Estimating spectrum density

Estimating spectral intensity is done by

$$\hat{I}(\beta) \equiv \frac{\left[\sum_{n=1}^{N} X_n \cdot \sin(\beta\, n)\right]^2 + \left[\sum_{n=1}^{N} X_n \cdot \cos(\beta\, n)\right]^2}{N \cdot \pi}$$

For white noise, we have

$$\begin{aligned} A &\equiv \frac{\sum_{n=1}^{N} X_n \cdot \sin(\beta\, n)}{\sqrt{N \cdot \pi}} \in \mathcal{N}\left(0, \sigma^2 \cdot \frac{\sum_{n=1}^{N} \sin^2(\beta\, n)}{N \cdot \pi} \simeq \frac{\sigma^2}{2\pi}\right) \\ B &\equiv \frac{\sum_{n=1}^{N} X_n \cdot \cos(\beta\, n)}{\sqrt{N \cdot \pi}} \in \mathcal{N}\left(0, \sigma^2 \cdot \frac{\sum_{n=1}^{N} \cos^2(\beta\, n)}{N \cdot \pi} \simeq \frac{\sigma^2}{2\pi}\right) \end{aligned}$$

(unless $\beta = 0$ or π), and their covariance is

$$\sigma^2 \cdot \frac{\sum_{n=1}^{N} \sin(\beta\, n) \cos(\beta\, n)}{N \cdot \pi} \simeq 0$$

This implies that A^2+B^2 has the χ_2^2 distribution, multiplied by $\frac{\sigma^2}{2\pi}$ (exponential, with the mean of $\frac{\sigma^2}{\pi}$). Its standard deviation is thus as large as its expected value, regardless of how large an N we use! (When $\beta = 0$ or π, we get χ_1^2 multiplied by $\frac{\sigma^2}{\pi}$ - the standard deviation is $\sqrt{2}$ times bigger than the expected value). Furthermore, when you take the $\hat{I}$ estimates at two different values of β (say β_1 and β_2), they will be practically uncorrelated (independent), since

$$\begin{aligned} 2\sin(\beta_1\, n)\sin(\beta_2\, n) &= \cos[(\beta_1 + \beta_2)n] - \cos[(\beta_1 - \beta_2)n] \\ 2\cos(\beta_1\, n)\cos(\beta_2\, n) &= \cos[(\beta_1 + \beta_2)n] + \cos[(\beta_1 - \beta_2)n] \\ 2\sin(\beta_1\, n)\cos(\beta_2\, n) &= \sin[(\beta_1 + \beta_2)n] + \sin[(\beta_1 - \beta_2)n] \end{aligned}$$

and all of these add up (to a good approximation) to zero. One can thus see that estimating a spectrum (of even the simplest model) is going to be very difficult.

Furthermore, one can easily show that when

$$X_n \equiv \sum_{j=0}^{\infty} f_j \varepsilon_{n-j}$$

(and all our models can be expressed in this form), then the spectral intensity of the X sequence is

$$I_x(\beta) = f(e^{i\beta}) f(e^{-i\beta}) I_\varepsilon(\beta)$$

where $I_\varepsilon(\beta)$ is the spectral intensity of the ε sequence, and

$$f(x) \equiv \sum_{j=0}^{\infty} f_j x^j$$

The same relationship holds also for the corresponding $\hat{I}_x(\beta)$ and $\hat{I}_\varepsilon(\beta)$ estimators; this implies that we encounter the same problem when estimating *any* spectrum - for each β, the standard deviation of the estimator will be as large as the corresponding expected value (regardless of N), and these huge fluctuations will be practically uncorrelated (for two distinct values of β).

Due to this, the actual empirical spectrum is a collection of literally hundreds of 'spikes' which, at their peak, highly overshoot the correct value of $\omega(\beta)$, and at their bottom are practically equal to zero. Clearly something needs to be done to alleviate this problem.

Windowing

It appears (and is quite logical) that the main contribution to the 'wild' behavior of empirical spectra comes from (highly oscillating) terms with large values of k in

$$\hat{\omega}(\beta) = 1 + 2 \sum_{k=1}^{N-1} \hat{\rho}_k \cos(\beta\, k) \tag{1}$$

At the same time, to estimate a spectrum expected to be reasonably 'smooth', we should not need more than a handful (say M) of the small-k terms.

This leads to the idea of simply discarding the large-k terms of the previous formula, with the objective of eliminating the random 'spikes' without flattening the real peaks (of the actual spectrum). Clearly, these two criteria work against each other, and we have to be rather careful to reach a good (never perfect) compromise: we should start with a relatively large M and keep on decreasing it - as soon as we see the 'real' peaks shrinking, we must stop (even though some residual of the random 'jerkiness' will still be present). To reconcile these two opposite trends (to diminish the real effect, and to introduce random spikes) as much as possible, we have to truncate the summation of the previous formula in a less abrupt manner - this is the task of the so called windowing. It works as follows: instead of a simple truncation, we gradually decrease the contribution of each term (by a factor λ_k), till we reach zero (at $k = M$). There are several 'popular' ways of doing this, namely (from best to worst):

Tukey:

$$\lambda_k = \frac{1}{2}\left(1 + \cos\frac{\pi k}{M}\right) \qquad k \le M$$

Parzen:

$$\lambda_k = \begin{cases} 1 - 6\left(\frac{k}{M}\right)^2 (1 - \frac{k}{M}) & 0 \le k \le \frac{M}{2} \\ 2\left((1 - \frac{k}{M}\right)^3 & \frac{M}{2} \le k \le M \end{cases}$$

Bartlett:

$$\lambda_k = 1 - \frac{k}{M} \qquad k \le M$$

Smoothing

There is also a somehow less successful (but, in a sense, more natural) idea of simply 'smoothing' the spectrum, by first discretizing the β scale (dividing it into exactly $\frac{N}{2}$ subintervals), evaluating (1) at each such β, and then taking as the final estimate of $\omega(\beta_j)$ - let us denote it $\bar{\omega}(\beta_j)$ - the simple average of $2\ell+1$ consecutive values of $\hat{\omega}(\beta_m)$, centered on $\hat{\omega}(\beta_j)$, i.e.

$$\bar{\omega}(\beta_j) = \frac{\sum_{m=j-\ell}^{j+\ell} \hat{\omega}(\beta_m)}{2\ell+1}$$

When close to $\beta = 0$ and $\beta = \pi$, we have to utilize the following symmetry: $\omega(-\beta) = \omega(\beta)$ and $\omega(\pi+\beta) = \omega(\pi-\beta)$.

One can show that this way of smoothing the spectrum is equivalent to using the following window

$$\lambda_k = \frac{\sin \frac{(2\ell+1)\pi k}{N}}{(2\ell+1)\sin \frac{\pi k}{N}} \qquad k \le N$$

This is based on (see the next proof to understand why):

$$\sum_{m=-\ell}^{\ell} \cos[(\beta + \frac{2m\pi}{N})k] = \mathrm{Re} \sum_{m=-\ell}^{\ell} \exp[i(\beta + \frac{2m\pi}{N})k] = \mathrm{Re}\left[\exp(i\beta k) \sum_{m=-\ell}^{\ell} a^m\right]$$

where $a \equiv \exp(i\frac{2\pi k}{N})$. Since

$$\sum_{m=-\ell}^{\ell} a^m = a^{-\ell}(1+a+a^2+a^{2\ell}) = a^{-\ell}\frac{a^{2\ell+1}-1}{a-1} = \frac{a^{\ell+1/2} - a^{-\ell-1/2}}{a^{1/2} - a^{-1/2}} = \frac{\sin \frac{\pi k(2\ell+1)}{N}}{\sin \frac{\pi k}{N}}$$

Hanning is a combination of windowing and smoothing: we first do the regular truncation of the sum in (1) to M terms (this is effectively taking $\lambda_k = 1$ for $k \le M$), and then replacing each $\hat{\omega}(\beta_j)$ by the following weighted average:

$$\bar{\omega}(\beta_j) = \frac{1}{4}\hat{\omega}(\beta_{j-1}) + \frac{1}{2}\hat{\omega}(\beta_j) + \frac{1}{4}\hat{\omega}(\beta_{j+1})$$

where the $[0, \pi]$ interval must be now divided into M subintervals (with β discretized accordingly).

One can show that this (rather elaborate) procedure is equivalent to Tukey's windowing.

`Proof:`

$$1 + 2\sum_{k=1}^{M} \hat{\rho}_k \left[\frac{1}{2}\cos(\beta\, k) + \frac{\cos[(\beta + \frac{\pi}{M})k] + \cos[(\beta - \frac{\pi}{M})k]}{4}\right] =$$

$$1 + 2\sum_{k=1}^{M} \hat{\rho}_k \cos(\beta\, k) \cdot \frac{1 + \cos(\frac{\pi}{M}k)}{2}$$

Hamming is the same as Hanning, except for the weighing scheme: 0.23, 0.54, 0.23 instead of $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{1}{4}$ (someone was clearly doing a lot of tinkering).