Classic Methods for Spectral Estimation

John MacLaren Walsh, Ph.D.

ECES 632, Winter Quarter, 2007

1 Methods for Spectral Estimation

As covered in the statistical DSP problem families lecture, in spectral estimation, we have a block of contiguous samples

$$X[\mathtt{n}], \quad \mathtt{n} \in \{0, 1, \ldots, \mathtt{N}-1\}$$

of a wide sense stationary random process $X[\mathtt{n}]$, $\mathtt{n} \in \mathbb{Z}$ (of doubly infinite support in time). The goal is to estimate the power spectral density of the WSSRP which is the DTFT $\mathsf{S}_X(f)$ of the auto-correlation function

$$\mathsf{R}_X[\mathtt{n}] := \mathbb{E}[X[0]X^*[\mathtt{n}]]$$

so that

$$\mathsf{S}_X(f) := \sum_{\mathtt{n}=-\infty}^{\infty} \mathsf{R}_X[\mathtt{n}] \exp\left(-j2\pi f \mathtt{n}\right)$$

We will consider four "conventional" methods for spectral estimation: (1) the periodogram, which is a scaled magnitude squared of a DTFT of the available data, (2) its smooth version, (3) the correlogram, which computes the DTFT of an estimated auto-correlation, and (4) its smooth version. **Caveat Emptor**: a quote from the Marple text regarding this (somewhat dated) material *Other optimal methods of smoothing the periodogram have appeared in the literature, but the methods are usually optimal only with respect to some restrictive class of signals. The three methods presented in this section may not always be optimum, but they have been proven by usage to be robust for many classes of signals.*

1.1 Periodogram: Schuster (Marple 5.7)

The periodogram is just the (scaled) magnitude squared of the DTFT of the available data

$$\hat{\mathsf{S}}_{\mathrm{PER}}(f) := \frac{1}{\mathtt{N}} \left| \sum_{\mathtt{n}=0}^{\mathtt{N}-1} X[\mathtt{n}] \exp(-j2\pi f \mathtt{n}) \right|^2$$

We related the mean of the periodogram to the PSD in class, showed the resolution $\propto \frac{1}{N}$, and wrote an equation for the variance of the periodogram for zero mean white Gaussian noise.

1.1.1 Smooth Version(s) (Marple 5.7.2, 5.7.3)

The **Bartlett segment and average periodogram** breaks the available data $X[\mathtt{n}], \quad \mathtt{n} \in \{0, 1 \ldots, \mathtt{N}-1\}$ up into $\mathtt{P}$ blocks of length $\mathtt{L}$, with $\mathtt{P} = \frac{\mathtt{N}}{\mathtt{L}}$, calculates the scaled magnitude squared of the DTFT of each block, and then averages the result. That is,

$$\hat{\mathsf{S}}_{\mathrm{SAP}}(f) := \frac{1}{\mathtt{P}} \sum_{\mathtt{p}=0}^{\mathtt{P}-1} \hat{\mathsf{S}}^{(\mathtt{p})}(f)$$

where

$$\hat{\mathsf{S}}^{(\mathtt{p})}(f) := \frac{1}{\mathtt{L}} \left| \sum_{\mathtt{n}=0}^{\mathtt{L}-1} X[\mathtt{p}\mathtt{L} + \mathtt{n}] \exp(-j2\pi f \mathtt{n}) \right|^2$$

We related the mean of the BSAP to the PSD in class, showing also that it trades lower variance for worse resolution.

The **Welch periodogram** changes the previous scheme by allowing the block to overlap and using data windows. We consider segments of the data of $\mathtt{L}$ samples each, with a shift of $\mathtt{S}$ samples between adjacent segments, so that the maximum number of segments $\mathtt{P}$ is the integer part of $(\mathtt{N}-\mathtt{L})/\mathtt{S}+1$. The $\mathtt{p}$th segement is then weighted with a window to produce the signal

$$X(\mathtt{p})[\mathtt{n}] = w[\mathtt{n}]X[\mathtt{pS}+\mathtt{n}], \quad \mathtt{n} \in \{1,\ldots,\mathtt{L}\}$$

the scaled magnitude of the DTFT is then calculated for each segment

$$\hat{\mathsf{S}}^{(\mathtt{p})}(f) := \frac{1}{U\mathtt{L}}\left|\sum_{\mathtt{n}=0}^{\mathtt{L}-1} X(\mathtt{p})[\mathtt{n}]\exp(-j2\pi f\mathtt{n})\right|^2$$

where U is a constant included in case the window did not have unit energy

$$U := \frac{1}{\mathtt{L}}\sum_{\mathtt{n}=0}^{\mathtt{L}-1}|w[\mathtt{n}]|^2$$

The blocks are then averaged to get the

$$\hat{\mathsf{S}}_{\mathrm{WEL}}(f) := \frac{1}{\mathtt{P}}\sum_{\mathtt{p}=0}^{\mathtt{P}-1}\hat{\mathsf{S}}^{(\mathtt{p})}(f)$$

1.2 Correlogram & Blackman- Tukey

The correlogram operates by first calculating a sample auto-correlation estimate and then taking the DTFT of that. To do this we need estimators for the auto-correlation based on the finite collection of samples $X[\mathtt{n}]$, $\mathtt{n} \in \{0,\ldots,\mathtt{N}-1\}$. We consider two such estimators here. An unbiased estimator for the autocorrelation is

$$\hat{\mathsf{R}}[\mathtt{k}] := \frac{1}{\mathtt{N}-\mathtt{k}}\sum_{\mathtt{n}=0}^{\mathtt{N}-\mathtt{k}-1} X[\mathtt{n}]X^*[\mathtt{n}+\mathtt{k}], \quad \mathtt{k} \in \{0,1,\ldots,\mathtt{N}-1\}$$

and $\hat{\mathsf{R}}[-\mathtt{k}] := \hat{\mathsf{R}}^*[\mathtt{k}]$ for $\mathtt{k} \in \{1,\ldots,\mathtt{N}-1\}$.

An alternative (biased) autocorrelation estimate is

$$\tilde{\mathsf{R}}[\mathtt{k}] := \frac{1}{\mathtt{N}}\sum_{\mathtt{n}=0}^{\mathtt{N}-\mathtt{k}-1} X[\mathtt{n}]X^*[\mathtt{n}+\mathtt{k}], \quad \mathtt{k} \in \{0,1,\ldots,\mathtt{N}-1\}$$

and $\tilde{\mathsf{R}}[-\mathtt{k}] := \tilde{\mathsf{R}}^*[\mathtt{k}]$ for $\mathtt{k} \in \{1,\ldots,\mathtt{N}-1\}$.

It is important to note that, since we have only a finite number of samples, the auto-correlation estimates for higher lags $|\mathtt{k}|$ are of poorer quality than lower lags $|\mathtt{k}|$. For this reason, Blackman-Tukey (heuristically — remember the Caveat Emptor) recommended using only the auto-correlation estimates for $|\mathtt{k}| \leq \mathtt{K} \approx \frac{\mathtt{N}}{10}$, yielding a spectral estimate

$$\hat{\mathsf{S}}_{\mathrm{BT}}(f) := \sum_{\mathtt{k}=-\mathtt{K}}^{\mathtt{K}} w[\mathtt{k}]\tilde{\mathsf{R}}[\mathtt{k}]\exp(-j2\pi f\mathtt{k})$$

where $w[\mathtt{k}]$ is an optional (heuristic) window (whose energy ought to be normalized to unity).

Note that we could have just as well substituted in the unbiased auto-correlation estimate in place of the biased auto-correlation estimate.

1.2.1 Smooth Version: Bartlett

The Bartlett spectrum is formed by using yet another biased auto-correlation estimate. This auto-correlation estimate is formed by breaking the observed data up into $\mathtt{P}$ blocks of length $\mathtt{L}$, so that $\mathtt{P} = \frac{\mathtt{N}}{\mathtt{L}}$, getting a (biased) auto-correlation estimate for each block and then averaging the results to get

$$\bar{\mathsf{R}}[\mathtt{k}] := \frac{1}{\mathtt{P}} \sum_{\mathtt{p}=0}^{\mathtt{P}-1} \tilde{\mathsf{R}}^{(\mathtt{p})}[\mathtt{k}], \quad \mathtt{k} \in \{0, \pm 1, \ldots, \pm \mathtt{K}\}$$

where

$$\tilde{\mathsf{R}}^{(\mathtt{p})}[\mathtt{k}] := \frac{1}{\mathtt{L}} \sum_{\mathtt{n}=0}^{\mathtt{L}-\mathtt{k}-1} X[\mathtt{pL}+\mathtt{n}] X^*[\mathtt{pL}+\mathtt{n}+\mathtt{k}], \quad \mathtt{k} \in \{0, 1, \ldots, \mathtt{K}\}$$

and $\tilde{\mathsf{R}}^{(\mathtt{p})}[-\mathtt{k}] = \tilde{\mathsf{R}}^{(\mathtt{p})}[\mathtt{k}]$ for all $\mathtt{k} \in \{0, 1, \ldots, \mathtt{K}\}$. The estimated auto-correlation is then windowed and transformed to get the estimated power spectral density

$$\hat{\mathsf{S}}_{\mathrm{BAR}}(f) := \sum_{\mathtt{k}=-\mathtt{K}}^{\mathtt{K}} \bar{\mathsf{R}}[\mathtt{k}] \exp(-j2\pi f \mathtt{k})$$

Again, discussed how trades better variance for worse resolution relative to BT.

1.3 Maximum Entropy: Burg

Burg's method assumes that the signal $X[\mathtt{n}]$, $\mathtt{n} \in \mathbb{Z}$ is a Gaussian random process. This extra assumption opens a whole host of possibilities. One possibility concerns a situation in which we are given only the first $\mathtt{K}$ samples of the auto-correlation function and we would like to guess the PSD from these known samples of the ACF. Burg's method performs this by choosing the PSD which maximizes the (Shannon) entropy (rate) subject to the (known auto-correlation) constraints

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \mathsf{S}(f) \exp(j2\pi f \mathtt{k}) df = \mathsf{R}[\mathtt{k}], \quad \mathtt{k} \in \{0, 1, \ldots, \mathtt{K}\} \tag{1}$$

It can be shown that entropy rate of a Gaussian process is proportional to

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln(\mathsf{S}(f)) df \tag{2}$$

The objective function (2) can be maximized subject to the constraints (1) by forming a Lagrangian (objective function minus sum of constraints multiplied by Lagrange multipliers) and setting its variation equal to zero. The solution may be written as

$$\mathsf{S}(f) := \frac{\sigma^2}{1 + \sum_{\mathtt{k}=1}^{\mathtt{K}} \lambda_{\mathtt{k}} \exp(-j2\pi f \mathtt{k})}$$

where the Lagrange multipliers may be solved by substituting into the constraints, or equivalently, by solving

$$\begin{bmatrix} \mathsf{R}[0] & \mathsf{R}[-1] & \cdots & \mathsf{R}[-\mathtt{K}] \\ \mathsf{R}[1] & & & \\ \vdots & & & \\ \mathsf{R}[\mathtt{K}] & \cdots & \cdots & \mathsf{R}[0] \end{bmatrix} \begin{bmatrix} 1 \\ \lambda_1 \\ \vdots \\ \lambda_{\mathtt{K}} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$