

TD1: Parameter and Spectrum Estimation

1 Parameter Estimation

Problem 1. Poisson Process - Bayesian Parameter Setting

Consider a network node where messages are passing and let the single observation y be the time between the passage of two consecutive messages. y has an exponential distribution

$$f(y|\theta) = \begin{cases} \theta e^{-\theta y} & , y \geq 0 \\ 0 & , y < 0 \end{cases} \quad (1)$$

where the parameter θ represents the average frequency of messages passing by the node we are considering. This average frequency has itself an a priori distribution (over the different nodes in the network) which we take to be exponential with average value $1/\lambda$:

$$f(\theta) = \begin{cases} \lambda e^{-\lambda \theta} & , \theta \geq 0 \\ 0 & , \theta < 0 . \end{cases} \quad (2)$$

- (a) Find the a posteriori distribution of θ given the measurement y and then find the estimators $\hat{\theta}_{MMSE}$, $\hat{\theta}_{MAP}$, $\hat{\theta}_{ABS}$ which correspond to the mean, the mode and the median of this distribution (to find the last one, the root of a transcendental equation will have to be found (by e.g. Matlab)). Observe how these three Bayes estimators are similar, but not identical.

Note: The formula $\int_0^{+\infty} \theta^n e^{-\lambda \theta} d\theta = \frac{n!}{\lambda^{n+1}}$ may be useful.

- (b) Evaluate the performances of the estimators.

Problem 2. MAP and MMSE Estimation - Hard and Soft Decisions for Gaussian Noise

Assume we send a binary symbol θ (θ takes the values ± 1 with equal probability) over an additive Gaussian noise channel so that we receive

$$y = h\theta + v$$

where $v \sim \mathcal{N}(0, \sigma_v^2)$ and $h > 0$.

(a) Find the posterior distribution of θ given y .

(b) Find the MAP estimate of θ given y .

Due to the fact that θ takes on discrete values, the MAP estimation problem becomes a detection problem and the MAP estimate yields a hard decision.

(c) Show that the MMSE estimate of the symbol θ in terms of y is

$$\hat{\theta}_{MMSE} = \tanh\left(\frac{hy}{\sigma_v^2}\right) \quad \text{where} \quad \tanh(y) = \frac{e^y - e^{-y}}{e^y + e^{-y}}.$$

This estimate can be interpreted as a *soft decision*, in contrast to the MAP hard decision: $\hat{\theta}_{MMSE}$ does not correspond to one of the possible values of θ but to a weighted average of all possible values, weighted according to their posterior probability.

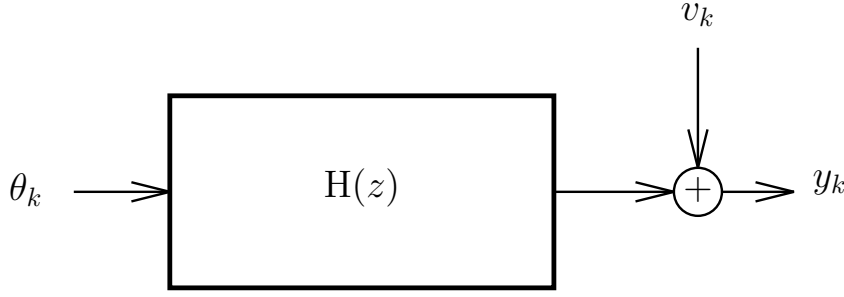
Problem 3. Bayesian Fourier Analysis

Many signals exhibit cyclical behavior. It is common practice to determine the presence of strong cyclical components by employing Fourier analysis. Large Fourier coefficients are indicative of strong components. Here we show how Fourier analysis can result from a well-founded estimation approach. Assume we have the following signal model

$$y_k = a \cos(2\pi f_0 k) + b \sin(2\pi f_0 k) + v_k, \quad k = 1, \dots, n \quad (3)$$

where f_0 is assumed to be an integer multiple of $\frac{1}{n}$ (as in the DFT), excepting 0 or $\frac{1}{2}$ (for which $\sin(2\pi f_0 k)$ would be identically zero), and v_k is WGN (white Gaussian noise): $V = [v_1 \dots v_n]^T \sim \mathcal{N}(0, \sigma_v^2 I_n)$. It is desired to estimate $\theta = [a \ b]^T$. We assume that a and b are random variables with prior pdf $f(\theta) \leftrightarrow \mathcal{N}(0, \sigma_\theta^2 I_2)$ and θ is assumed independent of V . This type of model is referred to as a *Rayleigh fading sinusoid* because the sinusoid amplitude $\sqrt{a^2 + b^2}$ has a Rayleigh distribution (the phase is uniform in $[0, 2\pi)$). This model is frequently used to model a sinusoid that has propagated through a dispersive medium.

- i) Find the AMMSE estimator $\hat{\theta}_{AMMSE}$ and its associated covariance matrix.
- ii) What are the MMSE and MAP estimators?
- iii) Compute the Fisher Information Matrix $J(\theta)$. Is $\hat{\theta}_{AMMSE}$ efficient?



Problem 4. Deconvolution

Consider a signal θ_k filtered by a channel with transfer function $H(z)$ and observed through additive noise as y_k .

This set-up could correspond to a communication problem in which we use Pulse Amplitude Modulation (PAM). So let the θ_k be an independent and identically distributed (i.i.d.) sequence of symbols $\pm\sigma_\theta$:

$$\Pr[\theta_k = \sigma_\theta] = \Pr[\theta_k = -\sigma_\theta] = \frac{1}{2}, \quad k = 1, \dots, m \quad (4)$$

and assume $\theta_k = 0$ for $k \leq 0$ or $k > m$. For the communications channel, we shall find inspiration in the telephone line channel, which, considering a sampling frequency of 8 kHz, has a bandpass characteristic that cuts off at dc and half the sampling frequency (passband: [300,3400] Hz). A simple model for such a characteristic could be $H(z) = 1 - z^{-2}$. We assume the additive noise samples v_k to be i.i.d., Gaussian: $v_k \sim \mathcal{N}(0, \sigma_v^2)$. We shall consider the observations y_k for $k = 1, \dots, n = m+2$. And we shall at first ignore the statistical description of the θ_k , and rather treat the θ_k as deterministic.

(a) Find the linear model representation $Y = H\theta + V$, where $Y = [y_1 \dots y_n]^T$, $\theta = [\theta_1 \dots \theta_m]^T$, $V = [v_1 \dots v_n]^T$, and H is a $n \times m$ matrix. That is, find H given $H(z) = 1 - z^{-2}$.

The classical approach to recovering θ would be to apply a linear (filtering) operation to the received signal

$$\hat{\theta}_{zd} = F Y \quad (5)$$

where F is a $m \times n$ matrix, so that the signal part θ passes undistortedly: $F H = I_m$ (*zero distortion* or *zero forcing equalization* (ZFE)).

(b) Verify that in this case, the (conditional on θ) bias $b_{\hat{\theta}_{zd}}(\theta) = 0$ ($\hat{\theta}_{zd}$ is conditionally unbiased). Note that requiring $b_{\hat{\theta}_{zd}}(\theta) = 0$ is not enough to specify F completely.

(c) Find the Weighted Least-Squares (WLS) estimator $\hat{\theta}_{WLS}$ (expressed in terms of H). Show that $\hat{\theta}_{WLS}$ is a possible $\hat{\theta}_{zd}$.

(d) For the (unweighted) Least-Squares (LS) estimator $\hat{\theta}_{LS}$, write a difference equation linking the $\{\hat{\theta}_k, k = 1, \dots, m\}$ to the y_k . Assuming for a moment that this difference equation were valid $\forall k$, interpret this difference equation as a certain filter $F(z)$ taking y_k as input signal and producing $\hat{\theta}_k$ as output signal. Is this filter $F(z)$ stable? What does this suggest to you about the performance of $\hat{\theta}_{LS}$?

(e) Find the Maximum Likelihood estimator $\hat{\theta}_{ML}$. Is $\hat{\theta}_{ML}$ unbiased?

$\hat{\theta}_{LS}$ and $\hat{\theta}_{ML}$ do not constrain a priori $\hat{\theta}$ to be linear, but they turn out to be linear. Now, if the estimator is to be linear, a better approach is to exploit the prior statistical information about the θ_k and to use the LMMSE estimator $\hat{\theta}_{LMMSE} = R_{\theta Y} R_{YY}^{-1} Y$.

(f) Compute m_θ , $R_{\theta\theta}$ and $C_{\theta\theta}$. Justify that we also have $\hat{\theta}_{LMMSE} = C_{\theta Y} C_{YY}^{-1} Y$.

(g) Show that $C_{\theta Y} = \sigma_\theta^2 H^T$ and that $C_{YY} = \sigma_\theta^2 H H^T + \sigma_v^2 I$ and find an expression for $C_{\theta Y} C_{YY}^{-1}$ in terms of H , σ_v^2 and σ_θ^2 . Compare to $\hat{\theta}_{ML}$.

(h) Find the conditional bias $b_{\hat{\theta}_{LMMSE}}(\theta)$ (and you will see that $\hat{\theta}_{LMMSE}$ is conditionally biased).

Even though the LMMSE estimator is conditionally biased, it will lead to better MSE performance.

(i) Show the following two expressions for both $\hat{\theta}_{ML}$ and $\hat{\theta}_{LMMSE}$:

$$\begin{aligned} R_{\theta\theta}^{-1} &= \left(E(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T \right)^{-1} = \frac{1}{\sigma_\theta^2} I_m + \frac{1}{\sigma_v^2} H^T H \\ &= \frac{1}{\sigma_v^2} \left(\left(2 + \frac{\sigma_v^2}{\sigma_\theta^2} \right) I_m - Z^2 - Z^{2T} \right) \end{aligned} \quad (6)$$

with $\begin{cases} \sigma_\theta^2 \text{ finite,} & \text{for } \hat{\theta}_{LMMSE} \\ \sigma_\theta^2 = \infty, & \text{for } \hat{\theta}_{ML} \end{cases}$

where Z is a shift matrix (ones on the first subdiagonal, zeros elsewhere) and $Z^{2T} = (Z^2)^T = (Z^T)^2$. The diagonal element (k, k) of $R_{\theta\theta}^{-1}$ gives the MSE for $\hat{\theta}_k$. Using the difference equation found earlier, it is possible to show that the estimation problem splits into two decoupled problems, one for the even k and one for the odd k . For each of these subproblems, the corresponding $H^T H$ can be written as the product of two bidiagonal triangular matrices plus a rank one modification. Using the matrix inversion lemma, it is then possible to find $(H^T H)^{-1}$ analytically. To find the inverse of $H^T H + \frac{\sigma_v^2}{\sigma_\theta^2} I_m$ requires a bit more work along similar lines. Using these results or using Matlab, it is possible to plot the MSE $E \hat{\theta}_k^2$ as a function of k . One can observe the following. First of all, the MSE is lower near the edges than in the middle of the packet of θ 's. Can you explain this? And for $\hat{\theta}_{LMMSE}$, the MSE remains bounded (and approaches a limit) as m increases, while for $\hat{\theta}_{ML}$, the MSE grows unboundedly.

Problem 5. Poisson Process - Deterministic Parameter Setting

Consider a network node where messages are passing and let the observation y_k be the time between the passage of two consecutive messages. y_k has an exponential distribution

$$f(y_k|\theta) = \begin{cases} \theta e^{-\theta y_k} & , y_k \geq 0 \\ 0 & , y_k < 0 \end{cases} \quad (7)$$

where the parameter θ represents the average frequency of messages passing by the node we are considering. We make n such observations y_1, \dots, y_n and, given θ , these observations y_k are i.i.d. Let $Y = [y_1 \cdots y_n]^T$ and we shall denote the sample average as $\bar{y} = \frac{1}{n} \sum_{k=1}^n y_k$.

- (a) Find the distribution $f(Y|\theta)$ and express it in terms of \bar{y} .
- (b) Find the Maximum Likelihood estimate $\hat{\theta}_{ML}$.
- (c) Show consistency of $\hat{\theta}_{ML}$ (hint: law of large numbers).

Problem 6. Variance Estimation of Zero-Mean Gaussian i.i.d. Variables

We observe the i.i.d. $y_k \sim \mathcal{N}(0, \sigma^2)$, $k = 1, \dots, n$ which have zero mean and unknown variance $\theta = \sigma^2$.

- (a) Find $\hat{\theta}_{ML}$. Interpret it in terms of the sample mean \bar{y}^2 .
- (b) Show that $\hat{\theta}_{ML}$ is unbiased.
- (c) Show that the MSE $E \tilde{\theta}_{ML}^2 = \frac{2}{n} \theta^2$. (Note: $E y_1^4 = 3\sigma^4$). Is $\hat{\theta}_{ML}$ consistent?

Now let θ be random with prior distribution

$$f(\theta) = \begin{cases} \lambda e^{-\lambda \theta} & , \theta \geq 0 \\ 0 & , \theta < 0 \end{cases}$$

and let $Y = [y_1 \cdots y_n]^T$.

- (d) Find the LMMSE estimator $\hat{\theta}_{LMMSE} = FY$ (hint: $E = E_\theta E_{Y|\theta}$). Compute the corresponding $E \tilde{\theta}^2$.

Since linear estimation in terms of Y doesn't work well, we propose to perform a fixed non-linear transformation on the data before using them in a linear estimator. Specifically, we shall work with $x_k = y_k^2$ and $X = [x_1 \cdots x_n]^T$.

- (e) Find the LMMSE estimator $\hat{\theta}_{LMMSE} = FX$ and express it in terms of \bar{y}^2 (with again $E = E_\theta E_{Y|\theta}$).
- (f) Compute the bias of $\hat{\theta}_{LMMSE} = FX$. Is $\hat{\theta}_{LMMSE}$ asymptotically unbiased?
- (g) Compute the conditional MSE $E_{Y|\theta} \tilde{\theta}^2$ of $\hat{\theta}_{LMMSE} = FX$ can compare to $\hat{\theta}_{ML}$. Then compute the average MSE.

Problem 7. Signal Amplitude Estimation in White Noise

Amplitude modulation consists essentially of multiplying a known signal $\{s_k\}$ with an amplitude θ that contains the information. Due to perturbations in the transmission medium, the received signal y_k can be written as

$$y_k = \theta s_k + v_k \quad , \quad (8)$$

where v_k is some zero-mean noise, and θ and s_k are deterministic quantities. The task of the receiver consists in determining the amplitude θ from a number of observations $\{y_1, y_2, \dots, y_n\}$. We can write (8) for $k = 1, \dots, n$ in vector form as

$$Y = \theta S + V \quad , \quad (9)$$

where $Y = [y_1 \ y_2 \ \dots \ y_n]^T$ and similarly for S and V . We consider $V \sim \mathcal{N}(0, C_{VV})$.

- (a) Give the maximum likelihood estimator $\hat{\theta}_{ML}$ (in terms of Y , S and C_{VV}).
- (b) Is $\hat{\theta}_{ML}$ biased? (justify your answer).
- (c) Compute the MSE $E\tilde{\theta}_{ML}^2$ (in terms of S and C_{VV}).
- (d) Consider $C_{VV} = \sigma_v^2 I_n$. Express now $\hat{\theta}_{ML}$, $b_{\hat{\theta}_{ML}}(\theta)$ and $E\tilde{\theta}_{ML}^2$ in terms of Y , S and σ_v^2 .
- (e) What is the condition on the signal s_k for $\hat{\theta}_{ML}$ to be (mean-square) consistent.
- (f) In addition to the statistical assumptions on V , $V \sim \mathcal{N}(0, \sigma_v^2 I_n)$, let θ now be random with prior distribution $\theta \sim \mathcal{N}(0, \sigma_\theta^2)$ and independent of V . Give $\hat{\theta}_{LMMSE}$ in terms of Y , S , σ_v^2 and σ_θ^2 .
- (g) Compute the conditional bias $b_{\hat{\theta}_{LMMSE}}(\theta)$. Is $\hat{\theta}_{LMMSE}$ conditionally (on θ) unbiased? Is $\hat{\theta}_{LMMSE}$ unbiased on the average (over θ)?
- (h) Compute the MSE $E\tilde{\theta}_{LMMSE}^2$. Compare $\hat{\theta}_{LMMSE}$ and $\hat{\theta}_{ML}$ in terms of MSE.

2 Spectrum Estimation

2.1 Non-Parametric Spectrum Estimation

Problem 8. Periodogram spectral leakage in the case of a sinusoid

Assume we have $y_k = A \cos(2\pi f_0 k + \phi)$, $k = 0, \dots, N-1$. We compute the periodogram (with a rectangular window) and we use the DFT to evaluate the periodogram at the frequencies $f_n = \frac{n}{N}$, $n = 0, 1, \dots, N-1$ (no zero padding). Explain why the (discrete) periodogram shows exactly one non-zero component (in the range $f \in [0, \frac{1}{2}]$) when f_0 is a multiple of $\frac{1}{N}$, and multiple components otherwise.

Problem 9. Bias and Variance of the (averaged) Periodogram on White Noise

We shall illustrate that the periodogram is an inconsistent estimator by examining the estimator at $f = 0$:

$$\hat{S}_{PER}(0) = \frac{1}{N} \left(\sum_{k=0}^{N-1} y_k \right)^2. \quad (10)$$

Let the y_k be i.i.d. and $y_k \sim \mathcal{N}(0, \sigma_y^2)$ (note that $E y_k^4 = 3\sigma_y^4$).

(a) What is the psdf $S_{yy}(f)$?

(b) Show that $\frac{1}{\sigma_y \sqrt{N}} \sum_{k=0}^{N-1} y_k \sim \mathcal{N}(0, 1)$.

(c) Find the mean and variance of $\hat{S}_{PER}(0)$. Does the variance go to zero as $N \rightarrow \infty$?

Consider the estimator

$$\hat{S}_{AVPER}(0) = \frac{1}{N} \sum_{n=0}^{N-1} \hat{S}_{PER}^{(n)}(0) \quad (11)$$

where

$$\hat{S}_{PER}^{(n)}(0) = y_n^2. \quad (12)$$

This estimator may be viewed as an averaged periodogram: the data record of length N is sectioned into blocks (in this case of length 1) and the periodograms for each block are averaged.

(d) Find the mean and variance of $\hat{S}_{AVPER}(0)$. Compare these results to those obtained in (c).

Problem 10. Periodogram Mean for Sinusoids in Colored Noise

The random process that we shall use is of the form

$$\begin{aligned}
 y_k &= s_k + v_k \\
 s_k &= A_1 \cos(2\pi f_1 k + \phi_1) + A_2 \cos(2\pi f_2 k + \phi_2) \\
 v_k &= H(q) e_k = \sum_{i=0}^4 h_i q^{-i} e_k = \sum_{i=0}^4 h_i e_{k-i} \\
 H(z) &= (1 - z^{-1})(1 + z^{-1})^3 = \sum_{i=0}^4 h_i z^{-i}
 \end{aligned} \tag{13}$$

where e_k is zero-mean unit variance white Gaussian noise which is filtered by $H(z)$ to obtain the colored noise v_k . The phases ϕ_i of the sinusoids are uniform over $[0, 2\pi]$ and independent of e_k and of each other, whereas the sinusoid amplitudes and frequencies are fixed.

- (i) Show that y_k is a stationary process.
- (ii) Compute the covariance sequence $r_{yy}(n)$.
- (iii) Compute the power spectral density function $S_{yy}(f)$.
- (iv) In the windowed periodogram

$$\hat{S}_{PER,w}(f) = \frac{c_{N,w}}{N} \left| \sum_{n=0}^{N-1} w_n y_n e^{-j2\pi f n} \right|^2 \tag{14}$$

the signal y_k is windowed by w_k before being used in the periodogram. The window is of duration N so that $w_n = 0$, $n < 0$ or $n \geq N$. $c_{N,w}$ is a normalization constant that depends on N and the window type. Show that we get for the mean of the windowed periodogram:

$$E \hat{S}_{PER,w}(f) = \frac{c_{N,w}}{N} |W(f)|^2 * S_{yy}(f) = \frac{c_{N,w}}{N} \mathcal{F} \{r_{ww}(k) r_{yy}(k)\} \tag{15}$$

where $r_{ww}(k) = w_k * w_{-k}$ and $W(f) = \mathcal{F} \{w_k\}$. Hint: introduce $w'_{-n} = w_n e^{-j2\pi f n}$.

- (v) How can $c_{N,w}$ be chosen so that the windowed periodogram becomes asymptotically unbiased?
- (vi) For the process y_k considered in (13), give an expression for the windowed periodogram mean that can be implemented straightforwardly in Matlab.

2.2 Parametric Spectrum Estimation

Problem 11. Autoregressive Processes

Let the AR(N) process y_k satisfy the recursion

$$y_k + a_1 y_{k-1} + \cdots + a_N y_{k-N} = e_k \quad (16)$$

where the e_k are i.i.d. with zero mean and variance σ_e^2 .

(a) Assume $N < 100$. Given $\{a_1, \dots, a_N, \sigma_e^2\}$, show how the correlation sequence $\{r_k, -100 \leq k \leq 100\}$ can be computed (give a realistic procedure that you are capable of programming in Matlab).

(b) Using your procedure in (a), compute $\{r_k, -\infty < k < \infty\}$ explicitly for an AR(1) process.

Problem 12. Linear Prediction and Triangular Matrix Factorization

(i) For a general stationary zero-mean process y_k , one can consider the orthogonalization scheme

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ A_{1,1} & 1 & 0 & \cdots \\ A_{2,2} & A_{2,1} & 1 & \ddots \\ \vdots & \vdots & \ddots & 0 \\ A_{n,n} & A_{n,n-1} & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} f_{0,0} \\ f_{1,1} \\ f_{2,2} \\ \vdots \\ f_{n,n} \end{bmatrix} \quad (17)$$

which we can rewrite as $LY = F$. When the forward prediction coefficients of the various orders minimize the MMSE, the orthogonality conditions $E(f_{i,i} y_{i-j}) = 0 \quad 1 \leq j \leq i$ are satisfied. Use this to show that

$$E(f_{i,i} f_{j,j}) = \sigma_{f,i}^2 \delta_{ij} \quad (18)$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. From (17) and (18), show that

$$LR_{n+1}L^T = D = \text{diag}\{\sigma_{f,0}^2, \sigma_{f,1}^2, \dots, \sigma_{f,n}^2\} \quad (19)$$

(ii) Show that (19) leads to the Upper-Diagonal-Lower Triangular factorization of R_{n+1}^{-1} :

$$R_{n+1}^{-1} = L^T D^{-1} L \quad (20)$$

and show also that

$$\det R_{n+1} = \prod_{i=0}^n \sigma_{f,i}^2 \quad (21)$$

(iii) Consider now an AR(1) process y_k satisfying $y_k + a y_{k-1} = e_k$ where e_k is white noise with variance σ^2 . Begin by finding $\sigma_{f,0}^2 = E(y_k^2)$ in terms of σ^2 and a . Then compute L , D , $\det R_{n+1}$ and R_{n+1}^{-1} for this AR(1) process.

Problem 13. Linear Prediction of a Constant Signal in White Noise

Consider the following random signal

$$y_k = s_k + v_k \quad . \quad (22)$$

The signal s_k is constant and random. This means that $s_k = C \quad \forall k$, where C is a random variable with zero mean and variance σ_s^2 . The signal v_k is a white noise with (zero mean and) variance σ_v^2 , and is independent of C .

(i) Give the autocorrelation function, $r_{yy}(i) = E(y_k y_{k-i})$, in terms of σ_s^2 and the parameter $\rho = \sigma_v^2 / \sigma_s^2$.

(ii) Express the autocorrelation matrix, R_n , as the sum of a diagonal matrix and a constant matrix.

(iii) Consider linear prediction of y_k of order n . Compute the prediction coefficients $A_{n,i}$ as a function of n and ρ . (Hint: use the matrix inversion lemma to invert R_n).

(iv) Give an expression for the predicted value $\hat{y}_k = y_k - f_{n,k}$ and interpret the result obtained.

(v) Compute the prediction error variance $\sigma_{f,n}^2$ and study its evolution as a function of n . Interpret the result for the limiting case as $n \rightarrow \infty$.

(vi) Consider the power spectral density estimate thus obtained by autoregressive modeling

$$\hat{S}_{yy}(f) = \frac{\sigma_{f,n}^2}{|A_n(f)|^2} \quad . \quad (23)$$

Compute $\hat{S}_{yy}(0)$. Is $\hat{S}_{yy}(0)$ a good estimator of σ_s^2 ?

(vii) Compute $\hat{S}_{yy}(\frac{1}{2})$ and $\lim_{n \rightarrow \infty} \hat{S}_{yy}(\frac{1}{2})$. As n gets large, is $\hat{S}_{yy}(f)$ a good estimator for $S_{yy}(f)$ for f near $\frac{1}{2}$ (i.e. for f far away from 0)?