25 Spring 439/639 TSA: Lecture 18

Dr Sergey Kushnarev

Table of contents

Forecasting

If we observe $Y_1, ..., Y_t$, what is going to be $Y_{t+1}, Y_{t+2}, ...$?

Example: if $Y_1, ..., Y_t \sim \text{iid}(\mu, \sigma^2)$, how should we predict Y_{t+1}? The "best" prediction is μ

Exercise: Show that for a random variable X, $\arg\min_c \ \mathbb{E}[(X-c)^2] = \mathbb{E}X$. In other words, $\mathbb{E}[(X-\mathbb{E}X)^2] \leq \mathbb{E}[(X-c)^2]$ for any real number c.

The key idea to forecasting is **conditional expectation**. In general, the prediction is the conditional expectation. For example, given observations $(Y_1, ..., Y_t)$, the predicted value for Y_{t+1} should be

$$\mathbb{E}[Y_{t+1} \mid Y_1, ..., Y_t].$$

Basics

We first define some notations.

- t is the **forecast origin**.
- h is the **lead time**.
- $\widehat{Y}_t(h)$ is the predicted value at lead time h, i.e., prediction of Y_{t+h}.
- $e_t(h)$ is the forecast/prediction error, defined as

$$e_t(h) = \underbrace{Y_{t+h}}_{\text{actual value}} - \underbrace{\widehat{Y}_t(h)}_{\text{predicted value}}.$$

If we assume (Y_t) is a normal process, then the $(1-\alpha)100\%$ prediction interval for Y_{t+h} is

$$\left[\widehat{Y}_t(h) \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\text{Var}(e_t(h))}\right],$$

which means, with probability $1-\alpha$, Y_{t+h} will be in this interval.

Conditional expectation

The main tool for forecasting is conditional expectation. The predicted value $\widehat{Y}_t(h)$ is defined as

$$\widehat{Y}_t(h) \overset{\text{def}}{=} \mathbb{E}[Y_{t+h} \mid Y_1, ..., Y_t].$$

This is the "min squared error prediction", since it satisfies

$$\widehat{Y}_t(h) = \arg\min_c \ \mathbb{E}[(Y_{t+h} - c)^2 \mid Y_1, ..., Y_t].$$

Properties of conditional expectation:

- For a random variable X, a fixed number x, and a function g,

$$\mathbb{E}[g(X) \mid X = x] = \underbrace{g(x)}_{\text{a fixed number}}.$$

- For a random variable X and a function g,

$$\mathbb{E}[g(X) \mid X] = \underbrace{g(X)}_{\text{a random variable}}.$$

- If random variables X, Y are independent, then

$$\mathbb{E}[X \mid Y] = \mathbb{E}[X].$$

Example 1: trend plus noise

Consider the trend + noise model

$$Y_t = \underbrace{\mu_t}_{\text{deterministic}} + \underbrace{X_t}_{\text{noise}}, \quad X_t \sim \text{iid}(0, \sigma^2).$$

Then the prediction $\widehat{Y}_t(h)$ is

$$\begin{aligned}
\widehat{Y}_t(h) &= \mathbb{E}[Y_{t+h} \mid Y_1, ..., Y_t] = \mathbb{E}[\mu_{t+h} + X_{t+h} \mid Y_{1,...,t}] \\
&= \mathbb{E}[\mu_{t+h} \mid Y_{1,...,t}] + \mathbb{E}[X_{t+h} \mid Y_{1,...,t}] \\
&= \mu_{t+h} + \mathbb{E}[X_{t+h}] \\
&= \mu_{t+h}.
\end{aligned}$$

So the forecast error $e_t(h)$ is

$$e_t(h) = Y_t(h) - \widehat{Y}_t(h) = (\mu_{t+h} + X_{t+h}) - \mu_{t+h} = X_{t+h}.$$

And we can see that all the $e_t(h)$ have the same distribution that does not depend on h, since they are $\text{iid}(0, \sigma^2)$. If we further assume $X_t \overset{\text{iid}}{\sim} N(0, \sigma^2)$, then the 95% prediction interval for Y_{t+h} is

$$\left[\widehat{Y}_t(h) \pm 2\sigma\right] = [\mu_{t+h} - 2\sigma,\ \mu_{t+h} + 2\sigma],$$

and the prediction intervals have the same width for all h.

Example 2: AR(1)

Consider an AR(1) model with mean μ, (assume it is causal and stationary)

$$Y_t - \mu = \phi\,(Y_{t-1} - \mu) + e_t, \quad e_t \sim \text{iid}(0, \sigma_e^2).$$

The prediction $\widehat{Y}_t(1)$ is

$$\begin{aligned}
\widehat{Y}_t(1) &= \mathbb{E}\left[Y_{t+1} \mid Y_{1,...,t}\right] = \mathbb{E}\left[\mu + \phi\,(Y_t - \mu) + e_{t+1} \mid Y_{1,...,t}\right] \\
&= \mathbb{E}[\mu \mid Y_{1,...,t}] + \phi\,\mathbb{E}[Y_t - \mu \mid Y_{1,...,t}] + \mathbb{E}[e_{t+1} \mid Y_{1,...,t}] \\
&= \mu + \phi(Y_t - \mu) + \mathbb{E}[e_{t+1}] \\
&= \mu + \phi(Y_t - \mu).
\end{aligned}$$

For $h \geq 2$, we have

$$\begin{aligned}
\widehat{Y}_t(h) &= \mathbb{E}\left[Y_{t+h} \mid Y_{1,\ldots,t}\right] = \mathbb{E}\left[\mu + \phi\left(Y_{t+h-1} - \mu\right) + e_{t+h} \mid Y_{1,\ldots,t}\right] \\
&= \mu + \phi\left(\mathbb{E}\left[Y_{t+h-1} \mid Y_{1,\ldots,t}\right] - \mu\right) + \mathbb{E}\left[e_{t+h} \mid Y_{1,\ldots,t}\right] \\
&= \mu + \phi\left(\widehat{Y}_t(h-1) - \mu\right).
\end{aligned}$$

Then we can get the following recursive result (for any $h \geq 1$)

$$\widehat{Y}_t(h) - \mu = \phi\left(\widehat{Y}_t(h-1) - \mu\right) = \phi^2\left(\widehat{Y}_t(h-2) - \mu\right) = \cdots = \phi^h(Y_t - \mu).$$

So the prediction $\widehat{Y}_t(h)$ is

$$\widehat{Y}_t(h) = \mu + \phi^h(Y_t - \mu).$$

Under the causality condition ($|\phi| < 1$): as $h \to \infty$, $\widehat{Y}_t(h) \to \mu$.

Next, we derive the prediction error $e_t(h)$.

$$e_t(h) = Y_{t+h} - \widehat{Y}_t(h) = Y_{t+h} - \mu - \phi^h(Y_t - \mu).$$

To deal with the $Y_{t+h} - \phi^h Y_t$ above, recall the GLP representation for this model is

$$Y_t = \mu + \sum_{j=0}^{\infty} \phi^j e_{t-j}.$$

Then we have

$$\begin{aligned}
e_t(h) = Y_{t+h} - \mu - \phi^h(Y_t - \mu) &= \sum_{j=0}^{\infty} \phi^j e_{t+h-j} - \phi^h \sum_{j=0}^{\infty} \phi^j e_{t-j} \\
&= \left(e_{t+h} + \phi e_{t+h-1} + \phi^2 e_{t+h-2} + \cdots\right) - \phi^h\left(e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \cdots\right) \\
&= e_{t+h} + \phi e_{t+h-1} + \cdots + \phi^{h-1} e_{t+1}.
\end{aligned}$$

From this we can see

$$\mathbb{E}[e_t(h)] = 0,$$

so the forecast is unbiased (*Note:* we always have $\mathbb{E}[e_t(h)] = 0$ under our definition of prediction, since $\mathbb{E}[Y_t(h)] = \mathbb{E}[\mathbb{E}[Y_t(h) \mid Y_{1,\ldots,t}]]$), and

$$\begin{aligned}
\operatorname{Var}(e_t(h)) &= \sigma_e^2(1 + \phi^2 + \cdots + \phi^{2h-2}) = \sigma_e^2 \frac{1 - \phi^{2h}}{1 - \phi^2} \quad \text{(using } \phi\text{)} \\
&= \sigma_e^2(\psi_0^2 + \psi_1^2 + \cdots + \psi_{h-1}^2) = \sigma_e^2 \sum_{j=0}^{h-1} \psi_j^2. \quad \text{(using GLP coefficients } \psi_j\text{)}
\end{aligned}$$

As $h \to \infty$, $\operatorname{Var}(e_t(h))$ increases and converges to γ_0 since

$$\operatorname{Var}(e_t(h)) = \sigma_e^2 \sum_{j=0}^{h-1} \psi_j^2 \to \sigma_e^2 \sum_{j=0}^{\infty} \psi_j^2 = \operatorname{Var}(Y_t) = \gamma_0.$$

If we assume the process is normal, 95% prediction interval for Y_{t+h} is

$$\left[\widehat{Y}_t(h) \pm 2\sigma_e \sqrt{\sum_{j=0}^{h-1} \psi_j^2}\right].$$

The width of prediction intervals increase and converge to some fixed number ($4\sqrt{\gamma_0}$).

Example 3: MA(1)

Consider an MA(1) model with mean μ, (assume this MA(1) is invertible)

$$Y_t = \mu + e_t - \theta e_{t-1}, \quad e_t \sim \text{iid}(0, \sigma_e^2).$$

The prediction $\widehat{Y}_t(1)$ is

$$\begin{aligned}\widehat{Y}_t(1) &= \mathbb{E}\left[Y_{t+1} \mid Y_{1,\ldots,t}\right] = \mathbb{E}\left[\mu + e_{t+1} - \theta e_t \mid Y_{1,\ldots,t}\right] \\ &= \mu - \theta\ \mathbb{E}\left[e_t \mid Y_{1,\ldots,t}\right] \\ &= \mu - \theta\ e_t,\end{aligned}$$

and the last step is because $e_t = g(Y_{1,\ldots,t})$ is a function of $Y_{1,\ldots,t}$, where the function g can be found from the invertible representation (AR(∞)) of the MA(1). *Note:* the AR(∞) representation for invertible MA(1) is $e_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j} = \sum_{j=0}^{\infty} \theta^j Y_{t-j}$ and we think it is approximately a function of $Y_{1,\ldots,t}$ since we assume t is large in practice.

After we got $\widehat{Y}_t(1) = \mu - \theta\ e_t$, the prediction error $e_t(1)$ is

$$e_t(1) = Y_{t+1} - \widehat{Y}_t(1) = Y_{t+1} - \mu + \theta\ e_t = e_{t+1}.$$

So $\mathbb{E}[e_t(1)] = 0$ and $\text{Var}(e_t(1)) = \sigma_e^2$. If we assume (Y_t) is a normal process, then the 95% prediction interval for Y_{t+1} is $\left[\widehat{Y}_t(1) \pm 2\sigma_e\right]$.

For lead time $h \geq 2$, we can show that $\widehat{Y}_t(h) = \mu$.

Exercise: verify $\widehat{Y}_t(h) = \mu$ for $h \geq 2$.

Then the prediction error $e_t(h)$ for $h \geq 2$ is

$$e_t(h) = Y_{t+h} - \widehat{Y}_t(h) = Y_{t+h} - \mu = e_{t+h} - \theta e_{t+h-1}.$$

So for $h \geq 2$, we have $\mathbb{E}[e_t(h)] = 0$ and $\text{Var}(e_t(h)) = (1 + \theta^2)\sigma_e^2 = \gamma_0$. This variance is a constant for any $h \geq 2$, does not depend on h.