

MATH 545: INTRODUCTION TO TIME SERIES ANALYSIS

WINTER 2025

MINIMUM MSE FORECASTING

We wish to make optimal predictions for future values of a time series process. In the simplest case, we presume that the series is stationary and Gaussian, and consider the prediction of X_{n+h} from an observed value of X_n . That is, we assume

$$\begin{bmatrix} X_n \\ X_{n+h} \end{bmatrix} \sim \mathcal{N}_2 \left(\begin{bmatrix} \mu \\ \mu \end{bmatrix}, \begin{bmatrix} \gamma_X(0) & \gamma_X(h) \\ \gamma_X(h) & \gamma_X(0) \end{bmatrix} \right)$$

Using the previously introduced theory for the multivariate Gaussian distribution, we deduce that

$$X_{n+h}|X_n = x_n \sim \text{Normal}(\mu + \gamma_X(h)\{\gamma_X(0)\}^{-1}(x_n - \mu), \gamma_X(0) - \gamma_X(h)\{\gamma_X(0)\}^{-1}\gamma_X(h))$$

and as $\gamma_X(h)/\gamma_X(0) = \rho_X(h)$, we have that

$$X_{n+h}|X_n = x_n \sim \text{Normal}(\mu + \rho_X(h)(x_n - \mu), \gamma_X(0)(1 - \rho_X^2(h))).$$

Minimum mean-square prediction error: Under the *minimum mean-square error* (minimum MSE) criterion, we attempt to make the prediction of X_{n+h} , denoted $\hat{x}_{n+h}$, based on x_n , such that

$$\mathbb{E}[(X_{n+h} - \hat{x}_{n+h})^2 | X_n = x_n]$$

is *minimized*. Here the expectation is taken with respect to the conditional distribution $X_{n+h}|X_n = x_n$.

This minimization can be carried out analytically: we have that

$$\hat{x}_{n+h} = \arg \min_{\eta} \int (x - \eta)^2 f_{X_{n+h}|X_n}(x) dx \quad (1)$$

and to achieve the minimization of the integral wrt η , we may differentiate under the integral, and equate the resulting form to zero. That is

$$-2 \int (x - \eta) f_{X_{n+h}|X_n}(x) dx = 0$$

so the solution is

$$\hat{x}_{n+h} = \int x f_{X_{n+h}|X_n}(x) dx = \mathbb{E}[X_{n+h}|X_n = x_n].$$

Therefore we have

$$\hat{x}_{n+h} = \mathbb{E}[X_{n+h}|X_n = x_n] = m_h(x) = \mu + \rho_X(h)(x_n - \mu) = m_X(h), \quad (2)$$

say. That is, the best prediction (measured in minimum prediction MSE terms) is the conditional expectation. The minimum prediction MSE corresponding to this prediction is

$$v_h(x) = \mathbb{E}[(X_{n+h} - m_X(h))^2 | X_n = x_n] = \text{Var}[X_{n+h}|X_n = x_n] = \gamma_X(0)(1 - \rho_X^2(h)) \equiv \sigma_X^2(1 - \rho_X^2(h))$$

say. Note that the minimization in (1) **always** leads to the same conclusion in (2), that is, that the minimum MSE prediction is obtained by using the conditional expectation. In the Gaussian case, this minimizer is **linear** in x_n .

We can extend this calculation to consider X_n as a random variable, and consider the minimization of

$$\mathbb{E}_{X_n, X_{n+h}} \left[(X_{n+h} - \hat{X}_{n+h})^2 \right]$$

where the expectation is taken over the joint distribution, and where $\hat{X}_{n+h}$ is a function, g say, of X_n . We have that

$$\begin{aligned} \mathbb{E}_{\hat{X}_{n+h}, X_{n+h}} \left[(X_{n+h} - \hat{X}_{n+h})^2 \right] &= \iint (x - u)^2 f_{\hat{X}_{n+h}, X_{n+h}}(u, x) dx du \\ &\equiv \iint (x - g(v))^2 f_{X_n, X_{n+h}}(v, x) dx dv \quad \hat{X}_{n+h} = g(X_n). \end{aligned}$$

Again assuming that g is linear, that is, $g(v) = av + b$, we can minimize the MSE analytically by choosing the optimal a and b . We may write the double integral

$$\iint (x - g(v))^2 f_{X_n, X_{n+h}}(v, x) dx dv = \iint (x - av - b)^2 f_{X_n, X_{n+h}}(v, x) dx dv$$

That is, we must solve simultaneously

$$\begin{aligned} a : \quad \frac{\partial}{\partial a} : \quad & -2 \iint v(x - av - b) f_{X_n, X_{n+h}}(v, x) dx dv = 0 \\ b : \quad \frac{\partial}{\partial b} : \quad & -2 \iint (x - av - b) f_{X_n, X_{n+h}}(v, x) dx dv = 0 \end{aligned}$$

that is

$$\begin{aligned} \mathbb{E}[X_n X_{n+h}] &= a\mathbb{E}[X_n^2] + b\mathbb{E}[X_n] \\ \mathbb{E}[X_{n+h}] &= a\mathbb{E}[X_n] + b \end{aligned}$$

Under stationarity, we know that $\mathbb{E}[X_{n+h}] = \mathbb{E}[X_n] = \mu$, so the second equation asserts that

$$\mu = a\mu + b \quad \implies \quad b = (1 - a)\mu$$

Substituting into the first equation, we have

$$\mathbb{E}[X_n X_{n+h}] = a\mathbb{E}[X_n^2] + (1 - a)\mu^2$$

or equivalently

$$\mathbb{E}[X_n X_{n+h}] - \mu^2 = a(\mathbb{E}[X_n^2] - \mu^2).$$

That is, we have that

$$a = \frac{\mathbb{E}[X_n X_{n+h}] - \mu^2}{\mathbb{E}[X_n^2] - \mu^2} = \frac{\gamma_X(h)}{\gamma_X(0)} = \rho_X(h)$$

and the optimal prediction is given by

$$aX_n + b = \rho_X(h)X_n + (1 - \rho_X(h))\mu = \mu + \rho_X(h)(X_n - \mu)$$

as per the original formula.