

Arima Model - 6

강의 : 김성범 교수님

정리 : 김선중

December 14, 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 간략하게 적어보았다. 지난 번에는 영어로 적었었는데, 이번에는 다시 한글로 적었다.

Contents

| | |
|--|----------|
| 1 Preliminaries | 3 |
| 1.1 Conditional Expectations | 3 |
| 1.1.1 Probability Spaces | 3 |
| 1.1.2 Expectations | 3 |
| 1.1.3 Joint and Marginal Distributions | 4 |
| 1.1.4 Conditional Probabilities and Conditinoal Expectations | 5 |
| 1.2 Prediction Intervals | 5 |

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

Seoung Bum Kim - Copyright © All rights reserved.

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

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

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

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

$$= \mu + E [a_{t+1} | X_1, \dots, X_t] + \frac{\phi}{1-\phi} E [a_t | X_1, \dots, X_t] + \phi_2 E [a_{t-1} | X_1, \dots, X_t] + \dots$$

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

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

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

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

$(1-\alpha) \cdot 100\%$ 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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

Seoung Bum Kim - Copyright © All rights reserved.

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

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

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

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

$$= \mu + E [a_{t+2} | X_1, \dots, X_t] + \frac{\phi}{1-\phi} E [a_t | X_1, \dots, X_t] + \phi_2 E [a_{t-1} | X_1, \dots, X_t] + \dots$$

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

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

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

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

$(1-\alpha) \cdot 100\%$ 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.

Prediction (Forecasting)

Given $X_1, X_2, \dots, \hat{X}_t$, want to predict \hat{X}_{t+3} .

$\hat{X}_{t+3} - \hat{\hat{X}}_{t+3}$: error

$$\min E [(\hat{X}_{t+3} - \hat{\hat{X}}_{t+3})^2]$$

$$\hat{X}_{t+3} = E [X_{t+3} | X_1, \dots, X_t] \rightarrow \text{conditional expectation.}$$

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

$$\hat{X}_{t+3} = \phi X_t + a_{t+3} + \mu$$

$$\begin{aligned} \hat{X}_{t+3} &= E [X_{t+3} | X_1, \dots, X_t] \\ &= \phi E [X_t | X_1, \dots, X_t] + E [a_{t+3} | X_1, \dots, X_t] + E [\mu | X_1, \dots, X_t] \\ &= \phi X_t + \mu \end{aligned}$$

$$\therefore \hat{X}_{t+3} = \phi X_t + \mu$$

Seoung Bum Kim - Copyright © All rights reserved.

$$X_{t+3} \rightarrow \hat{X}_{t+3} \quad (\text{P.I.})$$

$(1-\alpha) \cdot 100\%$ prediction interval for X_{t+3} ,

$$\hat{X}_{t+3} \pm c.v.(\alpha) \sqrt{\text{Var}(\hat{X}_{t+3})}$$

$$X_{t+3} = \phi X_t + a_{t+3} + \mu$$

$$\hat{X}_{t+3} = \phi X_t + a_{t+3} + \mu$$

$$X_{t+3} - \hat{X}_{t+3} = a_{t+3}$$

$$V(X_{t+3} - \hat{X}_{t+3}) = V(a_{t+3}) = \tilde{\sigma}^2$$

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

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

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

Seoung Bum Kim - Copyright © All rights reserved.

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

$$\hat{X}_{t+3} = \mu + \phi_1 X_t + \phi_2 X_{t-1} + a_{t+3}.$$

$$\begin{aligned} X_{t+3} &= E [X_{t+3} | X_1, \dots, X_t] \\ &= E[\mu | X_1, \dots, X_t] + \phi_1 E[X_t | X_1, \dots, X_t] + \phi_2 E[X_{t-1} | X_1, \dots, X_t] + E[a_{t+3} | X_1, \dots, X_t] \\ &= \mu + \phi_1 X_t + \phi_2 X_{t-1} \end{aligned}$$

$$(1-\alpha) \cdot 100\% \text{ P.I. for } X_{t+3} \Rightarrow \hat{X}_{t+3} \pm c.v.(\alpha) \sqrt{V(\hat{X}_{t+3})}$$

$$X_{t+3} = \mu + \phi_1 X_t + \phi_2 X_{t-1} + a_{t+3}$$

$$\hat{X}_{t+3} = \mu + \phi_1 X_t + \phi_2 X_{t-1}$$

$$X_{t+3} - \hat{X}_{t+3} = a_{t+3}$$

$$V(X_{t+3} - \hat{X}_{t+3}) = \tilde{\sigma}^2$$

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

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

Seoung Bum Kim - Copyright © All rights reserved.

P.I. for X_{t+3} , $\phi \neq 0$: (부사1)

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

$$X_{t+3} = \mu + a_{t+3} + \phi_1 a_t + \phi_2 a_{t-1} + \dots$$

$$\hat{X}_{t+3} = E [X_{t+3} | X_1, \dots, X_t]$$

$$= \mu + E [a_{t+3} | X_1, \dots, X_t] + \frac{\phi}{1-\phi} E [a_t | X_1, \dots, X_t] + \phi_2 E [a_{t-1} | X_1, \dots, X_t] + \dots$$

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

$$X_{t+3} = \mu + a_{t+3} + \phi_1 a_t + \phi_2 a_{t-1} + \dots$$

$$\hat{X}_{t+3} = \mu + \phi_1 a_t + \phi_2 a_{t-1} + \dots$$

$$V(X_{t+3} - \hat{X}_{t+3}) = \theta V(a_{t+3}) = \tilde{\sigma}^2$$

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

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

Seoung Bum Kim - Copyright © All rights reserved.

Prediction (Forecasting)

Given $X_1, X_2, \dots, \hat{X}_t$, want to predict \hat{X}_{t+4} .

$\hat{X}_{t+4} - \hat{\hat{X}}_{t+4}$: error

$$\min E [(\hat{X}_{t+4} - \hat{\hat{X}}_{t+4})^2]$$

$$\hat{X}_{t+4} = E [X_{t+4} | X_1, \dots, X_t] \rightarrow \text{conditional expectation.}$$

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

$$\hat{X}_{t+4} = \phi X_t + a_{t+4} + \mu$$

$$\begin{aligned} \hat{X}_{t+4} &= E [X_{t+4} | X_1, \dots, X_t] \\ &= \phi E [X_t | X_1, \dots, X_t] + E [a_{t+4} | X_1, \dots, X_t] + E [\mu | X_1, \dots, X_t] \\ &= \phi X_t + \mu \end{aligned}$$

$$\therefore \hat{X}_{t+4} = \phi X_t + \mu$$

Seoung Bum Kim - Copyright © All rights reserved.

$$X_{t+4} \rightarrow \hat{X}_{t+4} \quad (\text{P.I.})$$

$(1-\alpha) \cdot 100\%$ prediction interval for X_{t+4} ,

$$\hat{X}_{t+4} \pm c.v.(\alpha) \sqrt{\text{Var}(\hat{X}_{t+4})}$$

$$X_{t+4} = \phi X_t + a_{t+4} + \mu$$

$$\hat{X}_{t+4} = \phi X_t + a_{t+4} + \mu$$

$$X_{t+4} - \hat{X}_{t+4} = a_{t+4}$$

$$V(X_{t+4} - \hat{X}_{t+4}) = V(a_{t+4}) = \tilde{\sigma}^2$$

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

$(1-\alpha) \cdot 100\%$ P.I. for X_{t+4}

$$\hat{X}_{t+4} \pm Z(1-\frac{\alpha}{2}) \cdot \tilde{\sigma}$$

Seoung Bum Kim - Copyright © All rights reserved.

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

$$\hat{X}_{t+4} = \mu + \phi_1 X_t + \phi_2 X_{t-1} + a_{t+4}.$$

<math display="

Example) ARMA(1,1)

$$X_t = 1 + 0.4X_{t-1} + \alpha_t + 0.4\alpha_{t-1}$$

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

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

$$X_6 = 1 + 0.4X_5 + \alpha_6 + 0.4\alpha_5$$

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

$$\rightarrow \alpha_t = X_t - 0.4X_{t-1} - 0.4\alpha_{t-1} - 1. \quad 1 + 0.4 \cdot 1 + 0.4 \cdot (-2.05) \\ \hat{\alpha}_1 = 0 \\ \hat{\alpha}_2 = X_2 - 0.4X_1 - 0.4\hat{\alpha}_1 - 1 = -1 - 0.4(3) - 1 = -3.2$$

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

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

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

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

Seoung Bum Kim - Copyright © All rights reserved.

Partial autocorrelation function (PACF)
PA.

- 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.

P.I. for X_6 : $\hat{X}_6 \pm 1.96 \sqrt{\text{S}_{\alpha}^2}$ $\text{S}_{\alpha} = 1$

$$0.58 \pm 1.96$$

P.I. for X_7 : $\hat{X}_7 \pm 1.96 \sqrt{(1 + \phi_1^2) \text{S}_{\alpha}^2} \quad \phi = \phi + \theta = 0.4 + 0.4 = 0.8$

$$1.232 \pm 1.96 \sqrt{(1 + 0.8^2) \cdot 1}$$

P.I. for X_8 : $\hat{X}_8 \pm 1.96 \sqrt{(1 + \phi_1^2 + \phi_2^2) \text{S}_{\alpha}^2}$

$$-1.493 \pm 1.96 \sqrt{(1 + 0.8^2 + 0.32^2) \cdot 1}$$

$$\begin{aligned} \phi_2 &= \phi + \theta \\ &= 0.4 + 0.4 \cdot 0.4 \\ &= 0.32 \end{aligned}$$

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)참조했습니다. 해당 자료의 A.1, A.2를 읽어서 적었습니다.

1.1.1 Probability Spaces

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 확률 측도(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 역시 세가지 정도의 성질을 만족시킨다.

1.1.2 Expectations

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

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

이다. 이때, $x \mapsto \mathbb{P}(X = x)$ 는 확률질량함수(probability mass function)이다. 연속확률변수(continuous random variable) X 에 대하여는

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

와 같은 정의가 있다. 이때에는 $f(x)$ 는 확률밀도함수(probability density function)로서

$$\int_a^b f(x) dx = \mathbb{P}(a \leq X \leq b)$$

를 만족하는 함수이다.

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 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

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

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

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.¹ 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

(X, Y) 의 결합확률분포에서의 사건 A, B 를 고려하자. 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.

Y 가 주어졌을 때 X 의 확률밀도함수 $f_{X|Y}(x|y)$ 는 다음과 같이 정의된다.

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

¹그런데 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.

이때, $f_{X,Y}(x,y)$ 는 X, Y 에 대한 결합확률분포(joint distribution)에 대한 확률밀도함수이다. $f_Y(y)$ 는 Y 에 대한 확률밀도함수이면서, $f_{X,Y}(x,y)$ 에 대한 주변확률(marginal dis

1.2 Prediction Intervals