Estimation of mean and autocovariance function

Ergodic property

Methods we shall look at for estimating quantities such as the autocovariance function will use observations from a single realization. Such methods are based on the strategy of replacing ensemble averages by their corresponding time averages.

Sample mean:

Given a stationary time series $X_1, X_2, \ldots, X_N$. Let

$$\bar{X} = \frac{1}{N} \sum X_t.$$

Then,

$$\mathrm{E}\{\bar{X}\} = \frac{1}{N} \sum_{t=1}^{n} \mathrm{E}\{X_t\} = \frac{1}{N} N\mu = \mu$$

so $\bar{X}$ is an unbiased estimator of μ. Hence, $\bar{X}$ converges to μ in mean square if

$$\lim_{N \to \infty} \mathrm{var}\{\bar{X}\} = 0.$$

Now,

$$\begin{aligned}
\mathrm{var}\{\bar{X}\} &= \mathrm{E}\{(\bar{X}-\mu)^2\} \\
&= \mathrm{E}\left\{\left(\frac{1}{N}\sum_{i=1}^{N}(X_i-\mu)\right)^2\right\} \\
&= \frac{1}{N^2}\sum_{t=1}^{N}\sum_{u=1}^{N}\mathrm{E}\{(X_t-\mu)(X_u-\mu)\} \\
&= \frac{1}{N^2}\sum_{t=1}^{N}\sum_{u=1}^{N} s_{u-t} \\
&= \frac{1}{N^2}\sum_{\tau=-(N-1)}^{N-1}\;\sum_{k=1}^{N-|\tau|} s_\tau \\
&= \frac{1}{N^2}\sum_{\tau=-(N-1)}^{N-1}(N-|\tau|)s_\tau \\
&= \frac{1}{N}\sum_{\tau=-(N-1)}^{N-1}\left(1-\frac{|\tau|}{N}\right)s_\tau\,.
\end{aligned}$$

The summation interchange merely swaps row sums for diagonal sums.

To make further progress we need the condition $\sum_{\tau=-\infty}^{\infty} |s_\tau| < \infty$.
By the Cesàro summability theorem, if $\sum_{\tau=-(N-1)}^{N-1} s_\tau$ converges to a limit as $N \to \infty$,

$$\left[\text{it must since} \quad \left|\sum_{\tau=-(N-1)}^{N-1} s_\tau\right| \leq \sum_{\tau=-(N-1)}^{N-1} |s_\tau| < \infty \;\; \forall\, N\right]$$

then $\sum_{\tau=-(N-1)}^{N-1} \left(1 - \frac{|\tau|}{N}\right) s_\tau$ converges to the same limit.
We can thus conclude that,

$$\begin{aligned}
\lim_{N\to\infty} N\text{var}\{\bar{X}\} &= \lim_{N\to\infty} \sum_{\tau=-(N-1)}^{N-1} \left(1 - \frac{|\tau|}{N}\right) s_\tau \\
&= \lim_{N\to\infty} \sum_{\tau=-(N-1)}^{N-1} s_\tau = \sum_{\tau=-\infty}^{\infty} s_\tau .
\end{aligned}$$

The assumption of absolute summability of $\{s_\tau\}$ also implies that $\{X_t\}$ has a purely continuous spectrum with sdf

$$S(f) = \sum_{\tau=-\infty}^{\infty} s_\tau e^{-i2\pi f\tau},$$

so that

$$S(0) = \sum_{\tau=-\infty}^{\infty} s_\tau .$$

Thus

$$\lim_{N\to\infty} N\text{var}\{\bar{X}\} = S(0),$$

i.e.,

$$\text{var}\{\bar{X}\} \approx \frac{S(0)}{N} \quad \text{for large } N,$$

and therefore, $\text{var}\{\bar{X}\} \to 0$. Note (i) that the convergence of $\bar{X}$ depends only on the spectrum at $S(0)$, i.e. at $f = 0$, and (ii) $\bar{X}$ is a consistent estimator for μ.

Autocovariance Sequence:

Now,

$$s_\tau = \mathrm{E}\{(X_t - \mu)(X_{t+\tau} - \mu)\}$$

so that a natural estimator for the acvs is

$$\hat{s}_\tau^{(u)} = \frac{1}{N-|\tau|} \sum_{t=1}^{N-|\tau|} (X_t - \bar{X})(X_{t+|\tau|} - \bar{X}) \quad \tau = 0, \pm 1, \ldots, \pm(N-1).$$

Note $\hat{s}_{-\tau}^{(u)} = \hat{s}_\tau^{(u)}$ as it should.

If we replace $\bar{X}$ by μ:

$$\begin{aligned} \mathrm{E}\{\hat{s}_\tau^{(u)}\} &= \frac{1}{N-|\tau|} \sum_{t=1}^{N-|\tau|} \mathrm{E}\{(X_t - \mu)(X_{t+|\tau|} - \mu)\} \\ &= \frac{1}{N-|\tau|} \sum_{t=1}^{N-|\tau|} s_\tau = s_\tau, \quad \tau = 0, \pm 1, \ldots, \pm(N-1). \end{aligned}$$

Thus, $\hat{s}_{\tau}^{(u)}$ is an unbiased estimator of s_{τ} <u>when μ is known</u>. (Hence the (u) – for unbiased). Most texts refer to $\hat{s}_{\tau}^{(u)}$ as unbiased – however, if μ is estimated by $\bar{X}$, $\hat{s}_{\tau}^{(u)}$ is typically a biased estimator of s_{τ}!!!

A second estimator of s_τ is typically preferred:

$$\hat{s}_\tau^{(p)} = \frac{1}{N} \sum_{t=1}^{N-|\tau|} (X_t - \bar{X})(X_{t+|\tau|} - \bar{X}) \quad \tau = 0, \pm 1, \ldots, \pm(N-1).$$

With $\bar{X}$ replaced by μ:

$$\mathrm{E}\{\hat{s}_\tau^{(p)}\} = \frac{1}{N} \sum_{t=1}^{N-|\tau|} s_\tau = \left(1 - \frac{|\tau|}{N}\right) s_\tau,$$

so that $\hat{s}_\tau^{(p)}$ is a biased estimator, and the magnitude of its bias increases as $|\tau|$ increases. Most texts refer to $\hat{s}_\tau^{(p)}$ as biased.

Why should we prefer the "biased" estimator $\hat{s}_\tau^{(p)}$ to the "unbiased" estimator $\hat{s}_\tau^{(u)}$?

[1] For many stationary processes of practical interest

$$\mathrm{mse}\{\hat{s}_\tau^{(p)}\} < \mathrm{mse}\{\hat{s}_\tau^{(u)}\},$$

where

$$\begin{aligned}
\mathrm{mse}\{\hat{s}_\tau\} &= \mathrm{E}\{(\hat{s}_\tau - s_\tau)^2\} \\
&= \mathrm{E}\{\hat{s}_\tau^2\} - 2s_\tau \mathrm{E}\{\hat{s}_\tau\} + s_\tau^2 \\
&= (\mathrm{E}\{\hat{s}_\tau^2\} - E^2\{\hat{s}_\tau\}) + E^2\{\hat{s}_\tau\} - 2s_\tau \mathrm{E}\{\hat{s}_\tau\} + s_\tau^2 \\
&= \mathrm{var}\{\hat{s}_\tau\} + (s_\tau - \mathrm{E}\{\hat{s}_\tau\})^2 \\
&= \text{variance} + (\text{bias})^2
\end{aligned}$$

[2] If $\{X_t\}$ has a purely continuous spectrum we know that $s_\tau \to 0$ as $|\tau| \to \infty$. It therefore makes sense to choose an estimator that decreases nicely as $|\tau| \to N-1$ (i.e. choose $\hat{s}_\tau^{(p)}$).

[3] We know that the acvs must be positive semidefinite, the sequence $\{\hat{s}_\tau^{(p)}\}$ has this property, whereas the sequence $\{\hat{s}_\tau^{(u)}\}$ may not.

A naïve non-parametric spectral estimator – the periodogram

Suppose the zero mean discrete stationary process $\{X_t\}$ has a purely continuous spectrum with sdf $S(f)$. We have,

$$S(f) = \sum_{\tau=-\infty}^{\infty} s_\tau e^{-i2\pi f\tau} \quad |f| \leq \frac{1}{2}.$$

With $\mu = 0$, we can use the biased estimator of s_τ:

$$\hat{s}_\tau^{(p)} = \frac{1}{N} \sum_{t=1}^{N-|\tau|} X_t X_{t+|\tau|}$$

for $|\tau| \leq N-1$, but not for $|\tau| \geq N$. Hence we could replace s_τ by $\hat{s}_\tau^{(p)}$ for $|\tau| \leq N-1$ and assume $s_\tau = 0$ for $|\tau| \geq N$.

Periodogram

Then a spectrum estimate could be

$$\begin{aligned}\hat{S}^{(p)}(f) = \sum_{\tau=-(N-1)}^{(N-1)} \hat{s}_\tau^{(p)} e^{-i2\pi f\tau} &= \frac{1}{N} \sum_{\tau=-(N-1)}^{(N-1)} \sum_{t=1}^{N-|\tau|} X_t X_{t+|\tau|} e^{-i2\pi f\tau} \\ &= \frac{1}{N} \sum_{j=1}^{N} \sum_{k=1}^{N} X_j X_k e^{-i2\pi f(k-j)} \\ &= \frac{1}{N} \left| \sum_{t=1}^{N} X_t e^{-i2\pi ft} \right|^2,\end{aligned}$$

where the summation interchange has merely swapped diagonal sums for row sums. $\hat{S}^{(p)}(f)$ defined above is known as the periodogram, and is defined over $[-1/2, 1/2]$.

Periodogram

Note that $\{\hat{s}_\tau^{(p)}\}$ and $\hat{S}^{(p)}(f)$ are a FT pair:

$$\{\hat{s}_\tau^{(p)}\} \longleftrightarrow \hat{S}^{(p)}(f)$$

(hence the (p) for periodogram), just like the population quantities

$$\{s_\tau\} \longleftrightarrow S(f).$$

Hence, $\{s_\tau^{(p)}\}$ can be written as

$$s_\tau^{(p)} = \int_{-1/2}^{1/2} \hat{S}^{(p)}(f) e^{i2\pi f\tau}\, df \quad |\tau| \le N-1.$$

Periodogram

If $\hat{S}^{(p)}(f)$ were an ideal estimator of $S(f)$ we would have

[1] $\mathrm{E}\{\hat{S}^{(p)}(f)\} \approx S(f) \quad \forall f.$

[2] $\mathrm{var}\{\hat{S}^{(p)}(f)\} \to 0$ as $N \to \infty$ and,

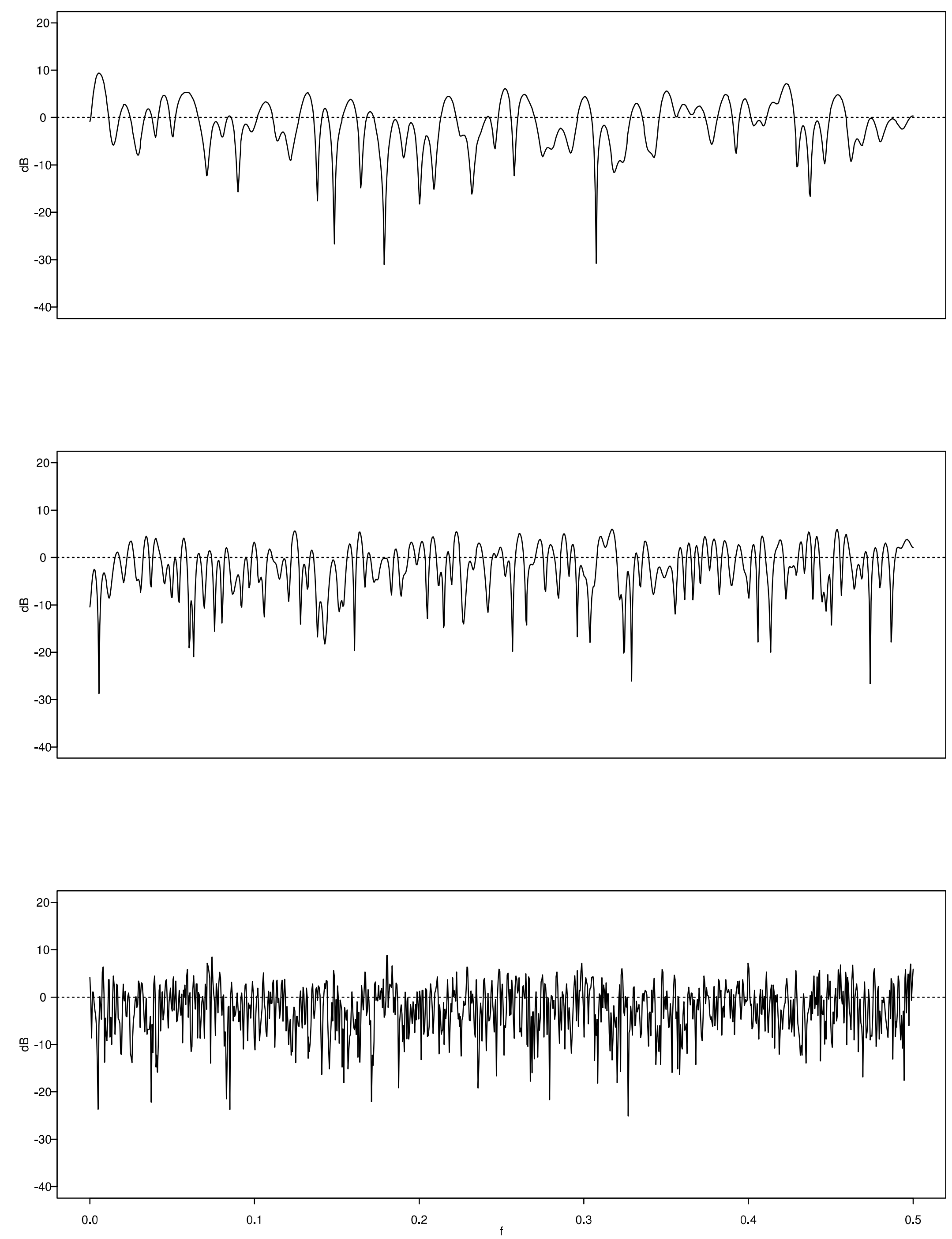
[3] $\mathrm{cov}\{\hat{S}^{(p)}(f), \hat{S}^{(p)}(f')\} \approx 0$ for $f \neq f'$.

We find that

[1] is a good approximation for some processes,

[2] is blatantly false (see Figure 26),

[3] holds if f and f' are certain distinct frequencies, namely, the Fourier frequencies $f_k = k/N \quad (\Delta t = 1)$.



Inconsistency of the periodogram. The plots show the periodogram (on a decibel scale) of a unit variance white noise process of length (from top to bottom) $N = 128, 256$ and 1024. The horizontal dashed line indicates the true sdf.

Periodogram - Statistical properties

We firstly look at the expectation in [1] (assuming $\mu = 0$).

$$\begin{aligned} \mathrm{E}\{\hat{S}^{(p)}(f)\} &= \sum_{\tau=-(N-1)}^{(N-1)} \mathrm{E}\{s_\tau^{(p)}\} e^{-i2\pi f\tau} \\ &= \sum_{\tau=-(N-1)}^{(N-1)} \left(1 - \frac{|\tau|}{N}\right) s_\tau e^{-i2\pi f\tau}. \end{aligned}$$

Hence, if we know the acvs $\{s_\tau\}$ we can work out from this what $\mathrm{E}\{\hat{S}^{(p)}(f)\}$ will be. We can obtain much more insight by considering:

$$\mathrm{E}\{|J(f)|^2\} \quad \text{where} \quad J(f) = \frac{1}{\sqrt{N}} \sum_{t=1}^{N} X_t e^{-i2\pi ft}, \quad |f| \leq \frac{1}{2}.$$

$[\hat{S}^{(p)}(f) = |J(f)|^2.]$

Periodogram - Statistical properties

We know from the spectral representation theorem that,

$$X_t = \int_{-1/2}^{1/2} e^{i2\pi f' t} \, dZ(f'),$$

so that,

$$\begin{aligned} J(f) &= \sum_{t=1}^{N} \left(\int_{-1/2}^{1/2} \frac{1}{\sqrt{N}} e^{i2\pi f' t} \, dZ(f') \right) e^{-i2\pi f t} \\ &= \int_{-1/2}^{1/2} \sum_{t=1}^{N} \frac{1}{\sqrt{N}} e^{-i2\pi (f - f') t} \, dZ(f') \end{aligned}$$

Then

$$
\begin{aligned}
\mathrm{E}\{\hat{S}^{(p)}(f)\} &= \mathrm{E}\{|J(f)|^2\} = \mathrm{E}\{J^*(f)J(f)\} \\
&= \mathrm{E}\left\{\int_{-1/2}^{1/2} \sum_{t=1}^{N} \frac{1}{\sqrt{N}} e^{i2\pi(f-f')t}\, dZ^*(f') \right. \\
&\qquad \left. \times \int_{-1/2}^{1/2} \sum_{s=1}^{N} \frac{1}{\sqrt{N}} e^{-i2\pi(f-f'')s}\, dZ(f'') \right\} \\
&= \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \sum_{t=1}^{N} \frac{1}{\sqrt{N}} e^{i2\pi(f-f')t} \sum_{s=1}^{N} \frac{1}{\sqrt{N}} e^{-i2\pi(f-f'')s} \mathrm{E}\{dZ^*(f')\, dZ(f'')\} \\
&= \int_{-1/2}^{1/2} \mathcal{F}(f-f') S(f')\, df',
\end{aligned}
$$

where $\mathcal{F}$ is Féjer's kernel defined by

$$
\mathcal{F}(f) = \left| \sum_{t=1}^{N} \frac{1}{\sqrt{N}} e^{-i2\pi ft} \right|^2 = \frac{\sin^2(N\pi f)}{N \sin^2(\pi f)}.
$$

Periodogram - Statistical properties

This result tells us that the expected value of $\hat{S}^{(p)}(f)$ is the true spectrum convolved with Féjer's kernel. To understand the implications of this we need to know the properties of Féjer's kernel:

[1] For all integers $N \geq 1, \mathcal{F}(f) \to N$ as $f \to 0$.

[2] For $N \geq 1$, $f \in [-1/2, 1/2]$ and $f \neq 0$, $\mathcal{F}(f) < \mathcal{F}(0)$.

[3] For $f \in [-1/2, 1/2]$, $f \neq 0$, $\mathcal{F}(f) \to 0$ as $N \to \infty$.

[4] For any integer $k \neq 0$ such that $f_k = k/N \in [-1/2, 1/2]$, $\mathcal{F}(f_k) = 0$.

[5] $\int_{-1/2}^{1/2} \mathcal{F}(f)\, df = 1$.

Periodogram - Statistical properties

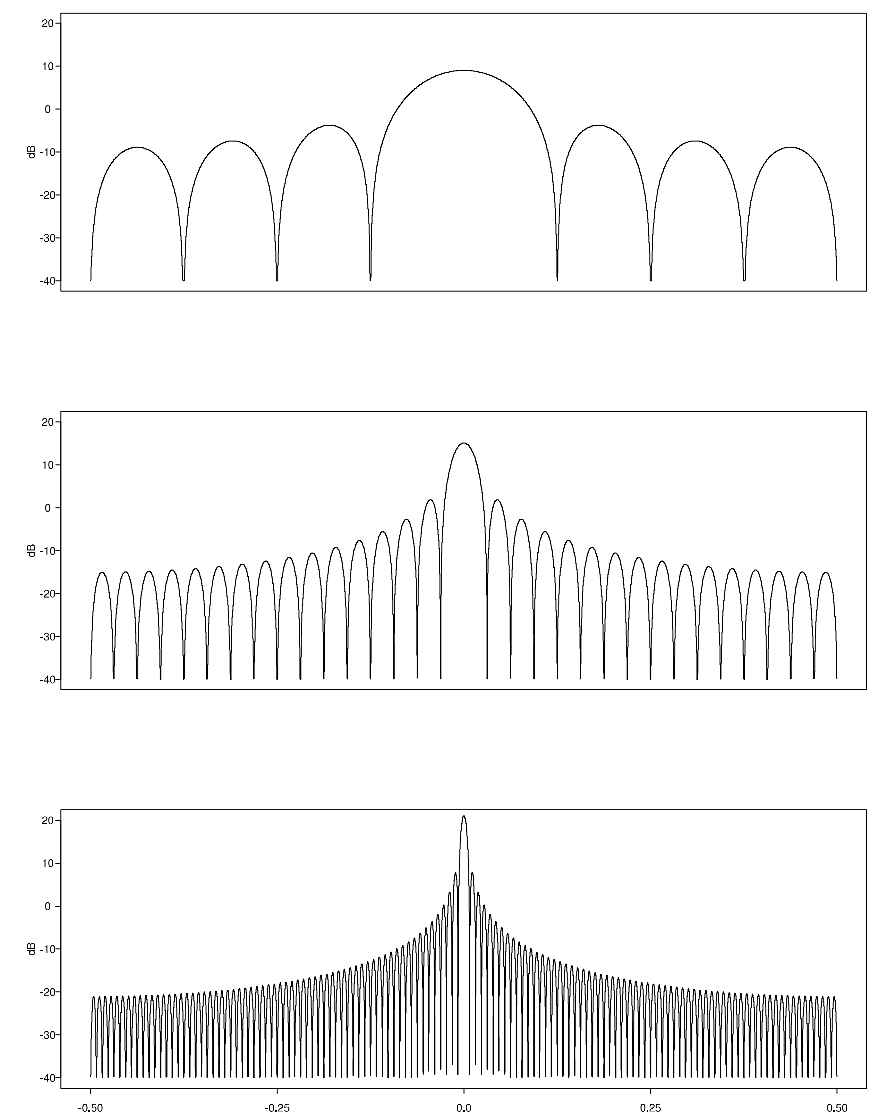


Figure 27 shows Féjer's kernel on a $10\log_{10}$ scale (dBs) for $N = 8, 32$ and 128. $\mathcal{F}(f)$ is symmetric about the origin and consists of a broad central peak ("lobe") and $N-2$ sidelobes which decrease as $|f|$ increases.

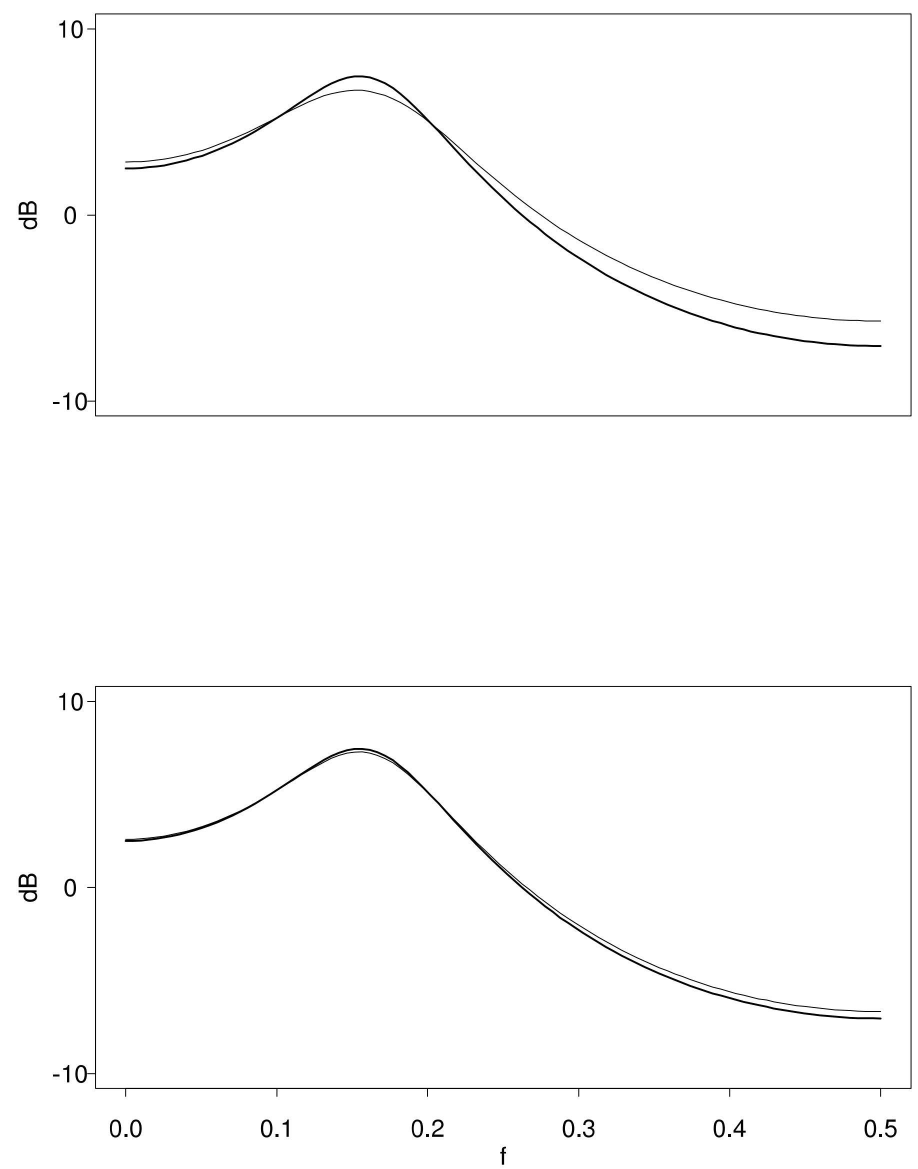
From [1], [3] and [5] it follows that as $N \to \infty$, $\mathcal{F}(f)$ acts as a Dirac δ function with an infinite spike at $f = 0$.

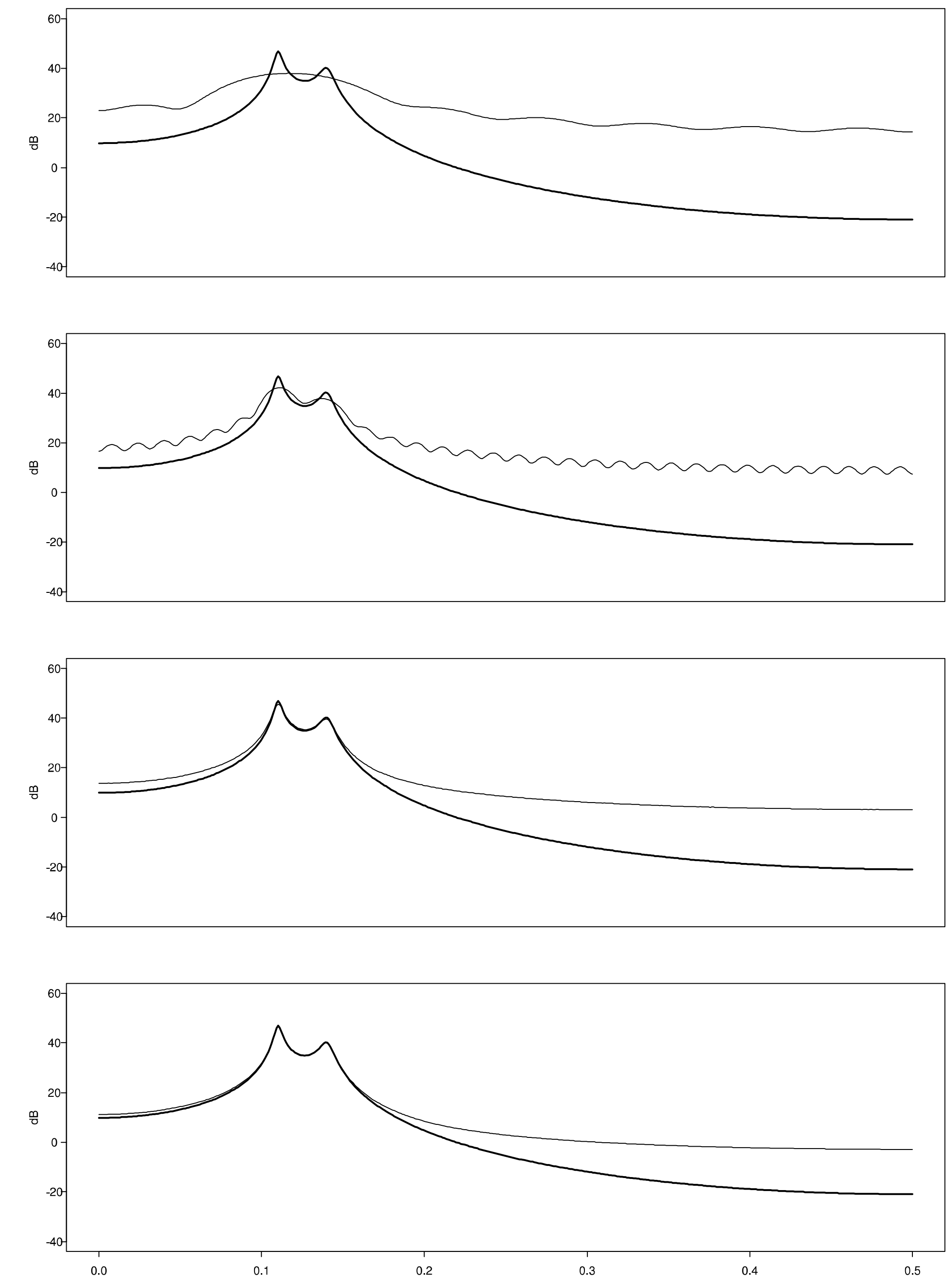
Periodogram - Statistical properties

So

$$\lim_{N\to\infty} \mathrm{E}\{\hat{S}^{(p)}(f)\} = \int_{-1/2}^{1/2} \delta(f - f')S(f')\,df' = S(f),$$

i.e., $\hat{S}^{(p)}(f)$ is *asymptotically* unbiased as an estimator of $S(f)$.





Figures 28 and 29 demonstrate this “sidelobe leakage” for two processes, the first with a dynamic range of 14 dB, the second with a dynamic range of 65 dB.

For a process with large dynamic range, defined as

$$10 \log_{10} \left(\frac{\max_f S(f)}{\min_f S(f)} \right),$$

since the expected value of the periodogram is a convolution of Féjer's kernel and the true spectrum, power from parts of the spectrum where $S(f)$ is large can "leak" via the sidelobes to other frequencies where $S(f)$ is small.

Bias reduction – Tapering

Much of the bias in the periodogram can be attributed to sidelobe leakage due to the presence of Féjer's kernel. Tapering is a technique which reduces the sidelobes associated with Féjer's kernel.

Let $X_1, X_2, \ldots, X_N$ be a portion of length N of a zero mean stationary process with sdf $S(f)$. We form the product $\{h_t X_t\}$ where $\{h_t\}$ is a sequence of real-valued constants called a <u>data taper</u> normalized so that

$$\sum_{t=1}^{N} h_t^2 = 1.$$

Define

$$J(f) = \sum_{t=1}^{N} h_t X_t e^{-i2\pi f t} \qquad |f| \leq 1/2.$$

By the spectral representation theorem,

$$X_t = \int_{-1/2}^{1/2} e^{i2\pi f' t}\, dZ(f'),$$

so that,

$$\begin{aligned} J(f) &= \sum_{t=1}^{N} h_t \left(\int_{-1/2}^{1/2} e^{i2\pi f' t}\, dZ(f') \right) e^{-i2\pi f t} \\ &= \int_{-1/2}^{1/2} \sum_{t=1}^{N} h_t e^{-i2\pi (f-f')t}\, dZ(f') \\ &= \int_{-1/2}^{1/2} H(f-f')\, dZ(f'), \end{aligned}$$

where,

$$H(f) = \sum_{t=1}^{N} h_t e^{-i2\pi f t} \quad \text{i.e.,} \quad \{h_t\} \longleftrightarrow H(f).$$

Let,

$$\hat{S}^{(d)}(f) = |J(f)|^2 = \left| \sum_{t=1}^{N} h_t X_t e^{-i2\pi f t} \right|^2 .$$

Then,

$$|J(f)|^2 = J^*(f)J(f) = \int_{-1/2}^{1/2} H^*(f-f')\, dZ^*(f') \int_{-1/2}^{1/2} H(f-f'')\, dZ(f''),$$

and hence,

$$\begin{aligned} \mathrm{E}\{\hat{S}^{(d)}(f)\} &= \mathrm{E}\{|J(f)|^2\} \\ &= \int_{-1/2}^{1/2} |H(f-f')|^2 S(f')\, df' \\ &= \int_{-1/2}^{1/2} \mathcal{H}(f-f') S(f')\, df', \end{aligned}$$

where $\mathcal{H}(f) = |H(f)|^2$, i.e.,

$$\mathcal{H}(f) = \left| \sum_{t=1}^{N} h_t e^{-i2\pi f t} \right|^2 .$$

Direct Spectral Estimator

A spectral estimator of the form of $\hat{S}^{(d)}(f)$ is called a direct spectral estimator (hence the (d)). Note, if $h_t = \frac{1}{\sqrt{N}}$ for $1 \leq t \leq N$, then

$$\hat{S}^{(d)}(f) = \hat{S}^{(p)}(f) \quad \text{and} \quad \mathcal{H}(f) = \mathcal{F}(f),$$

- i.e., $= \hat{S}^{(d)}(f)$ is the same as the periodogram,
- $\mathcal{H}(f)$ is the same as Fejer's kernel.

Choosing the taper

The key idea behind tapering is to select $\{h_t\}$ so that $\mathcal{H}(f)$ has much lower sidelobes that $\mathcal{F}(f)$.

Recall that $\mathcal{F}(f)$ corresponds to a rectangular taper

$$h_t = \begin{cases} \frac{1}{\sqrt{N}} & \text{for } 1 \leq t \leq N, \\ 0 & \text{otherwise.} \end{cases}$$

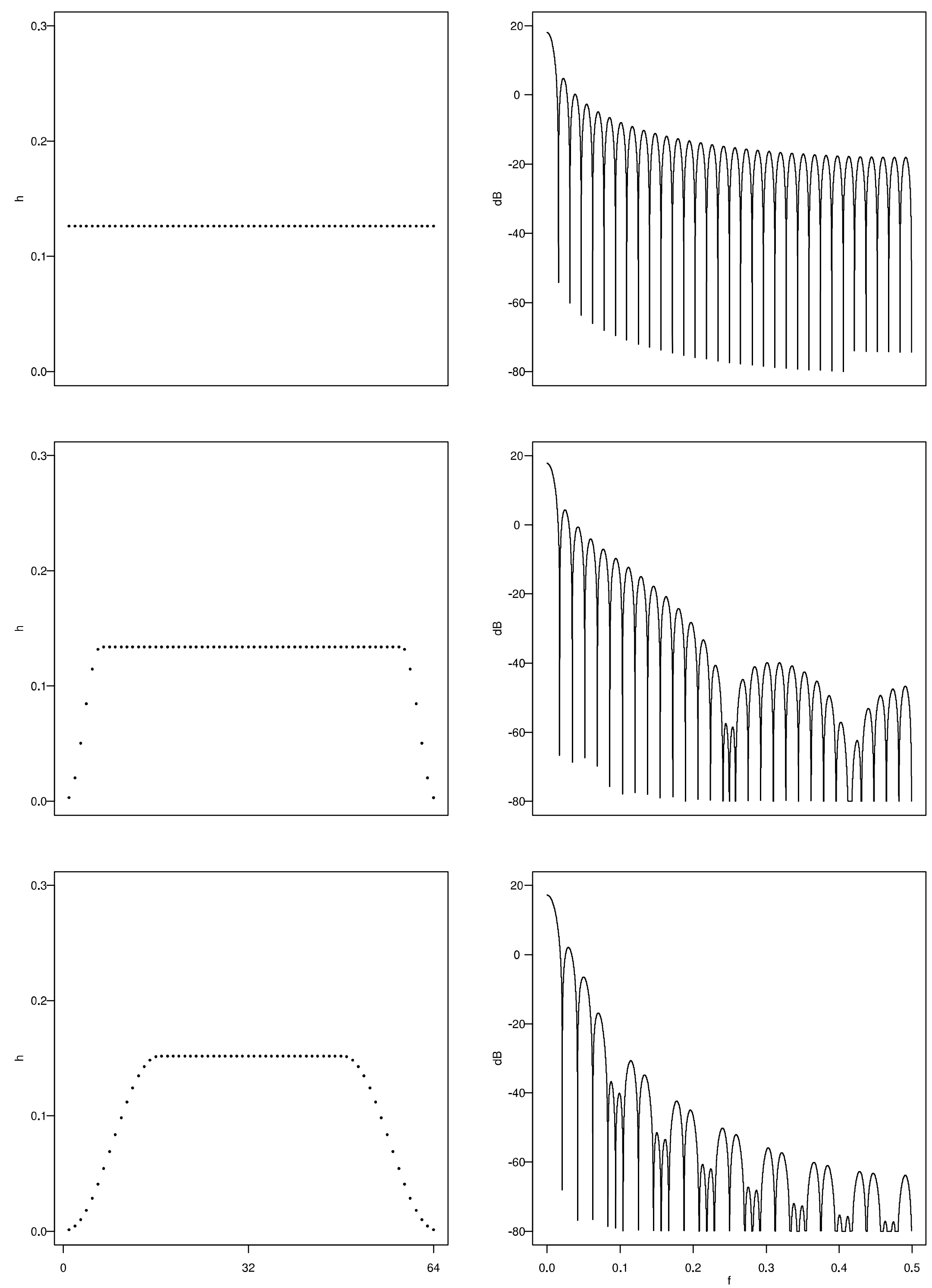
There is thus a sharp discontinuity between where the taper is "ON" $(1 \leq t \leq N)$ and where it is "OFF" $(N < t < 0)$. Tapering effectively creates a smooth transition at the ends of the data.

Choosing the taper

Figure 30 shows the effect of tapering on the shape of the spectral window $\mathcal{H}(f)$. The $p \times 100\%$ cosine taper is defined by

$$h_t = \begin{cases} \frac{C}{2}\left[1 - \cos\left(\frac{2\pi t}{\lfloor pN \rfloor + 1}\right)\right], & 1 \leq t \leq \frac{\lfloor pN \rfloor}{2}; \\ C, & \frac{\lfloor pN \rfloor}{2} < t < N + 1 - \frac{\lfloor pN \rfloor}{2}; \\ \frac{C}{2}\left[1 - \cos\left(\frac{2\pi(N+1-t)}{\lfloor pN \rfloor + 1}\right)\right], & N + 1 - \frac{\lfloor pN \rfloor}{2} \leq t \leq N, \end{cases}$$

where C is a normalizing constant that forces $\sum_{t=1}^{N} h_t^2 = 1$.

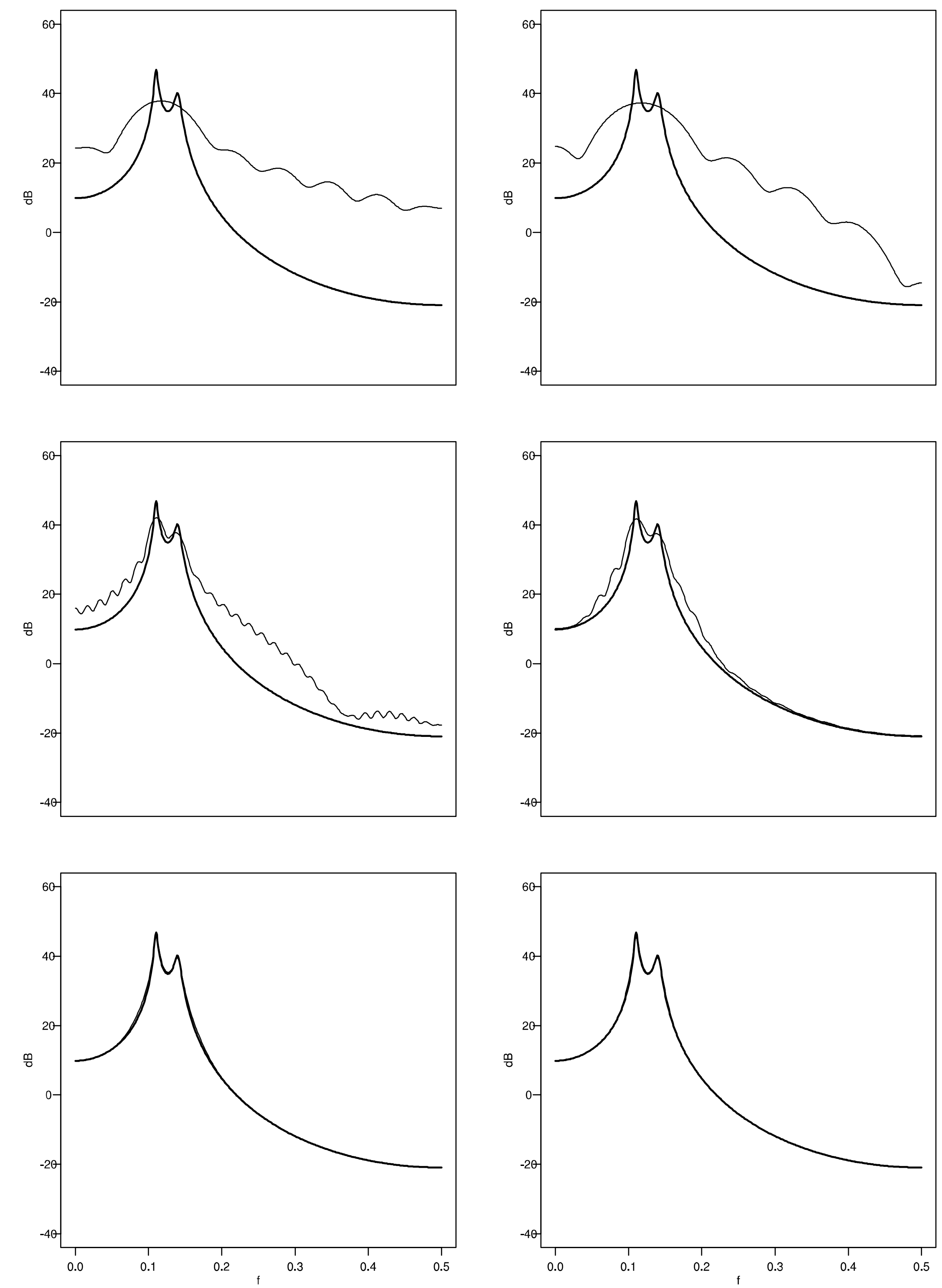


Different data tapers (left column) and associated spectral windows $\mathcal{H}(f)$ (right column), for $N = 64$.
The tapers are a rectangular taper (top), a 20% (middle) and 50% (bottom) split cosine bell taper.

As we perform more tapering, the main lobe of $\mathcal{H}(f)$ gets <u>wider</u>, but the sidelobes get <u>lower</u>. This means that the more tapering we perform:

Resolution of the spectrum DECREASES (bad!)
Sidelobe leakage DECREASES (good!).

Figure 31 demonstrates how the modification of the spectral window inherent in tapering reduces the sidelobe leakage at the expense of widening the main lobe (this results in smoothing bias) for the AR(4) process with high dynamic range.



Bias properties of direct spectral estimators for an AR(4) process with high dynamic range, using a 20% (left column) and 50% (right column) split cosine bell taper. The thick curves are the true sdf $S(f)$, while the thin curves are $\mathrm{E}\{\hat{S}^{(p)}(f)\}$ for sample sizes (from top to bottom) $N = 16, 64$ and 256.

Tapered AR(4) time series