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 적어보았다. 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

아리마 모델 (ARIMA Model) – Part 6

Prediction (Forecasting)

Given $X_1, X_2, \dots, X_t$, want to predict X_{t+1}.

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

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

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

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

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

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

$= \phi E[X_t | X_1, \dots, X_t] + \underbrace{E[a_{t+1} | X_1, \dots, X_t]}_{=0} + E[\mu | X_1, \dots, X_t]$

$= \phi X_t + \mu$

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

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

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

$\boxed{\hat{X}_{t+1} \pm C.V.(\alpha)\sqrt{Var(\hat{X}_{t+1})}}$

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

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

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

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

Assume $a_t \sim N(0, \sigma_a^2)$

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

$\boxed{\hat{X}_{t+1} \pm Z_{(1-\frac{\alpha}{2})} \cdot \sigma_a}$

AR(1) $X_{t+2} = \phi X_{t+1} + a_{t+2} + \mu$.

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

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

$= \phi \hat{X}_{t+1} + \mu$, → 점예측값

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

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

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

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

$V(\quad '' \quad) = \phi^2 V[(X_{t+1} - \hat{X}_{t+1})] + V(a_{t+2})$

$= \phi^2 \cdot \sigma_a^2 + \sigma_a^2$

$= (1+\phi^2)\sigma_a^2$

$\therefore (1-\alpha)\cdot 100\%$ P.I. for X_{t+2}, $a_t \sim N(0, \sigma_a^2)$

$\boxed{\hat{X}_{t+2} \pm Z_{(1-\frac{\alpha}{2})} \cdot \sqrt{(1+\phi^2)\sigma_a^2}}$,,

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

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

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

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

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

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

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

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

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

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

$a_t \sim N(0, \sigma_a^2)$.

$\boxed{\hat{X}_{t+1} \pm Z_{(1-\frac{\alpha}{2})} \cdot \sigma_a}$,,

$\boxed{X_{t+2} \sim}$

$X_{t+3} = \mu + \phi_1 X_{t+2} + \phi_2 X_{t+1} + a_{t+3}$　　추정 vs. 예측.

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

$= \mu + \phi_1 E[X_{t+2} | X_1, \dots, X_t] + \phi_2 E[X_{t+1} | X_1, \dots, X_t] + E[a_{t+3} | X_1, \dots, X_t]$

$= \mu + \phi_1 \hat{X}_{t+2} + \phi_2 \hat{X}_{t+1}$ → Point predicted value.

☆ $\boxed{95\%\ \text{P.I. for } X_{t+3}}$ ⇒ $\hat{X}_{t+3} \pm \underbrace{Z_{(1-\frac{0.05}{2})}}_{1.96} \cdot \sqrt{V(\hat{X}_{t+3})}$

$X_{t+3} = \mu + \phi_1 X_{t+2} + \phi_2 X_{t+1} + a_{t+3}$

$-\ \hat{X}_{t+3} = \mu + \phi_1 \hat{X}_{t+2} + \phi_2 \hat{X}_{t+1}$

$V(X_{t+3} - \hat{X}_{t+3}) = \phi_1^2 V(X_{t+2} - \hat{X}_{t+2}) + \phi_2^2 V(X_{t+1} - \hat{X}_{t+1}) + V(a_{t+3})$

$= \phi_1^2 (1+\phi_1^2)\sigma_a^2 + \phi_2^2 \sigma_a^2 + \sigma_a^2$

$= \{1 + \phi_1^2(1+\phi_1^2) + \phi_2^2\}\sigma_a^2$.

95% PI for X_{t+3}. $a_t \sim N(0, \sigma_a^2)$

$\boxed{\hat{X}_{t+3} \pm 1.96\sqrt{\{1+\phi_1^2(1+\phi_1^2)+\phi_2^2\}\cdot\sigma_a^2}}$,,

P.I. for X_{t+1},　ψ ⇒ psi (프사이)

☆ $X_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots$,　$\psi_1 = \phi + \theta$, $\psi_2 = \phi^2 + \phi\cdot\theta, \dots$

$X_{t+1} = \mu + a_{t+1} + \psi_1 a_t + \psi_2 a_{t-1} + \cdots$

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

$= \mu + E[a_{t+1} | X_1, \dots, X_t] + \psi_1 E[a_t | X_1, \dots, X_t] + \psi_2 E[a_{t-1} | X_1, \dots, X_t] + \cdots$

$= \mu + \psi_1 a_t + \psi_2 a_{t-1} + \cdots$

$X_{t+1} = \mu + a_{t+1} + \psi_1 a_t + \psi_2 a_{t-1} + \cdots$

$-\ \hat{X}_{t+1} = \mu + \psi_1 a_t + \psi_2 a_{t-1} + \cdots$

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

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

$\hat{X}_{t+1} \pm Z_{(1-\frac{\alpha}{2})} \cdot \sigma_a$.

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

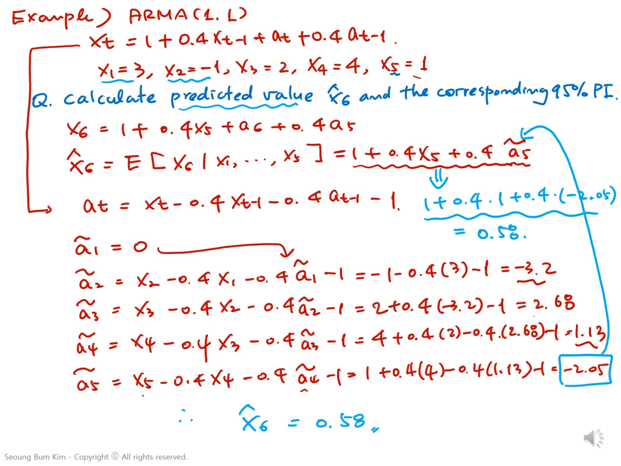
$\hat{X}_{t+2} \pm Z_{(1-\frac{\alpha}{2})}\sqrt{(1+\psi_1^2)\cdot\sigma_a^2}$; $\psi_1 = \phi + \theta$,

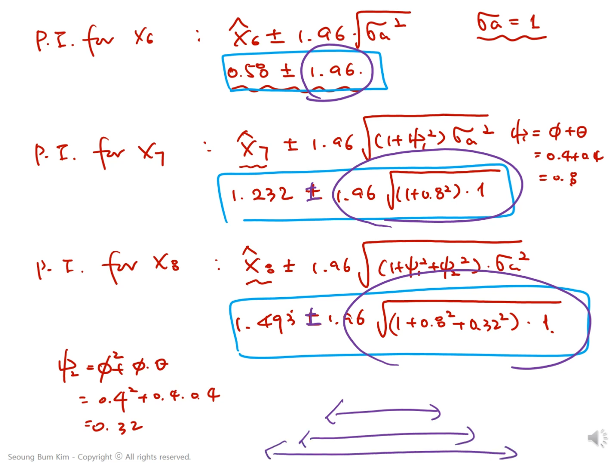
$(1-\alpha)\cdot 100\%$ P.T. for X_{t+3}.

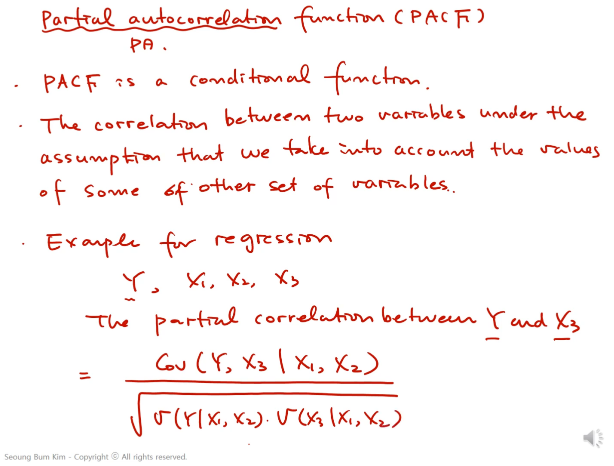
$\hat{X}_{t+3} \pm Z_{(1-\frac{\alpha}{2})}\sqrt{(1+\psi_1^2+\psi_2^2)\sigma_a^2}$.

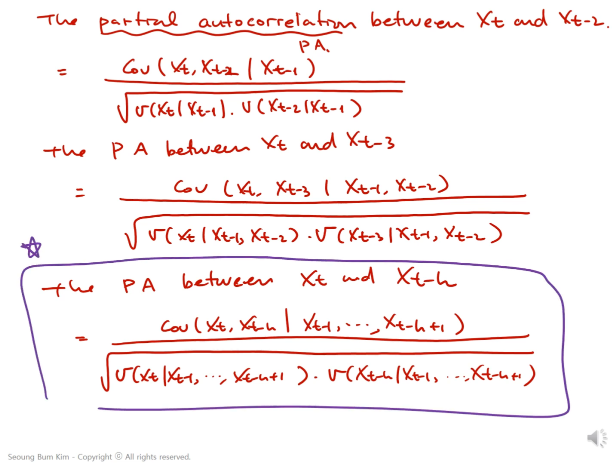
$\psi_2 = \phi^2 + \phi\cdot\theta$,,

⋮









1 Preliminaries

1.1 Conditional Expectations

Conditional expectation에 관해서는 이 자료(A Conditional expectation - Arizona Math)참조했다. 해당 자료의 A.1, A.2를 옮겨서 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 안적었고 전체확률의 법칙(the law of total probability)같은 것도 안 적었다. 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 Ω는 표본공간(sample space)이고, $\mathcal{F}$는 사건공간(event space)이고, $\mathbb{P}$는 확률측도(probability measure)이다. 사건(event)이란, 표본 공간 Ω의 부분집합이고, $\mathcal{F}$는 사건들의 집합이지만 모든 사건들의 집합일 필요는 없다. 그보다는, $\mathcal{F}$는 Ω에 대한 σ-algebra이라서 세 가지 정도의 성질을 만족시키는 집합이다. $\mathbb{P}$는 유한측도(finite measure)로서, $\mathbb{P} : \mathcal{F} \to [0, 1]$인 함수이고 역시 몇가지 성질을 만족시킨다.

확률변수란 Ω에서 $\mathbb{R}$로 가는 함수이다. 즉, X가 확률변수이면 $X : \Omega \to \mathbb{R}$이다. 이때 X의 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 X의 치역이 이살 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

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

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

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

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 Ω의 멱집합, 즉 Ω의 모든 부분집합들의 집합으로 둘 수 있다. 즉,

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

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 X라고 하면, 이 확률변수는 하나의 함수 $X : \{1, 2, 3, 4, 5, 6\} \to \mathbb{R}$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. X의 역은 $\{0, 1\}$이므로 X는 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 Ω의 모든 부분집합들의 집합일 수는 없다. [1] $\mathcal{F}$는 Ω의 보렐집합(borel set)이어야 한다. 다시 말해서 $\Omega = [0, 6]$의 standard topology를 $\mathcal{T}$라고 할 때,

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

[1] 아마 그러면 안될 것이다. 이유가 뭐였더라, 아마도 cantor set 같은게 문제가 됐던 것 같은데.

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 A의 원소의 개수(cardinality)를 의미했다면, 이번의 $|A|$는 집합 A의 길이, 즉, $\mathbb{R}$의 Lebesgue measure 이다. 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 X라고 하면, 이 확률변수는 하나의 함수 $X : [0, 6] \to \mathbb{R}$로서

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

로 정의된다. X의 치역은 $[0, 3]$이므로 X는 연속확률변수이고, 확률질량함수는

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

이 된다.

1.1.2 Expectations

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

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

이다. 연속확률변수 X에 대하여는

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

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 X는 평균적으로 $\frac{1}{2}$의 값으로 기대된다.' 두번째 예에서는

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

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, Y에 대하여는 이 확률변수들의 순서쌍 (X, Y)이 하나의 확률변수가 된다. 이 확률변수의 분포를 결합확률분포(joint distribution)라고 한다. 만약 X와 Y가 모두 이산확률변수이면, (X, Y) 또한 이산확률변수이고, 이때의 확률질량함수는

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

와 같은 이변수함수가 된다. X와 Y에 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^2Y에 대한 기댓값 $\mathbb{E}[X^2Y]$은 $(g(x, Y) = X^2Y)$

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

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

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

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

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

이다. 위에서와 같은 예를 들어서 설명하면

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

이 된다.

결합확률분포의 관점에서 보면 X의 확률분포(혹은 Y에 대한 확률분포)는 (X, Y)에 대한 주변확률분포(marginal distribution)이라고 말할 수 있다. X, Y 대한 확률질량함수[혹은 확률밀도함수]를 X에 대하여 (혹은 Y에 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 $X_1, X_2, \cdots, X_N$에 대해서도 새로운 확률변수 $(X_1, X_2, \cdots, X_N)$를 생각할 수 있다. 이것은, 그러니까 확률변수들의 tuple인 것이다. 확률변수 X_n $(1 \le n \le N)$들이 모두 이산확률변수일 경우, 확률질량함수

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

을 생각할 수 있고, 모두 연속확률변수인 경우, 확률밀도함수

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

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

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

혹은

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

이다.

1.1.4 Conditional Probabilities and Conditinoal Expectations

두 사건 A, B에 대하여, [3] 사건 A가 일어났다고 가정했을 때 사건 B가 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 X가 연속확률변수이면, 분모는 0이 될 수밖에 없다. 그래서 참고한 자료에서는, X가 연속확률변수일 때에 한해서 ϵ과 극한을 사용해서 정의하고 있다. 이런 여러가지 어려움들이 있지만, 그런 어려움들이 다 극복되고, 저 식이 그저 잘 정의되어 있다고 가정해보자. 그리고, 이제부터는 이산확률변수만 가정하자, 연속확률변수인 경우에는 summation이 integral로 적절히 바뀌면 될 것이다.

[2] 그런데 위의 표현은 너무 복잡하므로, 그냥 간단하게

$$f_{\boldsymbol{X}}(x_1, x_2, \cdots, x_N)$$

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

[3] 정확하게는, (X, Y)의 결합확률분포에서의 사건 A, B에 대하여

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, Y를 주사위를 던진 횟수, X를 1이 나온 횟수라고 하자. 그러면

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

이다. 왜냐하면, $Y = 1$이라는 뜻은, 한번에 6이 나왔다는 뜻이고, 그때의 X의 값은 0일 수밖에 없기 때문이다. 즉, 조건부 $Y = 1$ 하에서 X의 확률분포는 확률질량함수 $\mathbb{P}(X = x|Y = 1)$로 표현될 수 있는데, 가능한 x는 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 X의 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 X로 가능한 값은 0, 1, 2이고 (이제 감이 잡힌다.) X는 $B(2, \frac{1}{5})$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 y에 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, Y에 의존하는 값으로 나온다는 것이다. $\mathbb{E}[X|Y]$는 X에 대한 기댓값을 계산하는 것이기 때문에, X에 관한 식이 나오지 않는다. 다시 말해서, 그 결과값에 X가 포함되지 않는다. 그런데 조건부에 걸린 Y의 값이 특정되지 않았기 때문에 $\mathbb{E}[X|Y]$는 Y에 대한 함수가 되는 것이다. 이에 대한 설명이 참고자료에 적혀있으니 잘 따라가보자.

아까, 확률변수가 Ω에서 $\mathbb{R}$로 가는 함수였던 것을 상기해보자. 표본공간인 Ω는 주사위를 여러 번 던졌을 때 나올 수 있는 모든 경우들에 대한 집합이다. 만약 주사위의 눈이 차례로 $1, 3, 2, 3, 6$이 나왔다면, 이것은 $X = 1$, $Y = 5$인 상황을 나타내는데, $\omega = (1, 3, 2, 3, 6)$으로 표현할 수 있을 것이다. 그러니까, Ω를 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 $\mathbb{E}[X|Y] : \Omega \to \mathbb{R}$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 $\omega = (a_1, a_2, \cdots, a_{n-1}, 6)$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 ω를 $\mathbb{E}[X|Y=y$로 대응시킨다. $(y = Y(\omega))$ 그런데 그 값은 $\frac{1}{5}(y-1)$이라고 계산되었었으니까,

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

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

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

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

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

과 같은 확률과정을 고려하는 것이다.

이번 강의에서

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

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

기억해야 하는 것은 다음 두가지이다.

(1) $\mathbb{E}[X_{t+1}|X_1, X_2, \cdots, X_t]$는 확률변수이다.

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

1.2 Prediction Intervals