{K} \omega_i a \cos(2\pi f n_i + \phi) \right)^k p(f) p_A(a) p_\Phi(\phi) d\phi \, da \, df \\
&= \int_0^{1/2} \int_0^{\infty} \int_0^{2\pi} \frac{1}{2!} \left(j\sqrt{2} \sum_{i=1}^{K} \omega_i a \cos(2\pi f n_i + \phi) \right)^2 p(f) p_A(a) p_\Phi(\phi) d\phi \, da \, df + O(1/\sqrt{\lambda_0})
\end{aligned}$$

and as $\lambda_0 \to \infty$

$$\begin{aligned}
\ln \psi_{\mathbf{z}}(\boldsymbol{\omega}) &\to -E[A^2] \int_0^{1/2} \int_0^{2\pi} \left(\sum_{i=1}^{K} \omega_i \cos(2\pi f n_i + \phi) \right)^2 p(f) p_\Phi(\phi) d\phi \, df \\
&= -\frac{1}{2} \sum_{i=1}^{K} \sum_{j=1}^{K} \omega_i \omega_j [\mathbf{A}]_{ij}
\end{aligned}$$

where

$$\begin{aligned}
[\mathbf{A}]_{ij} &= 2E[A^2] \int_0^{1/2} \int_0^{2\pi} \cos(2\pi f n_i + \phi) \cos(2\pi f n_j + \phi) p(f) p_\Phi(\phi) d\phi \, df \\
&= E[A^2] \int_0^{1/2} \cos(2\pi f (n_i - n_j)) p(f) df.
\end{aligned}$$

Therefore, we have that

$$\begin{aligned}
[\mathbf{A}]_{ij} &= \int_0^{1/2} \cos(2\pi f (n_i - n_j)) p(f) df \\
&= \int_0^{1/2} \cos(2\pi f (n_i - n_j)) E[A^2] p(f) df \\
&= \int_{-1/2}^{1/2} \cos(2\pi f (n_i - n_j)) \left(\frac{E[A^2]}{2} p(|f|) \right) df \\
&= r_X[n_i - n_j].
\end{aligned}$$

Thus,

$$\psi_{\mathbf{z}}(\boldsymbol{\omega}) \to \exp\left(-\frac{1}{2} \boldsymbol{\omega}^T \mathbf{C} \boldsymbol{\omega} \right)$$

where

$$[\mathbf{C}]_{ij} = r_X[n_i - n_j]$$

and $\mathbf{C}$ is recognized as the covariance matrix, from which we can conclude that the random process approaches a Gaussian random process as $\lambda_0 \to \infty$.