

TIME SERIES ANALYSIS

Defⁿ: $\{X_t\}$ said to be COMPLETELY / STRICT STATIONARY if $\forall n \geq 1$, $\forall t_1, \dots, t_n \in T$ and $\forall s$ s.t. $t_1+s, \dots, t_n+s \in T$, joint cdf of $\{X_{t_1}, \dots, X_{t_n}\}$ is same as joint cdf of $\{X_{t_1+s}, \dots, X_{t_n+s}\}$.

(i.e. probabilistic structure of process invariant under shift in time)

Defⁿ: $\{X_t\}$ said to be SECOND-ORDER STATIONARY if $\forall n \geq 1$, $\forall t_1, \dots, t_n \in T$ and $\forall s$ s.t. $t_1+s, \dots, t_n+s \in T$, joint moments of order 1 and 2 of $\{X_{t_1}, \dots, X_{t_n}\}$ exist and equal to joint moments of $\{X_{t_1+s}, \dots, X_{t_n+s}\}$.

e.g. $E\{X_{t_1}\} = E\{X_{t_1+s}\}$, $E\{X_{t_1}^2\} = E\{X_{t_1+s}^2\}$. $E\{X_{t_1}X_{t_2}\} = E\{X_{t_1+s}X_{t_2+s}\}$

$\Rightarrow$ if second-order stationary then $E(X_t)$, $\text{var}(X_t)$ constant indep. of t .

Defⁿ: AUTOCOVARANCE SEQUENCE (ACVS), $S_\tau = \text{cov}\{X_t, X_{t+\tau}\}$
 $= \text{cov}\{X_0, X_\tau\}$.

Defⁿ 2: $\{X_t\}$ said to be SECOND-ORDER STATIONARY if $\forall t \in T$, $\forall \tau$ s.t. $t+\tau \in T$, $E(X_t)$ constant $\forall t$, and $S_\tau \equiv \text{cov}\{X_t, X_{t+\tau}\}$ finite and depends only on τ and not on t .

e.g. $S_0 = \sigma^2$, $S_{-\tau} = S_\tau$, $\text{ACS} = \rho_\tau = \frac{S_\tau}{S_0} = \frac{\text{cov}\{X_t, X_{t+\tau}\}}{\sigma^2}$.

N/B: $|S_\tau| \leq S_0$ so max. value of $S_\tau = S_0 = \sigma^2$.

EX: WHITE NOISE PROCESS, $\{X_t\}$ sequence of uncorrelated r.v.s

s.t. $E\{X_t\} = \mu$, $\text{var}(X_t) = \sigma^2 \forall t$ and $S_\tau = \begin{cases} \sigma^2 & \tau = 0 \\ 0 & \tau \neq 0 \end{cases}$.

Different realisations of white noise can be obtained for different distributions.

EX: q -th ORDER MOVING AVERAGE PROCESS, $MA(q)$:

$$X_t = \mu - \theta_{0q} \varepsilon_t - \theta_{1q} \varepsilon_{t-1} - \dots - \theta_{qq} \varepsilon_{t-q} = \mu - \sum_{j=0}^q \theta_{jq} \varepsilon_{t-j}$$

where $\theta_{0q} = -1$ and $\{\varepsilon_t\}$ zero-mean white noise process w/ variance σ^2 .

e.g. $MA(1) \rightarrow X_t = \varepsilon_t + \theta \varepsilon_{t-1}$

For MA(q), assume $E(X_t) = \mu = 0$.

$$\begin{aligned} \text{Then } S_\tau &= \text{cov}\{X_t, X_{t+\tau}\} = E(X_t \cdot X_{t+\tau}) \quad \text{only non-zero when } t-j = t+\tau-k \\ &= \sum_{j=0}^q \sum_{k=0}^q \theta_j \theta_k E\{\varepsilon_{t-j} \varepsilon_{t+\tau-k}\} \\ &= \sigma_\varepsilon^2 \sum_{j=0}^{q-\tau} \theta_j \theta_{j+\tau} \quad (k=j+\tau) \rightarrow \text{not depend on } t. \end{aligned}$$

so for MA(q) process, $S_\tau = \begin{cases} \sigma_\varepsilon^2 \sum_{j=0}^{q-|\tau|} \theta_j \theta_{j+|\tau|} & , |\tau| \leq q \\ 0 & , |\tau| > q \end{cases}$
 ↳ so MA(q) stationary

EX: p -th ORDER AUTOREGRESSIVE PROCESS, AR(p):

$$X_t = \phi_{1p} X_{t-1} + \phi_{2p} X_{t-2} + \dots + \phi_{pp} X_{t-p} + \varepsilon_t = \sum_{j=1}^p \phi_{jp} X_{t-j} + \varepsilon_t$$

$\{\varepsilon_t\}$ zero mean white noise process and unlike MA(q), parameters $\{\phi_{kp}\}$ must satisfy certain conditions for $\{X_t\}$ to be stationary, so not all AR(p) processes stationary.

MA(∞) $\rightarrow X_t = \sum_{j=0}^{\infty} \phi_{11}^j \varepsilon_{t-j} = \phi_{11} X_{t-1} + \varepsilon_t$ where $X_{t-1} = \phi_{11} X_{t-2} + \varepsilon_{t-1}$ etc.
 ↳ $S_\tau = \phi_{11}^\tau S_0$.

EX: (p, q)-th AUTOREGRESSIVE-MOVING AVERAGE PROCESS, ARMA(p, q):

$$X_t = \underbrace{\phi_{1p} X_{t-1} + \dots + \phi_{pp} X_{t-p}}_{\text{AR}(p)} + \underbrace{\varepsilon_t - \theta_{1q} \varepsilon_{t-1} - \dots - \theta_{qq} \varepsilon_{t-q}}_{\text{MA}(q)}$$

so: AR(p) = ARMA($p, 0$), MA(q) = ARMA($0, q$).

ARMA has $p + q + 1$ parameters $\rightarrow$ prone to overfitting

$\uparrow$ AR $\uparrow$ MA $\wedge$ σ_ε^2

EX: p -th AUTOREGRESSIVE CONDITIONALLY HETEROSCEDASTIC, ARCH(p):

assume $\{X_t\}$ has variance which changes through time, $X_t = \sigma_t \varepsilon_t$.

X_t is ARCH(p) if it satisfies prev. and: $\sigma_t^2 = \text{var}\{X_t | X_{t-1}, \dots, X_{t-p}\}$

e.g. ARCH(1) $\rightarrow \sigma_t^2 = \alpha + \beta_{11} X_{t-1}^2$, assume $\{X_t\}$ zero-mean, then:
 defining $V_t = X_t^2 - \sigma_t^2 \Rightarrow X_t^2 = \alpha + \beta_{11} X_{t-1}^2 + V_t$.

TREND ADJUSTMENT $\rightarrow$ represent simple linear trend by $\alpha + \beta t$

so take $X_t = \underbrace{\alpha + \beta t}_{\text{want to remove}} + Y_t$.

① Estimate α and β by least squares and use residuals $\hat{Y}_t = X_t - \hat{\alpha} - \hat{\beta}t$
 ↳ benefit is attempt to recover Y_t directly; but drawback its only an estimate of Y_t so will contain errors.

② Take First differences $\rightarrow X_t^{(1)} = X_t - X_{t-1} = \beta + Y_t - Y_{t-1} = \beta + Y_t^{(1)}$
 N/B: if Y_t statⁿary, so is $Y_t^{(1)}$.

In case of linear trend, differencing again: $X_t^{(2)} = X_t^{(1)} - X_{t-1}^{(1)} = Y_t^{(1)} - Y_{t-1}^{(1)} = Y_t^{(2)}$
 ↳ so effect of $\alpha + \beta t$ completely removed.

defⁿ: difference operator, $\rightarrow \Delta = (1 - B)$ where $BX_t = X_{t-1}$ $\rightarrow$ back shift operator

so: $X_t^{(d)} = \Delta^d X_t = \Delta^d Y_t$, $\Delta^2 X_t = \Delta X_t^{(1)} = X_t^{(1)} - X_{t-1}^{(1)} = X_t^{(2)}$.

incorporate into ARMA(p, q):

$$X_t = \phi_{1p} X_{t-1} + \dots + \phi_{pp} X_{t-p} + \varepsilon_t - \theta_{1q} \varepsilon_{t-1} - \dots - \theta_{qq} \varepsilon_{t-q}$$

$$\Rightarrow \underbrace{(1 - \phi_{1p} B - \dots - \phi_{pp} B^p)}_{\Phi(B)} X_t = \underbrace{(1 - \theta_{1q} B - \dots - \theta_{qq} B^q)}_{\Theta(B)} \varepsilon_t \Rightarrow \Phi(B) X_t = \Theta(B) \varepsilon_t$$

EX: include differencing in ARMA; ARIMA(p, d, q) so d-th differenced process is ARMA(p, q).

SEASONAL ADJUSTMENT $\rightarrow X_t = v_t + Y_t$, v_t seasonal component
 Y_t zero-mean statⁿary process.
 'Box+ Jenkins', use operator $(1 - B^s)$ (where v_t has period s)

$$X_t^{(s)} = (1 - B^s) X_t = X_t - X_{t-s} = v_t + Y_t - v_{t-s} - Y_{t-s} = Y_t - Y_{t-s}$$

defⁿ: GENERAL LINEAR PROCESS $\rightarrow X_t = \sum_{k=0}^{\infty} g_k \varepsilon_{t-k}$ $\rightarrow$ seq. of constants satisfying $\sum g_k^2 < \infty$
 $\rightarrow$ white-noise process.

CLP is statⁿary since $E(X_t) = \sum_{k=0}^{\infty} g_k E(\varepsilon_{t-k}) = 0$.

$$\text{Cov}(X_t, X_{t+\tau}) = \sigma_\varepsilon^2 \sum_{k=0}^{\infty} g_k g_{k+\tau}$$

For statⁿarity, if can write AR or ARMA in ACP form then we know it is stationary.

$\Rightarrow$ For AR(p), $\Phi(B)X_t = \varepsilon_t \Rightarrow X_t = \Phi^{-1}(B)\varepsilon_t$.

If all roots of $\Phi(z)$ satisfy $|z| > 1 \Rightarrow$ AR(p) statⁿary.

(same for ARMA(p, q), all roots of $\Phi(z)$ must lie outside unit circle for statⁿarity)

N/B: MA(q) always stationary.

process invertible if it can be written in AR form

$\Rightarrow$ For MA(q), $X_t = \Theta(B)\varepsilon_t$, if all roots of $\Theta(z)$ satisfy $|z| > 1 \Rightarrow$ MA(q) is invertible. (same for ARM(p, q))

N/B: AR(p) always invertible.

Defⁿ: COMPLEX-VALUED JUMP PROCESS $\{z(f)\}$, $f \in [-1/2, 1/2]$:

$$dz(f) \equiv \begin{cases} z(f+df) - z(f) & 0 \leq f < 1/2 \\ 0 & f = 1/2 \\ dz^*(-f) & -1/2 \leq f < 0. \end{cases}$$

orthogonal = uncorrelated

orthogonal process if orthogonal increments $\Rightarrow dz(f), dz(f')$ uncorrelated.

Thm (SPECTRAL REPRESENTATION THM): $\{X_t\}$ zero-mean statⁿary process,

Then $\exists$ an orthogonal process $\{z(f)\}$ s.t. $X_t = \int_{-1/2}^{1/2} e^{2\pi i f t} dz(f)$.

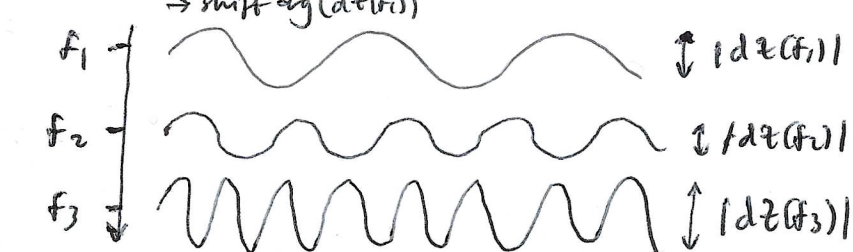
$\{z(f)\}$ has properties - $E\{dz(f)\} = 0$, $E\{|dz(f)|^2\} = dS^{(C)}(f)$

$\wedge S^{(C)}(f)$ is integrated spectrum.

For $2 f \neq f'$, $\text{cov}(dz(f), dz(f')) = 0$.

N/B: for complex r.v.s, $\text{cov}(W, Z) = E(W^* Z) - E(W^*) E(Z)$.

$\hookrightarrow$ spectral representation means we can write any discrete statⁿary process as an 'infinite' sum of complex exponentials at frequencies f w/ associated random amplitudes $|dz(f)|$ and random phases $\arg\{dz(f)\}$.



$\Rightarrow X_t$

The orthogonal increments property can be used to define relationships between a.c.v.s $\{S_\tau\}$ and integrated spectrum $S^{(I)}(f)$:

$$S_\tau = \int_{-1/2}^{1/2} e^{2\pi i f \tau} dS^{(I)}(f).$$

If in fact $S^{(I)}(f)$ differentiable with derivative $S(f)$, can write:

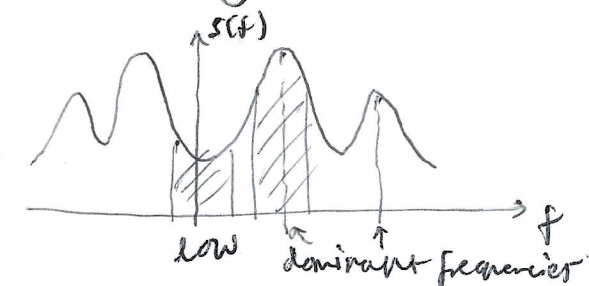
$$S_\tau = \int_{-1/2}^{1/2} S(f) e^{2\pi i f \tau} df. \rightarrow S(f) \text{ spectral density funct}^n.$$

$$\Rightarrow S(f) = \sum_{\tau=-\infty}^{\infty} S_\tau e^{-2\pi i f \tau} \quad (\text{so } \{S_\tau\}, S(f) \text{ are a FT pair})$$

$S^{(I)}$ similar to c.d.f. and $S(f)$ similar to p.d.f.

N/B: $S(-f) = S(f)$ since $S_\tau = S_{-\tau}$.

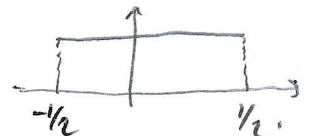
Interpreting SDF:



e.g. WHITE NOISE SPECTRUM:

$$S_\tau = \begin{cases} \sigma_\varepsilon^2 & \tau=0 \\ 0 & \tau \neq 0 \end{cases}$$

$$\Rightarrow S_\varepsilon(f) = S_0 = \sigma_\varepsilon^2$$



Any integrated spectrum can be written: $S^{(I)}(f) = S_1^{(I)}(f) + S_2^{(I)}(f)$

where $S_1^{(I)}$ absolutely continuous $\rightarrow S_1^{(I)}(f) = \int_{-1/2}^f S(f') df'$ and

$S_2^{(I)}$ step functⁿ with jumps at points $\{f_\ell: \ell=1, 2, \dots\}$.

① if $S_1^{(I)}(f) \geq 0$, $S_2^{(I)}(f) = 0 \rightarrow$ purely continuous spectrum, and

$S_\tau \rightarrow 0$ as $|\tau| \rightarrow \infty$. e.g. white noise / ARMA process

② if $S_1^{(I)}(f) = 0$, $S_2^{(I)}(f) \geq 0 \rightarrow$ pure discrete spectrum / line spectrum

e.g. harmonic process, $X_t = A \cos(2\pi f_0 t + \phi)$ random component.

and a.c.v.s NEVER drop down to 0.

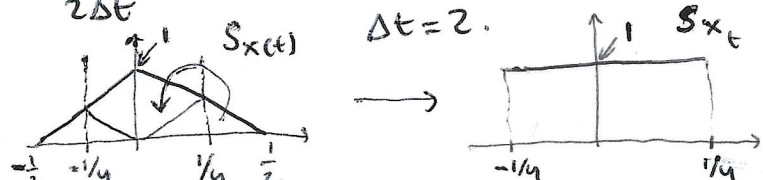
Sdf and a.c.v.s contain same amount of information but sdf proves to be more informative, far more sensitive + interpretable diagnostic or explanatory tool.

N/B: only necessary to have $f \in [-1/2, 1/2]$ because the values of $\exp\{2\pi i f t\}$ and $\exp\{2\pi i (f \pm k)t\}$, $k \in \mathbb{Z}$ are identical.

SAMPLING $\rightarrow$ e.g. $\Delta t = 1s$, then $f \in [-1/2, 1/2] Hz$, or observation every minute $\Delta t = 60s$, then $f \in [-1/120, 1/120] Hz$.
 General version of spectral representation $\rightarrow X_t = \int_{-f_N}^{f_N} e^{2\pi i f t} dZ(f)$.
 where $f_N = \frac{1}{2\Delta t}$, Nyquist frequency and $\frac{1}{\Delta t}$ sampling rate.
 • We can define a discrete time process through $X_t \equiv X(t_0 + t \Delta t)$,
 where t is integer, Δt sampling interval, $t = 0, \pm 1, \pm 2, \dots$

ALIASEING $\rightarrow S_{X_t}(f) = \sum_{k=-\infty}^{\infty} S_{X(t)}(f + \frac{k}{\Delta t})$ for $|f| \leq \frac{1}{2\Delta t}$.
 discrete time SDF (observed) $\uparrow$ cts time SDF (hidden process) $\leftarrow$ all frequencies that look like f for a sampling interval Δt .

$\Rightarrow$ the discrete time SDF at f is sum of continuous time SDF at frequencies $f \pm \frac{k}{\Delta t}$, $k = 0, 1, 2, \dots$
 i.e. fold $S_{X(t)}(f)$ about $f_N = \frac{1}{2\Delta t}$, and add.

i.e. if $S_{X(t)}$ close to zero for $|f| > \frac{1}{2\Delta t}$ then good correspondence between $S_{X_t}(f)$ and $S_{X(t)}(f)$ for $|f| \leq \frac{1}{2\Delta t}$.
 e.g. $S_{X(t)}(f) = \begin{cases} 1 - 2|f|, & |f| \leq 1/2 \\ 0, & \text{otherwise} \end{cases}$


LINEAR FILTERING:

Defⁿ: ~~analog~~ DIGITAL FILTER L transforms an input sequence $\{x_t\}$ into output sequence $\{y_t\}$.
 This is called linear time invariant (LTI) if:

$$L(\{\alpha x_t\}) = \alpha L(\{x_t\}), \quad L(\{x_{t_1} + x_{t_2}\}) = L(\{x_{t_1}\}) + L(\{x_{t_2}\}),$$

$$\text{if } L(\{x_t\}) = \{y_t\} \Rightarrow L(\{x_{t+\tau}\}) = \{y_{t+\tau}\}.$$

e.g. consider input sequence $e^{2\pi i f t}$, then $y_{f,t} = L\{e^{2\pi i f t}\}$.
 then $y_{f,t+\tau} = e^{2\pi i f \tau} y_{f,t} \Rightarrow y_{f,t} = e^{2\pi i f t} y_{f,0}$.
 constant indep. of time but depend on f , $g(f)$.

$\Rightarrow L\{e^{2\pi i f t}\} = e^{2\pi i f t} g(f)$ transfer functⁿ of L , $g: \mathbb{R} \rightarrow \mathbb{C}$.

Any LTI digital filter can be expressed in form: $L\{x_t\} = \sum_{u=-\infty}^{\infty} g_u x_{t-u}$

Any LTI can be written $L\{X_t\} = \sum_{n=-\infty}^{\infty} g_n X_{t-n}$ but comparing this with $L\{e^{2\pi i f t}\} = e^{2\pi i f t} g(f) \Rightarrow g(f) = \sum_{n=-\infty}^{\infty} g_n e^{-2\pi i f n}$ for $|f| \leq \frac{1}{2}$.
 so $\{g_n\} \leftrightarrow g(f)$ FT pair.

$$dZ_Y(f) = g(f) dZ_X(f) \text{ where } Y_t = L\{X_t\} = \sum_{n=-\infty}^{\infty} g_n X_{t-n}.$$

$$\Rightarrow \underline{S_Y(f) = |g(f)|^2 \cdot S_X(f)}.$$

e.g. Find SDF of MA(q) process by LTI digital filtering:

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \text{ so define } L\{\varepsilon_t\} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}.$$

$\therefore X_t = L\{\varepsilon_t\}$. To determine $g(f)$, input $e^{2\pi i f t}$.

$$\Rightarrow L\{e^{2\pi i f t}\} = e^{2\pi i f t} \underbrace{[1 - \theta_1 e^{-2\pi i f} - \dots - \theta_q e^{-2\pi i f q}]}_{g_\theta(f)} \quad g_0 = 1, g_1 = -\theta_1, \dots, g_q = -\theta_q.$$

$$\text{Since } S_X(f) = |g_\theta(f)|^2 \underbrace{S_\varepsilon(f)}_{\sigma^2} \Rightarrow S_X(f) = \sigma^2 |1 - \theta_1 e^{-2\pi i f} - \dots - \theta_q e^{-2\pi i f q}|^2.$$

$$\text{We can use this to compute } S_\tau \text{ e.g. since } S_\tau = \int_{-1/2}^{1/2} S_X(f) e^{2\pi i f \tau} df.$$

$$\text{CHAPTER 4, estimator } \bar{X} \rightarrow \text{var}(\bar{X}) = \frac{1}{N} \sum_{\tau=-(N-1)}^{N-1} \left(1 - \frac{|\tau|}{N}\right) S_\tau \approx \frac{S(0)}{N} \left(\sum_{\tau=-\infty}^{\infty} S_\tau \right) \text{ for large } N.$$

$$\text{ESTIMATOR for ACVS: } S_\tau = E[(X_t - \mu)(X_{t+\tau} - \mu)].$$

$$\text{If } \mu \text{ known, } \hat{S}_\tau^{(u)} = \frac{1}{N - |\tau|} \sum_{t=1}^{N-|\tau|} (X_t - \mu)(X_{t+|\tau|} - \mu), \tau = 0, \pm 1, \dots, \pm(N-1).$$

$\hookrightarrow$ this is unbiased.

$$\text{If } \mu \text{ not known, replace } \mu \text{ with } \bar{X}: \hat{S}_\tau^{(p)} = \frac{1}{N} \sum_{t=1}^{N-|\tau|} (X_t - \bar{X})(X_{t+|\tau|} - \bar{X}) \quad \sim \text{biased}.$$

$$\text{BUT } \text{mse}\{\hat{S}_\tau^{(p)}\} < \text{mse}\{\hat{S}_\tau^{(u)}\}.$$

$$\text{Recall } S(f) = \sum_{\tau=-\infty}^{\infty} S_\tau e^{-2\pi i f \tau}, \quad |f| \leq \frac{1}{2}.$$

$$\text{For zero-mean process, use } \hat{S}_\tau^{(p)} = \frac{1}{N} \sum_{t=1}^{N-|\tau|} X_t X_{t+|\tau|}, \text{ then:}$$

$$\hat{S}^{(p)}(f) = \frac{1}{N} \sum_{\tau=-(N-1)}^{N-1} \sum_{t=1}^{N-|\tau|} X_t X_{t+|\tau|} e^{-2\pi i f \tau} = \frac{1}{N} \left| \sum_{t=1}^N X_t e^{-2\pi i f t} \right|^2 \quad \xrightarrow{\text{PERIODOGRAM}}$$

$$\text{so } \{\hat{S}_\tau^{(p)}\} \leftrightarrow \{\hat{S}^{(p)}(f)\} \text{ are FT pair}$$

$$\Rightarrow \hat{S}_T^{(p)} = \int_{-1/2}^{1/2} \hat{S}^{(p)}(f) e^{2\pi i f T} df, \quad |T| \leq N-1.$$

$F * S(f)$ convolution.

consider expectation: $E[\hat{S}^{(p)}(f)] = \int_{-1/2}^{1/2} F(f-f') \cdot S(f') df'$

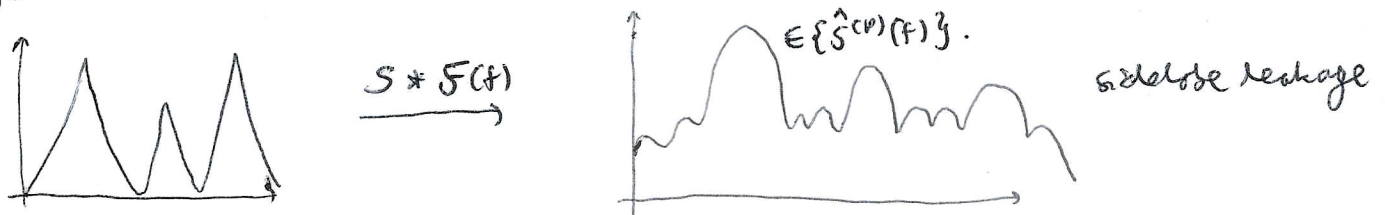
where F Féjer's kernel $\rightarrow F(f) = \frac{\sin^2(N\pi f)}{N\sin^2(\pi f)}$

so expected value of periodogram $\hat{S}^{(p)}(f)$ is true spectrum CONVOLVED with Féjer's kernel.

N/B: convolving with $\delta(x)$ preserves function. convolving w/ something like  blurs the function.

N/B: for $N \rightarrow \infty$, $F(f)$ acts as $\delta(x)$ so $\hat{S}^{(p)}(f)$ asymptotically unbiased.

SIDE-LOBE LEAKAGE: power from part of the spectrum where $S(f)$ is large can 'leak' via sidelobes to other frequencies where $S(f)$ is small.



much of bias in periodogram can be attributed to sidelobe leakage due to presence of Féjer's kernel. TAPERING technique reduces sidelobe.

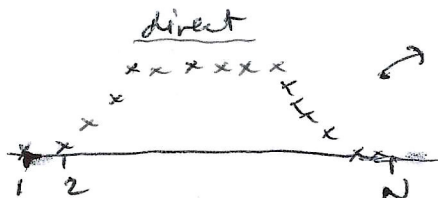
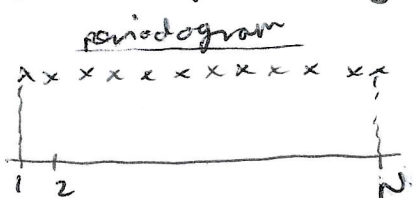
- Form the product $\{h_t X_t\}$ where $\{h_t\}$ sequence called data taper
s.t. $\sum_{t=1}^N h_t^2 = 1$ (normalised).

consider $\hat{S}^{(d)}(f) = \left| \sum_{t=1}^N h_t X_t e^{-2\pi i f t} \right|^2 \rightarrow$ direct spectral estimator.

here, $E[\hat{S}^{(d)}(f)] = \int_{-1/2}^{1/2} f f(f-f') S(f') df'$, where $f(f) = \left| \sum_{t=1}^N h_t e^{-2\pi i f t} \right|^2$

N/B: if $h_t = \frac{1}{\sqrt{N}} \forall 1 \leq t \leq N$ then $\hat{S}^{(d)}(f) = \hat{S}^{(p)}(f)$ and $f(f) = F(f)$.

Key idea of tapering to select $\{h_t\}$ so that $f(f)$ has lower sidelobes than $F(f)$.



$\rightarrow$ tapering creates a smooth transition at ends of data, instead of sharp discontinuity like in periodogram.

With more tapering, main lobe gets wider, sidelobes get lower.

⇒ resolution of spectrum DECREASES (bad!) but sidelobe leakage DECREASES (good!)

PARAMETRIC MODEL FITTING (for AR(p)):

$$x_t - \phi_1 x_{t-1} - \dots - \phi_p x_{t-p} = \varepsilon_t, \quad S(f) = \frac{\sigma_\varepsilon^2}{|1 - \phi_1 e^{-2\pi i f} - \dots - \phi_p e^{-2\pi i f p}|^2}$$

YULE-WALKER METHOD → start by multiplying defining eqⁿ by x_{t-k} :

$$x_t x_{t-k} = \sum_{j=1}^p \phi_j x_{t-j} x_{t-k} + \underbrace{\varepsilon_t x_{t-k}}_{E(\varepsilon_t x_{t-k}) = 0 \text{ since } x_{t-k} = \sum_{l=0}^{\infty} g_l \varepsilon_{t-k-l} \text{ (GLP)}}$$

Taking expectation of above: $S_k = \sum_{j=1}^p \phi_j S_{k-j}$
(use $S_{-k} = S_k$)

$$\Rightarrow \begin{cases} S_1 = \phi_1 S_0 + \dots + \phi_p S_{p-1} \\ \vdots \\ S_p = \phi_1 S_{p-1} + \dots + \phi_p S_0 \end{cases} \quad \left\{ \begin{array}{l} \hat{\Phi}_p = \hat{\Gamma}^{-1} \hat{\gamma}_p, \quad \hat{\Gamma} = \begin{pmatrix} S_0 & S_1 & \dots & S_{p-1} \\ S_1 & S_0 & \dots & S_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p-1} & S_{p-2} & \dots & S_0 \end{pmatrix} \\ \gamma = (S_1, \dots, S_p)^T, \quad \Phi = (\phi_1, \dots, \phi_p)^T. \end{array} \right.$$

→ Toeplitz matrix.

$$\text{For } \hat{\sigma}_\varepsilon^2 = \hat{S}_0 - \sum_{j=1}^p \hat{\phi}_j \hat{S}_j.$$

FORECASTING $\rightarrow$ predict value of X_{t+l} given $X_t, X_{t-1}, X_{t-2}, \dots$

ARMA(p, q): $\Phi(B) X_t = \Theta(B) \varepsilon_t$.

consider a forecast of X_{t+l} , $X_t(l) = \sum_{k=0}^{\infty} \pi_k X_{t-k}$. $\nearrow$ in AR form.

Write: $X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} = \Psi(B) \varepsilon_t \Rightarrow X_{t+l} = \Psi(B) \varepsilon_{t+l}$.

Then: $X_t(l) = \sum_{k=0}^{\infty} \pi_k \underbrace{X_{t-k}}_{\Psi(B) \varepsilon_{t-k}} = \sum_{k=0}^{\infty} \pi_k \Psi(B) \varepsilon_{t-k} = \underbrace{\Pi(B) \Psi(B)}_{\text{call this } \delta(B)} \varepsilon_t$.

want to learn the $\pi_0, \pi_1, \dots$ so the $\Pi(B)$.

and so $\rightarrow X_{t+l} = \underbrace{\sum_{k=0}^{l-1} \psi_k \varepsilon_{t+l-k}}_{\text{involves future } \varepsilon_t\text{'s so represents unpredictable part of } X_{t+l}} + \underbrace{\sum_{k=l}^{\infty} \psi_k \varepsilon_{t+l-k}}_{\text{depends only on past + present } \varepsilon_t, \text{ predictable part of } X_t}$.

Writing $X_t(l) = \sum_{k=0}^{\infty} \delta_k \varepsilon_{t-k}$ $\downarrow$ BUT we expect $X_t(l) = \sum_{k=l}^{\infty} \psi_k \varepsilon_{t+l-k} = \sum_{j=0}^{\infty} \psi_{j+l} \varepsilon_{t-j}$

$\Rightarrow \underline{\delta_k \equiv \psi_{k+l}}$ Since we don't know the 'unpredictable' part. minimizing of $E((X_{t+l} - X_t(l))^2)$ $\nearrow$ Variance of 'unpredictable' part.

(can be shown w/ least squares expectation)

choosing $\delta_k = \psi_{k+l} \rightarrow \sigma^2(l) = E[(X_{t+l} - X_t(l))^2] = \sigma_\varepsilon^2 \underbrace{\sum_{k=0}^{l-1} \psi_k^2}_{(l\text{-step prediction variance})}$

e.g. when $l=1$, $\delta_k = \psi_{k+1}$ so $X_t(1) = \sum_{j=0}^{\infty} \delta_j \varepsilon_{t-j} = \psi_1 \varepsilon_t + \psi_2 \varepsilon_{t-1} + \dots$ $\psi_0 = 1$

$X_{t+1} = \psi_0 \varepsilon_{t+1} + \psi_1 \varepsilon_t + \psi_2 \varepsilon_{t-1} + \dots \Rightarrow X_{t+1} - X_t(1) = \psi_0 \varepsilon_{t+1} = \varepsilon_{t+1}$

e.g. AR(1): $X_t = \underbrace{(1 - \phi_1 B)^{-1}}_{\Psi(B)} \varepsilon_t$, $\Psi^{-1}(B) = 1 - \phi_1 B$.

$\Psi(z) = 1 + \phi_1 z + \phi_1^2 z^2 + \dots = \psi_0 + \psi_1 z + \psi_2 z^2 + \dots \Rightarrow \psi_k = \phi_1^k$.

l -step prediction variance $\rightarrow \sigma^2(l) = \sigma_\varepsilon^2 \left(\sum_{k=0}^{l-1} \psi_k^2 \right) = \sigma_\varepsilon^2 \frac{(1 - \phi_1^{2l})}{(1 - \phi_1^2)}$ variance of forecast error.

If we wish to write $X_t(l)$ explicitly as functⁿ of $X_t, X_{t-1}, \dots$

then note: $X_t(l) = \sum_{k=0}^{\infty} \psi_{k+l} \varepsilon_{t-k} \quad (\varepsilon_k = \psi_{k+l})$

$\Rightarrow X_t(l) = \Psi^{(l)}(B) \varepsilon_t$. Assume $\Psi(z)$ statⁿary + invertible:

we can write $X_t = \Psi(B) \varepsilon_t \Rightarrow \varepsilon_t = \Psi^{-1}(B) X_t$.

$\therefore X_t(l) = \Psi^{(l)}(B) \varepsilon_t = \underbrace{\Psi^{(l)}(B) \Psi^{-1}(B)}_{G^{(l)}(B)} X_t$
 $= G^{(l)}(B) X_t$ where $G^{(l)}(z) = \Psi^{(l)}(z) \Psi^{-1}(z)$.

For the AR(1) example: $X_t(l) = G^{(l)}(B) X_t$:

$\Psi^{(l)}(z) = \sum_{k=0}^{\infty} \psi_{k+l} z^k = \sum_{k=0}^{\infty} \phi_1^{k+l} z^k$ (as before found - prev. page)

$\Psi^{-1}(z) = 1 - \phi_1 z \Rightarrow G^{(l)}(z) = (\phi_1^l + \phi_1^{l+1} z + \phi_1^{l+2} z^2 + \dots) (1 - \phi_1 z)$

$\hookrightarrow$ so $X_t(l) = \phi_1^l X_t$ as before. $= \phi_1^l$.

$\Rightarrow$ for AR(1) linear least squares predictor of X_{t+l} depends only on the most recent observation X_t and not involve $X_{t-1}, X_{t-2}, \dots$

for general AR(p) $\rightarrow X_t(l)$ depends only on last p observed values of $\{X_t\}$

BIVARIATE TIME SERIES: $\underline{X}_t = \begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix}$

defⁿ: $\{X_{1,t}\}, \{X_{2,t}\}$ are JOINTLY STATIONARY stochastic process if
 $\{X_{1,t}\}, \{X_{2,t}\}$ each separately second-order stationary and $\text{cov}\{X_{1,t}, X_{2,t+\tau}\}$
 is a functⁿ of τ only.

$S_{X_2 X_1, \tau} = \text{cov}\{X_{2,t}, X_{1,t+\tau}\} = E[(X_{2,t} - \mu_{X_2})(X_{1,t+\tau} - \mu_{X_1})]$.

$\nearrow$ ccvs

$S_{X_1 X_2, \tau} = S_{X_2 X_1, -\tau}$ but $S_{X_1 X_2, \tau} \neq S_{X_1 X_2, -\tau}$ (unlike aovs)

Given $X_{11}, X_{12}, \dots, X_{1N}; X_{21}, X_{22}, \dots, X_{2N}$, natural estimator for ccvs:

$$S_{X_1 X_2, \tau} = \begin{cases} \frac{1}{N} \sum_{t=1}^{N-\tau} (X_{1t} - \bar{X}_1)(X_{2,t+\tau} - \bar{X}_2) & , \tau = 0, 1, \dots, N-1 \\ \frac{1}{N} \sum_{t=1-\tau}^N (X_{1t} - \bar{X}_1)(X_{2,t+\tau} - \bar{X}_2) & , \tau = -1, -2, \dots, -(N-1) \end{cases}$$

CROSS-SPECTRA: $\{x_{1,t}\}, \{x_{2,t}\}$ have SDFs: $S_{x_j}(f) = \sum_{\tau=-\infty}^{\infty} S_{x_j,\tau} e^{-2\pi i f \tau}$
 $|f| \leq 1/2, j=1,2$

Then cross-spectra are:

$$S_{x_j x_k}(f) = \sum_{\tau=-\infty}^{\infty} S_{x_j x_k, \tau} e^{-2\pi i f \tau}; \quad |f| \leq \frac{1}{2}, j \neq k = 1, 2.$$

N/B: this is complex-valued since eqs are large symmetric so imag parts do not cancel.

inverse FT $\rightarrow S_{x_j x_k, \tau} = \int_{-1/2}^{1/2} S_{x_j x_k}(f) e^{2\pi i f \tau} df.$

spectral matrix: $S = \begin{pmatrix} S_{x_1}(f) & S_{x_1 x_2}(f) \\ S_{x_2 x_1}(f) & S_{x_2}(f) \end{pmatrix}.$

Defⁿ: COHERENCE: $\gamma_{x_j x_k}^2(f) = \frac{|S_{x_j x_k}(f)|^2}{S_{x_j}(f) \cdot S_{x_k}(f)}, \quad 0 \leq \gamma_{x_j x_k}^2(f) \leq 1.$

LINEAR FILTERING w/ NOISE: $X_{2,t} = \sum_{u=-K}^K g_u X_{1,t-u} + \eta_t.$

then: $S_{x_1 x_2}(f) = \sum_{\tau=-\infty}^{\infty} S_{x_1 x_2, \tau} e^{-2\pi i f \tau}$
 $= \sum_{u=-K}^K g_u \sum_{\tau=-\infty}^{\infty} S_{x_1, \tau-u} e^{-2\pi i f \tau}$
 $= \sum_{u=-K}^K g_u e^{-2\pi i f u} \sum_{\tau=-\infty}^{\infty} S_{x_1, \tau-u} e^{-2\pi i f (\tau-u)} = G(f) S_{x_1}(f).$

model can be written as: $dZ_{x_2}(f) = \sum_{u=-K}^K g_u e^{-2\pi i f u} dZ_{x_1}(f) + dZ_{\eta}(f)$

taking expectations $\rightarrow S_{x_2}(f) = |G(f)|^2 S_{x_1}(f) + S_{\eta}(f).$

BIVARIATE AR(p), $\begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \underline{X}_t = \Phi_1 \underline{X}_{t-1} + \dots + \Phi_p \underline{X}_{t-p} + \underline{\varepsilon}_t$
 $\nwarrow \text{2x2 matrices} \nearrow$

e.g. VAR(1) $\rightarrow x_{1,t} = \frac{1}{2} x_{1,t-1} + \frac{1}{10} x_{2,t-1} + \varepsilon_{1,t}$

$x_{2,t} = \frac{1}{2} x_{1,t-1} + \frac{1}{2} x_{2,t-1} + \varepsilon_{2,t}$

$\Rightarrow \begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \begin{pmatrix} 1/2 & 1/10 \\ 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \Rightarrow (I - \Phi B) \underline{X}_t = \underline{\varepsilon}_t$

for stationarity, check $\det(I - \Phi z) = \det\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1/2 & 1/10 \\ 1/2 & 1/2 \end{pmatrix} z\right) = 0$

if $|z| > 1$ for both roots $\Rightarrow$ stationary.