Time Series

Sudipta Das

Assistant Professor,
Department of Computer Science,
Ramakrishna Mission Vivekananda Educational & Research Institute

1. Modeling and Forecasting with ARMA Processes
   - Estimation of parameters of ARMA(p, q)
     - Initial Order Selection
     - Maximum Likelihood Estimation
     - Order Selections
   - Forecasting

Estimation of parameters of ARMA(p, q) I

- Steps to fit a time series $\{X_n\}$, by an ARMA(p, q) model as

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \text{ where } \{Z_t\} \sim WN(0, \sigma^2)$$

  - Make an initial guess of the orders p and q from the sample ACF and PACF plots
  - Perform a preliminary estimation of the parameters $\phi = (\phi_1, \ldots, \phi_p)', \theta = (\theta_1, \ldots, \phi_q)'$,and σ^2 from the sample observations $x_1, \ldots, x_n$.
  - Perform the final estimation of the parameters by maximum likelihood estimators
  - Recheck the orders p and q by calculating some metrics like AICC
  - Diagnostics checking of the residuals

Initial Order Selection I

- Behavior of the ACF and PACF for Causal and Invertible ARMA Models

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

Maximum Likelihood Estimation I

- Assume that $\{X_t\}$ is a Gaussian time series with mean zero and autocovariance function $\gamma(|i-j|) = E(X_iX_j)$
  - We assume that the sample size is large
- From the n data sample $\mathbf{x_n} = (x_1, x_2, \ldots, x_n)'$, form the likelihood

$$\mathcal{L}(\Gamma_n) = \frac{1}{(2\pi)^{n/2}|\Gamma_n|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{x_n}'\Gamma_n^{-1}\mathbf{x_n}\right\} = \mathcal{L}(\phi, \theta),$$

where $\Gamma_n = E[\mathbf{x}_n\mathbf{x}_n']$

  - Likelihood of a single data point of n dimensional random vector.

Maximum Likelihood Estimation II

- Matrix inversion (Γ_n^{-1}) can be avoided by the use of following identity

$$\mathbf{x_n} = C_n(\mathbf{x_n} - \hat{\mathbf{x}}_\mathbf{n}),$$

where $\hat{\mathbf{x}}_\mathbf{n} = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)'$ and

$$C_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \theta_{11} & 1 & 0 & \cdots & 0 \\ \theta_{22} & \theta_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 1 \end{bmatrix}$$

i.e. for all $k = 1, \ldots, n$,

$$x_k = \sum_{j=0}^{k-1} \theta_{k-1,j} \left(x_{k-j} - \hat{x}_{k-j} \right),$$

with $\theta_{k-1,0} = 1$ and $\hat{x}_1 = 0$.

Maximum Likelihood Estimation III

- Note that,
  - $x_n - \hat{x}_n$, (residuals) is uncorrelated with $x_1, \ldots, x_{n-1}$, (predictors).
  - It concludes that $(x_n - \hat{x}_n)$ is uncorrelated with the innovations $(x_1 - \hat{x}_1), \ldots, (x_{n-1} - \hat{x}_{n-1})$.
- As the components of $\mathbf{x_n} - \hat{\mathbf{x}}_\mathbf{n}$ are uncorrelated, $\mathbf{x_n} - \hat{\mathbf{x}}_\mathbf{n}$ has the diagonal covariance matrix, i.e.,

$$Var(\mathbf{x_n} - \hat{\mathbf{x}}_\mathbf{n}) = D_n = diag\{v_0, v_1, \ldots, v_{n-1}\},$$

  where $v_i = E(x_{i+1} - \hat{x}_{i+1})^2$.
- Therefore,

$$\Gamma_n = Var(\mathbf{x_n}) = C_n Var(\mathbf{x_n} - \hat{\mathbf{x}}_\mathbf{n}) C_n' = C_n D_n C_n'$$

Maximum Likelihood Estimation IV

- Thus,

$$\begin{aligned}\mathbf{x_n}'\Gamma_n^{-1}\mathbf{x_n} &= \mathbf{x_n}'(C_n')^{-1}D_n^{-1}C_n^{-1}\mathbf{x_n}\\ &= (\mathbf{x_n}-\hat{\mathbf{x}}_\mathbf{n})'\,D_n^{-1}\,(\mathbf{x_n}-\hat{\mathbf{x}}_\mathbf{n})\\ &= \sum_{j=1}^{n}(x_j-\hat{x}_j)^2/v_{j-1}\end{aligned}$$

  and $|\Gamma_n| = |C_n||D_n||C_n| = v_0 v_1 \cdots v_{n-1}$.

- Simplified likelihood

$$\mathcal{L}(\phi,\theta) = \frac{1}{(2\pi)^{n/2}\sqrt{v_0 v_1 \cdots v_{n-1}}}\exp\left\{-\frac{1}{2}\sum_{j=1}^{n}(x_j-\hat{x}_j)^2/v_{j-1}\right\}$$

Maximum Likelihood Estimation V

- For more simplification, consider a transformed process $\{W_t\}$ as

$$W_t = \begin{cases} \sigma^{-1} X_t, & t = 1, \ldots, m = max(p,q) \\ \sigma^{-1}\phi(B) X_t, & t > m \end{cases},$$

  whose MSE

$$r_j = E(W_{j+1} - \hat{W}_{j+1})^2 = \sigma^{-2} E(X_{j+1} - \hat{X}_{j+1})^2 = v_j/\sigma^2$$

- Thus replacing v_j by $\sigma^2 r_j$, we get the Gaussian Likelihood for an ARMA Process as

$$\mathcal{L}(\phi, \theta, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^n r_0 r_1 \cdots r_{n-1}}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^{n} (x_j - \hat{x}_j)^2 / r_{j-1} \right\}$$

Maximum Likelihood Estimation VI

- Differentiating $\log \mathcal{L}(\phi, \theta, \sigma^2)$ partially with respect to σ^2 and noting that $\hat{x}_j$ and r_j are independent of σ^2, we find that the maximum likelihood estimators ϕ, θ and σ^2 from the following equations

$$\hat{\sigma}^2 = n^{-1} S(\hat{\phi}, \hat{\theta})$$

where

$$S(\hat{\phi}, \hat{\theta}) = \sum_{j=1}^{n} (x_j - \hat{x}_j)^2 / r_{j-1}$$

and $\hat{\phi}, \hat{\theta}$ are the values of ϕ, θ that minimize

$$l(\phi, \theta) = \ln(n^{-1} S(\phi, \theta)) + n^{-1} \sum_{j=1}^{n} \log r_{j-1}$$

Order Selections I

- Akaike Information criterion
  $AIC(\beta, p, q) :=$
  $$-2\ln \mathcal{L}_X(\beta, n^{-1}S_X(\beta)) + 2(p+q+1),$$
  where $\beta = [\phi_1, \ldots, \phi_p, \theta_1 \ldots, \theta_q]'$
  - It is an estimate of the *Kullback-Leibler* (KL) index of the fitted model relative to the true model.
  - The KL index of $f(\cdot;\psi)$ relative to $f(\cdot;\theta)$
    $$\Delta(\psi|\theta) = E_\theta\left[-2\ln f(\mathbf{x};\psi)\right] = \int_{\mathcal{R}^n} -2\ln\left(f(\mathbf{x};\psi)\right) f(\mathbf{x};\theta)d\mathbf{x}$$
  - Note that the KL discrepancy between $f(\cdot;\psi)$ and $f(\cdot;\theta)$ is defined as
    $$d(\psi|\theta) = \Delta(\psi|\theta) - \Delta(\theta|\theta) = \int_{\mathcal{R}^n} -2\ln\left(\frac{f(\mathbf{x};\psi)}{f(\mathbf{x};\theta)}\right) f(\mathbf{x};\theta)d\mathbf{x} \overset{Jensen}{\geq} 0.$$

Order Selections II

- Akaike Information criterion corrected
  $AICc(\beta, p, q) :=$

$$-2 \ln \mathcal{L}_X(\beta, n^{-1} S_X(\beta)) + 2(p+q+1)n/(n-p-q-2)$$

  - It is a bias-corrected version of the AIC.

- Note
  - Suppose that our observations $X_1, \ldots, X_n$ are from a Gaussian ARMA process with parameter vector $\theta = [\beta, \sigma^2]$ and assume for the moment that the true order is (p, q).
  - Let $\hat{\theta} = [\hat{\beta}, \hat{\sigma}^2]$ be the mle of θ based on $X_1, \ldots, X_n$ and let $Y_1, \ldots, Y_n$ be an independent realization of the true process (with parameter θ).
  - Then,

$$\begin{aligned} -2 \ln \mathcal{L}_Y(\hat{\beta}, \hat{\sigma}^2) &= -\ln \left[2\pi (\hat{\sigma}^2)^n \sum_{j=1}^{n} r_{j-1} \right] + \hat{\sigma}^{-2} S_Y(\hat{\beta}) \\ &= -2 \ln \mathcal{L}_X(\hat{\beta}, \hat{\sigma}^2) + \hat{\sigma}^{-2} S_Y(\hat{\beta}) - n \end{aligned}$$

  - Thus,

$$\begin{aligned} \Delta(\hat{\theta}|\theta) &= E_\theta \left[-2 \ln \mathcal{L}_Y(\hat{\beta}, \hat{\sigma}^2) \right] \\ &= E_\theta \left[-2 \ln \mathcal{L}_X(\hat{\beta}, \hat{\sigma}^2) \right] + E_\theta \left[\hat{\sigma}^{-2} S_Y(\hat{\beta}) \right] - n \\ &\approx E_\theta \left[-2 \ln \mathcal{L}_X(\hat{\beta}, \hat{\sigma}^2) \right] + 2(p+q+1)n/(n-p-q-2) \end{aligned}$$

Order Selections IV

- Comments
  - For fitting autoregressive models, Monte Carlo studies the AIC has a tendency to overestimate p.
  - The AICC statistic has a more extreme penalty for large-order models, which counteracts the overfitting tendency of the AIC.
  - However, the penalty factors for the AICC and AIC statistics are asymptotically equivalent as $n \to \infty$.

Order Selections V

- Bayesian formation criterion
  $BIC(p,q) :=$

$$(n-p-q)\ln\left[\frac{n\hat{\sigma}^2}{n-p-q}\right]+n\left(1+\ln\sqrt{2\pi}\right)+(p+q)\ln\left[\left(\sum_{t=1}^{n}X_t^2-n\hat{\sigma}^2\right)/(p+q)\right],$$

  where $\hat{\sigma}^2$ is the mle of the variance of the white noise process.
  - Another criterion that attempts to correct the overfitting nature of the AIC.

Order Selections VI

- Comments
  - The BIC is a consistent order-selection criterion, i.e.,
    - If the data $\{X_1, \ldots, X_n\}$ are in fact observations of an $ARMA(p, q)$ process, and if $\hat{p}$ and $\hat{q}$ are the estimated orders found by minimizing the BIC, then $\hat{p} \to p$ and $\hat{q} \to q$ wp 1 as $n \to \infty$.
  - The AICC and AIC are not consistent.
  - On the other hand, order selection by minimization of the AICC or AIC is asymptotically efficient for autoregressive processes, while order selection by BIC minimization is not.

Order Selections VII

- In the modeling of real data there is rarely such a thing as the “true order."
  - For the process X_t there may be many polynomials $\theta(z), \phi(z)$ such that the coefficients of z_{t-j} in $\theta(z)/\phi(z)$ closely approximate $\psi(j)$ for moderately small values of j.
  - Correspondingly, there may be many ARMA processes with properties similar to $\{X_t\}$.
- The AICC criterion does, however, provide us with a rational criterion for choosing among competing models.
- It has been suggested that models with AIC values within c of the minimum value should be considered competitive (with $c = 2$ as a typical value).
- Selection from among the competitive models can then be based on such factors as whiteness of the residuals and model simplicity.

Forecasting I

- Once, we find the fitted model as

$$X_t - \hat{\phi}_1 X_{t-1} - \cdots - \hat{\phi}_p X_{t-p} = Z_t + \hat{\theta}_1 Z_{t-1} + \cdots + \hat{\theta}_q Z_{t-q}$$

with $Z \sim N(0, \hat{\sigma}^2)$, we can go for forecasting as mentioned below.

Forecasting II

- One step forecast

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{nj} \left(X_{n+1-j} - \hat{X}_{n+1-j} \right), & 1 \leq n < m \\ \phi_1 X_n + \cdots + \phi_p X_{n+1-p} + \sum_{j=1}^{q} \theta_{nj} \left(X_{n+1-j} - \hat{X}_{n+1-j} \right), & n \geq m \end{cases}$$

- Mean square error

$$E\left(X_{n+1} - \hat{X}_{n+1} \right)^2 = v_n^2 = \sigma^2 r_n$$

  - Parameters ϕ_is, θ_is and σ will be replaced by the corresponding estimates $\hat{\phi}_i$s, $\hat{\theta}_i$s and $\hat{\sigma}$, respectively.