

Arima Model - 6

강의 : 김성범 교수님

정리 : 김선중

December 16, 2022

이것은 ARIMA 모델 개요 - Part 6 강의에 대한 노트이다. 이번 내용은 계산이 많으므로, 설명을 길게 쓰는 것보다는 계산을 나열하는 것 위주로 정리했다. 다만, conditional expectation과 prediction interval에 대해서는 조금 공부해야 할 것 같아서 먼저 간략하게 적어보려고 했다. conditional expectation에 관해서는 적다보니 상당히 많은 양을 적게 되었고, 그래서 덕분에 확률론 전반에 대해 간략하게 복습을 한 셈이 되었다. prediction interval에 관해서는 wikipedia만 간략하게 읽어봤는데, 경우에 따라 prediction interval을 구하는 방법이 다른 것 같았다. 그래서 따로 정리하진 않았다. 나중에 시간이 나면 정리해보자.

지난 번에는 영어로 적었었는데, 이번에는 다시 한글로 적었다.

Contents

1 Preliminaries	3
1.1 Conditional Expectations	3
1.1.1 Probability Spaces and Random Variables	3
1.1.2 Expectations	5
1.1.3 Joint and Marginal Distributions	5
1.1.4 Conditional Probabilities and Conditinoal Expectations	7
1.1.5 Stochastic Processes and Conditional Expectations	10
1.2 Prediction Intervals	10
2 Prediction(Forecasting)	10
2.1 AR(1) model	11
2.1.1 Prediction on X_{t+1}	12
2.1.2 Prediction on X_{t+2}	13
2.1.3 Prediction on X_{t+3}	14
2.2 AR(2) model	14
2.2.1 Prediction on X_{t+1}	14
2.2.2 Prediction on X_{t+2}	15

아리마 모델 (ARIMA Model) - Part 6

Prediction (Forecasting)

Given $X_1, X_2, \dots, \hat{X}_t$, want to predict \hat{X}_{t+1} .

$\hat{X}_{t+1} - \hat{\hat{X}}_{t+1}$: error

$$\min E[(\hat{X}_{t+1} - \hat{\hat{X}}_{t+1})^2]$$

$$\hat{X}_{t+1} = E[X_{t+1} | X_1, \dots, X_t] \rightarrow \text{conditional expectation.}$$

$$AR(1): X_t = \phi X_{t-1} + a_t + \mu$$

$$\hat{X}_{t+1} = \phi X_t + a_{t+1} + \mu$$

$$\begin{aligned} \hat{X}_{t+1} &= E[\hat{X}_{t+1} | X_1, \dots, X_t] \\ &= \phi E[X_t | X_1, \dots, X_t] + E[a_{t+1} | X_1, \dots, X_t] + E[\mu | X_1, \dots, X_t] \\ &= \phi X_t + \mu \end{aligned}$$

$$\therefore \hat{X}_{t+1} = \phi X_t + \mu$$

Seoung Bum Kim - Copyright © All rights reserved.

$$X_{t+1} \rightarrow \hat{X}_{t+1} \quad (\text{P.I.})$$

$(1-\alpha) \cdot 100\%$ prediction interval for X_{t+1} ,

$$\hat{X}_{t+1} \pm c.v.(\alpha) \sqrt{\text{Var}(\hat{X}_{t+1})}$$

$$X_{t+1} = \phi X_t + a_{t+1} + \mu$$

$$\hat{X}_{t+1} = \phi X_t + a_{t+1} + \mu$$

$$X_{t+1} - \hat{X}_{t+1} = a_{t+1}$$

$$V(X_{t+1} - \hat{X}_{t+1}) = V(a_{t+1}) = \tilde{\sigma}^2$$

Assume $a_t \sim N(0, \tilde{\sigma}^2)$

$(1-\alpha) \cdot 100\%$ P.I. for X_{t+1}

$$\hat{X}_{t+1} \pm Z(1-\frac{\alpha}{2}) \cdot \tilde{\sigma}$$

Seoung Bum Kim - Copyright © All rights reserved.

$$AR(2): X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t.$$

$$\hat{X}_{t+1} = \mu + \phi_1 X_t + \phi_2 X_{t-1} + a_{t+1}.$$

$$\begin{aligned} \hat{X}_{t+1} &= E[X_{t+1} | X_1, \dots, X_t] \\ &= E[\mu | X_1, \dots, X_t] + \phi_1 E[X_t | X_1, \dots, X_t] + \phi_2 E[X_{t-1} | X_1, \dots, X_t] + E[a_{t+1} | X_1, \dots, X_t] \\ &= \mu + \phi_1 X_t + \phi_2 X_{t-1} \end{aligned}$$

$(1-\alpha) \cdot 100\%$ P.I. for $X_{t+1} \Rightarrow \hat{X}_{t+1} \pm c.v.(k) \sqrt{V(\hat{X}_{t+1})}$

$$X_{t+1} = \mu + \phi_1 X_t + \phi_2 X_{t-1} + a_{t+1}$$

$$\hat{X}_{t+1} = \mu + \phi_1 X_t + \phi_2 X_{t-1}$$

$$X_{t+1} - \hat{X}_{t+1} = a_{t+1}$$

$$V(X_{t+1} - \hat{X}_{t+1}) = \tilde{\sigma}^2$$

$\Rightarrow a_t \sim N(0, \tilde{\sigma}^2)$

$$\hat{X}_{t+1} \pm Z(1-\frac{\alpha}{2}) \cdot \tilde{\sigma}$$

Seoung Bum Kim - Copyright © All rights reserved.

P.I. for X_{t+1} , $\phi \neq 0$: (부사1)

$$\bullet X_t = \mu + a_t + \frac{1}{1-\phi} a_{t-1} + \frac{\phi}{1-\phi} a_{t-2} + \dots, \quad \phi_1 = \phi + \theta, \phi_2 = \phi^2 + \phi \cdot \theta, \dots$$

$$\hat{X}_{t+1} = \mu + a_{t+1} + \phi_1 a_t + \phi_2 a_{t-1} + \dots$$

$$\hat{X}_{t+1} = E[X_{t+1} | X_1, \dots, X_t]$$

$$= \mu + E[a_{t+1} | X_1, \dots, X_t] + \frac{\phi}{1-\phi} E[a_t | X_1, \dots, X_t] + \frac{\phi^2}{1-\phi} E[a_{t-1} | X_1, \dots, X_t] + \dots$$

$$= \mu + \phi_1 a_t + \phi_2 a_{t-1} + \dots$$

$$\hat{X}_{t+1} = \mu + a_{t+1} + \phi_1 a_t + \phi_2 a_{t-1} + \dots$$

$$\hat{X}_{t+1} = \mu + \phi_1 a_t + \phi_2 a_{t-1} + \dots$$

$$V(X_{t+1} - \hat{X}_{t+1}) = \theta V(a_{t+1}) = \tilde{\sigma}^2$$

$(1-\alpha) \cdot 100\%$ P.I. for X_{t+1}

$$\hat{X}_{t+1} \pm Z(1-\frac{\alpha}{2}) \cdot \tilde{\sigma}$$

Seoung Bum Kim - Copyright © All rights reserved.

Prediction (Forecasting)

Given $X_1, X_2, \dots, \hat{X}_t$, want to predict \hat{X}_{t+1} .

$\hat{X}_{t+1} - \hat{\hat{X}}_{t+1}$: error

$$\min E[(\hat{X}_{t+1} - \hat{\hat{X}}_{t+1})^2]$$

$$\hat{X}_{t+1} = E[X_{t+1} | X_1, \dots, X_t] \rightarrow \text{conditional expectation.}$$

$$AR(1): X_t = \phi X_{t-1} + a_t + \mu$$

$$\hat{X}_{t+1} = \phi X_t + a_{t+1} + \mu$$

$$\begin{aligned} \hat{X}_{t+1} &= E[\hat{X}_{t+1} | X_1, \dots, X_t] \\ &= \phi E[X_t | X_1, \dots, X_t] + E[a_{t+1} | X_1, \dots, X_t] + E[\mu | X_1, \dots, X_t] \\ &= \phi X_t + \mu \end{aligned}$$

$$\therefore \hat{X}_{t+1} = \phi X_t + \mu$$

Seoung Bum Kim - Copyright © All rights reserved.

$$X_{t+2} \rightarrow \hat{X}_{t+2} \quad (\text{P.I.})$$

$(1-\alpha) \cdot 100\%$ prediction interval for X_{t+2} ,

$$\hat{X}_{t+2} \pm c.v.(\alpha) \sqrt{\text{Var}(\hat{X}_{t+2})}$$

$$X_{t+2} = \phi X_t + a_{t+1} + \mu$$

$$\hat{X}_{t+2} = \phi X_t + a_{t+1} + \mu$$

$$X_{t+2} - \hat{X}_{t+2} = a_{t+1}$$

$$V(X_{t+2} - \hat{X}_{t+2}) = V(a_{t+1}) = \tilde{\sigma}^2$$

Assume $a_t \sim N(0, \tilde{\sigma}^2)$

$(1-\alpha) \cdot 100\%$ P.I. for X_{t+2}

$$\hat{X}_{t+2} \pm Z(1-\frac{\alpha}{2}) \cdot \tilde{\sigma}$$

Seoung Bum Kim - Copyright © All rights reserved.

$$AR(2): X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t.$$

$$\hat{X}_{t+2} = \mu + \phi_1 X_t + \phi_2 X_{t-1} + a_{t+2}.$$

$$\begin{aligned} \hat{X}_{t+2} &= E[X_{t+2} | X_1, \dots, X_t] \\ &= E[\mu | X_1, \dots, X_t] + \phi_1 E[X_t | X_1, \dots, X_t] + \phi_2 E[X_{t-1} | X_1, \dots, X_t] + E[a_{t+2} | X_1, \dots, X_t] \\ &= \mu + \phi_1 X_t + \phi_2 X_{t-1} \end{aligned}$$

$(1-\alpha) \cdot 100\%$ P.I. for $X_{t+2} \Rightarrow \hat{X}_{t+2} \pm c.v.(k) \sqrt{V(\hat{X}_{t+2})}$

$$X_{t+2} = \mu + \phi_1 X_t + \phi_2 X_{t-1} + a_{t+2}$$

$$\hat{X}_{t+2} = \mu + \phi_1 X_t + \phi_2 X_{t-1}$$

$$X_{t+2} - \hat{X}_{t+2} = a_{t+2}$$

$$V(X_{t+2} - \hat{X}_{t+2}) = \tilde{\sigma}^2$$

$\Rightarrow a_t \sim N(0, \tilde{\sigma}^2)$

$$\hat{X}_{t+2} \pm Z(1-\frac{\alpha}{2}) \cdot \tilde{\sigma}$$

Seoung Bum Kim - Copyright © All rights reserved.

P.I. for X_{t+2} , $\phi \neq 0$: (부사1)

$$\bullet X_t = \mu + a_t + \frac{1}{1-\phi} a_{t-1} + \frac{\phi}{1-\phi} a_{t-2} + \dots, \quad \phi_1 = \phi + \theta, \phi_2 = \phi^2 + \phi \cdot \theta, \dots$$

$$\hat{X}_{t+2} = \mu + a_{t+2} + \phi_1 a_t + \phi_2 a_{t-1} + \dots$$

$$\hat{X}_{t+2} = E[X_{t+2} | X_1, \dots, X_t]$$

$$= \mu + E[a_{t+2} | X_1, \dots, X_t] + \frac{\phi}{1-\phi} E[a_t | X_1, \dots, X_t] + \frac{\phi^2}{1-\phi} E[a_{t-1} | X_1, \dots, X_t] + \dots$$

$$= \mu + \phi_1 a_t + \phi_2 a_{t-1} + \dots$$

$$\hat{X}_{t+2} = \mu + a_{t+2} + \phi_1 a_t + \phi_2 a_{t-1} + \dots$$

$$\hat{X}_{t+2} = \mu + \phi_1 a_t + \phi_2 a_{t-1} + \dots$$

$$V(X_{t+2} - \hat{X}_{t+2}) = \theta V(a_{t+2}) = \tilde{\sigma}^2$$

$(1-\alpha) \cdot 100\%$ P.I. for X_{t+2}

$$\hat{X}_{t+2} \pm Z(1-\frac{\alpha}{2}) \cdot \tilde{\sigma}$$

Seoung Bum Kim - Copyright © All rights reserved.

Seoung Bum Kim - Copyright © All rights reserved.

ARMA(1,1)

$(1-\alpha) \cdot 100\%$ P.I. for X_{t+3}

$$\hat{X}_{t+3} \pm Z(1-\frac{\alpha}{2}) \cdot \sqrt{(1+\phi^2) \cdot \tilde{\sigma}^2}; \quad \phi = \phi + \theta,$$

$(1-\alpha) \cdot 100\%$ P.I. for X_{t+3} .

$$\hat{X}_{t+3} \pm Z(1-\frac{\alpha}{2}) \sqrt{(1+\phi^2 + \phi_2^2) \cdot \tilde{\sigma}^2}.$$

$$\phi_2 = \phi^2 + \phi \cdot \theta$$

Example) ARMA(1,1)

$X_t = 1 + 0.4X_{t-1} + \epsilon_t + 0.4\epsilon_{t-1}$

$X_1 = 3, X_2 = -1, X_3 = 2, X_4 = 4, X_5 = 1$

Q. calculate predicted value \hat{X}_6 and the corresponding 95% PI.

$X_6 = 1 + 0.4X_5 + \epsilon_6 + 0.4\epsilon_5$

$\hat{X}_6 = E[X_6 | X_1, \dots, X_5] = 1 + 0.4X_5 + 0.4\hat{\epsilon}_5$

$\hat{\epsilon}_6 = X_6 - \hat{X}_6 = 0.4X_5 + \epsilon_6 - 1 = 0.4 \cdot 1 + 0.4 \cdot (-2.05) = 0.58$

$\hat{\epsilon}_1 = 0$

$\hat{\epsilon}_2 = X_2 - 0.4X_1 - 0.4\hat{\epsilon}_1 - 1 = -1 - 0.4(3) - 1 = -3.2$

$\hat{\epsilon}_3 = X_3 - 0.4X_2 - 0.4\hat{\epsilon}_2 - 1 = 2 + 0.4(-3.2) - 1 = 2.68$

$\hat{\epsilon}_4 = X_4 - 0.4X_3 - 0.4\hat{\epsilon}_3 - 1 = 4 + 0.4(2.68) - 1 - 1.13$

$\hat{\epsilon}_5 = X_5 - 0.4X_4 - 0.4\hat{\epsilon}_4 - 1 = 1 + 0.4(4) - 0.4(1.13) - 1 = -2.05$

$\therefore \hat{X}_6 = 0.58$

Seoung Bum Kim - Copyright © All rights reserved.

P.I. for X_6 : $\hat{X}_6 \pm 1.96\sqrt{\hat{\sigma}^2} = 0.58 \pm 1.96$

P.I. for X_7 : $\hat{X}_7 \pm 1.96\sqrt{(1+\phi^2)\hat{\sigma}^2} = \phi + \theta = 0.4 + 0.4 = 0.8$

P.I. for X_8 : $\hat{X}_8 \pm 1.96\sqrt{(1+\phi^2+\phi^2)\hat{\sigma}^2} = 1.493 \pm 1.96\sqrt{(1+0.8^2+0.32^2) \cdot 1}$

$\phi_1 = \phi_2 = 0$
 $= 0.4 + 0.4 = 0.8$
 $= 0.32$

Seoung Bum Kim - Copyright © All rights reserved.

Partial autocorrelation function (PACF)

- PACF is a conditional function.
- The correlation between two variables under the assumption that we take into account the values of some of other set of variables.
- Example for regression

Y, X_1, X_2, X_3

The partial correlation between Y and X_3

$$= \frac{\text{Cov}(Y, X_3 | X_1, X_2)}{\sqrt{\text{V}(Y|X_1, X_2) \cdot \text{V}(X_3|X_1, X_2)}}$$

Seoung Bum Kim - Copyright © All rights reserved.

The partial autocorrelation between X_t and X_{t-2} .

$$= \frac{\text{Cov}(X_t, X_{t-2} | X_{t-1})}{\sqrt{\text{V}(X_t|X_{t-1}) \cdot \text{V}(X_{t-2}|X_{t-1})}}$$

the PA between X_t and X_{t-3}

$$= \frac{\text{Cov}(X_t, X_{t-3} | X_{t-1}, X_{t-2})}{\sqrt{\text{V}(X_t|X_{t-1}, X_{t-2}) \cdot \text{V}(X_{t-3}|X_{t-1}, X_{t-2})}}$$

the PA between X_t and X_{t-h}

$$= \frac{\text{Cov}(X_t, X_{t-h} | X_{t-1}, \dots, X_{t-h+1})}{\sqrt{\text{V}(X_t|X_{t-1}, \dots, X_{t-h+1}) \cdot \text{V}(X_{t-h}|X_{t-1}, \dots, X_{t-h+1})}}$$

Seoung Bum Kim - Copyright © All rights reserved.

1 Preliminaries

1.1 Conditional Expectations

Conditional expectation에 관해서는 이 자료(A Conditional expectation - Arizona Math)참조했다. 해당 자료의 A.1, A.2를 읽겨서 적으려고 했다. 그런데, 적다보니 확률론에 관한 일반적인 사실을 많이 적게 되었다. 사실 하고 싶었던 건 conditional expectation에 대해 정확하게 이해하는 거였고, 또 강의내용을 잘 이해하는 거였는데, 그러자면 확률과정(stochastic process)에 대해 이해할 수 있어야 했다. 확률과정에 대해서 배운 적은 없긴 하지만, 아마 이와 비슷한 걸 알고 있으면 되는 게 아닐까 하는 생각이 있다.

그렇다고 확률론의 모든 내용을 적은 건 당연히 아니다. 예를 들어 두 사건의 독립/종속에 대해서는 안 적었고 전체확률의 법칙(the law of total probability)같은 것도 안 적었다. 분산과 표준편차에 대해서도 일절 적지 않았다. 아무튼, 적다보니, 확률론에 대해 너무 많은 걸 적게 되었다. 그래도 이번 기회에 헷갈려했던 것들을 깔끔하게 정리하게 된 것 같다는 느낌이 든다.

1.1.1 Probability Spaces and Random Variables

확률공간 $(\Omega, \mathcal{F}, \mathbb{P})$ 을 고려하자. 늘 그렇듯이 Ω 는 표본공간(sample space)이고, \mathcal{F} 는 사건공간(event space)이고, \mathbb{P} 는 확률측도(probability measure)이다. 사건(event)이란, 표본 공간 Ω 의 부분집합이고, \mathcal{F} 는 사건들의 집합이지만 모든 사건들의 집합일 필요는 없다. 그보다는, \mathcal{F} 는 Ω 에 대한 σ -algebra이라서 세 가지 정도의 성질을 만족시키는 집합이다. \mathbb{P} 는 유한측도(finite measure)로서, $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ 인 함수이고 역시 몇가지 성질을 만족시킨다.

확률변수란 Ω 에서 \mathbb{R} 로 가는 함수이다. 즉, X 가 확률변수이면 $X : \Omega \rightarrow \mathbb{R}$ 이다. 이때 X 의 공역이 꼭 \mathbb{R} 일 필요는 없지만, 많은 경우에는 그렇게 하므로, 그냥 공역을 \mathbb{R} 이라고 하겠다.

확률변수에는 이산확률변수(discrete random variable)과 연속확률변수(continuous random variable)이 있다. 확률변수

X 의 치역이 이산적(discrete)이면 이산확률변수라고 하고 연속적(continuous)이면 연속확률변수라고 한다.

각각의 확률변수들은 확률분포(random distribution)을 가지고 있다. 이산확률변수의 확률분포는 이산확률분포(discrete random variable)라고 부르고, 이때의 확률분포는 확률질량함수 $x \mapsto \mathbb{P}(X = x)$ 로서 표현된다. 이때 확률질량함수는

- $0 \leq \mathbb{P}(X = x) \leq 1$
- $\sum_x \mathbb{P}(X = x) = 1$

을 만족하는 함수여야 한다. 연속확률변수의 확률분포는 연속확률분포(continuous random variable)라고 부르고, 이때의 확률분포는 확률밀도함수 $f(x)$ 로서 표현된다. 이때 확률밀도함수는

- $f(x) \geq 0$
- $\int_x f(x) dx = 1$
- $\int_a^b f(x) dx = \mathbb{P}(a \leq X \leq b)$

를 만족하는 함수여야 한다.

예를 들어, 주사위를 하나 던지는 시행을 생각하자. 그러면 표본공간은 $\Omega = \{1, 2, 3, 4, 5, 6\}$ 이고, 사건공간은 \mathcal{F} 는 Ω 의 면집합, 즉 Ω 의 모든 부분집합들의 집합으로 둘 수 있다. 즉,

$$\mathcal{F} = 2^\Omega = \{\emptyset, \{1\}, \{2\}, \dots, \{1, 2, 3, 4, 5, 6\}\}$$

이다. 그리고, 확률측도 \mathbb{P} 는, $A \in \mathcal{F}$ 에 대하여

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$$

이다. 다시 말해서, $A = \{1, 2\}$ 이면

$$\mathbb{P}(A) = \frac{|\{1, 2\}|}{|\{1, 2, 3, 4, 5, 6\}|} = \frac{2}{6} = \frac{1}{3}$$

이 된다. 주사위를 하나 던졌을 때 나온 눈을 2로 나누었을 때의 나머지를 X 라고 하면, 이 확률변수는 하나의 함수 $X : \{1, 2, 3, 4, 5, 6\} \rightarrow \mathbb{R}$ 로서

$$\begin{cases} X(1) = 1 \\ X(2) = 0 \\ X(3) = 1 \\ X(4) = 0 \\ X(5) = 1 \\ X(6) = 0 \end{cases}$$

로 정의된다. X 의 역은 $\{0, 1\}$ 이므로 X 는 이산확률변수이고, 확률질량함수는

$$\begin{cases} \mathbb{P}(X = 0) = \frac{1}{2} \\ \mathbb{P}(X = 1) = \frac{1}{2} \end{cases}$$

가 된다.

두번째 예로서, 수직선 위의 구간 $[0, 6]$ 에서 임의로 하나의 숫자를 선택하는 시행을 고려하자. 그러면 표본공간은 $\Omega = [0, 6]$ 이다. 사건공간 \mathcal{F} 에 대해 설명하는 건 좀 복잡하다. 아까처럼 \mathcal{F} 가 Ω 의 모든 부분집합들의 집합일 수는 없다. 아마 그러면 안될 것이다. 이유가 뭐였더라, 아마도 cantor set 같은게 문제가 됐던 것 같은데. 아, 그이유를 찾았다.¹ \mathcal{F} 는 Ω 의

¹<https://www.cantorsparadise.com/probability-and-a-little-of-borel-sets-a9bdc7be26c> 요약하면 singleton들의 확률값은 0이어야 하는데, 만약 \mathcal{F} 를 Ω 의 면집합이라고 둘 경우 모든 singleton들을 합집합한 집합($= \Omega$)에 대한 확률값이 0이 되어, 모순이 발생한다.

보렐집합(borel set)이어야 한다. 다시 말해서 $\Omega = [0, 6]$ 의 standard topology를 \mathcal{T} 라고 할 때,

$$\mathcal{F} = \bigcap \{\mathcal{S} \supset \mathcal{T} : \mathcal{S} \text{ is a } \sigma\text{-algebra of } \Omega\}$$

이다. 확률측도 \mathbb{P} 는 아까와 비슷하게 정의된다. $A \in \mathcal{F}$ 에 대하여

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$$

다만 아까 $|A|$ 가 집합 A 의 원소의 개수(cardinality)를 의미했다면, 이번의 $|A|$ 는 집합 A 의 길이, 즉, \mathbb{R} 의 Lebesgue measure이다. 다시 말해서 $A = [0, 2] \in \mathcal{F}$ 에 대하여

$$\mathbb{P}(A) = \frac{|[0, 2]|}{|[0, 6]|} = \frac{2}{6} = \frac{1}{3}$$

이다. $[0, 6]$ 에서 임의로 하나 고른 숫자를 2로 나눈 값을 X 라고 하면, 이 확률변수는 하나의 함수 $X : [0, 6] \rightarrow \mathbb{R}$ 로서

$$X(t) = \frac{t}{2} \quad (0 \leq t \leq 6)$$

로 정의된다. X 의 치역은 $[0, 3]$ 이므로 X 는 연속확률변수이고, 확률질량함수는

$$f(x) = \frac{1}{3} \quad (0 \leq x \leq 3)$$

이 된다.

1.1.2 Expectations

이산확률변수 X 에 대하여 기댓값(expectation) $\mathbb{E}[X]$ 은

$$\mathbb{E}[X] = \sum_x x \mathbb{P}(X = x)$$

이다. 연속확률변수 X 에 대하여는

$$\mathbb{E}[X] = \int_x x f(x) dx$$

와 같은 정의가 있다.

아까의 첫번째 예에서는

$$\mathbb{E}[X] = \sum_{x=0}^1 x \mathbb{P}(X = x) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2}$$

이 된다. 너무 장황하게 써서 어렵게 보이지만, 간단하게 쓰면 이렇다. ‘주사위를 한 개 던져서 나온 눈이 짝수이면 $X = 0$, 홀수이면 $X = 1$ 이라고 하자. 그러면 X 는 평균적으로 $\frac{1}{2}$ 의 값으로 기대된다.’ 두번째 예에서는

$$\mathbb{E}[X] = \int_0^3 x f(x) dx = \int_0^3 \frac{x}{3} dx = \left[\frac{1}{6} x^2 \right]_0^3 = \frac{3}{2}$$

가 된다. 이것도 간단히 묘사될 수 있다. ‘0부터 6까지의 실수 중 하나를 임의로 뽑은 다음 2로 나누면, 그것은 곧 0부터 3까지의 실수 중 하나를 임의로 뽑는 것과 같다. 그 숫자들의 평균은 당연히 $\frac{3}{2}$ 가 되어야 한다.’

1.1.3 Joint and Marginal Distributions

두 확률변수 X, Y 에 대하여는 이 확률변수들의 순서쌍 (X, Y) 이 하나의 확률변수가 된다. 이 확률변수의 분포를 결합 확률분포(joint distribution)라고 한다. 만약 X 와 Y 가 모두 이산확률변수이면, (X, Y) 또한 이산확률변수이고, 이때의

확률질량함수는

$$\mathbb{P}(X = x, Y = y)$$

와 같은 이변수함수가 된다. X 와 Y 에 대한 함수 $g(X, Y)$ 가 있을 때, $g(X, Y)$ 에 대한 기댓값 $\mathbb{E}[g(X, Y)]$ 은

$$\mathbb{E}[g(X, Y)] = \sum_x g(x, y) \mathbb{P}(X = x, Y = y)$$

가 된다. 간단한 예로, 확률변수 $X + 3Y$ 에 대한 기댓값 $\mathbb{E}[X + 3Y]$ 은 ($g(X, Y) = X + 3Y$)

$$\mathbb{E}[X + 3Y] = \sum_{x,y} (x + 3y) \mathbb{P}(X = x, Y = y)$$

이고, 확률변수 X^2Y 에 대한 기댓값 $\mathbb{E}[X^2Y]$ 은 ($g(x, Y) = X^2Y$)

$$\mathbb{E}[X^2Y] = \sum_x x^2 y \mathbb{P}(X = x, Y = y)$$

이다. 만약 X 와 Y 가 모두 연속확률변수이면 (X, Y) 또한 연속확률변수가 된다. X 가 연속확률변수라고 할 때에는 $X : \Omega_X \rightarrow \mathbb{R}$ 이라는 의미였지만, (X, Y) 가 연속확률변수라고 말할 때에는 $(X, Y) : \Omega_X \times \Omega_Y \rightarrow \mathbb{R}^2$ 이라는 의미가 된다. 확률변수 (X, Y) 의 확률밀도함수 $f_{X,Y}$ 는

$$\int_a^b \int_c^d f_{X,Y}(x, y) dx dy = \mathbb{P}(a \leq x \leq b, c \leq y \leq d)$$

를 만족시키는 이변수 함수이다. 함수 $g(X, Y)$ 에 대한 기댓값 $\mathbb{E}[g(X, Y)]$ 은

$$\mathbb{E}[g(X, Y)] = \iint_{x,y} g(x, y) f_{X,Y}(x, y) dx dy$$

이다. 위에서와 같은 예를 들어서 설명하면

$$\begin{aligned} \mathbb{E}[X + 3Y] &= \iint_{x,y} (x + 3y) f_{X,Y}(x, y) dx dy \\ \mathbb{E}[X^2Y] &= \iint_{x,y} x^2 y f_{X,Y}(x, y) dx dy \end{aligned}$$

이 된다.

결합확률분포의 관점에서 보면 X 의 확률분포(혹은 Y 에 대한 확률분포)는 (X, Y) 에 대한 주변확률분포(marginal distribution)이라고 말할 수 있다. X, Y 에 대한 확률질량함수[혹은 확률밀도함수]를 X 에 대하여 (혹은 Y 에 대하여) marginalize 하면 Y (혹은 X)에 대한 확률질량함수 [혹은 확률밀도함수]가 얻어지는 것이다. 이것을 수식으로 쓰면, 이산확률변수에 대해서는

$$\begin{aligned} \sum_x \mathbb{P}(X = x, Y = y) &= \mathbb{P}(Y = y) \\ \sum_y \mathbb{P}(X = x, Y = y) &= \mathbb{P}(X = x) \end{aligned}$$

이 성립하고, 연속확률변수에 대해서는

$$\int_x f_{X,Y}(x,y) = f_Y(y)$$

$$\int_y f_{X,Y}(x,y) = f_X(x)$$

이 성립한다는 말이 된다.

여러 개의 확률변수 X_1, X_2, \dots, X_N 에 대해서도 새로운 확률변수 (X_1, X_2, \dots, X_N) 를 생각할 수 있다. 이것은, 그러니까 확률변수들의 tuple인 것이다. 확률변수 X_n ($1 \leq n \leq N$)들이 모두 이산확률변수일 경우, 확률질량함수

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_N = x_N)$$

을 생각할 수 있고, 모두 연속확률변수인 경우, 확률밀도함수

$$f_{X_1, X_2, \dots, X_N}(x_1, x_2, \dots, x_N)$$

을 생각하게 된다.² 이것들은 당연히 이변수일 때와 비슷한 성질을 만족시킬 것이지만 여기에 적지는 않겠다. 함수 $g(X_1, X_2, \dots, X_N)$ 에 대한 기댓값은

$$\mathbb{E}[g(X_1, X_2, \dots, X_N)] = \sum_{x_1, x_2, \dots, x_N} g(x_1, x_2, \dots, x_N) \mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_N = x_N)$$

혹은

$$\mathbb{E}[g(X_1, X_2, \dots, X_N)] = \int_{x_1, x_2, \dots, x_N} g(x_1, x_2, \dots, x_N) f_{X_1, X_2, \dots, X_N}(x_1, x_2, \dots, x_N) dx_1 dx_2 \dots dx_N$$

이다.

1.1.4 Conditional Probabilities and Conditionoal Expectations

두 사건 A, B 에 대하여, ³ 사건 A 가 일어났다고 가정했을 때 사건 B 가 일어날 확률을 $\mathbb{P}(B|A)$ 라고 쓰고

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

라고 정의한다. 이런 확률을 조건부확률(conditional probability)이라고 부른다.

위의 정의는 고등학교때 이미 배우는 아주 기본적인 개념이긴 하지만, 조금 들어가면 헷갈린다. A 대신에 $X = x, B$ 대신에 $Y = y$ 같은 걸 넣는 것이다. 그러면 위의 식은

$$\mathbb{P}(Y = y | X = x) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)}$$

와 같이 쓰게 되는데 위의 식과는 꽤 달라보인다. 교집합이었던 것이 comma로 바뀌었고 집합이었던 것이 (A) 등식으로 바뀌었다($X = x$). 이건, 정확하게 말하면 $A = \{x\} \times Y, B = X \times \{y\}$ 로 보는 것이다. 그러면 당연히 $A \cap B = \{(x, y)\}$ 가 되어 딱 맞아떨어진다.

위의 식은 다른 의미에서도 상당히 까다로운 식이다. 저 식에서 분모는 0이 아니어야 하는데, 만약 X 가 연속확률변수이면, 분모는 0이 될 수밖에 없다. 그래서 참고한 자료에서는, X 가 연속확률변수일 때에 한해서 ϵ 과 극한을 사용해서

²그런데 위의 표현은 너무 복잡하므로, 그냥 간단하게

$$f_{\mathbf{X}}(x_1, x_2, \dots, x_N)$$

라고 쓸 수도 있을 것이다. 즉 $\mathbf{X} = (X_1, X_2, \dots, X_N)$ 이라고 생각하는 것이다.

³정확하게는, (X, Y) 의 결합확률분포에서의 사건 A, B 에 대하여

정의하고 있다. 이런 여러가지 어려움들이 있지만, 그런 어려움들이 다 극복되고, 저 식이 그저 잘 정의되어 있다고 가정해 보자. 그리고, 이제부터는 이산확률변수만 가정하자, 연속확률변수인 경우에는 summation이 integral로 적절히 바뀌면 될 것이다.

아까 기댓값 $\mathbb{E}[X]$ 을 다음과 같이 정의했었다.

$$\mathbb{E}[X] = \sum_x x\mathbb{P}(X = x).$$

이번에는 조건부 기댓값(conditional expectation) $\mathbb{E}[X|Y = y]$ 를 다음과 같이 정의한다.

$$\begin{aligned}\mathbb{E}[X|Y = y] &= \sum_x \mathbb{P}(X = x|Y = y) \\ &= \sum_x \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}\end{aligned}$$

예를 들어보자. 참고한 자료의 A.2 앞부분에 있는 예를 가져왔다. 거기서는 아주 빠르게 논의를 전개해나가고 있는데, 잘 이해가 되지 않으니, 하나하나 따져가면서 계산해보자. 주사위의 눈이 6으로 나올때까지 계속해서 주사위를 던질 때, Y 를 주사위를 던진 횟수, X 를 1이 나온 횟수라고 하자. 그러면

$$\mathbb{E}[X|Y = 1] = 0$$

이다. 왜냐하면, $Y = 1$ 이라는 뜻은, 한번에 6이 나왔다는 뜻이고, 그때의 X 의 값은 0일 수밖에 없기 때문이다. 즉, 조건부 $Y = 1$ 하에서 X 의 확률분포는 확률질량함수 $\mathbb{P}(X = x|Y = 1)$ 로 표현될 수 있는데, 가능한 x 는 오직 0 뿐이므로

$$\mathbb{P}(X = 0|Y = 1) = 1$$

이다. 따라서

$$\mathbb{E}[X|Y = 1] = \sum_x x\mathbb{P}(X = x|Y = 1) = 0 \times 1 = 0$$

인 것이다.

$Y = 2$ 인 경우는 어떨까? $\mathbb{E}[X|Y = 2]$ 의 값은 얼마일까? $Y = 2$ 라는 뜻은, 처음 시행에서는 6이 나오지 않았다가, 두번째 시행에서 6이 나왔다는 것이다. 그러면, 처음 시행에서 6이 안나온 건 확실한데, 1이 나왔는지, 아니면 2와 5 사이의 값 중에서 나왔는지를 따져야 한다. 조건부 $Y = 2$ 하에서 X 의 확률분포는 확률질량함수 $\mathbb{P}(X = x|Y = 2)$ 로 표현될 수 있다. 첫번째 시행에서 2,3,4,5 중 하나의 값이 나왔으면 $X = 0$, 첫번째 시행에서 1이 나왔으면 $X = 1$ 인 것이므로

$$\begin{cases} \mathbb{P}(X = 0|Y = 2) = \frac{4}{5} \\ \mathbb{P}(X = 1|Y = 2) = \frac{1}{5} \end{cases}$$

이다. 따라서

$$\mathbb{E}[X|Y = 2] = \sum_{x=0}^1 x\mathbb{P}(X = x|Y = 2) = 0 \times \frac{4}{5} + 1 \times \frac{1}{5} = \frac{1}{5}$$

이다.

다음으로 $Y = 3$ 인 경우이다. 처음 두 시행에서는 6이 나오지 않았다가, 세번째 시행에서 6이 나오는 그런 상황이다.

그렇다면 X 로 가능한 값은 0, 1, 2이고 (이제 감이 잡힌다.) X 는 $B(2, \frac{1}{5})$ 인 이항분포를 따른다. 다시 말해서

$$\begin{cases} \mathbb{P}(X = 0|Y = 3) = \binom{2}{0} \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^2 = \frac{16}{25} \\ \mathbb{P}(X = 1|Y = 3) = \binom{2}{1} \left(\frac{1}{5}\right)^1 \left(\frac{4}{5}\right)^1 = \frac{8}{25} \\ \mathbb{P}(X = 2|Y = 3) = \binom{2}{2} \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^0 = \frac{1}{25} \end{cases}$$

이다. 이걸 식에다가 넣어서

$$\mathbb{E}[X|Y = 3] = \sum_{x=0}^2 x \mathbb{P}(X = x|Y = 3) = 0 \times \frac{16}{25} + 1 \times \frac{8}{25} + 2 \times \frac{1}{25} = \frac{2}{5}$$

로 계산해도 되고, 아니면 그냥 $\mathbb{E}[X|Y = 3] = 2 \times \frac{1}{5} = \frac{2}{5}$ 로 계산해도 되는 것이다.

$Y = 4$ 이면 처음 세 시행에서는 6이 나오지 않았다가, 네번째 시행에서 6이 나온다. $X \sim B(3, \frac{1}{5})$ 이고, 따라서 $\mathbb{E}[X|Y = 4] = \frac{3}{5}$ 가 된다. 마찬가지로 $\mathbb{E}[X|Y = 5] = \frac{4}{5}$, $\mathbb{E}[X|Y = 6] = 1$, $\mathbb{E}[X|Y = 7] = \frac{6}{5}$ 등이 될 것이다. 이걸 일반적으로

$$\mathbb{E}[X|Y = y] = \frac{1}{5}(y - 1)$$

이라고 쓸 수 있다.

여기까지 잘 이해가 된다. 잘 이해가 되나? 아니다, 잘 이해 된다. 여기서의 핵심은, $\mathbb{E}[X|Y = y]$ 를 계산한 것인데, y 값이 하나 주어질 때마다 $\mathbb{E}[X|Y = y]$ 값이 하나로 정해진다는 것이다. 각각의 y 에 대하여 $\mathbb{E}[X|Y = y]$ 의 값을 explicit하게 numerical value로 구할 수 있다.

문제는, $\mathbb{E}[X|Y = y]$ 와 같은 종류의, 비교적 간단한 conditional expectation 말고도 $\mathbb{E}[X|Y]$ 와 같은 종류의 conditional expectation도 있다는 것이다. 핵심만 간단히 먼저 말하면, 이번 종류의 conditional expectation은 하나의 값으로 딱 떨어지게 나오는 것이 아니라, Y 에 의존하는 값으로 나온다는 것이다. $\mathbb{E}[X|Y]$ 는 X 에 대한 기댓값을 계산하는 것이기 때문에, X 에 관한 식이 나오지 않는다. 다시 말해서, 그 결과값에 X 가 포함되지 않는다. 그런데 조건부에 걸린 Y 의 값이 특정되지 않았기 때문에 $\mathbb{E}[X|Y]$ 는 Y 에 대한 함수가 되는 것이다. 이에 대한 설명이 참고자료에 적혀있으니 잘 따라가보자.

아까, 확률변수가 Ω 에서 \mathbb{R} 로 가는 함수였던 것을 상기해보자. 표본공간인 Ω 는 주사위를 여러 번 던졌을 때 나올 수 있는 모든 경우들에 대한 집합이다. 만약 주사위의 눈이 차례로 1, 3, 2, 3, 6이 나왔다면, 이것은 $X = 1, Y = 5$ 인 상황을 나타내는데, $\omega = (1, 3, 2, 3, 6)$ 으로 표현할 수 있을 것이다. 그러니까, Ω 를 굳이 표현하자면

$$\Omega = \bigcup_{n=1}^{\infty} \{1, 2, 3, 4, 5, 6\}^{n-1} \times \{6\}$$

가 될 것이다. 참고자료에서는 ‘ ω would be a string of 1, 2, 3, 4, 5’s ending with a 6.’라고 되어 있다.

새로 정의할 conditional expectation인 $\mathbb{E}[X|Y]$ 는, 확률변수로서 해석된다. 다시 말해서 $\mathbb{E}[X|Y] : \Omega \rightarrow \mathbb{R}$ 인데,

$$\mathbb{E}[X|Y] : \omega \mapsto \mathbb{E}[X|Y = y]$$

로 정의되는 확률변수라는 것이다. 이때, $y = Y(\omega)$ 이다. 말이 너무 길었다. 다시 한 문장으로 표현하자. ‘ $\mathbb{E}[X|Y]$ 는 확률변수로서, string $\omega = (a_1, a_2, \dots, a_{n-1}, 6)$ 이 주어져 있을 때, 이 string을 $\mathbb{E}[X|Y = Y(\omega)]$ 으로 대응시키는 확률변수이다.’ 여전히 복잡해보인다. 하지만, 아까 언급한 예를 통해서 보면 좀 이해가 갈지도 모른다.

$\mathbb{E}[X|Y]$ 는 ω 를 $\mathbb{E}[X|Y = y]$ 로 대응시킨다. ($y = Y(\omega)$) 그런데 그 값은 $\frac{1}{5}(y - 1)$ 이라고 계산되었었으니까,

$$\mathbb{E}[X|Y](\omega) = \frac{1}{5}(Y(\omega) - 1)$$

라고 쓸 수 있다. 확률변수는 함수라고 했었다. 그러면 왼쪽도 함숫값 오른쪽도 함숫값이다. 그런데 임의의 ω 에 대하여 함숫값이 일치하므로, 두 함수가 같다는 말이 된다. 다시 말해서 좌변과 우변에 있는 확률변수가 일치한다는 말이다. 따라서 다음과 같이 쓸 수 있다.

$$\mathbb{E}[X|Y] = \frac{1}{5}(Y - 1)$$

완전히 이해가 된 듯하다. 따라서 $\mathbb{E}[X|Y]$ 라고 하는 이 새로운 대상은, 확률변수인데, 기존의 확률변수 Y 로 표현될 수 있다는 것이다.

1.1.5 Stochastic Processes and Conditional Expectations

시계열에서 다루는 대상은 기본적으로 확률과정(stochastic process)이다. 여기에서 말하는 확률과정이란, ‘확률변수들의 수열’을 의미한다. 다시 말해서

$$X_1, X_2, \dots, X_t, X_{t+1}$$

과 같은 확률과정을 고려하는 것이다.

이번 강의에서

$$\mathbb{E}[X_{t+1}|X_1, X_2, \dots, X_t]$$

와 같은 conditional expectation이 자주 등장한다. 아까의 $\mathbb{E}[X|Y]$ 보다도 더 복잡해졌지만, 기본적으로 같은 것이라고 생각하면 될 것 같다. X_1, X_2, \dots, X_t 라고 표시된 것을 그냥 t 개의 확률변수들이 결합된 새로운 확률변수(jointly distributed random variable)라고 생각하면 되는 것이다.

기억해야 하는 것은 다음 두 가지이다.

- (1) $\mathbb{E}[X_{t+1}|X_1, X_2, \dots, X_t]$ 는 확률변수이다.
- (2) $\mathbb{E}[X_{t+1}|X_1, X_2, \dots, X_t]$ 는 X_1, X_2, \dots, X_t 에 대한 함수로 나타날 것이다.

1.2 Prediction Intervals

2 Prediction(Forecasting)

이번 시간에 하고 싶은 것은 예측이다. 구체적으로는 X_1, X_2, \dots, X_t 를 가지고 X_{t+1} 을 예측하고 싶다. 예측값은 \hat{X}_{t+1} 으로 쓴다. 이때 $X_{t+1} - \hat{X}_{t+1}$ 을 오차라고 할 수 있는데 (오차가 맞나? 잔차가 아닌가?) 이 오차의 제곱에 대한 기댓값 $\mathbb{E}[(X_{t+1} - \hat{X}_{t+1})^2]$ 이 최소가 되는 \hat{X}_{t+1} 의 값을 \hat{X}_{t+1} 로 정한다. 단어 두 개가 한꺼번에 쓰여서 혼동된다. 정확하게 쓰면 다음과 같이 써야 할 것이다.

하고 싶은 것은 X_1, X_2, \dots, X_t 가 주어져 있을 때 예측값 \hat{X}_{t+1} 을 결정하는 것이다. 그런데 어떤 걸 예측값으로 할지는 아직 정하지 않았으므로, 예측값의 후보를 \tilde{X}_{t+1} 이라고 적자. (hat 대신 tilde를 썼다.) 여러 예측값들의 후보인 \tilde{X}_{t+1} 중에서 가장 괜찮은 것을 예측값 X_{t+1} 이라고 부를 것이다. 그러면 가장 괜찮은 예측값이라는 것은 무엇인가? 오차가 가장 적은 예측값을 말한다. 오차란 $X_{t+1} - \tilde{X}_{t+1}$ 를 뜻한다. 오차가 가장 작은 것보다는 오차의 제곱이 가장 작은 때가 좋겠다. 그런데 오차의 제곱인 $(X_{t+1} - \tilde{X}_{t+1})^2$ 은 확률변수로서 이해되어야 한다. 어떤 확률변수가 최소가 되는 상황을 고려하는 것보다는, 그 확률변수의 평균이 최소가 되는 상황을 고려하는 것이 더 말이 된다. 그래서 최적의 예측값 \hat{X}_{t+1} 을

$$\mathbb{E}[(X_{t+1} - \tilde{X}_{t+1})^2]$$

이 최소가 되는 \tilde{X}_{t+1} 의 값으로 정하는 것이다. 조금 더 정확하게는

$$\mathbb{E}[(X_{t+1} - \tilde{X}_{t+1})^2 | X_1, X_2, \dots, X_t]$$

가 최소가 되는 \tilde{X}_{t+1} 의 값이다. 기호로 한 번에 쓰면

$$\hat{X}_{t+1} = \arg \min_{\tilde{X}_{t+1}} \mathbb{E} [(X_{t+1} - \tilde{X}_{t+1})^2 | X_1, X_2, \dots, X_t] \quad (1)$$

이 될 것이다. 그리고 이 문제 (1)에 대한 답은

$$\hat{X}_{t+1} = \mathbb{E}[X_{t+1} | X_1, X_2, \dots, X_t] \quad (2)$$

이 된다고 한다.

교수님이 이에 대한 유도과정을 생략한다고 했는데, (1)에 대한 답이 (2)처럼 나온다는 것은, 분산과 평균의 관점에서 보면 당연하다. 확률변수 X 에 대한 평균을 m 이라고 해보자 ($\mathbb{E}[X] = m$). 이때, m 은 실수 n 에 대한 함수

$$\mathbb{E}[(X - n)^2] \quad (3)$$

이 최소가 되는 n 의 값이라고 생각할 수 있다. 아까 argmin를 사용해 표현한 방식으로 쓰면

$$m = \arg \max_n \mathbb{E}[(X - n)^2] \quad (4)$$

라고 쓸 수 있다. 왜냐하면 (3)을 전개하여

$$\begin{aligned} \mathbb{E}[(X - n)^2] &= n^2 - 2\mathbb{E}[X] + \mathbb{E}[X^2] \\ &= (n - \mathbb{E}[X])^2 + \mathbb{E}[X]^2 - \mathbb{E}[X^2] \end{aligned}$$

으로 쓸 수 있기 때문이다. 중학교 3학년때 배우는 식으로 해석하면 위의 (n 에 대한) 이차함수는 $n = \mathbb{E}[X]$ 일 때, 최솟값 $\mathbb{E}[X]^2 - \mathbb{E}[X^2]$ 을 가진다. 그리고 이 최솟값은 정확히 분산인 $\text{V}(X)$ 와 같다. 즉

$$m = \mathbb{E}[X] \quad (5)$$

이다. (1), (2)의 관계는 (4), (5)의 관계와 거의 같다. 차이는 조건부가 붙느냐 붙지 않았느냐의 차이밖에는 없는 것이다.

2.1 AR(1) model

이번에는 AR(1) 모델에 대하여 X_1, X_2, \dots, X_t 가 주어졌을 때, X_{t+1} 과 X_{t+2} 를 예측해보려고 한다.

이번 강의의 내용은 계산이 워낙 많아서, 설명은 최대한 배제하고 계산만 적어보려고 했었다. 그런데, 그래도 처음 계산에 대해서는 설명을 좀 적어야 할 것 같다.

AR(1) 모델은 원래

$$X_t = \phi X_{t-1} + a_t \quad (6)$$

라는 식이었다. 하지만 이번에는

$$X_t = \phi X_{t-1} + a_t + \mu \quad (7)$$

로 적고 있다. 새로 적혀 있는 이 μ 에 대해서 교수님은 ‘constant term’이라고만 말씀하고 계시다. 이에 관한 내 생각은 다음과 같다. (6)에서는 X_t 의 평균이 0임을 가정했었다. 그런데, 평균이 0이 아니고 μ_X 인 상황을 생각하면 (6)은

$$X_t - \mu_X = \phi(X_{t-1} - \mu_X) + a_t$$

와 같이 쓰게 된다. 이 식을 정리한 것이 (7)이다. 다시 말해서, $\mu = (1 - \phi)\mu_X$ 이다. 따라서 μ 는 t 에는 의존하지 않는 값이다. (그런 의미에서 constant term이다.) 이것은 Shumway의 책에서 (3.1), (3.2)의 식에 해당된다.

2.1.1 Prediction on X_{t+1}

식 (7)의 t 대신 $t + 1$ 을 대입하여

$$X_{t+1} = \phi X_t + a_{t+1} + \mu \quad (8)$$

을 만들고 여기에 $\mathbb{E}[\cdot | X_1, X_2, \dots, X_t]$ 를 취하여 (2)를 대입하면 X_{t+1} 에 대한 point estimate(점 예측값)인 \hat{X}_{t+1} 을 얻을 수 있다 :

$$\begin{aligned} \hat{X}_{t+1} &= \mathbb{E}[X_{t+1} | X_1, X_2, \dots, X_t] \\ &= \mathbb{E}[\phi X_t + a_{t+1} + \mu | X_1, X_2, \dots, X_t] \\ &= \phi [X_t | X_1, X_2, \dots, X_t] + \mathbb{E}[a_{t+1} | X_1, X_2, \dots, X_t] + \mathbb{E}[\mu | X_1, X_2, \dots, X_t] \\ &= \phi X_t + 0 + \mu \\ &= \phi X_t + \mu \end{aligned} \quad (9)$$

이번에는 $\mathbb{V}[\hat{X}_{t+1}]$ 을 구해야 한다. X_{t+1} 의 prediction interval(예측 구간)을 구하기 위함이다. 이를 위해서

$$\mathbb{V}[X_{t+1}] = \mathbb{V}[\hat{X}_{t+1} - X_{t+1}]$$

임을 사용한다. 다시 말해서, X_{t+1} 는 \hat{X}_{t+1} 의 관점에서 보면 상수라는 것을 가정하고 있다. 그건 맞는 말인 것 같다. X_{t+1} 의 범위를 추정하는 게 지금 하려는 것이기는 하지만, X_{t+1} 은 고정된 값이라고 가정할 수 있는 것이다. 물론, X_{t+1} 은 그 자체로 하나의 확률변수이기는 하지만, \hat{X}_{t+1} 을 추정하는 데 있어서는 X_1, X_2, \dots, X_t 를 가지고 추정하는 것인지 X_{t+1} 을 가지고 추정하지는 않는다. 그런 의미에서 X_{t+1} 은 \hat{X}_{t+1} 의 관점에서 상수이다.

(8)와 (9)을 빼서

$$X_{t+1} - \hat{X}_{t+1} = a_{t+1}$$

을 얻고 여기에 $\mathbb{V}[\cdot]$ 를 취하면

$$\mathbb{V}[\hat{X}_{t+1}] = \mathbb{V}[X_{t+1} - \hat{X}_{t+1}] = \mathbb{V}[a_{t+1}] = \sigma_a^2$$

가 된다.

X_{t+1} 에 대한 $(1 - \alpha) \cdot 100\%$ prediction interval은

$$\left[\hat{X}_{t+1} - z_{(1-\frac{\alpha}{2})} \cdot \sqrt{\mathbb{V}[\hat{X}_{t+1}]}, \hat{X}_{t+1} + z_{(1-\frac{\alpha}{2})} \cdot \sqrt{\mathbb{V}[\hat{X}_{t+1}]} \right]$$

로 주어진다는 것 같다. 근데 이게 너무 기니까 간단히

$$\hat{X}_{t+1} \pm z_{(1-\frac{\alpha}{2})} \cdot \sqrt{\mathbb{V}[\hat{X}_{t+1}]}$$

으로 표시한다는 것 같다. 이때, $z_{(1-\frac{\alpha}{2})}$ 는 α 에 대한 critical value를 의미한다고 한다. 따라서, X_{t+1} 에 대한 prediction interval은

$$\hat{X}_{t+1} \pm z_{(1-\frac{\alpha}{2})} \sigma_a \quad (10)$$

이 된다. 한번에 정리하면

$$\begin{aligned} \text{point estimate} : \hat{X}_{t+1} &= \phi X_t + \mu \\ \text{prediction interval} : \hat{X}_{t+1} &\pm z_{(1-\frac{\alpha}{2})} \sigma_a \end{aligned}$$

이 된다.

2.1.2 Prediction on X_{t+2}

식 (7)의 t 대신 $t+2$ 을 대입하고 똑같은 과정을 거치면 (여기서부터는 설명은 거의 적지 않고 계산식으로만 적겠다.)

$$\begin{aligned} X_{t+2} &= \phi X_{t+1} + a_{t+2} + \mu \\ \hat{X}_{t+2} &= \mathbb{E}[X_{t+2}|X_1, X_2, \dots, X_t] \\ &= \mathbb{E}[\phi X_{t+1} + a_{t+2} + \mu | X_1, X_2, \dots, X_t] \\ &= \phi \mathbb{E}[X_{t+1}|X_1, X_2, \dots, X_t] + \mu \\ &= \phi \hat{X}_{t+1} + \mu \\ &= \phi^2 X_t + \phi \mu + \mu \\ X_{t+2} - \hat{X}_{t+2} &= \phi (X_{t+1} - \hat{X}_{t+1}) + a_{t+2} \\ \mathbb{V}[\hat{X}_{t+2}] &= \mathbb{V}[X_{t+2} - \hat{X}_{t+2}] \\ &= \mathbb{V}[\phi (X_{t+1} - \hat{X}_{t+1}) + a_{t+2}] \\ &\stackrel{(*)}{=} \phi^2 \mathbb{V}[X_{t+1} - \hat{X}_{t+1}] + \mathbb{V}[a_{t+2}] \\ &= (\phi^2 + 1) \sigma_a^2 \end{aligned}$$

이다. 확률변수 X, Y 에 대해서 일반적으로 $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y]$ 는 성립하지 않는 식이다. 하지만 X 와 Y 가 서로 독립이면 이 식은 성립한다.⁴ 위의 식에서도 $X_{t+1} - \hat{X}_{t+1}$ 과 a_{t+2} 가 서로 독립이기 때문에 (*)가 성립한다. X 와 Y 가 서로 독립이면

$$\begin{aligned} \mathbb{V}[X + Y] &= \mathbb{E}[(X + Y - \mu_X - \mu_Y)^2] \\ &= \mathbb{E}[(X - \mu_X)^2 + (Y - \mu_Y)^2 + 2(X - \mu_X)(Y - \mu_Y)] \\ &= \mathbb{V}(X) + \mathbb{V}(Y) + \text{Cov}(X, Y) \\ &= \mathbb{V}(X) + \mathbb{V}(Y) \end{aligned}$$

이 되는 것이다. 따라서

$$\begin{aligned} \text{point estimate} : \hat{X}_{t+2} &= \phi^2 X_t + (\phi + 1)\mu \\ \text{prediction interval} : \hat{X}_{t+2} &\pm z_{(1-\frac{\alpha}{2})} \sqrt{\phi^2 + 1} \sigma_a \end{aligned}$$

이 된다. X_{t+2} 에 대한 prediction interval은 X_{t+1} 에 대한 prediction interval에 비해서는 조금 더 넓어졌다.

2.1.3 Prediction on X_{t+3}

강의에서는 설명되지 않았지만 X_{t+3} 에 대해서도 해보자,

$$\begin{aligned}
X_{t+3} &= \phi X_{t+2} + a_{t+3} + \mu \\
\hat{X}_{t+3} &= \mathbb{E}[X_{t+3}|X_1, X_2, \dots, X_t] \\
&= \mathbb{E}[\phi X_{t+2} + a_{t+3} + \mu|X_1, X_2, \dots, X_t] \\
&= \phi \mathbb{E}[X_{t+2}|X_1, X_2, \dots, X_t] + \mu \\
&= \phi \hat{X}_{t+2} + \mu \\
X_{t+3} - \hat{X}_{t+3} &= \phi(X_{t+2} - \hat{X}_{t+2}) + a_{t+3} \\
\mathbb{V}[\hat{X}_{t+3}] &= \mathbb{V}[X_{t+3} - \hat{X}_{t+3}] \\
&= \mathbb{V}[\phi(X_{t+2} - \hat{X}_{t+2}) + a_{t+3}] \\
&= \phi^2 \mathbb{V}[X_{t+2} - \hat{X}_{t+2}] + \mathbb{V}[a_{t+3}] \\
&= \phi^2(\phi^2 + 1)\sigma_a^2 + \sigma_a^2 \\
&= (\phi^4 + \phi^2 + 1)\sigma_a^2
\end{aligned}$$

따라서

$$\begin{aligned}
\text{point estimate} : \hat{X}_{t+3} &= \phi^3 X_t + (\phi^2 + \phi + 1)\mu \\
\text{prediction interval} : \hat{X}_{t+3} &\pm z_{(1-\frac{\alpha}{2})} \sqrt{\phi^4 + \phi^2 + 1}\sigma_a
\end{aligned}$$

이 된다.

2.2 AR(2) model

이번에는 AR(2) 모델에 대하여 X_1, X_2, \dots, X_t 가 주어졌을 때, X_{t+1} 과 X_{t+2} 를 예측해보려고 한다.

AR(2) 모델도 (7)에서와 마찬가지로 μ 를 추가하여

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t + \mu \quad (11)$$

로 쓸 수 있다.

2.2.1 Prediction on X_{t+1}

식 (11)의 t 대신 $t+1$ 을 대입하여

$$X_{t+1} = \phi_1 X_t + \phi_2 X_{t-1} + a_{t+1} + \mu \quad (12)$$

을 만들고 여기에 $\mathbb{E}[\cdot|X_1, X_2, \dots, X_t]$ 를 취하여 (2)를 대입하면 X_{t+1} 에 대한 point estimate(점 예측값)인 \hat{X}_{t+1} 을 얻을 수 있다;

$$\begin{aligned}
\hat{X}_{t+1} &= \mathbb{E}[X_{t+1}|X_1, X_2, \dots, X_t] \\
&= \mathbb{E}[\phi_1 X_t + \phi_2 X_{t-1} + a_{t+1} + \mu|X_1, X_2, \dots, X_t] \\
&= \phi_1 [X_t|X_1, X_2, \dots, X_t] + \phi_2 [X_{t-1}|X_1, X_2, \dots, X_t] + \mathbb{E}[a_{t+1}|X_1, X_2, \dots, X_t] + \mathbb{E}[\mu|X_1, X_2, \dots, X_t] \\
&= \phi_1 X_t + \phi_2 X_{t-1} + \mu
\end{aligned} \tag{13}$$

마찬가지로 (12)와 (13)을 빼면 $X_{t+1} - \hat{X}_{t+1} = a_{t+1}$ 이고, 따라서

$$\mathbb{V}[\hat{X}_{t+1}] = \mathbb{V}[X_{t+1} - \hat{X}_{t+1}] = \mathbb{V}[a_{t+1}] = \sigma_a^2$$

이다. 따라서

$$\begin{aligned} \text{point estimate} : \hat{X}_{t+1} &= \phi_1 X_t + \phi_2 X_{t-1} + \mu \\ \text{prediction interval} : \hat{X}_{t+1} &\pm z_{(1-\frac{\alpha}{2})} \sigma_a \end{aligned}$$

이 된다.

2.2.2 Prediction on X_{t+2}

이 부분은 강의에는 생략되었던 부분이다.

$$\begin{aligned} X_{t+2} &= \phi_1 X_{t+1} + \phi_2 X_t + a_{t+2} + \mu \\ \hat{X}_{t+2} &= \mathbb{E}[X_{t+2}|X_1, X_2, \dots, X_t] \\ &= \mathbb{E}[\phi_1 X_{t+1} + \phi_2 X_t + a_{t+2} + \mu | X_1, X_2, \dots, X_t] \\ &= \phi_1 \mathbb{E}[X_{t+1}|X_1, X_2, \dots, X_t] + \phi_2 \mathbb{E}[X_t|X_1, X_2, \dots, X_t] + \mu \\ &= \phi_1 \hat{X}_{t+1} + \phi_2 X_t + \mu \\ &= \phi_1^2 X_t + \phi_1 \phi_2 X_{t-1} + (\phi_2 + 1)\mu \\ X_{t+2} - \hat{X}_{t+2} &= \phi_1 (X_{t+1} - \hat{X}_{t+1}) + a_{t+2} \\ \mathbb{V}[\hat{X}_{t+2}] &= \mathbb{V}[X_{t+2} - \hat{X}_{t+2}] \\ &= \mathbb{V}[\phi_1 (X_{t+1} - \hat{X}_{t+1}) + a_{t+2}] \\ &= \phi_1^2 \mathbb{V}[X_{t+1} - \hat{X}_{t+1}] + \mathbb{V}[a_{t+2}] \\ &= (\phi_1^2 + 1) \sigma_a^2 \end{aligned}$$

이다. 따라서

$$\begin{aligned} \text{point estimate} : \hat{X}_{t+2} &= \phi_1^2 X_t + \phi_1 \phi_2 X_{t-1} + (\phi_2 + 1)\mu \\ \text{prediction interval} : \hat{X}_{t+2} &\pm z_{(1-\frac{\alpha}{2})} \sqrt{\phi_1^2 + 1} \sigma_a \end{aligned}$$

이 된다.

$$\Phi, \phi, \Psi, \psi, \pi, \Pi, \sigma, \Sigma$$