ISMT S-136 Time Series Analysis with Python

Harvard Summer School

Dmitry Kurochkin

Summer 2021
Lecture 6

Contents

Estimation with Minimum Mean Square Error

Claim.
Given X, the best predictor $\hat{Y}$ of Y that minimizes[1]

$$\mathrm{E}\left[(Y - \hat{Y})^2\right]$$

is

$$\hat{Y} = \mathrm{E}\left[Y|X\right].$$

Remark: Generally, $\mathrm{E}\left[Y|X\right]$ could be non-linear function of X.

[1] X can be a vector-valued random variable.

Contents

Time Series Forecasting: Best Predictor

Claim.

Let's fix first n observations: $x_1, x_2, \ldots, x_n$.

Let x^n_{n+m}, where $m = 1, 2, \ldots$, denote the forecasts of x_{n+m} based on $x_1, x_2, \ldots, x_n$. Then the best predictor that minimizes

$$\mathrm{E}\left[(x_{n+m} - x^n_{n+m})^2\right]$$

is

$$\underbrace{x^n_{n+m}}_{\text{``}\hat{Y}\text{''}} = \mathrm{E}\Big[\underbrace{x_{n+m}}_{\text{``}Y\text{''}} \mid \underbrace{x_1, x_2, \ldots, x_n}_{\text{``}X\text{''}}\Big]$$

Contents

Definition of Mean Square Error

Def.

The *mean square m-step-ahead prediction error* P_{n+m}^n is defined as

$$P_{n+m}^n = \mathrm{E}\left[(x_{n+m} - x_{n+m}^n)^2\right],$$

where $m = 1, 2, \dots$

Contents

Time Series Forecasting: Best Linear Predictor

Assume that the predictor x_{n+m}^n is linear in $x_1, x_2, \ldots, x_n$:

$$
\begin{aligned}
x_{n+m}^n &= \sum_{k=1}^{n} \alpha_k x_k \\
&= \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n,
\end{aligned}
$$

then x_{n+m}^n satisfies

$$
\mathrm{E}\left[(x_{n+m} - x_{n+m}^n) x_k\right] = 0, \quad k = 1, 2, \ldots, n.
$$

Contents

Linear One-step-ahead Prediction: Prediction Equations

Let's focus on 1-step-ahead prediction:

$$
\begin{aligned}
x_{n+1}^{n} &= \sum_{j=1}^{n} \underbrace{\phi_{nj}}_{\alpha\text{'s}} x_{n+1-j} \\
&= \phi_{n1}x_n + \phi_{n2}x_{n-1} + \ldots + \phi_{nn}x_1,
\end{aligned}
$$

then

$$\mathrm{E}\left[(x_{n+1} - x_{n+1}^{n})x_{n+1-k}\right] = 0, \quad k = 1, 2, \ldots, n$$

becomes

$$\mathrm{E}\left[\left(x_{n+1} - \sum_{j=1}^{n} \phi_{nj}x_{n+1-j}\right)x_{n+1-k}\right] = 0, \quad k = 1, 2, \ldots, n.$$

Linear One-step-ahead Prediction: Prediction Equations

Assuming $\mathrm{E}[x_t] = 0$, we then notice that

$$\mathrm{E}\left[\left(x_{n+1} - \sum_{j=1}^{n} \phi_{nj} x_{n+1-j}\right) x_{n+1-k}\right] = 0, \quad k = 1, 2, \ldots, n.$$

is equivalent to

$$\gamma(k) - \sum_{j=1}^{n} \phi_{nj} \gamma(k-j) = 0, \quad k = 1, 2, \ldots, n$$

or

$$\sum_{j=1}^{n} \phi_{nj} \gamma(k-j) = \gamma(k), \quad k = 1, 2, \ldots, n.$$

Linear One-step-ahead Prediction: Prediction Equations

<u>Remark:</u> This system of linear equations

$$\phi_{n1}\gamma(k-1) + \phi_{n2}\gamma(k-2) + \dots \phi_{nn}\gamma(k-n) = \gamma(k), \quad k = 1, 2, \dots, n$$

can also be written in the matrix form:

$$\underbrace{\begin{bmatrix} \gamma(1-1) & \gamma(1-2) & \dots & \gamma(1-n) \\ \gamma(2-1) & \gamma(2-2) & \dots & \gamma(2-n) \\ \vdots & & \ddots & \\ \gamma(n-1) & \gamma(n-2) & \dots & \gamma(n-n) \end{bmatrix}}_{\Gamma_n} \underbrace{\begin{bmatrix} \phi_{n1} \\ \phi_{n2} \\ \vdots \\ \phi_{nn} \end{bmatrix}}_{\phi_n} = \underbrace{\begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(n) \end{bmatrix}}_{\gamma_n},$$

i.e.

$$\Gamma_n \phi_n = \gamma_n.$$

Linear One-step-ahead Prediction: Prediction Equations

If matrix Γ_n is invertible,[2] solution to

$$\Gamma_n \phi_n = \gamma_n$$

is given by

$$\phi_n = \Gamma_n^{-1} \gamma_n$$

and the prediction is

$$\begin{aligned} x_{n+1}^n &= \phi_{n1} x_n + \phi_{n2} x_{n-1} + \ldots + \phi_{nn} x_1 \\ &= \underbrace{\left[\phi_{n1}, \phi_{n2}, \ldots \phi_{nn}\right]}_{\phi_n'} \underbrace{\begin{bmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_1 \end{bmatrix}}_{x} = \phi_n' x. \end{aligned}$$

[2] If Γ_n is not invertible, x_{n+1}^n are still unique.

Mean Square Error of Linear One-step-ahead Predictor

Claim.
If Γ_n is invertible,
then the mean square 1-step-ahead prediction error obtained with a linear predictor is

$$P_{n+1}^{n} = \gamma(0) - \gamma_n' \Gamma_n^{-1} \gamma_n.$$

Example 3.19 Prediction for an AR(2)

Let's consider the AR(2) process: $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$, then

1. if $n = 1$ (one observation only),

$$x_2^1 = \phi_{11} x_1 = \ldots = \phi_1 x_1$$

2. if $n = 2$ (two observations),

$$x_3^2 = \phi_{21} x_2 + \phi_{22} x_1 = \ldots = \phi_1 x_2 + \phi_2 x_1$$

3. if $n \geq 2$ (at least two observations),

$$x_{n+1}^n = \phi_{n1} x_n + \phi_{n2} x_{n-1} + \ldots + \phi_{nn} x_1 = \ldots = \phi_1 x_n + \phi_2 x_{n-1}$$

The Durbin–Levinson Algorithm

Algorithm:

1. Set $\phi_{00} = 0$ and $P_1^0 = \gamma(0)$
2. For $n \geq 1$,

$$\phi_{nn} = \frac{\rho(n) - \sum_{k=1}^{n-1} \phi_{n-1,k}\rho(n-k)}{1 - \sum_{k=1}^{n-1} \phi_{n-1,k}\rho(k)}, \quad P_{n+1}^n = P_n^{n-1}(1 - \phi_{nn}^2),$$

where, for $n \geq 2$,

$$\phi_{nk} = \phi_{n-1,k} - \phi_{nn}\phi_{n-1,n-1}, \quad k = 1, 2, \ldots, n-1.$$

Contents

Method of Moments

<u>Method of Moments:</u>
Assume that a distribution can be parametrized by $\theta_1, \theta_2, \dots, \theta_d$, then given a sample of i.i.d. $X_1, X_2, \dots, X_n$ from this distribution, the unknown θ_k can be obtained as the solution to the following system:

$$\begin{aligned}
\mathrm{E}[X] &= m_1 = \frac{1}{n}\sum_{k=1}^{n} X_k, \\
\mathrm{E}[X^2] &= m_2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2, \\
&\vdots \\
\mathrm{E}[X^d] &= m_d = \frac{1}{n}\sum_{k=1}^{n} X_k^d,
\end{aligned}$$

where $m_j = \frac{1}{n}\sum_{k=1}^{n} X_k^j$ is called *j's moment.*

Contents

Maximum Likelihood Estimation (MLE)

Maximum likelihood estimation:

Assume that a distribution can be parametrized by $\theta_1, \theta_2, \ldots, \theta_d$, then given a sample of i.i.d. $X_1, X_2, \ldots, X_n$ from this distribution $f(x|\theta_1, \theta_2, \ldots, \theta_d)$,
the unknown θ_k can be obtained by maximizing the likelihood function:

$$
\begin{aligned}
L(\theta_1, \theta_2, \ldots, \theta_d | X_1, X_2, \ldots, X_n) &= \underbrace{f(X_1, X_2, \ldots, X_n | \theta_1, \theta_2, \ldots, \theta_d)}_{\text{joint distribution}} \\
&= \prod_{k=1}^{n} f(X_k | \theta_1, \theta_2, \ldots, \theta_d).
\end{aligned}
$$

Contents

Exponential Distribution

Consider a sample of i.i.d. $X_1, X_2, \ldots, X_n$ from Exponential distribution

$$f(x|\lambda) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0, \\ 0, & \text{otherwise} \end{cases},$$

then

1. Method of Moments:
$$\hat{\lambda}_{\text{mm}} = \frac{1}{\bar{X}}$$
2. MLE:
$$\hat{\lambda}_{\text{mle}} = \frac{1}{\bar{X}}$$

Exponential Distribution

Consider a sample of i.i.d. $X_1, X_2, \ldots, X_n$ from Normal distribution

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R},$$

then

1. Method of Moments:

$$\hat{\mu}_{\text{mm}} = \bar{X} \quad \text{and} \quad \hat{\sigma}^2_{\text{mm}} = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \bar{X}^2$$

2. MLE:

$$\hat{\mu}_{\text{mle}} = \bar{X} \quad \text{and} \quad \hat{\sigma}^2_{\text{mle}} = \frac{1}{n}\sum_{k=1}^{n} (X_k - \bar{X})^2$$