Chapter 1 - Introduction

Time series: $\{X_t\}$ is a collection of random variables, where t is an index of time

In this class, $t \in S$ where S is a discrete set of indices. discrete-time time series

(plotting) Describe — concise summaries of the data

(model) Explain

probability model

$\{X_t\}$ If X_t are mutually independent

$P(X_1 \le x_1, X_2 \le x_2 \cdots X_n \le x_n) = \prod_{i=1}^{n} Pr(X_i \le x_i)$

most complex model

$P(X_1 \le x_1, X_2 \le x_2 \cdots X_n \le x_n)$

$= P(X_1 \le x_1) P(X_2 \le x_2 \mid X_1 \le x_1) \cdots P(X_n \le x_n \mid X_1 \le x_1 \cdots)$

(more uncertainty) Predict

Some Simple Time Series Models

semi-parametric model

Don't specify the pdf, cdf of $\{X_t\}$

But specify $E(X_t)$, $Cov(X_t, X_{t+j})$

① Some zero-mean model

e.g. iid noise

Let $E(X_t) = 0 \quad \forall t$ and $P(X_1 \le x_1 \cdots X_n \le x_n) = \prod_{i=1}^{n} P(X_i \le x_i) = \prod_{i=1}^{n} F(x_k)$

where $F(\cdot)$ is the cumulative distribution function.

In this model, there is no dependence between observations. In particular, for all $h \ge 1$ and all $x, x_1 \cdots x_n$,

$$P[X_{n+h} \le x \mid X_1 = x_1 \cdots X_n = x_n] = P[X_{n+1} \le x]$$

e.g. Random walk

Let $\{X_t\}$ be iid noise. Let $S_t = X_1 + X_2 + \cdots + X_n$

$\{S_t\}$ is a random walk.

$E = 0 \quad Var \to \infty$

③ Model with trend and seasonality

Model with structure

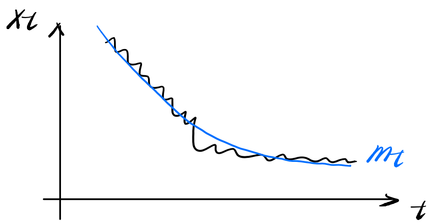
$X_t = m_t + Y_t$ where $E(Y_t) = 0 \ \forall t$

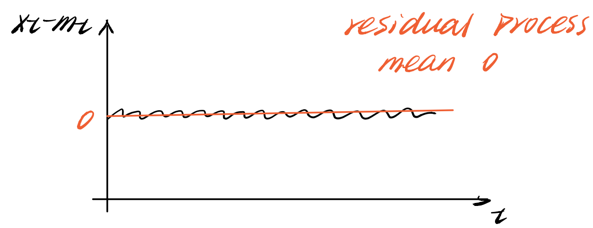
- m_t: slowly changing function of time (don't have high flexibility)

<u>choice for m_t</u> linear function of t ; polynomial function of t

- one estimation method: least square regression $\min_{\hat{m}_t} \sum_{i=1}^{n} (X_t - \hat{m}_t)^2$

Trend may be the source of dependence





Seasonal variation (Periodicity)

$X_t = S_t + Y_t$ where $E(Y_t) = 0 \ \forall t$

- S_t: periodic function with period d ($S_{t-d} = S_t$)

<u>common choice of S_t</u> sum of harmonic function

$S_t = a_0 + \sum_{j=1}^{k} (a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t))$

where a_j, b_j are estimated and λ_j are fixed frequencies

General strategy for analysis

① Plot the data : Identify potential signal (trend, seasonal)
Identify possible models for the residual process
Identify outliers and other wierd things

② Remove signals

③ Choose a model to fit the residuals and estimate the dependence

④ Forcast by inverting projected residuals.

Why focus on residuals $(X_t - \hat{m}_t, X_t - \hat{s}_t)$?

Let $W_i \overset{iid}{\sim} N(\mu, \sigma^2) \Rightarrow W_i - \mu \sim N(0, \sigma^2)$

① Estimate μ to remove the signals

② $W_i - \bar{W} \sim N(0, \sigma^2)$, now we can estimate σ^2.

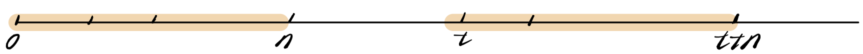
iid noise (special case of stationary)

$\{Z_t\}$ $E(Z_t) = 0$ $\Rightarrow Z_t \overset{iid}{\sim} F(z)$

$E(Z_t^2) = Var(Z_t) = \sigma^2 < \infty$

Stationary Models and the Autocorrelation Function

Let $\{X_s\}$ $s = 0 \cdots n$ has the same properties as $\{X_{t+s}\}$ $s = 0 \cdots n$



We will focus on the first and second order moments.

Note: iid noise is a special case of a stationary process
stronger assumption

Definition $\{X_t\}$ is weakly stationary if

① $E(X_t) = \mu_X(t)$ is independent of t

② For all h, $Cov(X_{t+h}, X_t) = \gamma_X(t+h, t)$ is independent of t

$\Rightarrow \gamma_X(t+h, t) = \gamma_X(h, 0) = Cov(X_h, X_0)$

$Cov(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))]$
$= \gamma_X(r, s)$
Covariance function of $\{X_t\}$.

Strictly stationary process

The joint distribution of $\{X_s\}$ $s = 0 \dots n$ is the same as $\{X_{s+t}\}$ $s = 0 \dots n$ $\forall t$

ACVF
① $\gamma_X(h, 0) = \gamma_X(h)$ Auto-covariance function of a stationary series at lag h.

ACF
② $\rho_X(h) = \dfrac{\gamma_X(h)}{\gamma_X(0)}$ Auto-correlation function at lag h.

$$Corr(X_h, X_0) = \frac{Cov(X_h, X_0)}{\sqrt{Var(X_h)}\sqrt{Var(X_0)}} = \frac{\gamma_X(h)}{\sqrt{\gamma_X(0)}\sqrt{\gamma_X(0)}}$$

Useful Identity If $E(Y^2)<\infty$, $E(X^2)<\infty$, $E(Z^2)<\infty$ and a,b,c are real constants, then $Cov(aX+bY+c, Z) = a\,Cov(X,Z) + b\,Cov(Y,Z)$

e.g. iid noise $X_t \overset{iid}{\sim} (0, \sigma^2)$

① $E(X_t)=0 \Rightarrow$ doesn't dependent

② If $E(X^2)=\sigma^2<\infty$ then

$$\gamma_X(h) = Cov(X_{t+h}, X_t) = \begin{cases} \sigma^2 & h=0 \\ 0 & h\neq 0 \end{cases} \quad \text{by independence}$$

$\Rightarrow$ doesn't dependent

Therefore, iid process is stationary.

e.g. white noise process If $\{X_t\}$ is a sequence of uncorrelated random variables with $E(X_t)=0$, $Var(X_t)=\sigma^2<\infty$, $\gamma_X(h)=0\ \forall\ h\neq 0$, then we refer to it as white noise.

Suppose $\{W_t\}$ and $\{Z_t\}$ are iid sequences.

$\{W_t\} \perp \{Z_t\}$ independent

$P(W_i=0) = P(W_i=1) = \frac{1}{2}$

$P(Z_i=-1) = P(Z_i=1) = \frac{1}{2}$

$X_t = W_t(1-W_{t-1})Z_t \in \{0, \pm 1\}$

W_{t-1}	W_t	X_t
1	0	0
1	1	0
0	0	0
0	1	Z_t

① $P(X_t=1 \mid X_{t-1}=1) = 0$ because if $X_{t-1}=1 \Rightarrow W_{t-1}=1$

$\Rightarrow X_t$ are not independent

② $E(X_t) = E(W_t)E(W_{t-1})E(Z_t) = \frac{1}{2}\times\frac{1}{2}\times 0 = 0$

$Cov(X_t, X_{t+h}) = E(X_t X_{t+h})$ zero mean (always transform)

$$= E(W_t(1-W_{t-1})Z_t\, W_{t+h}(1-W_{t+h-1})Z_{t+h})$$

For $h=0$, $Cov(X_t, X_{t+h}) = E(W_t^2(1-W_{t-1})^2 Z_t^2)$

$$= E(W_t^2)\,E[(1-W_{t-1})^2]\,E(Z_t^2)$$

$$= \frac{1}{2}\cdot\frac{1}{2}\cdot 1$$

$$= \frac{1}{4}$$

For $h\neq 0$ $Cov = E(W_t)E(1-W_{t-1})\underset{0}{E(Z_t)}E(W_{t+h})E(1-W_{t+h-1})\underset{0}{E(Z_{t+h})}$

$= 0$

$\Rightarrow X_t$ is white noise

(not linearly associated)

Note: X_t and X_{t-1} is dependent, but uncorrelated.

e.g. **The Random Walk**

Let $\{X_t\}$ be iid noise

$$S_t = X_1 + \cdots + X_t = \sum_{i=1}^{t} X_i$$

$\Rightarrow$ $E(S_t) = 0$

$Var(S_t) = t\sigma^2$ violation $\gamma_{S_t}(t,t)$ depends on t.

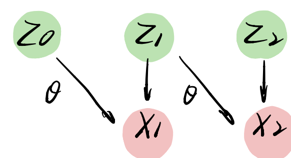
$$\begin{aligned} Cov(S_{t+h}, S_t) &= Cov(S_t + (X_{t+1} \cdots + X_{t+h}), S_t) \\ &= Cov(S_t, S_t) + \underbrace{Cov(X_{t+1} \cdots X_{t+h}, S_t)}_{\{X_t\}\text{ iid}} \quad 0 \\ &= t\sigma^2 \end{aligned}$$

$\Rightarrow$ NOT stationary

e.g. **First-Order Moving Average.** MA(1)

Let $\{Z_t\} \sim WN(0, \sigma^2)$ white noise

Let θ be a real-valued constant



$$X_t = Z_t + \theta Z_{t-1}, \quad t = 0, \pm 1, \pm 2 \cdots$$

① $E(X_t) = E(Z_t) + \theta E(Z_{t-1}) = 0 + \theta \cdot 0 = 0$

zero mean

② $$\begin{aligned} Var(X_t) &= E(X_t^2) = E((Z_t + \theta Z_{t-1})^2) \\ &= E(Z_t^2) + 2\theta E(Z_t Z_{t-1}) + E(\theta^2 Z_{t-1}^2) \\ &= \sigma^2 + 0 + \theta^2\sigma^2 = (1+\theta^2)\sigma^2 \end{aligned}$$

$$\gamma_X(t+h, t) = E(X_{t+h} X_t) = \begin{cases} \sigma^2(1+\theta^2) & h = 0 \\ \theta\sigma^2 & h = 1 \\ 0 & |h| > 1 \end{cases}$$

If $|h| > 1$, $E(X_{t+h} X_t) = E((Z_{t+h} + \theta Z_{t+h-1})(Z_t + \theta Z_{t-1})) = 0$

none of them are equal ($t \neq t-1 \neq t+h \neq t+h-1$)

If $h = 1$, $$\begin{aligned} E(X_{t+1} X_t) &= E((Z_{t+1} + \theta Z_t)(Z_t + \theta Z_{t-1})) \\ &= \underset{0}{E(Z_{t+1} Z_t)} + E(Z_t^2)\theta + \theta \underset{0}{E(Z_{t+1} Z_{t-1})} + \theta^2 \underset{0}{E(Z_t Z_{t-1})} \\ &= \sigma^2\theta \end{aligned}$$

Therefore, MA(1) is stationary.

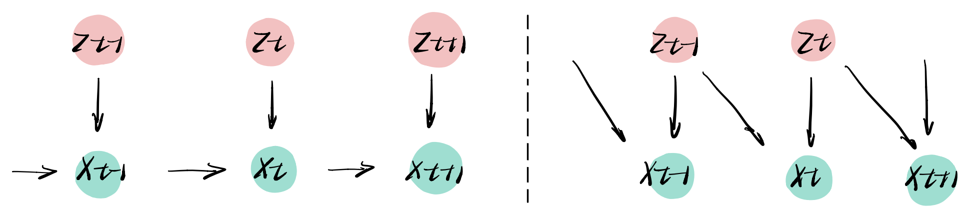
$$\rho_X(h) = \begin{cases} 1 & \text{for } h = 0 \\ \dfrac{\theta}{1+\theta^2} & \text{for } h = \pm 1 \\ 0 & \text{for } |h| > 1 \end{cases}$$

e.g. First-Order Autoregression AR(1)

Assume $\{X_t\}$ is a sequence of random variables that is stationary statisfying

$$X_t = \phi X_{t-1} + Z_t \quad \text{for } t = 0, \pm1, \pm2 \cdots$$

where $\{Z_t\}$ is a $WN(0, \sigma^2)$ process and $\{Z_t\}$ is uncorelated with $\{X_t\}$ for $\dots X_1, \dots X_s$ $(\forall s < t)$ and ϕ is a real-valued constant



① $\begin{cases} E(X_t) = \phi E(X_{t+1}) + E(Z_t) = \phi \mu_x + 0 \\ E(X_t) = \mu_x \end{cases} \Rightarrow \mu_x = 0$ or $\phi = 0$ or 1

② $E[(X_{t-h})X_t] = E[X_{t-h}(\phi X_{t-1} + Z_t)]$

$$= \phi E[X_{t-h} X_{t-1}] + \underline{E[X_{t-h} Z_t]}$$ 0 if $h \neq 0$

$$\gamma_x(h) = \phi \gamma_x(h-1)$$
$$= \phi[\phi \gamma_x(h-2)]$$ recursion
$$\vdots$$
$$= \phi^h \gamma_x(0)$$

$$\rho_x(h) = \frac{\gamma_x(h)}{\gamma_x(0)} = \frac{\phi^h \gamma_x(0)}{\gamma_x(0)} = \phi^h$$

By symmetry and stationary $\rho_x(h) = \phi^{|h|}$

$$\gamma_x(0) = Cov(X_t, X_t)$$
$$= E([\phi X_{t-1} + Z_t][\phi X_{t-1} + Z_t])$$
$$= \phi^2 E(X_{t-1}^2) + E(Z_t^2)$$
$$= \phi^2 \gamma_x(0) + \sigma^2$$

$\Rightarrow \gamma_x(0) = \frac{\sigma^2}{1-\phi^2} \rightarrow |\phi| < 1$ variance is strengthened.

Estimating Autocorrelation

Let $x_1, \cdots x_n$ be observed values for a stationary sequence.

① sample mean $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_n$ unbiased

② $Cov(V,W) = E[(V-E(V))(W-E(W))]$

sample covariance $\widehat{Cov}(V,W) = \frac{\sum_{i=1}^{n}(V_i-\bar{V})(W_i-\bar{W})}{n-1}$ unbiased

$\gamma_X(h) = Cov(X_{t+h}, X_t)$

e.g. $x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5$

$V = \begin{pmatrix} x_1 \\ \vdots \\ x_{n-2} \end{pmatrix} \qquad W = \begin{pmatrix} x_3 \\ \vdots \\ x_n \end{pmatrix}$

$\gamma(2) = Cov(X_t, X_{t+2})$

for $h \geq 0$, $\hat{\gamma}(h)$ is approximately equal to the sample covariance of the $n-h$ pairs of observations $(X_1, X_{1+h}), (X_2, X_{2+h}) \cdots (X_{n-h}, X_n)$

$\Rightarrow$ sample ACVF

$$\hat{\gamma}_X(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X})(X_t - \bar{X}) \qquad -n < h < n$$

use n as the divisor instead of $n-h$

subtracting overall mean

$\Rightarrow$ ensure sample covariance function $\hat{\Gamma}_n := [\hat{\gamma}(i-j)]_{i,j=1}^{n}$ is nonnegative definite.

sample ACF

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \qquad \text{where } \hat{\gamma}(0) = \frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X})^2$$

Estimation and Elimination of Trend and Seasonal Components

classical decomposition model

$$X_t = \underset{\text{trend}}{m_t} + \underset{\text{seasonal}}{S_t} + \underset{\text{random "noise" component}}{Y_t}$$

Remove m_t and S_t to estimate Y_t

① Estimate trend/seasonal using a "model" (filter)

② Differencing $\{X_t\}$ to estimate trend and seasonality (filter)

transform data

Estimating Trend in the absense of Seasonality

Nonseasonal model with trend

$X_t = m_t + Y_t$, $t = 1, \dots, n$ where $E(Y_t) = 0$

Method 1 : Trend Estimation

<Non-parametric methods>

- Advantage : flexible, fewer assumption
- Disadvantage : subjective

① Finite moving average filter

Consider the two-sided moving average

$$W_t = \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t-j} \quad \text{where } q \text{ is a positive integer}$$

"Local average"

$$\Rightarrow W_t = \frac{1}{2q+1} \sum_{j=-q}^{q} (m_{t-j} + Y_{t-j})$$

$$= \frac{1}{2q+1} \Big(\underbrace{\sum_{j=-q}^{q} m_{t-j}}_{\text{slowly changing}} \Big) + \frac{1}{2q+1} \Big(\underbrace{\sum_{j=-q}^{q} Y_{t-j}}_{\text{high-frequency variance}} \Big)$$

$$\approx m_t$$

Note that moving average is a linear filter

$$\hat{m}_t = \sum_{j=-\infty}^{\infty} a_j X_{t-j} \quad \text{where } a_j = \begin{cases} \frac{1}{2q+1} & \text{for } |j| \le q \\ 0 & \text{otherwise} \end{cases}$$

② Exponential Smoothing

Consider the one-sided moving average defined by the recursions

$$\hat{m}_t = aX_t + (1-a)\hat{m}_{t-1} \quad \text{and} \quad \hat{m}_1 = X_1$$

$$\Rightarrow \text{for } t \ge 2, \quad \hat{m}_t = \sum_{j=1}^{t-2} a(1-a)^j X_{t-j} + (1-a)^{t-1} X_1$$

influence decreases for earlier X_t

③ Parametric Smoothing (linear, polynomial, basic function)

④ High-frequency smoothing via Fourier Series.

Method 2: Differencing

Define the lag-1 difference as $\nabla(X_t) = X_t - X_{t-1} = (1-B)X_t$

backward shift operator

$BX_t = X_{t-1}$

We can generalize ∇ and B to general lags by taking powers.

- $B^j X_t = B^{j-1}(BX_t) = B^{j-1}X_{t-1} \cdots = X_{t-j}$

  $\Rightarrow X_t - X_{t-j} = (1-B^j)X_t$
- $\nabla^j(X_t) = \nabla(\nabla^{j-1}(X_t))$, $j \geq 1$ with $\nabla^0(X_t) = X_t$

e.g. $\nabla^2 X_t = \nabla(\nabla X_t) = \nabla((1-B)X_t) = (1-B)(1-B)X_t = (1-2B+B^2)X_t$

$$= X_t - 2BX_t + B^2 X_t = X_t - 2X_{t-1} + X_{t-2}$$

$$= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2})$$

Let $X_t = m_t + Y_t$ where $m_t = a + bt$ (linear trend)

$$\nabla X_t = \nabla(m_t + Y_t) = \nabla m_t + \nabla Y_t = m_t - m_{t-1} + Y_t - Y_{t-1}$$

$$= b + Y_t - Y_{t-1}$$

∇X_t will be stationary if $Y_t - Y_{t-1}$ is stationary

Any polynomial trend of degree k can be reduced to a constant by application of the operator ∇^k

If $X_t = m_t + Y_t$ where $m_t = \sum_{j=0}^{k} c_j t^j$ and Y_t is stationary with mean 0

$$\nabla^k X_t = k!c_k + \nabla^k Y_t.$$

is a stationary process with mean $k!c_k$.

Estimation and Elimination of Both Trend and Seasonality

Classical Decomposition Model

$$X_t = \underbrace{m_t}_{\text{trend}} + \underbrace{s_t}_{\text{season}} + \underbrace{Y_t}_{\text{noise}}$$

where $E(Y_t) = 0$ $\quad s_{t+d} = s_t \ \forall d, t$ and $\sum_{j=1}^{d} s_j = 0$

Method 1: Estimation of Trend and Seasonal Components:

① Estimate the trend and remove it

$$\tilde{X}_t = X_t - \hat{m}_t \quad \Rightarrow \quad \tilde{X}_t \approx S_t + Y_t$$

$E(\tilde{X}_t) \approx S_t$

② Estimate seasonal component

$$d = 4 \qquad \begin{array}{cccc|c} k=1 & k=2 & k=3 & k=4 & \\ \tilde{X}_1 & \tilde{X}_2 & \tilde{X}_3 & \tilde{X}_4 & j=0 \\ \tilde{X}_5 & \tilde{X}_6 & \tilde{X}_7 & \tilde{X}_8 & j=1 \\ \tilde{X}_9 & \tilde{X}_{10} & \tilde{X}_{11} & \tilde{X}_{12} & j=2 \\ \hat{S}_1 & \hat{S}_2 & \hat{S}_3 & \hat{S}_4 & \end{array}$$

$$W_k = \sum_{j=0}^{t/d-1} (X_{k+jd} - \hat{m}_{k+jd})$$

$$\hat{S}_k = W_k - \frac{1}{d}\sum_{i=1}^{d} W_i$$ (subtract mean to force zero mean)

$d_t = X_t - \hat{S}_t$ Deseasonalized data

Reestimate the trend from d_t, $\hat{m}_t$

$$\tilde{Y}_t = X_t - \hat{S}_t - \hat{m}_t$$

Method 2: Differencing

For trends, $\nabla^d X_t = (1-B)^d X_t$

For seasonal, $\nabla_d X_t = X_t - X_{t-d}$

$$= (1-B^d) X_t$$

Apply this to $X_t = m_t + S_t + Y_t$

$$\nabla_d X_t = \nabla_d (m_t + S_t + Y_t)$$
$$= (m_t - m_{t-d}) + \overset{0}{(S_t - S_{t-d})} + (Y_t - Y_{t-d})$$
$$= (m_t - m_{t-d}) + (Y_t - Y_{t-d})$$

If $m_t = a + bt$

$$\tilde{X}_t = (m_t - m_{t-d}) + (Y_t - Y_{t-d})$$
$$= (a + bt - (a + b(t-d))) + Y_t - Y_{t-d}$$
$$= bd + Y_t - Y_{t-d}$$

Testing the Estimated Noise Sequence

Test for weakly stationary

Based on $\begin{cases} \text{moment estimator} - \text{WN} \\ \text{distribution} - \text{iid}. \end{cases}$

Test whether any of a group of autocorrelations of the residual time series are different from zero.

① Sample autocorrelation   white noise

If $\rho(h)=0$, then $\hat{\rho}(h) \sim N(0, \frac{1}{n})$

95% CI $\hat{\rho}(h) \pm 1.96 \frac{1}{\sqrt{n}}$

If we reject $H_0: \rho(h)=0$, then we don't have a white noise process

$\Rightarrow$ Problem: Too many h's.

> If $Y_t \overset{iid}{\sim} (0, \sigma^2)$
>
> $\hat{\rho}(h) \sim N(0, \frac{1}{n})$ for large n or as $n \to \infty$
>
> Recall $\hat{\rho}(h) = \dfrac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \dfrac{\sum_{i=1}^{n-|h|} (X_i-\bar{X})(X_{i+h}-\bar{X})/n}{\sum_{i=1}^{n} (X_i-\bar{X})^2/n}$
>
> 95% CI $\hat{\rho}(h) \pm 1.96 \frac{1}{\sqrt{n}}$

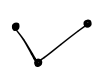
② Portmonteau test   white noise

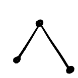
$Q_{BP} = n \sum_{j=1}^{h} \hat{\rho}^2(j)$ Box-Pierce

If $Y_i \overset{iid}{\sim} (0, \sigma^2)$, then $Q \sim \chi^2_h$

$Q_{LB} = n(n+2) \sum_{j=1}^{h} \dfrac{\hat{\rho}^2(j)}{n-j}$ Ljung-Box

③ Turning points test   iid

① $y_i < y_{i-1}$ and $y_i < y_{i+1}$ 

② $y_i > y_{i-1}$ and $y_i > y_{i+1}$ 

Let y_i be a turning point

For iid sequences, let T be the number of turning points.

The probability of turning point at time t is $\frac{2}{3}$

$$\Rightarrow \quad E(T) = (n-2)\frac{2}{3} \qquad V(T) = \frac{16n-29}{90}$$

$$\frac{T-E(T)}{\sqrt{V(T)}} \overset{\cdot}{\sim} N(0,1) \text{ for large } n$$

④ Sign test iid

Count $y_{i+1} - y_i > 0$

Exact Binomial hypothesis test for $H_0: p = 0.5$

⑤ Rank test iid

Compare ranks of y_t with t