Stochastic Hydroclimatic Modeling and Simulation (SHCMS)

Time series modeling of hydroclimatic processes

Professor Ke-Sheng Cheng

Department of Bioenvironmental Systems Engineering/Master Program in Statistics

National Taiwan University

Large-sample properties of the estimates (MLE & LSE)

- For large *n*, the estimators are **approximately unbiased and normally distributed**. The variances and correlations are as follows:

$$\text{AR}(1):\quad Var(\hat{\phi}) \approx \frac{1-\phi^2}{n}$$

$$\text{AR}(2):\ \begin{cases} Var(\hat{\phi}_1) \approx Var(\hat{\phi}_2) \approx \dfrac{1-\phi_2^2}{n} \\[2ex] \boxed{Corr(\hat{\phi}_1, \hat{\phi}_2) \approx -\dfrac{\phi_1}{1-\phi_2} = -\rho_1} \end{cases}$$

Jonathan D. Cryer • Kung-Sik Chan
Time Series Analysis
With Applications in R

- Asymptotic ($n \to \infty$) property of the AR(p) Yule-Walker estimators

$$\widehat{\overrightarrow{\Phi}_p} \xrightarrow{d} N\left(\overrightarrow{\Phi}_{p,}\, \sigma_\epsilon^2\, \Sigma_p^{-1}/n\right)$$

$$\hat{\sigma}_\epsilon^2 \to \sigma_\epsilon^2$$

$$\text{MA}(1):\quad Var(\hat{\theta}) \approx \frac{1-\theta^2}{n}$$

$$\text{MA}(2):\begin{cases} Var(\hat{\theta}_1) \approx Var(\hat{\theta}_2) \approx \dfrac{1-\theta_2^2}{n} \\ Corr(\hat{\theta}_1, \hat{\theta}_2) \approx -\dfrac{\theta_1}{1-\theta_2} \end{cases}$$

$$
\text{ARMA}(1,1): \begin{cases} Var(\hat{\phi}) \approx \left[\dfrac{1-\phi^2}{n}\right]\left[\dfrac{1-\phi\theta}{\phi-\theta}\right]^2 \\ Var(\hat{\theta}) \approx \left[\dfrac{1-\theta^2}{n}\right]\left[\dfrac{1-\phi\theta}{\phi-\theta}\right]^2 \\ Corr(\hat{\phi}, \hat{\theta}) \approx \dfrac{\sqrt{(1-\phi^2)(1-\theta^2)}}{1-\phi\theta} \end{cases}
$$

Jonathan D. Cryer • Kung-Sik Chan

Time Series Analysis

With Applications in R

Model diagnostics

- Model diagnostics is concerned with testing the goodness of fit of a model and, if the fit is poor, suggesting appropriate modifications.
- Residual Analysis
  - Standardized residuals time series plot
  - Normality of the residuals
    - Test of normality (GOF test, Q-Q plot)
  - Independence of the Residuals
    - Test of independence (autocorrelation function of the residuals)
- Overfitting

Test of the residuals independence

The Ljung–Box–Pierce Q-statistic given by

$$Q = n(n+2)\sum_{h=1}^{H} \frac{\widehat{\rho}_e^2(h)}{n-h}$$

can be used to perform such a test. The value H in (3.154) is chosen somewhat arbitrarily, typically, $H = 20$. Under the null hypothesis of model adequacy, asymptotically ($n \to \infty$), $Q \sim \chi^2_{H-p-q}$. Thus, we would reject the null hypothesis at level α if the value of Q exceeds the $(1-\alpha)$-quantile of the χ^2_{H-p-q} distribution.

Time Series Forecasting

- Time series forecasting is the process of analyzing time series data, using statistics and modeling, to make predictions and inform strategic decision-making.

Terminology

- **Forecasting**
  - Prediction of data at a specific future point in time.
  - Short lead time (short time horizon)
- **Prediction**
  - Prediction of the future data in general.
  - Longer lead time

- **Lead time (or time horizon)**: The length of time between the issuance of a forecast and the occurrence of the phenomena that were predicted.
- **Multistep-ahead forecast** – the task of forecasting a sequence of values in a time series.
    - 1-step ahead 3-hour forecast
    - 3-step ahead 3-hour forecast

Time Series Forecasting

- **The minimum MSE predictor**
- **Forecast error**
  - Expected value
  - Variance
- **Forecast intervals**

AR(1) Forecasting

$$X_{t+1} - \mu = \phi(X_t - \mu) + \varepsilon_{t+1}, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

$$X_{t+2} - \mu = \phi(X_{t+1} - \mu) + \varepsilon_{t+2} = \phi(\phi(x_t - \mu) + \varepsilon_{t+1}) + \varepsilon_{t+2}$$

$$= \phi^2(x_t - \mu) + \phi\varepsilon_{t+1} + \varepsilon_{t+2}$$

$$X_{t+3} - \mu = \phi(X_{t+2} - \mu) + \varepsilon_{t+3} = \phi(\phi^2(x_t - \mu) + \phi\varepsilon_{t+1} + \varepsilon_{t+2}) + \varepsilon_{t+3}$$

$$= \phi^3(x_t - \mu) + \phi^2\varepsilon_{t+1} + \phi\varepsilon_{t+2} + \varepsilon_{t+3}$$

$$x_{t+\ell} = \phi^\ell(x_t - \mu) + \sum_{i=1}^{\ell} \phi^{(\ell-i)}\varepsilon_{t+i}$$

Minimum MSE predictor

Given $(x_t, x_{t-1}, \cdots)$, we want to predict the values of $(x_{t+1}, x_{t+2}, \cdots, x_{t+p})$.

The minimum MSE predictor of x_{t+1} is

$$\hat{x}_{t+1} - \mu = E(X_{t+1} | x_t, x_{t-1}, \cdots) = \phi(x_t - \mu).$$

$$\hat{x}_{t+2} - \mu = \phi(\hat{x}_{t+1} - \mu) = \phi^2(x_t - \mu).$$

$$\hat{x}_{t+\ell} - \mu = \phi^\ell(x_t - \mu)$$

$$\hat{x}_{t+\ell} = (1 - \phi^\ell)\mu + \phi^\ell x_t$$

Forecast errors

$$e_t^1 = x_{t+1} - \hat{x}_{t+1} = \varepsilon_{t+1}$$

$$e_t^2 = x_{t+2} - \hat{x}_{t+2} = \phi\varepsilon_{t+1} + \varepsilon_{t+2}$$

$$e_t^\ell = x_{t+\ell} - \hat{x}_{t+\ell} = \sum_{i=1}^{\ell} \phi^{(\ell-i)} \varepsilon_{t+i}$$

$$E(e_t^\ell) = 0$$

$$Var(e_t^\ell) = \sigma_\varepsilon^2 \sum_{i=1}^{\ell} \phi^{2(\ell-1)} = \sigma_\varepsilon^2 \left(\frac{1-\phi^{2\ell}}{1-\phi^2}\right)$$

Forecast Intervals

$$P\left[-1.96 \leq \frac{e_t^{\ell}}{\sqrt{Var(e_t^{\ell})}} \ll 1.96\right] = 0.95$$

$$P\left[-1.96 \leq \frac{x_{t+\ell} - \hat{x}_{t+\ell}}{\sqrt{Var(e_t^{\ell})}} \leq 1.96\right] = 0.95$$

$$P\left[\hat{x}_{t+\ell} - 1.96\sqrt{Var\left(e_t^{\ell}\right)} \leq x_{t+\ell} \leq \hat{x}_{t+\ell} + 1.96\sqrt{Var(e_t^{\ell})}\right] = 0.95$$