Asymptotic bias

Assumptions:

- $w(\cdot)$ is a Lipschitz continuous function on $[-1, 1]$
- $\int_{\mathbb{R}} w(\alpha) \, d\alpha = 1$
- $\int_{\mathbb{R}} w(\alpha) \alpha \, d\alpha = 0$
- $\int_{\mathbb{R}} w(\alpha) \alpha^2 \, d\alpha \neq 0$

Then we have for the bias of $\hat{f}^{(T)}$

$$
\text{bias}\left( \hat{f}^{(T)}(\lambda) \right) = \int_{\mathbb{R}} \int_{\mathbb{R}} w^{(T)}(\lambda - \mu) \Phi_2^{(T)}(\mu - \eta) \left[ f(\eta) - f(\lambda) \right] \, d\eta \, d\mu \\
= \int_{\mathbb{R}} w^{(T)}(\lambda - \mu) \left[ f(\mu) - f(\lambda) \right] \, d\mu + O\left( \frac{\log T}{T} \right) \\
= \int_{\mathbb{R}} w(\mu) \left[ f(\lambda - b_T \mu) - f(\lambda) \right] \, d\mu + O\left( \frac{\log T}{T} \right) \\
= \int_{\mathbb{R}} w(\mu) \left[ - f'(\lambda) b_T \mu + \frac{1}{2} f''(\lambda) b_T^2 \mu^2 + o(b_T^2) \right] \, d\mu + O\left( \frac{\log T}{T} \right) \\
= \frac{b_T^2}{2} \int_{\mathbb{R}} w(\mu) \mu^2 \, d\mu f''(\lambda) + o(b_T^2) + O\left( \frac{\log T}{T} \right)
$$

Asymptotic bias of $\hat{f}^{(T)}$:

$$
\text{bias}\left( \hat{f}^{(T)}(\lambda) \right) \doteq \frac{b_T^2 \nu_2}{2} f''(\lambda)
$$

where

$$
\nu_2 = \int_{\mathbb{R}} w(\alpha) \alpha^2 \, d\alpha.
$$

Asymptotically

- the bias due to smoothing dominates the bias due to tapering
- the bias depends on the smoothing window only through $\nu_2$
- the bias depends on the curvature of $f$ and is larger at sharp peaks
Optimal bandwidth selection

A local distance measure: The Mean Square Error

\[
\text{MSE}(\hat{f}(T)(\lambda)) = \mathbb{E} \left[ \hat{f}(T)(\lambda) - f(\lambda) \right]^2 = \text{bias}(\hat{f}(T)(\lambda))^2 + \text{var}(\hat{f}(T)(\lambda)) = \frac{b_T^4}{4} \nu_2^2 f''(\lambda)^2 + \frac{1}{Tb_T} \frac{2\pi H_4}{H_2^2} \mu_2 f(\lambda)^2 + \text{terms of smaller order}
\]

where

\[
\mu_2 = \int_{\mathbb{R}} w(\alpha)^2 d\alpha, \quad \text{and} \quad \nu_2 = \int_{\mathbb{R}} w(\alpha) \alpha^2 d\alpha.
\]

The minimum of the asymptotic mean square error is attained at

\[
b_T(\lambda) = T^{-\frac{1}{5}} \left( \frac{2\pi H_4 \mu_2}{H_2^2 \nu_2^2} \right)^{\frac{1}{5}} \left( \frac{f(\lambda)}{f''(\lambda)} \right)^{\frac{2}{5}}.
\]

Global distance measures:
The Integrated Mean Square Error

\[
\text{IMSE}(\hat{f}(T)) = \int_{\Pi} \mathbb{E} \left( \hat{f}(T)(\lambda) - f(\lambda) \right)^2 d\lambda
\]

The minimum is attained at

\[
b_T = T^{-\frac{1}{5}} \left( \frac{2\pi H_4 \mu_2}{H_2^2 \nu_2^2} \int_{\Pi} f(\lambda)^2 d\lambda \right)^{\frac{1}{5}} \int_{\Pi} f''(\lambda)^2 d\lambda.
\]

The Integrated Mean Square Percentage Error

\[
\text{IMSPE}(\hat{f}(T)) = \int_{\Pi} \mathbb{E} \left( \frac{\hat{f}(T)(\lambda)}{f(\lambda)} - 1 \right)^2 d\lambda
\]

The minimum is attained at

\[
b_T = T^{-\frac{1}{5}} \left( \frac{4\pi^2 H_4 \mu_2}{H_2^2 \nu_2^2} \right)^{\frac{1}{5}} \left( \int_{\Pi} \left( \frac{f''(\lambda)}{f(\lambda)} \right)^2 d\lambda \right)^{-\frac{1}{5}}.
\]
Optimal bandwidth selection

Problem: Optimal bandwidth $b_T$ depends on unknown quantities $f$ and $f''$.

Idea: Use preliminary estimates to estimate optimal bandwidth.

Algorithm: (for optimal locally adaptive bandwidth $b_T(\lambda)$)

- Choose preliminary bandwidth $b_0$.
- Replace in expression for $b_T(\lambda)$ the ratio $\frac{f''(\lambda)}{f(\lambda)}$ by $|(\log f)''(\lambda)| + |(\log f)'(\lambda)|^2 + 1$.
- Estimate derivatives by
  
  $$(\log f)'(\lambda) = \frac{1}{b_0^2} \int_{\Pi} w_1(\frac{\lambda - \mu}{b_0}) \log I^{(T)}(\mu) d\mu$$
  
  $$(\log f)''(\lambda) = \frac{1}{b_0^3} \int_{\Pi} w_2(\frac{\lambda - \mu}{b_0}) \log I^{(T)}(\mu) d\mu$$

  where $w_1(\cdot)$ and $w_2(\cdot)$ are kernels such that

  $$\int_{\mathbb{R}} w_1(\alpha) d\alpha = \int_{\mathbb{R}} w_2(\alpha) d\alpha = \int_{\mathbb{R}} w_2(\alpha) \alpha d\alpha = 0$$

  and

  $$\int_{\mathbb{R}} w_1(\alpha) \alpha d\alpha = 1, \quad \int_{\mathbb{R}} w_2(\alpha) \alpha^2 d\alpha = 2.$$

- Compute $\hat{b}_T$ by substituting estimates for $(\log f)'$ and $(\log f)''$.
- To increase stability, the local bandwidth $\hat{b}_T(\lambda)$ can be smoothed.
- Compute final estimate $\hat{f}^{(T)}(\lambda)$ using $\hat{b}_T(\lambda)$.
- For computing the globally optimal bandwidth $b_T$, we can estimate $f$ and $f''$ directly (using preliminary bandwidth).
Optimal bandwidth selection

Example: ARMA(18,6) process
Optimal bandwidth selection

Example: AR(10) process
Optimal bandwidth selection

Example: ARMA(12,12) process
Nonparametric high resolution spectral estimation

Let $\mathcal{X}_T$ be the class of all fourth-order stationary processes $X$ of the form

$$X(t) = \sum_{j \in \mathbb{Z}} a(j) Y(t - j), \quad t \in \mathbb{Z}$$

such that

$$f_X(\lambda) = \frac{\prod_{j=1}^{r_1} \left( \frac{1}{T^2} + (\lambda - \lambda_1 j)^2 \right)^{s_1 j}}{\prod_{j=1}^{r_2} \left( \frac{1}{T^2} + (\lambda - \lambda_2 j)^2 \right)^{s_2 j}} f_Y(\lambda)$$

with $|\lambda_{ij} - \lambda_{i'j'}| \geq 2\delta_0$, and $f_Y(\lambda)$ satisfies

$$C^{-1} \leq f(\lambda) \leq C, \quad \lambda \in [-\pi, \pi].$$

Suppose that $Y$ satisfies further regularity conditions (smoothness of $f$ and bounded fourth-order spectrum)

**High resolution property:**

An estimator $\hat{f}^{(T)}$ is a high resolution spectral estimator if

$$\sup_{X \in \mathcal{X}_T} \text{IMSPE}(\hat{f}^{(T)}) = O(T^{-\frac{4}{5}}).$$

Let $\hat{f}^{(T)}$ be a kernel spectral estimator

- with kernel $w^{(T)}(\alpha) = b_T(\lambda)^{-1} w(b_T(\lambda)^{-1} \alpha)$ and local adaptive bandwidth
  $$b_T(\lambda) = c \left( T^{-\frac{1}{2}} \left\{ |(\log f)'(\lambda)| + |(\log f)''(\lambda)|^2 + 1 \right\}^{\frac{5}{2}} \right),$$

- with data tapers $h^{(T)}(t) = h(t/T)$ of degree 2, that is,
  - $h$ is continuously differentiable for all $t \in [0, 1]$ and
  - $h$ is three times continuously differentiable for all $t \in [0, 1]$ except for finitely many points $p_1, \ldots, p_r$.

Then $\hat{f}^{(T)}$ is a high resolution spectral estimator.

*Reference:* Dahlhaus (1990), *Probability Theory and Related Fields*
Optimality property of the quadratic window

Substituting the locally optimal bandwidth
\[ b_T(\lambda) = T^{-\frac{1}{8}} \left( \frac{2\pi H_4 \mu_2}{H_2^2 \nu_2^2} \right)^{\frac{1}{8}} |f(\lambda)|^\frac{2}{5} \left( \frac{f''(\lambda)}{f''(\lambda)} \right)^{\frac{2}{5}}. \]

into the expression for the asymptotic mean square error, we obtain
\[ \text{MSE}(\hat{f}^{(T)}(\lambda)) = T^{-\frac{4}{5}} \left( \frac{5\pi H_4 \mu_2 \nu_2^{1/2}}{2 H_2^2} f''(\lambda)^2 f(\lambda) \right)^{\frac{4}{5}}. \]

Thus the asymptotic mean square error depends on the kernel window $w$ only through the constant $\mu_2 \nu_2^{1/2}$.

Aim: Minimize $C_w = \mu_2 \nu_2^{1/2}$ with respect to $w$

Let
\[ w_0(\alpha) = \frac{3}{4} (1 - \alpha^2) 1_{[-1,1]}(\alpha) \]
be the Bartlett-Priestley (quadratic) window and let $w$ be any other non-negative kernel function satisfying
\[ \int_{\mathbb{R}} w(\alpha) \, d\alpha = \int_{\mathbb{R}} w_0(\alpha) \, d\alpha = 1 \]
and
\[ \int_{\mathbb{R}} w(\alpha) \alpha^2 \, d\alpha = \int_{\mathbb{R}} w_0(\alpha) \alpha^2 \, d\alpha. \]
Then
\[ \int_{\mathbb{R}} w(\alpha)^2 \, d\alpha \geq \int_{\mathbb{R}} w_0(\alpha)^2 \, d\alpha, \]
that is, $C_w$ is minimized by $w_0$. 

Confidence interval for the peak frequency

Aim: Confidence interval for peak frequency $\theta$ of $f$:

$$f(\theta) > f(\lambda) \quad \text{for all } \lambda \neq \theta.$$ 

An estimator for $\theta$ is

$$\theta^{(T)} = \inf\{\lambda \in [0, \pi] : \hat{f}^{(T)}(\lambda) \geq \hat{f}^{(T)}(\mu) \text{ for all } \mu\}$$

Idea: Bootstrap in the time domain

The periodogram has asymptotic distribution

$$I^{(T)}(\lambda) \sim \begin{cases} \frac{1}{2} f(\lambda) \chi^2_2 & \text{if } \lambda \not\equiv 0 \mod \pi \\ f(\lambda) \chi^2_1 & \text{otherwise} \end{cases}$$

This suggests the following algorithm:

- Estimate the spectrum.
- Simulate $B$ bootstrap samples of the periodogram
  $$I^*(b)\left(\frac{2\pi j}{T}\right) = \hat{f}^{(T)}\left(\frac{2\pi j}{T}\right) \varepsilon^*(j), \quad j = 0, \ldots, T - 1$$
  with $\varepsilon^*(j) = \varepsilon^*(T - j) \sim \frac{1}{2} \chi^2_2$ for $j = 1, \ldots, \frac{T}{2} - 1$ and $\varepsilon^*(j) \sim \chi^2_1$ for $j = 0, \frac{T}{2}$.
- For each of the bootstrap periodograms reestimate the spectrum
  $$\hat{f}^*(b)\left(\frac{2\pi j}{T}\right) = \sum_{k \in \mathbb{Z}} w^{(T)}\left(\frac{2\pi (j - k)}{T}\right) I^*(b)\left(\frac{2\pi k}{T}\right)$$
  and the peak frequency $\hat{\theta}^*(b)$.
- Obtain confidence region from the empirical distribution of the bootstrap peak frequencies $\hat{\theta}^*(b)$, $b = 1, \ldots, B$.

References:

- Timmer et al. (1997), *Biometrical Journal*