

Lecture 7

Table of Contents

- Basic terminology..... 1
 - Parameter vs. Statistic..... 1
 - Estimators..... 1
 - Estimator Evaluation..... 2
 - Bias..... 2
 - Variance..... 2
 - Mean Square Error..... 2
 - Bias-variance tradeoff..... 3
 - Consistency..... 3
- Power Spectrum Estimation..... 3
 - Periodogram..... 4
 - Statistical Properties of the Periodogram..... 4
 - Covariance..... 6
 - Modified Periodogram..... 6
 - Bartlett Window..... 6
- Welch's method..... 7
- Functions..... 8

Basic terminology

Parameter vs. Statistic

In statistics, a **parameter** denoted θ is a numerical value that states something about the entire population. Since a parameter is difficult to obtain, we use a statistic. A **statistic** is a numerical value that states something about a small sample of the entire population. The value of a parameter is a fixed number whereas a statistic can vary because it depends on the given sample.

Estimators

An **estimator** denoted $\hat{\theta}$ is a rule or a formula that can approximate some parameter θ given a sample of N observations.

For example, to estimate the mean value m of a random variable X , we can use the arithmetic average as an estimate:

$$\hat{m} = \frac{1}{N} \sum_{k=1}^N x_k.$$

This is called the **sample mean** estimator.

Sample variance

$$\hat{\sigma}^2 \triangleq \frac{1}{N} \sum_{k=1}^N (x_k - \hat{m})^2$$

Sample covariance

$$\hat{\sigma}_{xy} \triangleq \frac{1}{N} \sum_{k=1}^N (x_k - \hat{m}_x)(y_k - \hat{m}_y)$$

Estimator Evaluation

The quality of an estimator $\hat{\theta}$ of a parameter θ is evaluated based on three properties:

1. Bias
2. Variance
3. Mean square error

Bias

$$B(\hat{\theta}) \triangleq E(\hat{\theta}) - \theta. \quad (14.3)$$

When $B(\hat{\theta}) = 0$ then the estimator yields the correct value on average when used a large number of time. An estimator with $B = 0$ is called **unbiased** estimator.

Variance

$$\text{var}(\hat{\theta}) = E\{[\hat{\theta} - E(\hat{\theta})]^2\} \quad (14.4)$$

The variance describes the spread of the estimates about its expected value.

Mean Square Error

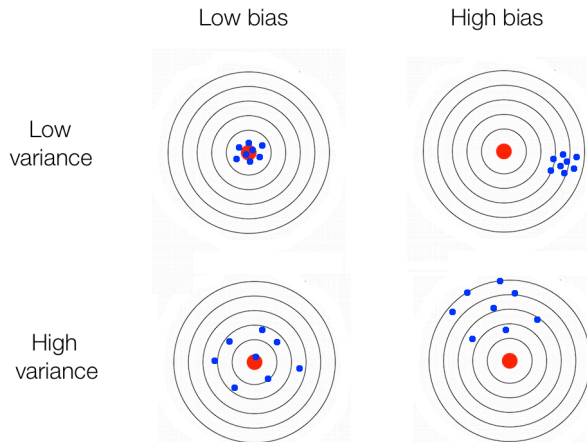
measures the average deviation of the estimator $\hat{\theta}$ from the true value θ :

$$\text{mse}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]. \quad (14.5)$$

By rewrite Eq. (14.5) we see that the mean square error takes into consideration both the variance and the bias of the estimator:

$$\text{mse}(\hat{\theta}) = \text{var}(\hat{\theta}) + B^2(\hat{\theta}). \quad (14.6)$$

Bias-variance tradeoff



A good estimator should have zero bias and the smallest possible variance. However, typically these two quantities conflict with each other. This is known as the bias-variance tradeoff.

Consistency

An estimator is said to be **consistent** if its estimate converges to the true value θ as the number of observations N approaches infinity.

From Eq. (14.6), we see that if variance and the bias of the estimator is zero then the estimator converges to the true value.

$$\text{mse}(\hat{\theta}) = \text{var}(\hat{\theta}) + B^2(\hat{\theta}). \quad (14.6)$$

Therefore, a sufficient condition for a consistent estimator is that both its variance and bias converge to zero as the number of observations becomes very large.

Power Spectrum Estimation

Recall power spectrum is defined as follows:

$$S_{xx}(\omega) = \sum_{\ell=-\infty}^{\infty} r_{xx}(\ell) e^{-j\omega\ell}$$

We want to estimate the PSD of a zero-mean wide-sense stationary process.

General approach for power spectrum estimation of a random signal from a finite set of N observations:

1) Window the observed signal $y(n)$

$$y_w(n) = w(n)y(n), 0 \leq n \leq N-1$$

2) Take the Fourier Transform:

$$y_w(n) \leftrightarrow X_w(e^{j\omega})$$

3) Estimate the power spectrum

$$\hat{S}_{xx}(\omega) = \frac{1}{N \cdot F} |X_w(e^{j\omega})|^2 \text{ where } F \text{ is constant that depends on the window}$$

Periodogram

A periodogram is a method for estimating the Power Spectrum Density of a random signal from a finite set of N observations given by:

$$I(\omega) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-j\omega n} \right|^2 \quad (14.34)$$

The periodogram uses a rectangular window $w(n) = 1$.

Statistical Properties of the Periodogram

Mean of the periodogram Taking the mathematical expectation of (14.33) and using (14.26) (recall that $\hat{c}[\ell] = \hat{r}[\ell]$ for $m = 0$), we obtain the relation

$$E[I(\omega)] = \sum_{\ell=-(N-1)}^{N-1} E[\hat{r}[\ell]] e^{-j\omega\ell} = \sum_{\ell=-(N-1)}^{N-1} \left(1 - \frac{|\ell|}{N}\right) r[\ell] e^{-j\omega\ell}. \quad (14.40)$$

Since $E[I(\omega)] \neq S(\omega)$, the periodogram is a *biased* estimator of $S(\omega)$. However, for each $|\ell|$, the factor $(1 - |\ell|/N) \rightarrow 1$ as $N \rightarrow \infty$. Hence, we have

$$\lim_{N \rightarrow \infty} E[I(\omega)] = S(\omega), \quad (14.41)$$

Eq. (14.41) says that as the window length N increases towards infinity, the mean of the periodogram converges to the true power spectrum. Asymptotically (as N increases) the periodogram is an unbiased estimator of $S(\omega)$.

For finite N , it has some bias which is reflected by the window function.

Variance

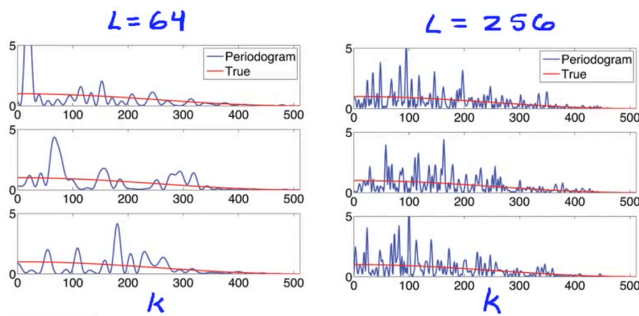
Under a variety of assumptions which holds for most random processes of practical interest, the variance of the periodogram estimator behaves like the square of the true PSD:

$$\text{var}[I(\omega)] \approx S^2(\omega), \quad 0 < \omega < \pi \quad (14.50)$$

The main implication of (14.50) is that the **periodogram is not a good estimator of PSD** because, independently of N , the standard deviation of the estimator is as large as the quantity to be estimated.

In other words, the periodogram is not a consistent estimator; that is, its distribution does not tend to cluster more closely around the true PSD as N increases.

Thus the definition of the PSD by $S(\omega) = \lim_{N \rightarrow \infty} I(\omega)$ is not valid because even if $\lim_{N \rightarrow \infty} E[I(\omega)] = S(\omega)$, the variance of $I(\omega)$ does not tend to zero as $N \rightarrow \infty$.



Note that the variance is high at lower frequencies where the true power spectrum is larger. As the true spectrum goes to zero, the variance gets smaller. This is in agreement with (14.50). If we get better resolution L goes from 64 to 256, the behaviour is still present.

Covariance

The covariance between values of the periodogram at harmonically spaced frequencies is given by

$$\text{cov}[I(2\pi k_1/N), I(2\pi k_2/N)] \approx 0, \quad k_1 \neq k_2 \quad (14.51)$$

which states that closely spaced values of the periodogram are uncorrelated.

Modified Periodogram

The modified periodogram uses a window $w(n)$ that is not rectangular.

$$\tilde{I}(\omega) \triangleq \frac{1}{N} \left| \sum_{n=0}^{N-1} w[n]x[n]e^{-j\omega n} \right|^2 \quad (14.52)$$

The normalisation factor has to be

$$F = \frac{1}{L} \sum_{n=0}^{L-1} (w[n])^2$$

Bartlett Window

We define the zero-phase triangular or Bartlett window

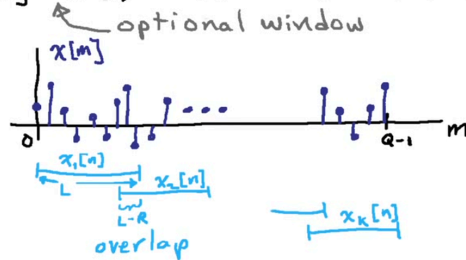
$$w_B[\ell] \triangleq \begin{cases} 1 - \frac{|\ell|}{N}, & 0 \leq |\ell| \leq N-1 \\ 0, & |\ell| \geq N \end{cases} \quad (14.42)$$

```
w = bartlett(64);  
plot(w)  
F = fft(w);  
plot(abs(fft(w)))
```

Welch's method

- average K independent random variables: reduce variance by factor of K
- Divide $x[m]$, $m=0,1,\dots,Q-1$ into (possibly) overlapping segments of length L . r th segment:

$$x_r[n] = x[rR+n]w[n], \quad n=0,1,\dots,L-1, \quad r=1,2,\dots,K$$

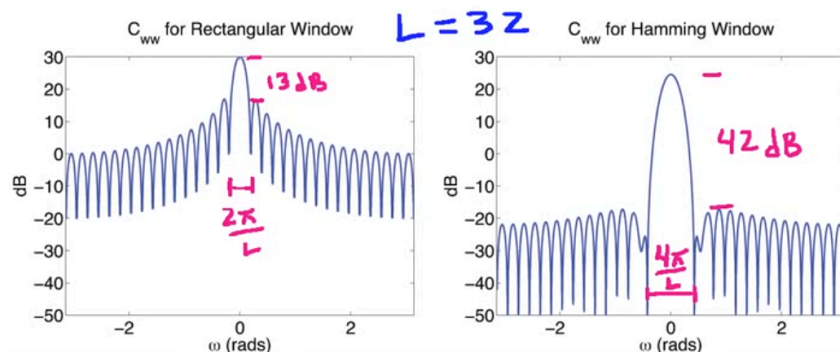


- Average periodograms from each segment

$$I_r(\omega) = \frac{1}{2\pi \sum_{n=0}^{L-1} w^2[n]} \left| \sum_{n=0}^{L-1} x_r[n] e^{-j\omega n} \right|^2$$

$$\hat{S}_{xx}^{av}(\omega) = \frac{1}{K} \sum_{k=1}^K I_r(\omega)$$

Variance: $\text{Var}\{\hat{S}_{xx}^{per}(\omega)\} \geq \text{Var}\{\hat{S}_{xx}^{av}(\omega)\} \geq \frac{1}{K} \text{Var}\{\hat{S}_{xx}^{per}(\omega)\}$
 $1/K$ reduction if $I_r(\omega)$ are independent



The nature of bias depends on the chosen window. There is a trade-off between main lobe width and side-lobe height.

- Rectangular window gives you the narrowest main lobe but a fairly large side-lobe height.
- Hamming window gives you twice the main lobe width but the side-lobes are much lower.

Hamming window sacrifices some resolution for increased dynamic range.

Dynamic range is the ratio of the largest signal amplitude and the smallest signal amplitude. How well can we reliably distinguish the largest from the smallest.

Windows are used to manage bias.

Typically, we will trade resolution (main-lobe width) for dynamic range (side-lobe height)

Functions

```
function r=acrs(x, L)
    % Computes the ACRS r[m] for 0 <= m <= L
    % r=acrs(x-mean(x),L) yields the ACVS
    N=length(x);
    x1=zeros(N+L-1,1);
    x2=x1;
    x1(1:N,1)=x;
    for m=1:L
        x2=zeros(N+L-1,1);
        x2(m:N+m-1,1)=x;
        r(m)=x1'*x2;
    end
    r=r(:)/N;
end

function I=psdper(x, K)
    % Compute periodogram I(w) using the FFT.
    % K-point FFT >= N
    N=length(x);
    X=fft(x,K);
    I=X.*conj(X)/N;
    I(1)=I(2); % Avoid DC bias
    I=I(:);
end

function r=acrsfft(x, L)
    % Compute the autocorrelation sequence using the FFT.
    % r=acrsfft(x-mean(x),L) yields the ACVS
    N=length(x);
    Q=nextpow2(N+L);
    X=fft(x,2^Q);
    r0=real(ifft(X.*conj(X)));
    r=r0(1:L)/N;
end
```



```

function I=psdmodper(x, K)
    % Compute the modified periodogram PSD estimate.
    % K-point FFT >= N
    N=length(x);
    w=hann(N); % choose window
    w=w/(norm(w)/sqrt(N)); % sum w^2[k]=N
    X=fft(x(:).*w(:),K);
    I=X.*conj(X)/N;
    I(1)=I(2); % Avoid DC bias
    I=I(:);
end

```

```

function psdbt(x, L, K)
    % Compute the Blackman-Tukey PSD estimate.
    % Blackman-Tukey PSD estimator of  $S(2\pi k/K)$ 
    N=length(x);
    w=hann(N); % Data window
    w=w/(norm(w)/sqrt(N)); % sum w^2[k]=N
    x=x.*w; % Data windowing
    r=acrsfft(x,L);
    wc=parzenwin(2*L-1); % Lag window
    rw=r.*wc(L:2*L-1); % Lag windowing
    g=zeros(K,1);
    g(1:L)=rw;
    g(K-L+2:K)=flipud(rw(2:L));
    G=fft(g,K); % f even => F real
    S=2*real(G(1:K/2)); S(1)=S(2);
end

```

```

function S=psdwelch(x, L, K)
    % Compute the Welch PSD estimate.
    % Welch PSD estimator of  $S(2\pi k/K)$ 
    M=fix((length(x)-L/2)/(L/2)) % 50% overlap
    time=(1:L)';
    I=zeros(K,1);
    w=hanning(L); % Choose window
    w=w/(norm(w)/sqrt(L)); % sum w^2[k]=L
    for m=1:M
        %xw=w.*detrend(x(time)); % detrending
        xw=w.*x(time); % data windowing
        X=fft(xw,K);
        I=I+X.*conj(X);
        time=time+L/2;
    end
    I=I/(M*L); S=2*I(1:K/2); S(1)=S(2);
end

```