System Identification

Supplementary notes: lecture 4

Roy Smith

4 Sampled dynamic models & frequency-domain analysis

4.1 Convergence of the periodogram to the spectral density

It is fairly straightforward to show that the periodogram is an asymptotically unbiased estimate of the power spectral density,

$$\lim_{N\longrightarrow\infty} E\left\{\frac{1}{N}|V_N(\mathrm{e}^{j\omega})|^2\right\} \;=\; \phi_v(\mathrm{e}^{j\omega})$$

We begin with a useful result from Stoica & Moses [1, Ex. 1.6, p. 17].

Lemma 4.1. *Given an arbitrary function of a discrete-variable, $f(\tau)$,*

$$\sum_{t=0}^{N-1}\sum_{s=0}^{N-1} f(t-s) \;=\; \sum_{\tau=-N+1}^{N-1} (N-|\tau|)f(\tau).$$

Proof of lemma: 4.1: Define $\tau = t - s$ and note that given the bounds on the left-hand summation τ can range from $-N+1$ (corresponding to $t = 0$, $s = N-1$) to $N-1$ (corresponding to $t = N-1$, $s = 0$).

For any given τ in the range $-N+1$ to $N-1$ we have multiple instances of $f(\tau)$ appearing in the left-hand summation. We can therefore write this as,

$$\sum_{t=0}^{N-1}\sum_{s=0}^{N-1} f(t-s) \;=\; \sum_{\tau=-N+1}^{N-1} \alpha(\tau)f(\tau),$$

where $\alpha(\tau)$ is the number of instances where $\tau = t - s$ for all $t = 1, \ldots, N$ and all $s = 1, \ldots, N$. It is easy to count up these instances to get,

$$\alpha(\tau) = N - |\tau|.$$

□

Now to show the periodogram convergence result (based on Stoica & Moses [1, p. 7]):

$$\begin{aligned}
&\lim_{N\longrightarrow\infty} E\left\{\frac{1}{N}\left|\sum_{k=0}^{N-1} v(k)\mathrm{e}^{-j\omega k}\right|^2\right\} \\
&= \lim_{N\longrightarrow\infty}\frac{1}{N}\sum_{t=0}^{N-1}\sum_{s=0}^{N-1} E\{v(t)v^*(s)\}\mathrm{e}^{-j\omega(t-s)} \\
&= \lim_{N\longrightarrow\infty}\frac{1}{N}\sum_{\tau=-N+1}^{N-1}(N-|\tau|)R_v(\tau)\mathrm{e}^{-j\omega\tau} \\
&\qquad = \underbrace{\lim_{N\longrightarrow\infty}\sum_{\tau=-N+1}^{N-1} R_v(\tau)\mathrm{e}^{-j\omega\tau)}}_{\phi_v(\mathrm{e}^{j\omega})} - \underbrace{\lim_{N\longrightarrow\infty}\frac{1}{N}\sum_{\tau=-N+1}^{N-1}|\tau|R_v(\tau)\mathrm{e}^{-j\omega\tau}}_{\longrightarrow 0 \text{ via Lemma 4.1}}
\end{aligned}$$

Note that this is only asymptotically unbiased. Simply averaging repeated periodograms will not converge to the spectrum. The error results from the Fourier Transform calculation truncation for finite N.

The $N-|\tau|$ term gives an indication of how this might be corrected in an estimate of $R_v(\tau)$.

Ljung [2, Section 2.3] introduces "quasi-stationarity" as a means of treating combinations of stochastic and deterministic signals. This essentially says that stationarity holds only in the limit as $N\longrightarrow\infty$. Under this assumption the above result holds only weakly; i.e.

$$\lim_{N\longrightarrow\infty}\int_{-\pi}^{\pi} E\left\{\frac{1}{N}\left|V_N(\mathrm{e}^{j\omega})\right|^2\right\}\psi(\mathrm{e}^{j\omega})\,d\omega = \int_{-\pi}^{\pi}\phi(\mathrm{e}^{j\omega})\psi(\mathrm{e}^{j\omega})\,d\omega,$$

for any smooth function, $\psi(\mathrm{e}^{j\omega})$. The above has been modified slightly to account for the $1/\sqrt{N}$ definition that Ljung uses for his DFT.

Obviously $\psi(\mathrm{e}^{j\omega})$ can be chosen to focus on a very narrow range of frequencies. This is a similar result to the convergence of the Fourier Transform—the Gibbs phenomenon illustrates that the Fourier Transform does not converge point-wise.

4.2 Finite-length data estimates of the power spectral density

In practice we can only estimate $R_x(\tau)$ from one or more data records, $x(k)$, $k=0,\ldots,N-1$. We will denote estimates of the autocorrelation by $\hat{R}_x(\tau)$. There are two common choices for the calculation of such estimates,

$$1.\quad \hat{R}_x(\tau) \;=\; \begin{cases} \dfrac{1}{N}\displaystyle\sum_{k=\tau}^{N-1} x(k)x(k-\tau), & \text{for } \tau \geqslant 0, \\ \dfrac{1}{N}\displaystyle\sum_{k=0}^{N+\tau-1} x(k)x(k-\tau), & \text{for } \tau < 0, \end{cases}$$

$$2.\quad \hat{R}_x(\tau) \;=\; \begin{cases} \dfrac{1}{N-|\tau|}\displaystyle\sum_{k=\tau}^{N-1} v(k)v(k-\tau), & \text{for } \tau \geqslant 0, \\ \dfrac{1}{N-|\tau|}\displaystyle\sum_{k=0}^{N+\tau-1} v(k)v(k-\tau), & \text{for } \tau < 0, \end{cases}$$

Note that using either of the above gives us estimates of $R_x(\tau)$ for $2N-1$ values of τ,

$$\hat{R}_x(\tau), \quad \text{with} \quad \tau = -N+1, \ldots, 0, \ldots, N-1.$$

The Fourier transform of $\hat{R}_x(\tau)$ is now our estimate of the power spectral density, $\phi_x(\mathrm{e}^{j\omega})$.

$$\hat{\phi}_x(\mathrm{e}^{j\omega}) \;=\; \sum_{\tau=-\infty}^{\infty} \hat{R}_x(\tau)\mathrm{e}^{j\omega_\tau \cdot}$$

This immediately gives a problem as we only have autocorrelation estimates, $\hat{R}_x(\tau)$, for $\tau \in \{-N+1, \ldots, N-1\}$. The obvious approximation is to use only those values in the calculation, giving,

$$\hat{\phi}_x(\mathrm{e}^{j\omega}) \;=\; \sum_{\tau=-N+1}^{N-1} \hat{R}_x(\tau)\mathrm{e}^{j\omega_\tau \cdot}$$

Keep in mind that there is an implicit assumption (which probably will not be satisfied exactly) that $R_x(\tau) = 0$ for $|\tau| \geqslant N$. This is an additional source of potential error in the finite data calculation.

We can compare this to the Fourier transform[1] of the sequence, $x(k)$, $k = 0, \ldots, N-1$, which is given by,

$$X(\omega_n) \;=\; \sum_{k=0}^{N-1} x(k)\mathrm{e}^{-j\omega_n k}, \quad \text{for} \quad \omega_n \;=\; \frac{2\pi n}{N}, \quad n = 0, \ldots, N-1.$$

In the case where we choose the first estimator estimator $\hat{R}_x(\tau)$, (i.e. with $1/N$ scaling), it is possible to show that,

$$\hat{\phi}_x(\omega_n) \;=\; \frac{1}{N}|X(\omega_n)|^2.$$

[1]The definition of the discrete Fourier Transform given here is standard but differs in scaling and indexing from that used by Lennart Ljung [2]. Ljung's definition of the periodogram therefore differs slightly.

This estimate of the power spectral density is also known as the periodogram.

We are interested in the "quality" of each of the $\hat{R}_x(\tau)$ estimates given above, and how the quality of the resulting estimates of the power spectral density, $\hat{\phi}_x(\mathrm{e}^{j\omega})$. First consider the case for a fixed value of τ and for an independent, identically distributed, stochastic signal. By definition, for stochastic signals,

$$R_x(\tau) = E\{x(k)x(k+\tau)\}.$$

In a finite data record of length N, we can find $N - |\tau|$ product pairs $x(k)x(k+\tau)$ upon which to base our estimate of $E\{x(k)x(k+\tau)\}$. One estimate of $E\{x(k)x(k+\tau)\}$ is simply the mean of the $N - |\tau|$ samples that can be calculated from our data. Note that each of these product pairs are independent (as the underlying samples $x(k)$ are assumed independent of one another) and so the sample estimate is unbiased. In other words, if

$$\hat{R}_x(\tau) \;=\; \begin{cases} \dfrac{1}{N-|\tau|} \displaystyle\sum_{k=\tau}^{N-1} v(k)v(k-\tau), & \text{for } \tau \geqslant 0, \\ \dfrac{1}{N-|\tau|} \displaystyle\sum_{k=0}^{N+\tau-1} v(k)v(k-\tau), & \text{for } \tau < 0, \end{cases}$$

then,

$$E\left\{\hat{R}_x(\tau)\right\} \;=\; R_x(\tau).$$

The $1/(N-|\tau|)$ scaling therefore gives an unbiased estimate of the autocorrelation function. In this case, this is the same as the autocovariance function.

Now look at the corresponding estimate of the power spectral density at the discrete frequency points.

$$\begin{aligned} E\left\{\hat{\phi}_x(\omega_n)\right\} &= E\left\{\sum_{\tau=-N+1}^{N-1} \hat{R}_x(\tau)\mathrm{e}^{j\omega_n\tau}\right\} \\ &= \sum_{\tau=-N+1}^{N-1} E\left\{\hat{R}_x(\tau)\right\}\mathrm{e}^{j\omega_n\tau} \\ &= \sum_{\tau=-N+1}^{N-1} R_x(\tau)\mathrm{e}^{j\omega_n\tau} \\ &\neq \sum_{\tau=-\infty}^{\infty} R_x(\tau)\mathrm{e}^{j\omega_n\tau} \quad = \quad \phi_x(\omega_n). \end{aligned}$$

So the power spectral density estimate is biased in this case, but as a result of the finite summation. If, for $|\tau| > N$, the autocorrelation function has decayed to essentially zero then there will be very little bias as a result of the finite summation.

Does the first autocorrelation estimate (with $1/N$ scaling) give a better result? Because the only difference is a scaling we can see that if,

$$\hat{R}_x(\tau) \;=\; \begin{cases} \dfrac{1}{N}\displaystyle\sum_{k=\tau}^{N-1} x(k)x(k-\tau), & \text{for } \tau \geqslant 0, \\ \dfrac{1}{N}\displaystyle\sum_{k=0}^{N+\tau-1} x(k)x(k-\tau), & \text{for } \tau < 0, \end{cases}$$

then,

$$E\left\{\hat{R}_x(\tau)\right\} \;=\; \frac{N-|\tau|}{N} R_x(\tau).$$

This is clearly a biased estimate of the autocorrelation for stochastic signals. The resulting power spectral density estimate is therefore also biased,

$$E\left\{\hat{\phi}_x(\omega_n)\right\} \;=\; E\left\{\sum_{\tau=-N+1}^{N-1} \hat{R}_x(\tau)\mathrm{e}^{j\omega_n\tau}\right\} \;=\; \sum_{\tau=-N+1}^{N-1} \frac{N-|\tau|}{N} R_x(\tau)\mathrm{e}^{j\omega_n\tau} \;\neq\; \phi_x(\omega_n).$$

As N becomes large, and τ/N becomes small, the distinction in bias error between the two methods lessens. Note that,

$$\lim_{\substack{N\longrightarrow\infty \\ \tau/N\longrightarrow 0}} E\left\{\hat{R}_x(\tau)\right\} \;=\; R_x(\tau),$$

for both autocorrelation estimation methods. The finite summation bias in the power spectral density estimate also decays for both methods,

$$\lim_{\substack{N\longrightarrow\infty \\ \tau/N\longrightarrow 0}} E\left\{\hat{\phi}_x(\mathrm{e}^{j\omega})\right\} \;=\; \phi_x(\mathrm{e}^{j\omega}).$$

Both estimation methods therefore give *asymptotically unbiased* estimates of the power spectral density.

The situation is reversed for a periodic signal. In the periodic case the autocorrelation of a signal $x(k)$ is defined as

$$R_x(\tau) = \frac{1}{N}\sum_{k=0}^{N-1} x(k)x(k-\tau), \qquad \tau = -N/2+1, \ldots, N/2.$$

The autocorrelation, $R_x(\tau)$, is also periodic of period N. If we apply the first of the autocorrelation estimation methods,

$$\hat{R}_x(\tau) \;=\; \begin{cases} \dfrac{1}{N}\displaystyle\sum_{k=\tau}^{N-1} x(k)x(k-\tau), & \text{for } \tau \geqslant 0, \\ \dfrac{1}{N}\displaystyle\sum_{k=0}^{N+\tau-1} x(k)x(k-\tau), & \text{for } \tau < 0, \end{cases}$$

for $\tau = 0, \ldots, N-1$, (or indeed for any choice of N consecutive τ values), then we will have exactly calculated one period of the true autocorrelation function, $R_x(\tau)$. So this estimate is exact in the case where there is no noise and no transient.

Furthermore, estimating the power spectral density, $\hat{\phi}_x(\omega_n)$, by taking the Fourier transform of the $1/N$ scaled estimate, is also exact. This is because the autocorrelation function is periodic and there is no approximation involved in using the DFT to calculate the Fourier transform on a single period. This means that

$$\hat{\phi}_x(\omega_n) \;=\; \phi_x(\omega_n).$$

Therefore in the case where $x(k)$ is a periodic signal, the $1/N$ scaling gives the best result.

This leads to a difficulty in deciding which method to use for estimating the autocorrelation of an arbitrary signal, $x(k)$. In system identification the measured output is almost always composed of the sum of a stochastic signal (from the noise) and a deterministic, or even periodic, signal from the convolution of the plant with a deterministic input signal, $u(k)$. This means that there is no obvious correct choice in deciding which autocorrelation estimation method to apply. The better method will inevitably be problem dependent.

So how important is the distinction between autocorrelation estimation methods for a practical (finite data length) problem?

One thing to observe is that if we scale the autocorrelation estimate by $1/N$ we make the estimate of the tail of $R_x(\tau)$ (i.e. for τ close to N) smaller than it would otherwise be in the case of a $1/(N - |\tau|)$ scaling. This reduces the truncation error which comes from calculating the spectral estimate over a finite range of autocorrelation lags. In contrast, scaling by $1/(N - \tau)$, can result in large values of $R_x(\tau)$ when τ is close to N. This can result in what appear to be "oscillations" on the spectral estimate.

One way of viewing this is to consider the $1/N$ scaling as time-domain multiplication by the function

$$f(\tau) \;:=\; \frac{N-\tau}{N}, \quad \text{for } 0 \leqslant \tau \leqslant N,$$

which decays linearly from 1 (at $\tau = 0$) to zero (at $\tau = N$). In the frequency domain the effect of the scaling by $f(\tau)$ corresponds to convolving $\phi_x(\mathrm{e}^{j\omega})$ with the Fourier transform of $f(\tau)$. By noting that $f(\tau)$ looks something like the impulse response of a low-pass filter, we see that this choice of scaling effectively smooths the frequency domain estimate of $\phi_x(\mathrm{e}^{j\omega})$. This will often make the $1/N$ based estimate, $\hat{\phi}_x(\mathrm{e}^{j\omega})$, appear more realistic. However, for finite N, it is still biased. And if the true spectral density is discontinuous—as might be the case if electric noise from the power grid is a significant component of the noise—then the low pass filtering of the spectral density will give a poor estimate of the true spectral density.

References

[1] P. Stoica and R. Moses, *Introduction to Spectral Analysis.* Prentice-Hall, 1997.

[2] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Prentice-Hall, 1999.