

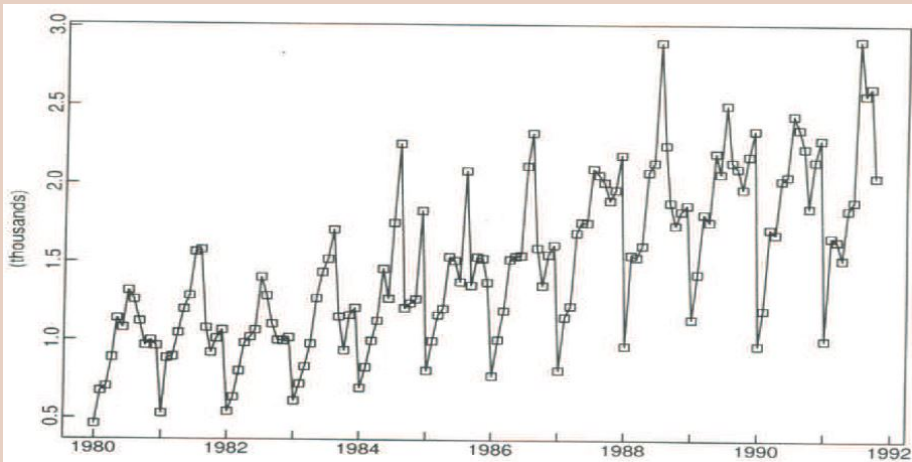
Time Series

Ching-Kang Ing (銀慶剛)

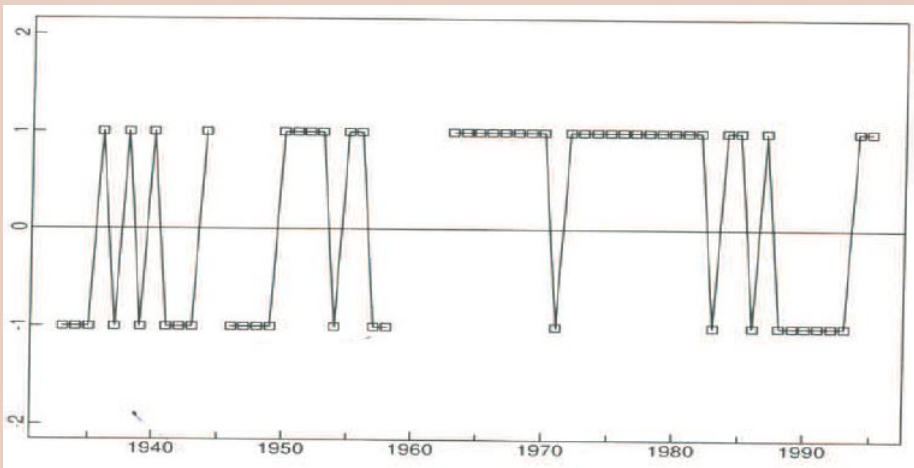
Institute of Statistics, National Tsing Hua University, Hsinchu, Taiwan

時間數列(隨著時間變動而變化的資料)的亂與序

The Australian red wine sales, Jan.1980–Oct.1991

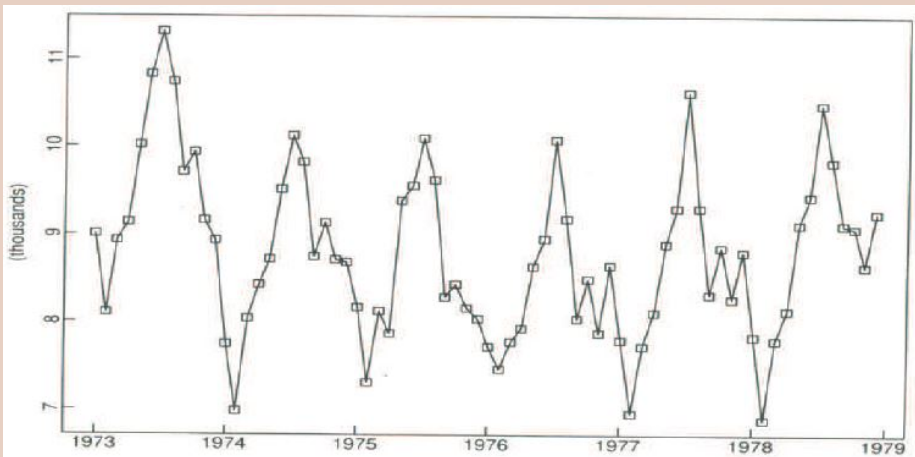


The results of All-star baseball games, 1933–1995



1: the National League won
-1: the American League won

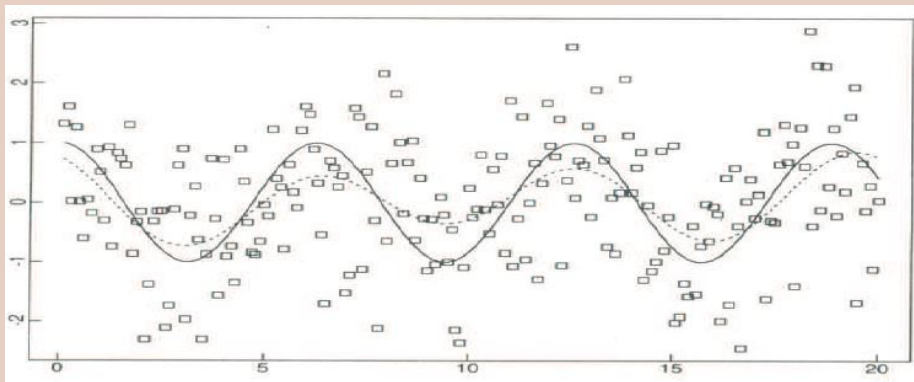
The monthly Accidental Deaths in the U.S.A., 1973–1978



The simulated values of the series

$$X_t = \cos(t) + N_t, \quad t = 0.1, 0.2, \dots, 20,$$

where $\{N_t\}$ is a sequence of independent normal random variables, with mean 0 and variance 0.25.

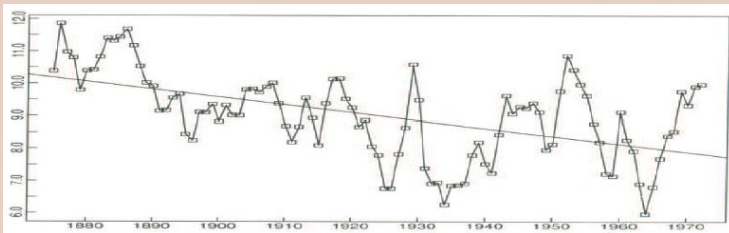


序：長期趨勢、週期性。
亂："序"之外的不規則變動。

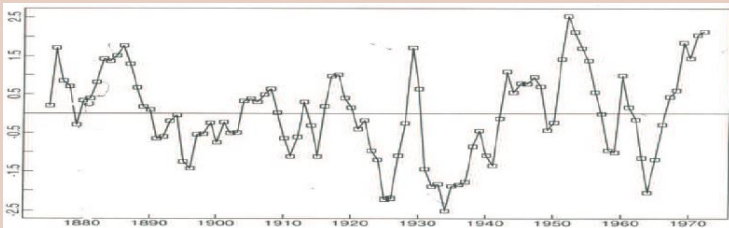
但，我們是否還遺漏了什麼？

Level of Lake Huron, 1875–1972

- The line fitted by least squared.



- Residuals from fitting a line to the Lake Huron data.



忽略了"亂"中的"序"所帶來的困境

(i) 有礙"訊息"偵測

例：

$$y_t = \alpha + \epsilon_t, \quad t = 1, \dots, n, \quad (1)$$

且

$$\epsilon_t = a\epsilon_{t-1} + \delta_t, \quad (2)$$

其中 $|a| < 1$ ， $\{\delta_t\}$ 是獨立且同分佈的隨機變數滿足

$$E(\delta_t) = 0 \text{ 及 } \sigma_\delta^2 := \text{Var}(\delta_t) = E(\delta_t^2) < \infty。$$

模型(2)被稱為穩定(stationary)一階自我迴歸(first-order autoregressive, AR(1))模型， δ_t 是 AR(1) 中的隨機干擾項，也是存在於(1)中真正的"無序"部分。

- 可證明 $E(\epsilon_t) = 0$ 且 $\text{Var}(\epsilon_t) := \sigma_\epsilon^2 = \frac{\sigma_\delta^2}{1 - a^2} > \sigma_\delta^2$ 。
- 若(2)被輕忽，則吾人將以 $\hat{\alpha} = \frac{1}{n} \sum_{i=1}^n y_i$ 估計 α ，且

$$\text{Var}(\hat{\alpha}) = E((\hat{\alpha} - \alpha)^2) = E \left[\left(\frac{1}{n} \sum_{i=1}^n (y_i - \alpha) \right)^2 \right] = E \left[\left(\frac{1}{n} \sum_{i=1}^n \epsilon_i \right)^2 \right]$$

將被誤認為 $\frac{1}{n} \sigma_\epsilon^2$ 。

- 但，我們可證明在模型(1)下，

$$\text{Var}(\hat{\alpha}) = \frac{1}{n} \sigma_\epsilon^2 \times \frac{1 + a}{1 - a},$$

此值可大於 $\frac{1}{n} \sigma_\epsilon^2$ (當 $0 < \alpha < 1$)，也可小於 $\frac{1}{n} \sigma_\epsilon^2$ (當 $-1 < \alpha < 0$)。

- 故，當我們以 $\hat{\alpha}$ 及其被誤認的變異數(或標準差)檢定 α 是否顯著時，將可能導致誤判(若 $a > 0$ ，可能將不顯著的 α 認定為顯著；若 $a < 0$ ，則可能反之)。

(ii) 放大預測誤差

- 在模型(1)及(2)下，若(2)被輕忽，則 $\hat{\alpha}$ 也將被用來預測 y_{n+1} 。此時，預測均方差(mean squared prediction error, MSPE)為：

$$E((y_{n+1} - \hat{\alpha})^2) = E \left[\left(\epsilon_{n+1} - \frac{1}{n} \sum_{i=1}^n \epsilon_i \right)^2 \right] \xrightarrow{n \rightarrow \infty} \sigma_{\epsilon}^2 = \frac{\sigma_{\delta}^2}{1 - a^2}。$$

- 我們希望找到一個預測方式 $\hat{y}_{n+1}$ ，能達到

$$E((y_{n+1} - \hat{y}_{n+1})^2) \xrightarrow{n \rightarrow \infty} \sigma_{\delta}^2。 \quad (3)$$

- 如何達到(3)?

考慮非線性估計式 $(\tilde{\alpha}, \tilde{a})$ 使得 $\sum_{t=1}^n ((y_t - \alpha) - a(y_{t-1} - \alpha))^2$ 最小。

令 $\hat{y}_{n+1} = \tilde{\alpha} + \tilde{a}(y_n - \tilde{\alpha})$ 。我們可證明 $E((y_{n+1} - \hat{y}_{n+1})^2) \xrightarrow{n \rightarrow \infty} \sigma_{\delta}^2$ 。

- 事實上，(1)可被稱為 regression model with time series error，是最常用的時間序列模型之一。

穩定時間序列(非精確定義)

$\{X_t\}$ 被稱為穩定的時間序列，如果對所有的整數 t 及 k ，

$$E(X_t) = \mu \quad \leftarrow \text{常數}$$

$$\text{Cov}(X_{t+k}, X_t) = \gamma_k \quad \leftarrow \text{只與 } k \text{ 有關的常數}$$

穩定時間序列的模型化(a brief review)

(1) 1920 年代： $X_t = \alpha \cos(wt) + \beta \sin(wt) + \delta_t, t = 1, 2, \dots$

(2) Yule(1927) 及 Walker(1931)：

$$X_t = a_1 X_{t-1} + \dots + a_p X_{t-p} + \delta_t \quad \leftarrow \text{M. Pourahmadi (2001)}$$

稱之為"monumental achievement"

(3) Wold 分解(1938)：

從 random disturbances, δ_t , 的角度來看不穩定時間序列，證明任何一個穩定的時間序列都可分解成所有過去 δ_t 的線性組合及一個命定(deterministic)過程。

穩定時間序列的模型化(a brief review) (Cont.)

(4) Kolmogorov 1940 年代：

建立了時間域(time domain)及頻譜域(frequency domain)間的同構性(isomorphism)；並以頻譜密度函數(spectral density)來表現預測方差。

(5) Wiener 及 Masani(1958)：

得到了目前的時間序列可用過去的時間序列來表現，亦即，

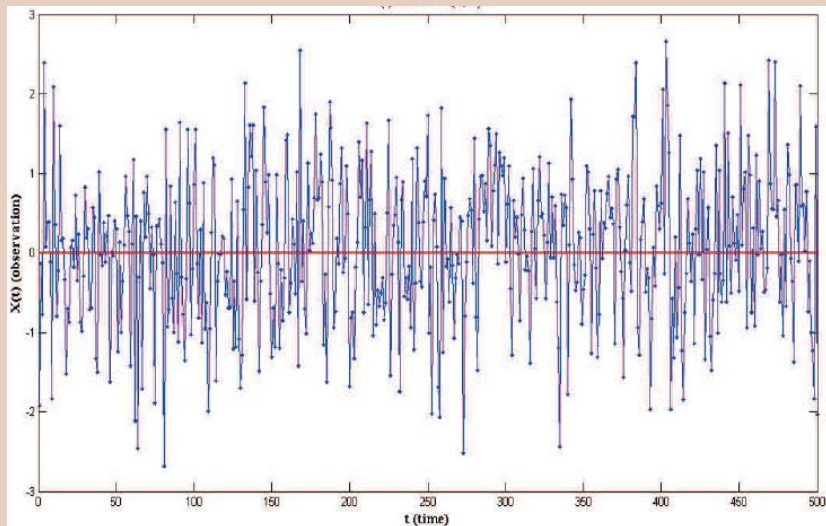
$$X_t = \sum_{k=1}^{\infty} a_k X_{t-k} + \delta_t$$

的充分條件。

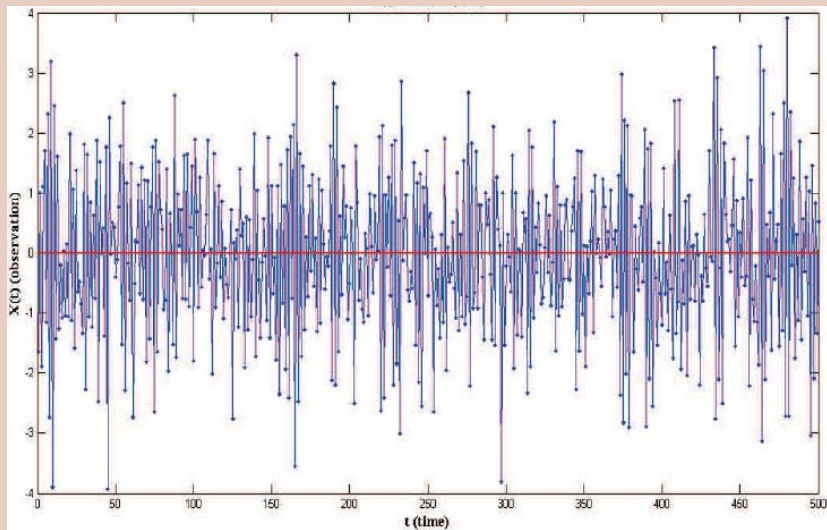
(6) Akaike(1974)：

以訊息理論來選擇時間序列模型。

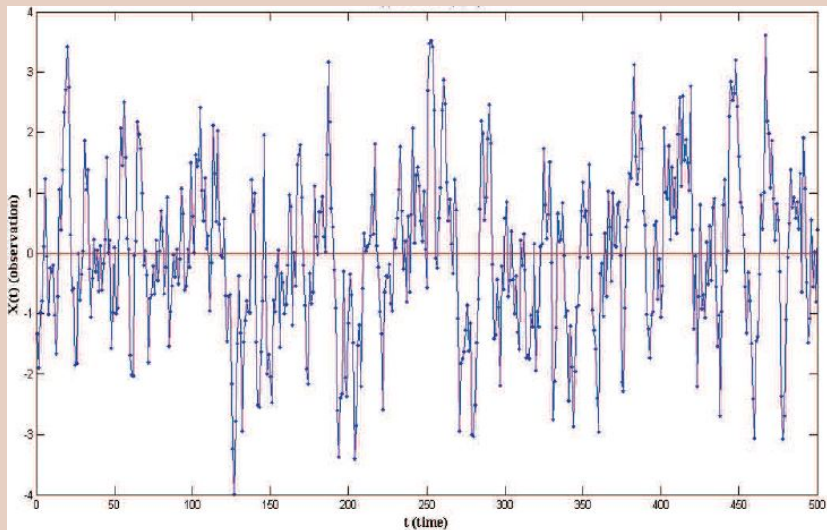
White noise: $X_t = \beta X_{t-1} + \epsilon_t$ where $\beta = 0$ and $\epsilon_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$



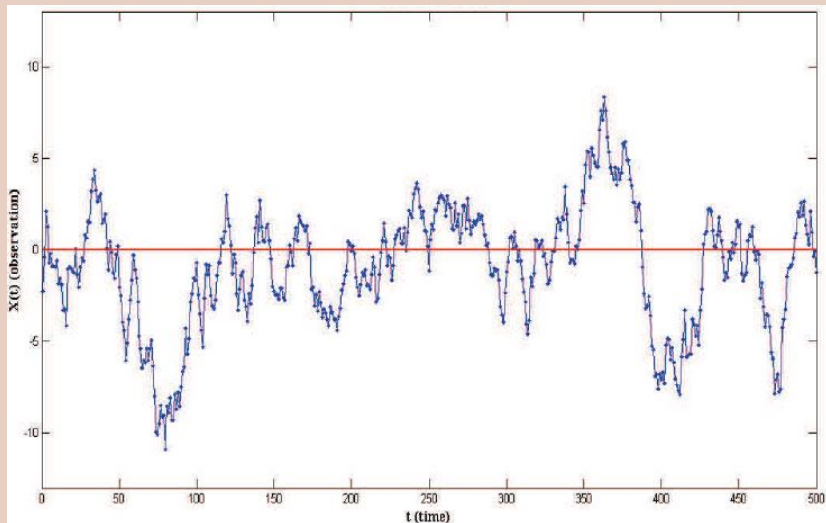
First-Order Autoregressive Process ($\beta = -0.7$)



First-Order Autoregressive Process ($\beta = 0.7$)



First-Order Autoregressive Process ($\beta = 0.95$)



Random Walk ($\beta = 1$)

