1 Introduction

Def. 1.1 *a* <u>*Time Series (TS)*</u> [1] X_T *is a sequence of r.v. that ordered in time T. i.e.,*

$$X_T = \{X_t, t \in T\}.$$

Def. 1.2 (sample ACVF) <u>*sample auto-covariance function*</u>

$$\widehat{\gamma}_X(h) := \frac{1}{n}\sum_{i=t}^{n-|h|}(X_{t+|h|} - \bar{X})(X_t - \bar{X}), \quad w.l \quad |h| < n.$$

Def. 1.3 (sample ACF) <u>*sample auto-correlation function*</u>

$$\widehat{\rho}_X(h) := \frac{\widehat{\gamma}_X(h)}{\widehat{\gamma}_X(0)}$$

Lemma 1.1 *If* $\gamma_{t,t+h} = \gamma_X(h), \forall\, t \in T$, *then* $\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}$.
Notice that $\{X_t\}$ *do not need to be weakly stationary.*

Def. 1.4 $f : \mathbb{Z} \to \mathbb{R}$ *is* <u>*non-negative definite*</u> *iff*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i f(i-j) a_j \geq 0, \forall\, n \in \mathbb{N}, \boldsymbol{a} \in \mathbb{R}^n.$$

prop. (Basic property of $\gamma(h)$)

1. $|\gamma(h)| \leq \gamma(0), \forall\, h$.
2. $\gamma(h) = \gamma(-h), \forall\, h$.
3. $\gamma(h)$ is non-negative definite.

   pf:

   $var(\sum a_i X_i) \geq 0 \Rightarrow \sum_i \sum_j a_i Cov(X_i, X_j) a_j \geq 0.$ ■

Thm 1.2 $\gamma : \mathbb{Z} \to \mathbb{R}$ *is ACVF of* X_T *iff symmetric & non-negative definite.*

2 Stationary

Def. 2.1 X_T *is* <u>*(weakly) stationary*</u> *iff*

- $\mu_i = \mu$, $\forall\, i$, *for some const.* $\mu \in R$.
- $\gamma_{i,i+h} = f(|h|), \forall\, i, h$, *for some fnc.* $f : \mathbb{R} \to \mathbb{R}$. [2]

Def. 2.2 X_T *is* <u>*strictly stationary*</u> *iff the joint distribution*

$$f_{X_i, X_{i+1}, \cdots, X_{i+h}} = f_{X_{i+k}, X_{i+k+1}, \cdots, X_{i+k+h}}, \forall\, i, k, h$$

Thm 2.1 *If* $\{X_t\}$ *is strictly stationary &* $\mu_i, \gamma_{i,j}$ *exists,* $\forall\, i, j$, *then* X_T *is weakly stationary.*

[1] In this chapter, T is always <u>discrete</u> and <u>equally spaced</u>.
[2] From now on, $\gamma_X(h) := f(|h|)$, and $\gamma(h)$ and $\rho(h)$ represent the Short cut of "ACVF" and "ACF" respectively.

E.X. 2.1 (strictly stationary $\nRightarrow$ weakly stationary) *Cauthy distribution* $F(x|x_o,\gamma)=\frac{1}{\pi}\arctan(\frac{x-x_o}{\gamma})+\frac{1}{2}$.

pf:

Let k = 2b, then

$$\begin{aligned} EX &= \int_{-\infty}^{\infty} xf(x)dx = \lim_{a,b\to\infty}\int_{-b}^{a} xf(x)dx \\ &= \lim_{k=a-b,b\to\infty}\int_{k}^{k+b} xf(x)dx + \cancelto{0}{\int_{-b}^{b} xf(x)dx} . \end{aligned}$$

That is, $EX\ DNE$.

Similarly, $EX^2 \propto \int_{-\infty}^{\infty}\frac{x^2}{1+x^2}dx = \int_{-\infty}^{\infty} 1dx - \int_{-\infty}^{\infty}\frac{1}{1+x^2}dx = DNE$.

By Holder Inequality, $\forall$ moments of it is DNE. ■

3 Processes

Def. 3.1 *$\{W_t\}$ is i.i.d noise iff $\{W_t\}$ is i.i.d & $E(W_t)=0, \forall\ t$.*

Def. 3.2 *$\{Z_t\}$ is White noise iff $E(Z_t)=0$ & $Cov(Z_t,Z_s)=\begin{cases}\sigma^2 & ,t=s\\ 0 & ,t\neq s\end{cases}$.*

noted as $WN(0,\sigma^2)$.

prop.

1. independent $\Rightarrow \gamma_\epsilon(h)=\begin{cases}0 & ,t\neq s\\ \sigma_i^2 & ,t=s\end{cases}$.
2. $i.i.d \Rightarrow$ independent + Homoscedasticity $\Rightarrow WN(0,\sigma^2)$.

   E.X. 3.1 ($WN(0,\sigma^2)\nRightarrow$ i.i.d.) *$Y_t=Z_tZ_{t-1}$, where $Z_t\sim N(0,\sigma^2)$.*

Def. 3.3 *$\{X_t\}$ is called random walk [1]. iff $X_t=\sum_{j=1}^{t}W_j$, w.l. $\{W_j\}$ be i.i.d noise.*

Def. 3.4 *$\{X_t\}$ is called simple random walk iff $(W_j+\frac{1}{2})\overset{i.i.d}{\sim} bern(\frac{1}{2})$.*

prop.

$Var(X_t)=i\sigma^2$, i.e., not stationary.

[1] origin: Brownian motion (布朗运动).

3.1 Linear process

Def. 3.5 *$\{X_t\}$ is linear process iff*

- $X_t = \sum_{j=-\infty}^{\infty} a_j Z_j \quad w.l \quad Z_j \sim WN(0, \sigma^2)$.
- $\sum_{j=-\infty}^{\infty} |a_j| < \infty$, *i.e., converge.*

That is, X_t is a L.C. of $\{Z_j\}$. Notice that a_j could be 0.

prop.

Let $\{Y_t\}$ is stationary, and $\{X_t\}$ is a L.C. of $\{Y_t\}$, and $\sum_{j=-\infty}^{\infty} |a_j| < \infty$. Then,

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_j a_k \gamma_Y(h+k-j).$$

If $Y_t \sim WN(0, \sigma^2)$, then $\gamma_X(h) = \sigma^2 \sum_{j=-\infty}^{\infty} a_j a_{j+h}$.

- q^{th} order moving average MA(q) process.

Def. 3.6 *$\{X_t\}$ is MA(1) process w.l coefficient a, σ^2 iff $X_t = Z_t + aZ_{t-1}$ w.l. $Z_t \sim WN(0, \sigma^2)$.*

Def. 3.7 *$\{X_t\}$ is MA(q) process iff*

$$X_t = \sum_{j=t-q}^{t} a_j Z_j \quad w.l \quad Z_j \sim WN(0, \sigma^2).$$

Def. 3.8 *$\{X_t\}$ is q-dependent iff $\forall\, k > q$, $X_t \perp\!\!\!\perp X_{t+k}$.*

prop.

$\{e_t\}$ is q-dependent $\Rightarrow$ for any fnc. $G : \mathbb{R}^q \to \mathbb{R}^q$, s.t., $X_t = G(e_t, \cdots, e_{t+q})$, $\{X_t\}$ is q-dependent.

Def. 3.9 *$\{X_t\}$ is q-correlated iff $\forall\, k > q$, $\rho_{X_t, X_{t+k}} = 0$.*

Thm 3.1 *$\{X_t\}$ can be written as a MA(q) process $\Leftrightarrow$ $\{X_t\}$ is q-correlated & stationary.*

pf:

($\Leftarrow$) Trivial, 略.
($\Rightarrow$) Too non-trivial, 略. ■

Def. 3.10 *$\{X_t\}$ is MA(∞) process w.l coefficient $\boldsymbol{a}$, σ^2 iff $X_t = \sum_{j=-\infty}^{t} a_j Z_j$ w.l. $Z_t \sim WN(0, \sigma^2)$ and $\sum_{j=-\infty}^{\infty} |a_j| < \infty$.*

- p^{th} order auto-regressive AR(p) process

Def. 3.11 *$\{X_t\}$ is a AR(1) process w.l coefficient a, σ^2 iff $X_t = Z_t + aX_{t-1}, \quad Z_t \sim WN(0, \sigma^2)$.*

Def. 3.12 $\{X_t\}$ *is causal process iff* $X_t = \sum_{j=-\infty}^{t} a_j Z_j$, *w.l.* $Z_t \sim WN(0, \sigma^2)$.
Notice that X_t *is a L.C. of historical pts. only.*

Lemma 3.2 *AR(1) is casual* $\Leftrightarrow$ *AR(1) is MA(*∞*)* $\Leftrightarrow |a| \leq 1 \Leftrightarrow \sum |a_j| < \infty$.

pf:

$X_t = \frac{1}{1-aB} Z_t = \sum_{j=-\infty}^{t} a^{t-j} Z_j$ is a kind of MA(∞), and MA(∞) is causal. ■

- The merger between AR(p) and MA(q) process

Def. 3.13 $\{X_t\}$ *is ARMA(1,1) process iff* $(1-aB)X_t = (1+\theta B)Z_t$, *w.l.* $\theta + a \neq 0$ [1] *and* $Z_t \sim WN(0, \sigma^2)$.

prop.

1. |a|< 1 $\Rightarrow$ ARMA(1,1) is MA(∞).
   pf:
   SInce |a| < 1, $X_t = \frac{1}{1-aB} Z_t = (1 + \frac{a+\theta}{1-aB}) Z_t = Z_t + \sum_{j=1}^{\infty} a^{j-1} Z_j$. [2] ■
2. |a|>1, $\exists$ a stationary non-causal solution to $(1-aB)X_t = (1+\theta B)Z_t$.
   XXX. (Unfinished)

Def. 3.14 $\{X_t\}$ *is invertable process iff* Z_t *can be written as a L.C. of* X_s, *where* $s \leq t$.

Lemma 3.3 $|\theta| < 1 \Rightarrow$ *ARMA(1,1) is invertable.*

pf:

$Z_t = \frac{1-aB}{1+\theta B} X_t = \cdots B^j X_t$. ■

4 Hypothesis tests of TS

4.1 i.i,d ?

Target:

- *before*: If true, the history is useless for forecasting.
- *after*: If false, $\hat{e}$ is not good enough.

1. **Non-rigorous test**

   XXX

   * *dis-adv*: Multiple testing problem [3].

[1] if $a = -\theta$, then $X_t = Z_t$, $WN(0, \sigma^2)$ は興味がありません!
[2] It is incorrect to write $(1 + \frac{a+\theta}{1-aB})$, but it is fine here.
[3] t-test v.s F-test.

2. **Ljung Box test**[1]

   $H_0 : (i.i.d \Rightarrow \rho = 0 \Rightarrow \rho^2 = 0)$.

$$\hat{Q}_{LB} = n \sum_{j=1}^{h} \hat{\rho}^2(j) \sim \chi^2_h.$$

$$modified - \hat{Q}_{LB} = n(n+2) \sum_{j=1}^{h} \frac{\hat{\rho}^2(j)}{n-j} \sim \chi^2_h.$$

   * *dis-adv*:
   When h is large, the # of Q is too small;
   When h is small, the lag h is not large enough.

   Rule of thumb: h $= 2\lfloor \ln n \rfloor$.

3. **McLeod & Li test**[2]

   $H_0 : (i.i.d \Rightarrow \rho = 0 \Rightarrow \rho^2 = 0)$.

$$\hat{Q}_{ML} = n(n+2) \sum_{j=1}^{h} \frac{\hat{\rho}^2_{\boldsymbol{x^2}}(j)}{n-j} \sim \chi^2_h.$$

4.2 Normal ?

Target:
XXX

1. **Rough test**: Q-Q plot.[3]

5 MSE & Linear forecast

Let $\{X_t\}$ be stationary.

Def. 5.1 *MSE of forecast is defined as* $E[X_{t+h} - f(X_h)]^2$, *where* $f(X_h)$ *is the forecast of* X_{t+h}.

Thm 5.1 $E(X_{t+h}|X_t)$ *minimize MSE.*

pf:

Let $f^*(X_t) = E(X_{t+h}|X_t)$, then

$$\begin{aligned} E(X_{t+h} - f)^2 &= E(X_{t+h} - f^* + f^* - f)^2 \\ &= E(X_{t+h} - f)^2 + 2E(X_{t+h} - f^*)(f^* - f) + E(f^* - f)^2. \end{aligned}$$

Recall that $E(Y) = E[E(Y|X)]$, then

$$\begin{aligned} 2E(X_{t+h} - f^*)(f^* - f) &= 2EE[(X_{t+h} - f^*)(f^* - f)|X_t] \\ &= E[(f^* - f)(E(X_{t+h}|X_t) - f^*)] = 0 \end{aligned}$$

$\blacksquare$

[1] modified Q_{LB} is better when $n \leq 100$, same when n is large.
[2] The ACF of x_t is not large enough, thus making it to be x^2.
[3] for STA457, this is enough.

5.1 Forecast the Future Given TS $\{X_t\}_{t=1}^n$.

Criterion:

Let $\boldsymbol{x} = (X_1, \cdots, X_n)^T$, $E(Y) = \mu$ and $E(X_t) = \mu_t$ for notation convenience. Then

$$\min\{\text{MSE of forecast}\} = \min_{(a_0, \boldsymbol{a})} E(Y - a_0 - a_1 X_1 - \cdots - a_n X_n)^2 = \min_{(a_0, \boldsymbol{a})} E(Y - a_0 - \boldsymbol{a} \cdot \boldsymbol{x})^2.$$

and view $X_0 = 1$, we have

$$\frac{\partial \mathcal{L}}{\partial a_t} = -2E(Y - a_0 - \boldsymbol{a} \cdot \boldsymbol{x})X_t = 0, \forall\ t \quad \Rightarrow \quad E(YX_t) = E(a_0 + \boldsymbol{a} \cdot \boldsymbol{x})X_t, \forall\ t.$$

Notice that when t = 0, define $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)$, we obtain

$$\mu = a_0 + \boldsymbol{a} \cdot \boldsymbol{\mu} \quad \Leftrightarrow \quad a_0 = \mu - \sum_{j=1}^n a_j \mu_j$$

Thus,

$$E(YX_t) = (\mu - \boldsymbol{a} \cdot \boldsymbol{\mu})\mu_t + E(\boldsymbol{a} \cdot \boldsymbol{x})X_t \quad \Rightarrow \quad E(YX_t) - \mu\mu_t = E(\boldsymbol{a} \cdot \boldsymbol{x})X_t - (\boldsymbol{a} \cdot \boldsymbol{x})\mu_t.$$

That is, $Cov(Y, X_t) = Cov(\boldsymbol{x}, X_t), \forall\ t$, or,

$$Cov(Y, \boldsymbol{x}) = \boldsymbol{Cov}(\boldsymbol{x}, \boldsymbol{x})\boldsymbol{a} \Leftrightarrow \gamma = \boldsymbol{\Gamma}\boldsymbol{a}.$$

If $\boldsymbol{\Gamma}$ is not singular, i.e., $det(\boldsymbol{\Gamma}) \neq 0$, then $\boldsymbol{a} = \boldsymbol{\Gamma}^{-1}\gamma$ is not only solvable, but also unique.

Thm 5.2 *The best linear forecast of Y based on* $\{X_t\}_{t=1}^n$ *is*

$$P(Y|\boldsymbol{x}) = a_0 + \boldsymbol{a} \cdot \boldsymbol{x},$$

where $a_0 = \mu - \boldsymbol{a} \cdot \boldsymbol{x}$ *and* $\boldsymbol{\gamma} = \boldsymbol{\Gamma}\boldsymbol{a}$.

Thm 5.3 *The MSE of forecast when predicting U from* $\{X_t\}_{t=1}^n$ *is* $Var(Y^*) - \boldsymbol{a} \cdot \gamma$, *where* $Y^* = Y - \mu$.

pf:

Let $Y^* = Y - \mu$. Then

$$\begin{aligned}
\text{MSE of forecast} &= E(Y - \mu + \boldsymbol{a} - \boldsymbol{a} \cdot \boldsymbol{x})^2 \\
&= E(Y - \mu)^2 - 2\boldsymbol{a} \cdot E(\boldsymbol{x})Y + E(\boldsymbol{a} \cdot \boldsymbol{x})^2 \\
&= Var(Y^*) - 2\boldsymbol{a} \cdot \boldsymbol{\gamma} + \boldsymbol{a} \cdot \boldsymbol{\Gamma}\boldsymbol{a} = Var(Y^*) - \boldsymbol{a} \cdot \boldsymbol{\gamma}.
\end{aligned}$$

■

prop.

1. $E(Y - P(Y|\boldsymbol{x}) = 0$, i.e., expected forecast error is 0.
2. $E(Y - P(Y|\boldsymbol{x})\boldsymbol{x} = \boldsymbol{0}$, i.e., forecast error is uncorrelated w.l predictors.
3. (Even if $\boldsymbol{\Gamma}$ is singular) $P(Y|\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x}$ is still unique.

   pf:

Assume that we have 2 best linear forecasters Q_1 and Q_2, s.t.,

$$Q_1 = a_0^1 + \boldsymbol{a}_n^1 \cdot \boldsymbol{x}$$
$$Q_2 = a_0^2 + \boldsymbol{a}_n^2 \cdot \boldsymbol{x}$$

Then, by prop. 2, we have

$$\begin{cases} E(Y - Q_1)\boldsymbol{x} = \boldsymbol{0} \\ E(Y - Q_2)\boldsymbol{x} = \boldsymbol{0} \end{cases} \Rightarrow E(Q_1 - Q_2)\boldsymbol{x} = \boldsymbol{0}.$$

Therefore,

$$E(Q_1 - Q_2)^2 = (a_0^1 - a_0^2)\cancelto{0}{E(Q_1 - Q_2)} + \boldsymbol{a} \cdot \cancelto{\boldsymbol{0}}{E(Q_1 - Q_2)\boldsymbol{x}} = 0.$$

That is, $Q_1 - Q_2 = 0$, w.l $Pr = 1$. ■

4. $P(*|\boldsymbol{x}) : \Omega \to \Omega$ is a linear operator [1] over field $\mathbb{R}$, s.t.,

$$Cov(Y, \boldsymbol{x}) = \boldsymbol{0} \Rightarrow P(Y|\boldsymbol{x}) = \mu \quad \& \quad P(X_t|\boldsymbol{x}) = \mu_t, \forall\, t.$$

5. (*Tower Law of Predictor*) $P(Y|\boldsymbol{u}_n) = P(P(Y|\boldsymbol{u}_n, \boldsymbol{v}_n)|\boldsymbol{u}_n)$, where $\boldsymbol{v}$ is r.v. s.t., $Cov(\boldsymbol{u}_n, \boldsymbol{v}_n) < \infty$. ***pf:***

   XXX ■

6 Inference - Given n obs. X_t

Note:

$$\boldsymbol{Cov(x, x)} = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n) \\ \gamma(-1) & \gamma(0) & \cdots & \gamma(n-1) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(-n) & \gamma(-(n-1)) & \cdots & \gamma(0) \end{bmatrix}$$

6.1 Mean μ

$$Var(\hat{\mu}_n) = E(\hat{\mu}_n - \mu)^2 = \frac{1}{n^2} E(X_t - \mu)^2 = \frac{1}{n^2} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \gamma(i - j)$$
$$= \frac{1}{n^2}(n\gamma(0) + (n-1) \cdot [\gamma(1) + \gamma(-1)] + (n-2) \cdot [\gamma(2) + \gamma(-2)] + \cdots)$$

Thm 6.1 (CLT[2]) *If $\{X_t\}$ is stationary & $\sum_{j=1}^{\infty} |\gamma(j)| < \infty$, then*

$$\sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow{D} \mathcal{N}(0, \sum_{h=-n}^{n} (1 - \frac{|h|}{n})\gamma(h)),$$

where $\gamma(h)$ is assumed to be known. [3]

[1] Cov is a linear operator.

[3] If $\gamma(h)$ is unknown, then

$$CI_{95\%} = \hat{\mu}_n \pm \frac{1.96}{\sqrt{n}} \sqrt{\hat{V}}, \quad w.l \quad \hat{V} = \sum_{h=-\sqrt{n}}^{\sqrt{n}} (1 - \frac{|h|}{n})\hat{\gamma}(h).$$

6.2 ACVF $\gamma(h)$ & ACF $\rho(h)$

$\hat{\boldsymbol{\rho}}_k := (\hat{\rho}(1), \hat{\rho}(2), \cdots, \hat{\rho}(k))^T$.

Thm 6.2 *If $\{X_t\}$ is stationary & $\sum_{j=0}^{\infty} \gamma(j) < \infty$, then $\hat{\boldsymbol{\rho}}_k \xrightarrow{d} \mathcal{N}(\boldsymbol{\rho}_k, \frac{1}{n}\boldsymbol{W}_k)$, where $\boldsymbol{W}_k$ is a $k \times k$ matrix, s.t.,*

$$w_{ij} = \sum_{k=1}^{\infty} [\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)] \cdot [\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)],$$

thus, $CI_{95\%} = \hat{\rho}(h) \pm \frac{1.96}{\sqrt{n}} \sqrt{w_{hh}}$.

7 Remove the trend

7.1 Remove the trend only

1. * Linear regression $X_t = \beta_0 + \beta_1 t + \epsilon_t$.

   $\Rightarrow \widehat{\epsilon}_t = X_t - \widehat{X}_t \perp\!\!\!\perp i$.

2. * Homonic regression $X_t = \sum_{j=1}^{k} [a_j(\cos \lambda_j t) + b_j(\sin \lambda_j t)] + \epsilon_t$.

   - λ_j are chosen manually as the potential frequency $\frac{2\pi}{T}$ (usually).
   - K is large enough.

   $\Rightarrow \widehat{\epsilon}_t = X_t - \widehat{X}_t \perp\!\!\!\perp i$.

3. * Exponential Smoothing $X_t = \mu_t + \epsilon_t$.

   $$\widehat{\mu}_t{}^{1} = \sum_{j=0}^{t-2} [\alpha(1-\alpha)^j X_{t-j}] + (1-\alpha)^{t-1} X_1.$$

   $\Rightarrow \widehat{\epsilon}_t = X_t - \widehat{\mu}_t$.

4. Moving average $X_t = \mu_t + \epsilon_t$.

   $$\widehat{\mu}_t = \frac{1}{2q+1} \sum_{j=t-q}^{t+q} X_j{}^{2},$$

   where q is manually[3] picked, usually $\lfloor n^{\frac{1}{3}} \rfloor$. $\Rightarrow \widehat{\epsilon}_t = X_t - \widehat{\mu}_t$.

5. Differencing[4] $X_t = \mu_t + s_t + \epsilon_t$.

   Def. 7.1 *Backward shift operator B is a operator s.t., B: $X_T \to X_T$, w.l. $x_t \to x_t$.*

   $\Rightarrow \Delta X_t{}^{5} = (I - B)X_t$.

[1]Esitmation of the trend. Weighted average of $\{X_t, X_{t-1}, \ldots, X_1\}$, where ω decaying exponentially & decaying speed $\uparrow$ as $\alpha \uparrow$.

[2]if $t - q < 0$ or $t + q > n + 1$, then let $x_j = x_1$, $x_j = x_n$ respectively.

[3]Non-trivial.

[4]dis: i) # data $\downarrow$; ii) loss the trend information.

[5]Assume that ΔX_t has no trend. Sometimes it still has, then do it again.

7.2 Remove trend & Seasonality

1. Estimation & Removal $X_t = \mu_t + s_t$(Seasonality) + ϵ_t.

$$\widehat{trend} \begin{cases} \widehat{\mu}_t = \frac{1}{2q+1} \sum_{j=t-q}^{t+q} x_j, \quad d = 2q+1(odd) \\ \\ \widehat{\mu}_t = \frac{1}{2q}[0.5x_{t-q} + \sum_{j=t-q}^{t+q} x_j + 0.5x_{t+q}], \quad d = 2q(even) \end{cases}$$

   $\Rightarrow \widehat{X_t} = X_t - \widehat{\mu}_t \rightarrow$ Only seasonal component left.

   $\widehat{seasonal}$: $\forall\ k \in \mathbb{N}, 0 \le k \le d$, $A_k := \{\widehat{X}_{k+d\cdot j} | j \in \mathbb{N}, k + d \cdot j \le n\} = \{\widehat{x}_k, \widehat{x}_{k+d}, \widehat{x}_{k+2d} \cdots\}$. $\widehat{s}_t :=$ average of obs. in A_k.

   Thus, $\forall\ i < d$, $\widehat{s}_t = \widehat{s}_{i-md}$, for some $m \in \mathbb{N}$.

   $\Rightarrow \widehat{\epsilon}_t = \widehat{X}_t - \widehat{s}_t$.

2. Differencing $X_t = \mu_t + s_t + \epsilon_t$.

$$\begin{aligned} Y_t &= (I - B_X^d)X_t = X_t - X_{t-d} = (\mu - \mu_{t-d} + \cancelto{0}{(s_t - s_{t-d})} + (\epsilon_t + \epsilon_{t-d}) \\ &= \widetilde{\mu}_t + \widetilde{\epsilon}_t. \end{aligned}$$

   $\Rightarrow \Delta Y_t = (I - B_Y)Y_t$.