

이공학도를 위한 확률 및 통계학

- 3장 확률변수와 확률분포 -

이 성 현(seonghyeon@pel.sejong.ac.kr)

세종대학교 프로토콜공학연구실

목 차

- 확률변수(Random Variable)의 개념
- 확률분포(Probability Distribution)와 분포함수(Distribution Function)
- 결합확률분포(Joint Probability Distribution)

목 차

- 확률변수(Random Variable)의 개념
- 확률분포(Probability Distribution)와
분포함수(Distribution Function)
- 결합확률분포(Joint Probability Distribution)

확률변수의 개념

- 확률변수 (1/9)

- 정의

- 표본공간 내 각 원소에 하나의 실수 값을 대응시키는 함수

- 특징

- 확률변수는 주로 대문자로, 그에 대응하는 하나의 값은 소문자로 나타냄
 - e.g., 주사위를 굴렀을 때 나오는 눈금의 수를 확률변수 X 라 하면, 대응되는 값 x 는 1,2,3,4,5,6이 됨
- 확률변수 X 의 가능한 모든 값들은 실험에서 표본공간의 부분집합이 되는 사상을 나타냄
 - e.g., 주사위를 굴렀을 때 나오는 눈금의 짝수 여부를 확률변수 X 라 하면, 눈금의 수가 2,4,6일 때 $X = 1$ 이고 1,3,5일 때 $X = 0$ 이 됨

확률변수의 개념

- 확률변수 (2/9)

- 예제 3.1

4개의 붉은 공(R)과 3개의 검은 공(B)이 들어있는 항아리에서 연속적으로 2개의 공을 비복원추출하는 실험에서 Y 를 붉은 공의 개수라 할 때, 출현가능한 결과와 확률변수 Y 의 값 y 를 구하시오.

표본공간	y
RR	2
RB	1
BR	1
BB	0

[표 3.1] 예제 3.1의 출현가능한 결과

- $Y = \{0, 1, 2\}$

확률변수의 개념

- 확률변수 (3/9)

- 예제 3.2

공구보관소 직원이 3명의 공장 종업원들에게 헬멧을 임의로 꺼내 주었다. 만일 스미스(S), 존스(J), 그리고 브라운(B)의 순서로 헬멧을 받을 때, 헬멧을 받는 가능한 순서들을 나열하고, M 을 원래 주인에게 지급된 헬멧의 수라 할 때, 확률변수 M 의 값 m 을 구하시오.

표본공간	m
SJB	3
SBJ	1
JSB	1
JBS	0
BSJ	0
BJS	1

[표 3.2] 예제 3.2의 표본공간과 원래 주인에게 지급된 헬멧의 수 m

- $M = \{0, 1, 3\}$

확률변수의 개념

- 확률변수 (4/9)
 - 이산표본공간(Discrete Sample Space)
 - 정의
 - 표본공간의 원소가 유한하거나 가산 무한(Countably Infinite)인 표본공간
 - e.g., 주사위 굴리기의 표본공간 $S = \{1, 2, 3, 4, 5, 6\}$, 어느 제품의 고장 발생 횟수의 표본공간 $S = \{0, 1, 2, 3, 4, \dots\}$
 - 연속표본공간(Continuous Sample Space)
 - 정의
 - 표본공간의 원소가 어떤 연속 구간을 이루는 표본공간
 - e.g., 어떤 작업이 끝나는 데 걸린 시간 $S = [0, \infty)$, 어느 공장에서 생산된 막대 부품의 길이 $S = [1990mm, 2010mm]$

확률변수의 개념

- 확률변수 (5/9)

- 이산형 확률변수(Discrete Random Variable) (1/2)

- 정의

- 가질 수 있는 값의 집합이 유한 또는 가산 무한인 확률변수

- 특징

- 주로 계수자료(Count Data)를 나타낼 때 사용함
 - e.g., 불량품의 개수, 연간 사고 횟수, 투표 참여자 수

- 예제 3.3

어느 공장의 100개의 제품 중 12개의 불량품이 존재하고 이 중 10개를 선택하였을 때, 10개의 제품 표본에서 발견되는 불량품의 수를 확률변수 X 라 할 때, X 가 가질 수 있는 값을 구하시오.

- $X = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

- 확률변수 X 가 가능한 값이 유한하므로 이산형 확률변수임

확률변수의 개념

- 확률변수 (6/9)
 - 이산형 확률변수(Discrete Random Variable) (2/2)
 - 예제 3.4

어느 공장에서 하나의 불량품이 발견될 때까지 공정으로부터 표본을 추출하면서 양품을 N , 불량품을 D , 불량품이 발견될 때까지 추출한 제품의 수를 나타내는 확률변수를 X 라고 할 때, 확률변수에 따른 표본공간 S 는 어떻게 되는가?

- $X = 1$ 이라면 표본공간은 $S = \{D\}$, $X = 2$ 라면 $S = \{ND\}$, $X = 3$ 이라면 $S = \{NND\}$ 와 같이 됨
- 확률변수 X 의 가능한 값 x 가 무한하나, 가산 무한이므로 X 는 이산형 확률변수임

확률변수의 개념

- 확률변수 (7/9)
 - 연속형 확률변수(Continuous Random Variable) (1/2)
 - 정의
 - 연속적인 구간 내의 값을 취하는 확률변수
 - 특징
 - 주로 측정자료(Measured Data)를 나타낼 때 사용함
 - e.g., 높이, 무게, 온도, 거리, 시간
 - 예제 3.5

통신판매 광고에 반응하는 사람의 비율을 X 라 할 때, 확률변수 X 의 값 x 의 범위는 어떻게 되는가?

- x 의 범위는 0% ~ 100%이므로 $0 \leq x \leq 1$ 이 됨
- 비율은 연속적인 구간 내의 값을 취하므로 X 는 연속형 확률변수임

확률변수의 개념

- 확률변수 (8/9)
 - 연속형 확률변수(Continuous Random Variable) (2/2)
 - 예제 3.6

과속탐지 카메라에 적발되는 과속 차량들 사이의 시간 간격을 확률변수 X 라 할 때, X 의 값 x 의 범위는 어떻게 되는가?

- 시간 간격이 음수가 될 수 없으므로 $x \geq 0$ 이 됨
- 가능한 시간 간격들은 연속적인 값을 가지고 있으므로 X 는 연속형 확률변수임

확률변수의 개념

- 확률변수 (9/9)
- 가변수(Dummy Variable)
 - 수치로 나타내기 어려운 범주형 확률변수를 임의의 숫자를 부여하여 나타내는 변수
 - e.g., 공장에서 생성된 부품이 불량품인 경우 1, 아닌 경우 0, 투표에 참여한 사람의 성별이 남성인 경우 1, 아닌 경우 0
 - 두 개의 가능한 값을 0과 1로 표현하는 확률변수를 베르누이 확률변수(Bernoulli Random Variable)라고도 함

목 차

- 확률변수(Random Variable)의 개념
- 확률분포(Probability Distribution)와 분포함수(Distribution Function)
- 결합확률분포(Joint Probability Distribution)

확률분포와 분포함수

- 확률분포

- 정의

- 표본공간에 정의된 확률을 통해 확률변수의 값 또는 해당 값들이 속하는 집합에 대한 확률을 표현한 것

- 구분

- 이산형 확률분포(Discrete Probability Distribution)
- 연속형 확률분포(Continuous Probability Distribution)

- 표현 방식

- 확률질량함수(PMF, Probability Mass Function)
- 확률밀도함수(PDF, Probability Density Function)
- 누적분포함수(CDF, Cumulative Distribution Function)

확률분포와 분포함수

- 이산형 확률분포 (1/3)

- 정의

- 이산형 확률변수가 가지는 확률분포로, 다음 조건을 만족함
 - $f(x) \geq 0$
 - $\sum_x f(x) = 1$
 - $P(X = x) = f(x)$

- 특징

- 이산형 확률분포를 표현하기 위해 사용하는 $f(x)$ 를 확률 질량함수(PMF)라고 함
- 확률변수가 특정 값을 가질 확률을 0 이상으로 보고, 확률 변수가 특정한 값을 가질 확률을 계산함

확률분포와 분포함수

- 이산형 확률분포 (2/3)

- 예제 3.7

상점에 진열된 20대의 노트북 중에 불량품이 3개 끼여 있다. 어느 학교에서 이 중 임의로 2대를 구입했을 때, 불량품 개수의 확률분포를 구하시오.

- 구입한 노트북 중 불량품의 수를 확률변수 X 라 하면, X 의 값 x 가 취할 수 있는 값은 0, 1, 2임

- $$f(0) = P(X = 0) = \frac{\binom{3}{0}\binom{17}{2}}{\binom{20}{2}} = \frac{\frac{3!}{0!3!} \times \frac{17!}{2!15!}}{\frac{20!}{2!18!}} = \frac{68}{95},$$

$$f(1) = P(X = 1) = \frac{\binom{3}{1}\binom{17}{1}}{\binom{20}{2}} = \frac{\frac{3!}{1!2!} \times \frac{17!}{1!16!}}{\frac{20!}{2!18!}} = \frac{51}{190},$$

$$f(2) = P(X = 2) = \frac{\binom{3}{2}\binom{17}{0}}{\binom{20}{2}} = \frac{\frac{3!}{2!1!} \times \frac{17!}{0!17!}}{\frac{20!}{2!18!}} = \frac{3}{190}$$

x	$f(x)$
0	$\frac{68}{95}$
1	$\frac{51}{190}$
2	$\frac{3}{190}$

[표 3.3] 예제 3.7의 가능한 값 x 와 확률분포 $f(x)$

확률분포와 분포함수

• 이산형 확률분포 (3/3)

• 예제 3.8

어느 대리점에서 판매된 외제차의 50%에 디젤 엔진이 장착됐다고 할 때, 이 대리점에서 다음에 판매될 4대의 외제차 가운데 디젤 엔진이 장착된 차의 수의 확률분포에 대한 식을 구하시오.

- 디젤 엔진인 경우를 D, 아닌 경우를 N이라고 하면 표본공간은 $S = \{DDDD, DDDN, DDND, \dots, NNNN\}$ 의 16개로 나타낼 수 있음
- 4대 중 디젤 엔진인 차량의 수를 확률변수 X 라고 하면 X 가 가질 수 있는 값 x 는 0, 1, 2, 3, 4의 값을 가질 수 있음
- 디젤 엔진일 확률과 아닐 확률이 50%로 동일하므로 $\binom{4}{x}$ 로 표현할 수 있음

$$\therefore \text{확률분포 } f(x) = \frac{\binom{4}{x}}{16}, x = 0, 1, 2, 3, 4$$

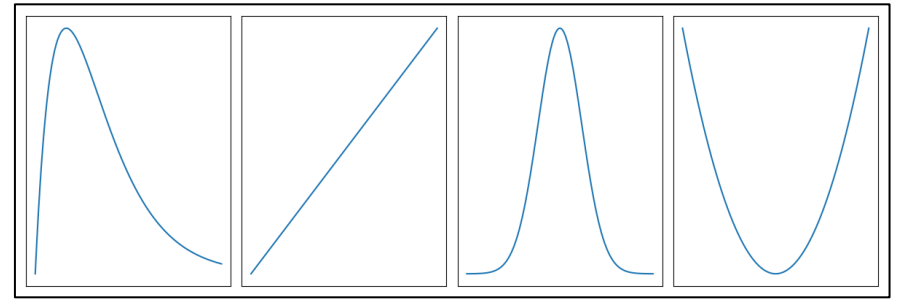
확률분포와 분포함수

• 연속형 확률분포 (1/2)

• 정의

- 연속형 확률변수가 가지는 확률분포로, 다음 조건을 만족함

- 모든 $x \in R$ 에 대해 $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x) dx = 1$
- $P(a < X < b) = \int_a^b f(x) dx$



<그림 1> 다양한 형태의 확률밀도함수

• 특징

- 연속형 확률분포를 표현하기 위해 사용하는 $f(x)$ 를 확률 밀도함수(PDF) 또는 간단하게 밀도함수(Density Function)라고 함
- 확률변수가 어떤 특정한 값을 가진 확률을 0으로 보고, 확률변수가 구간 내에 속할 확률을 계산함
- 밀도함수에 따라 여러 가지 형태의 그래프를 가질 수 있음

확률분포와 분포함수

• 연속형 확률분포 (2/2)

• 예제 3.9

제어실험에서 반응온도(°C)의 변화에 따른 오차는 다음과 같은 확률분포를 가지는 연속확률변수 X 라고 가정하자.

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2 \\ 0, & \text{그 외의 경우} \end{cases}$$

(a) 함수 $f(x)$ 가 확률밀도함수임을 증명하시오.

(b) 확률 $P(0 < X \leq 1)$ 을 구하시오.

- (a) $f(x) \geq 0$ 임은 명확하므로, 다음 식을 통해 확률밀도함수임을 보일 수 있음

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-1}^2 \frac{x^2}{3} dx = \frac{x^3}{9} \Big|_{-1}^2 = \frac{8}{9} + \frac{1}{9} = 1$$

- (b) $P(0 < x \leq 1) = \int_0^1 \frac{x^2}{3} dx = \frac{x^3}{9} \Big|_0^1 = \frac{1}{9}$

확률분포와 분포함수

- 누적분포함수(CDF) (1/5)

- 정의

- 실수 x 에 대해 확률변수 X 가 x 이하의 값을 가질 확률로 정의되는 함수

- 특징

- 누적분포함수는 비감소 함수이며, 주어진 확률변수가 취할 수 있는 값 뿐만 아니라 모든 실수에서 정의됨
- 이산형 확률분포에 대한 누적분포함수 $F(x)$ 는 다음과 같음
 - $F(x) = P(X \leq x) = \sum_{t \leq x} f(t), -\infty < x < \infty$
- 연속형 확률분포에 대한 누적분포함수 $F(x)$ 는 다음과 같음
 - $F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt, -\infty < x < \infty$

확률분포와 분포함수

• 누적분포함수(CDF) (2/5)

• 예제 3.10 (1/2)

예제 3.8에서 구한 확률질량함수 $f(x) = \frac{\binom{4}{x}}{16}$ 에 대해 확률변수 X 의 누적분포를 구하시오. 또한 $F(x)$ 를 사용하여 $f(2) = \frac{3}{8}$ 이 됨을 증명하시오.

• 예제 3.8에서 도출한 확률변수 X 의 확률분포

$$\bullet f(0) = \frac{1}{16}, f(1) = \frac{1}{4}, f(2) = \frac{3}{8}, f(3) = \frac{1}{4}, f(4) = \frac{1}{16}$$

• 이에 대한 누적분포함수를 구하면

$$\bullet F(0) = f(0) = \frac{1}{16}$$

$$\bullet F(1) = f(0) + f(1) = \frac{5}{16}$$

$$\bullet F(2) = f(0) + f(1) + f(2) = \frac{11}{16}$$

$$\bullet F(3) = f(0) + f(1) + f(2) + f(3) = \frac{15}{16}$$

$$\bullet F(4) = 1$$

x	$f(x)$	$F(x)$
0	$\frac{1}{16}$	$\frac{1}{16}$
1	$\frac{1}{4}$	$\frac{5}{16}$
2	$\frac{3}{8}$	$\frac{11}{16}$
3	$\frac{1}{4}$	$\frac{15}{16}$
4	$\frac{1}{16}$	1

[표 3.4] 예제 3.10의 가능한 값 x 에 따른 확률분포 $f(x)$ 와 누적분포함수 $F(x)$

확률분포와 분포함수

- 누적분포함수(CDF) (3/5)

- 예제 3.10 (2/2)

예제 3.8에서 구한 확률질량함수 $f(x) = \frac{\binom{4}{x}}{16}$ 에 대해 확률변수 X 의 누적분포를 구하시오. 또한 $F(x)$ 를 사용하여 $f(2) = \frac{3}{8}$ 이 됨을 증명하시오.

$$\therefore F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{16}, & 0 \leq x < 1 \\ \frac{5}{16}, & 1 \leq x < 2 \\ \frac{11}{16}, & 2 \leq x < 3 \\ \frac{15}{16}, & 3 \leq x < 4 \\ 1, & x \geq 4 \end{cases}$$

$$\text{또한 } f(2) = F(2) - f(0) - f(1) = F(2) - F(1) = \frac{3}{8}$$

x	$f(x)$	$F(x)$
0	$\frac{1}{16}$	$\frac{1}{16}$
1	$\frac{1}{4}$	$\frac{5}{16}$
2	$\frac{3}{8}$	$\frac{11}{16}$
3	$\frac{1}{4}$	$\frac{15}{16}$
4	$\frac{1}{16}$	1

[표 3.4] 예제 3.10의 가능한 값 x 에 따른 확률분포 $f(x)$ 와 누적분포함수 $F(x)$

확률분포와 분포함수

- 누적분포함수(CDF) (4/5)

- 예제 3.11

예제 3.9의 확률밀도함수 $f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2 \\ 0, & \text{그 외의 경우} \end{cases}$ 에 대해 $F(x)$ 를 구하고, 그것을 이용하여 $P(0 < X \leq 1)$ 을 구하시오.

- $-1 < x < 2$ 에 대해

$$F(x) = \int_{-\infty}^x f(t) dt = \int_{-1}^x \frac{t^2}{3} dt = \frac{t^3}{9} \Big|_{-1}^x = \frac{x^3+1}{9} \text{이므로}$$

$$F(x) = \begin{cases} 0, & x < -1 \\ \frac{x^3+1}{9}, & -1 \leq x < 2 \\ 1, & x \geq 2 \end{cases} \text{이고,}$$

$$P(0 < X \leq 1) = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9} \text{로 이전 결과와 동일함}$$

확률분포와 분포함수

- 누적분포함수(CDF) (5/5)

- 예제 3.12

어느 기관에서 프로젝트를 입찰에 부치고 적절한 입찰가를 b 로 예상한다고 할 때, 해당 기관은 낙찰가 y 의 밀도함수를 다음과 같이 생각하고 있다.

$$f(y) = \begin{cases} \frac{5}{8b}, & \frac{2}{5}b \leq y \leq 2b \\ 0, & \text{그 외의 경우} \end{cases}$$

$F(y)$ 를 구하고, 이것을 이용하여 낙찰가 y 가 해당 기관의 예상치 b 보다 작을 확률을 구하시오.

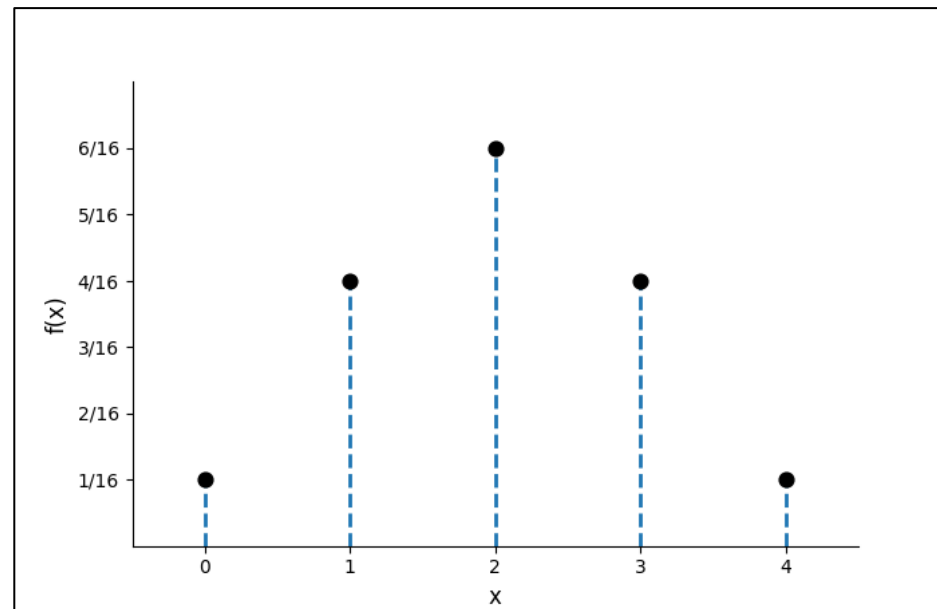
- $\frac{2b}{5} \leq y \leq 2b$ 이므로 $F(y) = \int_{\frac{2b}{5}}^y \frac{5}{8b} dt = \frac{5t}{8b} \Big|_{\frac{2b}{5}}^y = \frac{5y}{8b} - \frac{1}{4}$ 이 되어

$$F(y) = \begin{cases} 0, & y < \frac{2}{5}b \\ \frac{5y}{8b} - \frac{1}{4}, & \frac{2}{5}b \leq y < 2b \\ 1, & y \geq 2b \end{cases} \text{이며,}$$

낙찰가가 예상치 b 보다 작을 확률은 $P(Y \leq b) = F(b) = \frac{5}{8} - \frac{1}{4} = \frac{3}{8}$

확률분포와 분포함수

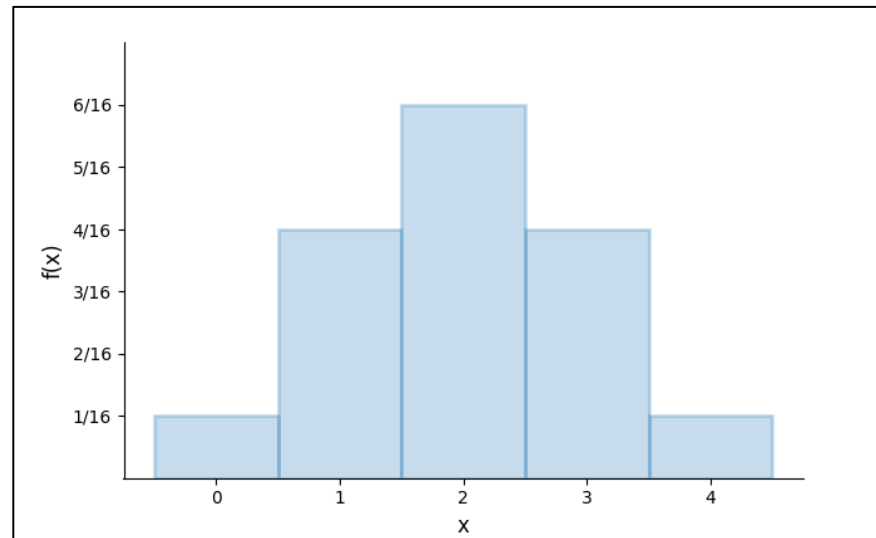
- 확률분포의 시각화 (1/4)
- 확률질량함수도(Probability Mass Function Plot)
 - 확률 값을 x 축과 점선이나 실선으로 연결하여 이산형 확률 분포를 나타낸 그림
 - 이산형 확률변수에서 각 값에 대응하는 확률을 점의 높이로써 직관적으로 파악할 수 있음



<그림 2> 예제 3.8의 확률질량함수도

확률분포와 분포함수

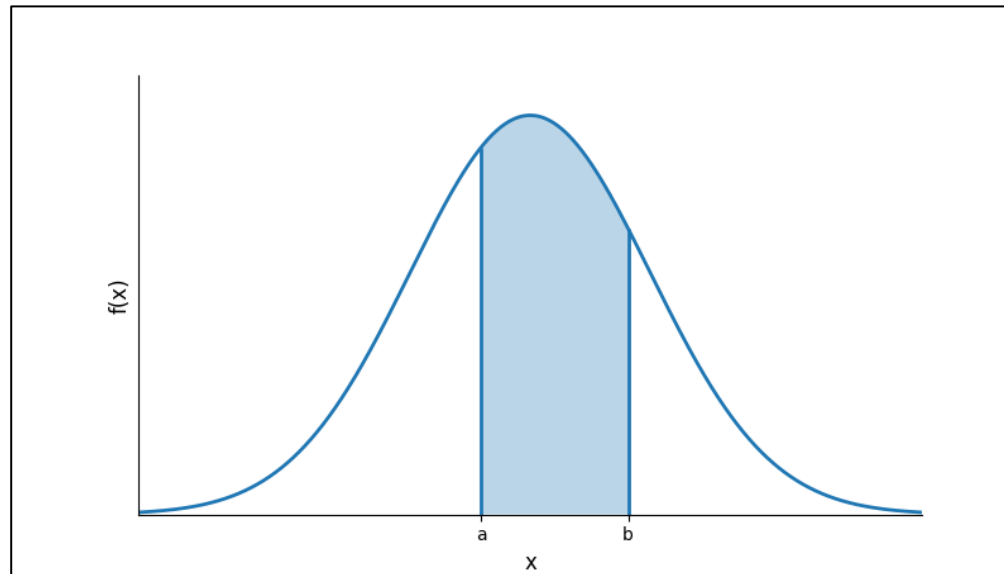
- 확률분포의 시각화 (2/4)
- 확률히스토그램(Probability Histogram)
 - 확률질량함수도에서 점 대신 직사각형을 사용해 나타낸 그림
 - 이산형 확률변수에서 확률을 막대의 면적으로써 표현하여, 확률을 구간 단위로 직관적으로 파악할 수 있음



<그림 3> 예제 3.8의 확률히스토그램

확률분포와 분포함수

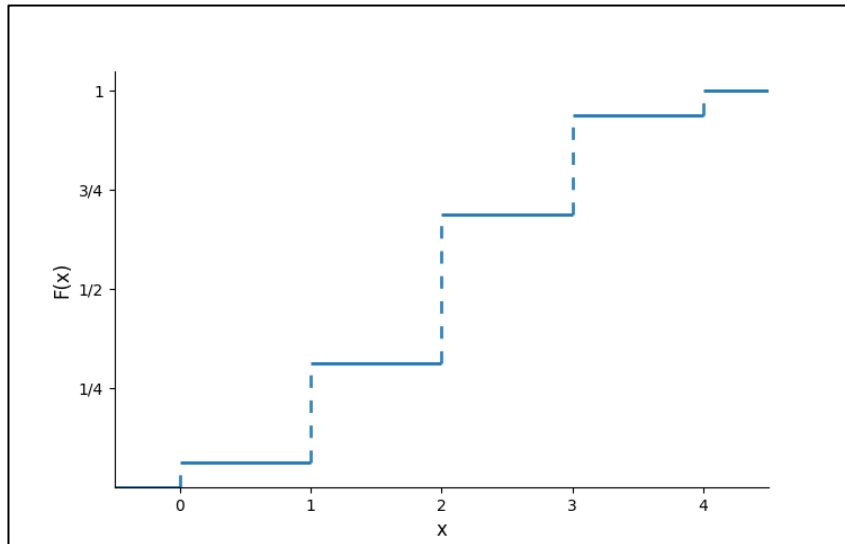
- 확률분포의 시각화 (3/4)
- 확률 밀도 함수도(Probability Mass Function Plot)
 - 연속형 확률분포를 확률 밀도 함수의 그래프로 나타낸 그림
 - 연속형 확률변수에서 확률을 구간의 면적으로써 직관적으로 파악할 수 있음



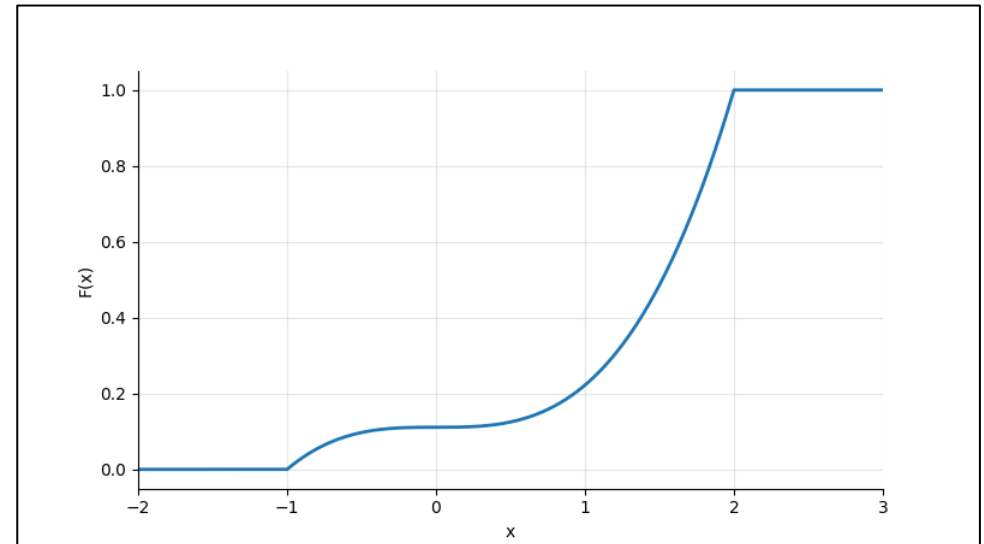
<그림 4> 정규분포에 대한 확률밀도함수도

확률분포와 분포함수

- 확률분포의 시각화 (4/4)
- 누적분포도(Discrete Cumulative Distribution Plot)
 - 확률분포의 누적분포를 그래프로 나타낸 그림
 - 확률변수가 특정 값 이하일 누적확률을 함수의 값으로써 직관적으로 파악할 수 있음
 - 확률을 누적된 함수값의 차이로써 표현하여 구간 확률을 쉽게 계산할 수 있음



<그림 5> 예제 3.10의 이산형 누적분포도



<그림 6> 예제 3.11의 연속형 누적분포도

목 차

- 확률변수(Random Variable)의 개념
- 확률분포(Probability Distribution)와 분포함수(Distribution Function)
- 결합확률분포(Joint Probability Distribution)

결합 확률 분포

• 결합 확률 분포 (Joint Probability Distribution) (1/6)

• 정의

- 두 개 이상의 확률변수가 동시에 취하는 값 또는 값들의 집합에 대한 확률을 표현한 것

• 필요성

- 여러 확률변수가 어떻게 분포하는지 나타내고 확률변수 간의 관계를 파악하는 데 사용할 수 있음
 - e.g., 공부 시간 x 와 시험 점수 y 를 통해 x 시간 공부했을 때의 시험 점수 y 의 분포를 구하여 공부 시간에 따른 시험 점수의 상관관계를 파악해볼 수 있음
- 확률변수 간의 독립성을 판단하는 데 사용될 수 있음
 - e.g., 동전 던지기와 주사위 굴리기를 동시에 진행할 때, 동전 던지기 결과와 주사위 눈이 동시에 특정 값을 가질 확률이 각각의 확률의 곱으로 표현될 수 있다면 두 확률변수는 통계적으로 독립임

결합 확률 분포

- 결합 확률 분포 (Joint Probability Distribution) (2/6)
 - 결합 확률 질량 함수 (Joint PMF, Joint Probability Mass Function)
 - 두 이산형 확률 변수에 대해 다음을 만족하는 확률 질량 함수
 - 모든 (x, y) 에 대해 $f(x, y) \geq 0$
 - $\sum_x \sum_y f(x, y) = 1$
 - $P(X = x, Y = y) = f(x, y)$
 - 결합 확률 밀도 함수 (Joint PDF, Joint Probability Density Function)
 - 두 연속형 확률 변수에 대해 다음을 만족하는 확률 밀도 함수
 - 모든 (x, y) 에 대해 $f(x, y) \geq 0$
 - $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = 1$
 - $P[(X, Y) \in A] = \int \int_A f(x, y) \, dx \, dy$ (A 는 xy 평면 상의 임의의 영역)

결합확률분포

• 결합확률분포(Joint Probability Distribution) (3/6)

• 예제 3.13

3개의 청색, 2개의 적색, 3개의 녹색 볼펜이 들어있는 상자에서 임의로 2개를 추출하고자 한다. X 를 청색 볼펜의 수, Y 를 적색 볼펜의 수라고 할 때, 다음을 구하시오.

(a) 결합확률분포 $f(x, y)$ 를 구하시오.

(b) $A = \{(x, y) | x + y \leq 1\}$ 라고 할 때, $P[(X, Y) \in A]$ 를 구하시오.

(a) 결합확률분포 $f(x, y)$

- 8개 볼펜 중 임의의 2개를 추출하는 경우의 수: $\binom{8}{2}$
- 청색 볼펜을 뽑는 경우의 수: $\binom{3}{x}$
- 적색 볼펜을 뽑는 경우의 수: $\binom{2}{y}$
- 녹색 볼펜을 뽑는 경우의 수: $\binom{3}{2-x-y}$

$$\therefore f(x, y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}}, \quad x = 0, 1, 2, y = 0, 1, 2, 0 \leq x + y \leq 2$$

$f(x, y)$		x			행의 합
		0	1	2	
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
열의 합		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

[표 3.5] 예제 3.13의 결합확률분포 표

$$(b) P[(X, Y) \in A] = P(X + Y \leq 1) = f(0, 0) + f(0, 1) + f(1, 0) = \frac{3}{28} + \frac{3}{14} + \frac{9}{28} = \frac{9}{14}$$

결합확률분포

- 결합확률분포(Joint Probability Distribution) (4/6)
- 예제 3.14 (1/3)

어느 과자회사에서는 연한 초콜릿과 진한 초콜릿을 입힌 과자상자를 취급하고 있다. 임의로 하나의 과자상자를 선택했을 때 X 와 Y 를 각각 연한 초콜릿과 진한 초콜릿의 비율이라 하면, 결합확률분포는 다음과 같이 주어진다.

$$f(x, y) = \begin{cases} \frac{2}{5}(2x + 3y), & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{그 외의 경우} \end{cases}$$

이때 다음을 구하시오.

- (a) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$ 을 증명하시오.
- (b) $A = \{(x, y) | 0 < x < \frac{1}{2}, \frac{1}{4} < y < \frac{1}{2}\}$ 일 때의 $P[(X, Y) \in A]$ 를 구하시오.

결합 확률 분포

- 결합 확률 분포 (Joint Probability Distribution) (5/6)
- 예제 3.14 (2/3)

어느 과자회사에서는 연한 초콜릿과 진한 초콜릿을 입힌 과자상자를 취급하고 있다. 임의로 하나의 과자상자를 선택했을 때 X 와 Y 를 각각 연한 초콜릿과 진한 초콜릿의 비율이라 하면, 결합 확률 분포는 다음과 같이 주어진다.

$$f(x, y) = \begin{cases} \frac{2}{5}(2x + 3y), & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{그 외의 경우} \end{cases}$$

이때 다음을 구하시오.

(a) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$ 을 증명하시오.

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy &= \int_0^1 \int_0^1 \frac{2}{5}(2x + 3y) dx dy = \int_0^1 \left(\frac{2x^2}{5} + \frac{6xy}{5} \right) \Big|_{x=0}^{x=1} dy \\ &= \int_0^1 \left(\frac{2}{5} + \frac{6y}{5} \right) dy = \left(\frac{2y}{5} + \frac{3y^2}{5} \right) \Big|_0^1 = \frac{2}{5} + \frac{3}{5} = 1 \end{aligned}$$

결합확률분포

- 결합확률분포(Joint Probability Distribution) (6/6)
- 예제 3.14 (3/3)

어느 과자회사에서는 연한 초콜릿과 진한 초콜릿을 입힌 과자상자를 취급하고 있다. 임의로 하나의 과자상자를 선택했을 때 X 와 Y 를 각각 연한 초콜릿과 진한 초콜릿의 비율이라 하면, 결합확률분포는 다음과 같이 주어진다.

$$f(x, y) = \begin{cases} \frac{2}{5}(2x + 3y), & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{그 외의 경우} \end{cases}$$

이때 다음을 구하시오.

(b) $A = \{(x, y) | 0 < x < \frac{1}{2}, \frac{1}{4} < y < \frac{1}{2}\}$ 일 때의 $P[(X, Y) \in A]$ 를 구하시오.

$$\begin{aligned} \bullet P[(X, Y) \in A] &= P\left(0 < X < \frac{1}{2}, \frac{1}{4} < Y < \frac{1}{2}\right) = \int_{\frac{1}{4}}^{\frac{1}{2}} \int_0^{\frac{1}{2}} \frac{2}{5}(2x + 3y) \, dx \, dy \\ &= \int_{\frac{1}{4}}^{\frac{1}{2}} \left(\frac{2x^2}{5} + \frac{6xy}{5}\right) \Big|_{x=0}^{x=\frac{1}{2}} dy = \int_{\frac{1}{4}}^{\frac{1}{2}} \left(\frac{1}{10} + \frac{3y}{5}\right) dy = \left(\frac{y}{10} + \frac{3y^2}{10}\right) \Big|_{\frac{1}{4}}^{\frac{1}{2}} \\ &= \frac{1}{10} \left[\left(\frac{1}{2} + \frac{3}{4}\right) - \left(\frac{1}{4} + \frac{3}{16}\right)\right] = \frac{1}{10} \left(\frac{5}{4} - \frac{7}{16}\right) = \frac{13}{160} \end{aligned}$$

결합 확률 분포

• 주변분포(Marginal Distribution) (1/4)

• 정의

- 결합 확률 분포에서 다른 확률 변수를 제거하여 얻어지는 확률 분포
- 이산형인 경우, $g(x) = \sum_y f(x, y), h(y) = \sum_x f(x, y)$
- 연속형인 경우, $g(x) = \int_{-\infty}^{\infty} f(x, y) dy, h(y) = \int_{-\infty}^{\infty} f(x, y) dx$

• 특징

- 확률 변수가 n 개일 때, 다음과 같이 확장하여 사용할 수 있음
 - $g(x_1) = \sum_{x_2} \cdots \sum_{x_n} f(x_1, x_2, \cdots, x_n),$ (이산형인 경우)
 - $g(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \cdots, x_n) dx_2 dx_3 \cdots dx_n,$ (연속형인 경우)
- 두 개 이상의 확률 변수에 대한 결합 주변분포(Joint Marginal Distribution)로 확장하여 사용할 수 있음
 - $g(x_1, x_2) = \sum_{x_3} \cdots \sum_{x_n} f(x_1, x_2, \cdots, x_n),$ (이산형인 경우)
 - $g(x_1, x_2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \cdots, x_n) dx_3 dx_4 \cdots dx_n,$ (연속형인 경우)

결합 확률 분포

- 주변분포(Marginal Distribution) (2/4)

- 필요성

- 여러 확률변수가 존재하는 데이터에서 특정 확률변수의 정보만을 얻어야 할 때 사용할 수 있음
 - e.g., 기말 점수 X , 학년 Y , 전공 Z 에 대한 결합확률분포가 주어졌을 때, 학년 Y 의 분포만 얻고자 하는 경우
- 통계를 분석하기 전에 각 확률변수에 대한 각각의 분포를 알아내야 할 경우에 사용할 수 있음
 - e.g., 소득 X 에 따른 소비 Y 에 대한 분석을 하기 이전에 소득 X 와 소비 Y 의 분포를 먼저 분석하고자 하는 경우

결합확률분포

- 주변분포(Marginal Distribution) (3/4)
- 예제 3.15

표 3.5의 열과 행의 합이 각각 x 와 y 의 주변분포가 됨을 증명하시오.

- 확률변수 x 에 대해

$$g(0) = f(0,0) + f(0,1) + f(0,2) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14}$$

$$g(1) = f(1,0) + f(1,1) + f(1,2) = \frac{9}{28} + \frac{3}{14} + 0 = \frac{15}{28}$$

$$g(2) = f(2,0) + f(2,1) + f(2,2) = \frac{3}{28} + 0 + 0 = \frac{3}{28}$$

$f(x,y)$		x			행의 합
		0	1	2	
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
열의 합		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

[표 3.5] 예제 3.13의 결합확률분포 표

위 결과는 표 3.1의 열의 합과 일치하며, 동일한 방법으로 $h(y)$ 의 값들이 행의 합과 일치함을 알 수 있음

결합 확률 분포

- 주변분포(Marginal Distribution) (4/4)

- 예제 3.16

예제 3.14의 결합분포 $f(x, y) = \begin{cases} \frac{2}{5}(2x + 3y), & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{그 외의 경우} \end{cases}$ 에 대해 $g(x)$ 와 $h(y)$ 를 구하시오.

- $0 \leq x \leq 1$ 구간에서

$$\begin{aligned} g(x) &= \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 \frac{2}{5}(2x + 3y) dy \\ &= \left(\frac{4xy}{5} + \frac{6y^2}{10} \right) \Big|_{y=0}^{y=1} = \frac{4x+3}{5} \text{이며,} \\ &\text{그 외의 경우에는 0이 됨} \end{aligned}$$

- $0 \leq y \leq 1$ 구간에서

$$\begin{aligned} h(y) &= \int_{-\infty}^{\infty} f(x, y) dx = \int_0^1 \frac{2}{5}(2x + 3y) dx \\ &= \left(\frac{2x^2}{5} + \frac{6xy}{5} \right) \Big|_{x=0}^{x=1} = \frac{2(1+3y)}{5} \text{이며,} \\ &\text{그 외의 경우에는 0이 됨} \end{aligned}$$

결합 확률 분포

• 조건부 분포(Conditional Distribution) (1/7)

• 정의

조건부 확률(Conditional Probability)
$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

- 한 확률변수가 특정 값 또는 구간을 가졌다는 조건하에서 다른 확률변수의 확률 분포

- $f(y|x) = \frac{f(x,y)}{g(x)} \quad (g(x) > 0, X = x), \quad f(x|y) = \frac{f(x,y)}{h(y)} \quad (h(y) > 0, Y = y)$

• 특징

- 두 개 이상의 확률변수에 대한 결합조건부 분포(Joint Conditional Distribution)로 확장하여 사용할 수 있음

- $f(x_1, x_2, \dots, x_{a-1} | x_a, x_{a+1}, \dots, x_n) = \frac{f(x_1, x_2, \dots, x_n)}{g(x_a, x_{a+1}, \dots, x_n)}$

- $Y = y$ 일 때 X 가 $a < X < b$ 일 확률은 다음과 같이 구할 수 있음

- $P(a < X < b | Y = y) = \sum_{a < X < b} f(x|y), \quad (\text{이산형인 경우})$

- $P(a < X < b | Y = y) = \int_a^b f(x|y) dx, \quad (\text{연속형인 경우})$

결합 확률 분포

- 조건부분포(Conditional Distribution) (2/7)

- 필요성

- 여러 확률변수가 존재하는 데이터에서 특정 조건에서의 확률분포가 필요한 경우에 사용할 수 있음
 - e.g., 공부 시간 X 에 따른 시험 점수 Y 에 대한 데이터가 있을 때, 공부 시간 X 가 5 시간일 때의 시험 점수 Y 에 대한 분포가 필요한 경우
- 예측, 의사결정 등을 할 때의 조건부 판단에 사용할 수 있음
 - e.g., 기계 과열 여부 X 와 고장 여부 Y 에서 과열 상태일 때의 고장날 확률을 계산하여 위험도 판단에 이용할 수 있음

결합확률분포

• 조건부분포(Conditional Distribution) (3/7)

• 예제 3.17

예제 3.13에서 $Y = 1$ 로 주어졌을 때 X 의 조건부분포를 구하고, 그것을 이용하여 $P(X = 0|Y = 1)$ 을 구하시오.

- $y = 1$ 일 때의 $f(x|y)$ 가 필요하므로 $h(1)$ 을 구해야 함

- $h(1) = \sum_{x=0}^2 f(x, 1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}$ 이므로
 $f(x|1) = \frac{f(x,1)}{h(1)} = \frac{7}{3} f(x, 1), x = 0, 1, 2$ 이 됨

- $f(0|1) = \left(\frac{7}{3}\right) f(0,1) = \left(\frac{7}{3}\right) \left(\frac{3}{14}\right) = \frac{1}{2},$

$$f(1|1) = \left(\frac{7}{3}\right) f(1,1) = \left(\frac{7}{3}\right) \left(\frac{3}{14}\right) = \frac{1}{2},$$

$$f(2|1) = \left(\frac{7}{3}\right) f(2,1) = \left(\frac{7}{3}\right) (0) = 0$$

- $P(X = 0|Y = 1) = f(0|1) = \frac{1}{2}$

$f(x,y)$		x			행의 합
		0	1	2	
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
열의 합		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

[표 3.5] 예제 3.13의 결합확률분포 표

결합 확률 분포

- 조건부분포(Conditional Distribution) (4/7)

- 예제 3.18 (1/3)

확률변수 X 와 Y 를 각각 단위 온도 변화량과 어떤 원자가 방출하는 스펙트럼 변화율을 나타내는 확률변수라 할 때, 확률변수 (X, Y) 에 대한 결합밀도함수는 다음과 같다.

$$f(x, y) = \begin{cases} 10xy^2, & 0 < x < y < 1 \\ 0, & \text{그 외의 경우} \end{cases}$$

이 때 다음을 구하시오.

(a) 주변밀도함수 $g(x), h(y)$ 와 조건부밀도함수 $f(y|x)$ 를 구하시오.

(b) 온도가 0.25 단위 높아졌을 때 스펙트럼 변화량이 $\frac{1}{2}$ 보다 클 확률을 구하시오.

결합 확률 분포

• 조건부분포(Conditional Distribution) (5/7)

• 예제 3.18 (2/3)

확률변수 X 와 Y 를 각각 단위 온도 변화량과 어떤 원자가 방출하는 스펙트럼 변화율을 나타내는 확률변수라 할 때, 확률변수 (X, Y) 에 대한 결합밀도함수는 다음과 같다.

$$f(x, y) = \begin{cases} 10xy^2, & 0 < x < y < 1 \\ 0, & \text{그 외의 경우} \end{cases}$$

(a) 주변밀도함수 $g(x), h(y)$ 와 조건부밀도함수 $f(y|x)$ 를 구하시오.

$$\begin{aligned} \bullet \quad g(x) &= \int_{-\infty}^{\infty} f(x, y) dy = \int_x^1 10xy^2 dy = \frac{10}{3} xy^3 \Big|_{y=x}^{y=1} \\ &= \frac{10}{3} x(1 - x^3), 0 < x < 1 \end{aligned}$$

$$\begin{aligned} \bullet \quad h(y) &= \int_{-\infty}^{\infty} f(x, y) dx = \int_0^y 10xy^2 dx = 5x^2 y^2 \Big|_{x=0}^{x=y} \\ &= 5y^4, 0 < y < 1 \end{aligned}$$

$$\bullet \quad f(y|x) = \frac{f(x, y)}{g(x)} = \frac{10xy^2}{\frac{10}{3}x(1-x^3)} = \frac{3y^2}{1-x^3}, 0 < x < y < 1$$

결합 확률 분포

- 조건부 분포(Conditional Distribution) (6/7)

- 예제 3.18 (3/3)

확률변수 X 와 Y 를 각각 단위 온도 변화량과 어떤 원자가 방출하는 스펙트럼 변화율을 나타내는 확률변수라 할 때, 확률변수 (X, Y) 에 대한 결합 밀도 함수는 다음과 같다.

$$f(x, y) = \begin{cases} 10xy^2, & 0 < x < y < 1 \\ 0, & \text{그 외의 경우} \end{cases}$$

(b) 온도가 0.25 단위 높아졌을 때 스펙트럼 변화량이 $\frac{1}{2}$ 보다 클 확률을 구하시오.

- $$\begin{aligned} P\left(Y > \frac{1}{2} \mid X = 0.25\right) &= \int_{\frac{1}{2}}^1 f(y \mid x = 0.25) dy \\ &= \int_{\frac{1}{2}}^1 \frac{3y^2}{1 - 0.25^3} dy \\ &= \frac{64}{63} y^3 \Big|_{\frac{1}{2}}^1 = \frac{8}{9} \end{aligned}$$

결합확률분포

• 조건부분포(Conditional Distribution) (7/7)

• 예제 3.19

결합확률분포가 다음과 같이 주어졌을 때 $g(x), h(x), f(x|y)$ 를 구하고, $P(\frac{1}{4} < X < \frac{1}{2} | Y = \frac{1}{3})$ 의 값을 구하시오.

$$f(x, y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2, 0 < y < 1 \\ 0, & \text{그 외의 경우} \end{cases}$$

- $g(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 \frac{x(1+3y^2)}{4} dy = \left(\frac{xy}{4} + \frac{xy^3}{4} \right) \Big|_{y=0}^{y=1} = \frac{x}{2}, \quad 0 < x < 2$
- $h(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_0^2 \frac{x(1+3y^2)}{4} dx = \left(\frac{x^2}{8} + \frac{3x^2y^2}{8} \right) \Big|_{x=0}^{x=2} = \frac{1+3y^2}{2}, \quad 0 < y < 1$
- $f(x|y) = \frac{f(x, y)}{h(y)} = \frac{\frac{x(1+3y^2)}{4}}{\frac{1+3y^2}{2}} = \frac{x}{2}, \quad 0 < x < 2$
- $P\left(\frac{1}{4} < X < \frac{1}{2} \mid Y = \frac{1}{3}\right) = \int_{\frac{1}{4}}^{\frac{1}{2}} \frac{x}{2} dx = \frac{x^2}{4} \Big|_{\frac{1}{4}}^{\frac{1}{2}} = \frac{3}{64}$

결합 확률 분포

• 통계적 독립(Statistically Independent) (1/3)

• 정의

독립사상(Independent Event)
 $P(B|A) = P(B)$

- 결합 확률 분포 $f(x, y)$, 주변 분포 $g(x), h(y)$ 를 가지는 확률 변수 X, Y 가 모든 (x, y) 에 대해 $f(x, y) = g(x)h(y)$ 이 성립하는 관계

• 특징

- $f(x|y)$ 에서 y 가 종속되지 않아 확률 변수 Y 의 발생이 확률 변수 X 의 발생으로부터 아무런 영향을 받지 않음을 의미함
- 두 개 이상의 확률 변수에 대해 다음과 같이 확장하여 상호 통계적 독립(Mutual Independent)으로 정의할 수 있음
 - 결합 확률 분포 $f(x_1, x_2, \dots, x_n)$ 과 주변 분포 $f_1(x_1), f_2(x_2), \dots, f_n(x_n)$ 을 가지는 확률 변수 $X_1, X_2, \dots, X_n$ 가 모든 $(x_1, x_2, \dots, x_n)$ 에 대해 $f(x_1, x_2, \dots, x_n) = f_1(x_1)f_2(x_2)\dots f_n(x_n)$ 인 관계

결합 확률 분포

- 통계적 독립(Statistically Independent) (2/3)
- 예제 3.20

예제 3.13의 확률변수들이 통계적으로 독립이 아님을 증명하시오.

- $f(0,1) = \frac{3}{14}$
- $g(0) = \sum_{y=0}^2 f(0,y) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14}$
- $h(1) = \sum_{x=0}^2 f(x,1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}$
- $f(0,1) \neq g(0)h(1) \therefore X$ 와 Y 는 통계적으로 독립이 아님

$f(x,y)$		x			행의 합
		0	1	2	
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
열의 합		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

[표 3.5] 예제 3.13의 결합확률분포 표

결합 확률 분포

• 통계적 독립(Statistically Independent) (3/3)

• 예제 3.21

종이팩으로 포장된 부패성 식품이 보존기간(단위: 년)이 다음과 같은 확률밀도함수를 가지는 확률변수라고 하자.

$$f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & \text{그 외의 경우} \end{cases}$$

X_1, X_2, X_3 가 독립적으로 추출된 3개의 포장된 음식의 보존기간을 나타낸다고 할 때, $P(X_1 < 2, 1 < X_2 < 3, X_3 > 2)$ 를 구하시오.

- 3개의 포장단위가 독립적으로 추출됐으므로 세 확률변수가 통계적으로 독립이라 할 수 있음
- $f(x_1, x_2, x_3) = f(x_1)f(x_2)f(x_3) = e^{-x_1}e^{-x_2}e^{-x_3} = e^{-x_1-x_2-x_3}$
- $$P(X_1 < 2, 1 < X_2 < 3, X_3 > 2) = \int_2^\infty \int_1^3 \int_0^2 e^{-x_1-x_2-x_3} dx_1 dx_2 dx_3$$
$$= (1 - e^{-2})(e^{-1} - e^{-3})e^{-2} \approx 0.03722$$

Thanks!

이 성 현(seonghyeon@pel.sejong.ac.kr)

부록#1

• <그림 1> 다양한 형태의 확률밀도함수 그래프 코드

```
1. import numpy as np
2. import matplotlib.pyplot as plt
3.
4. fig, axs = plt.subplots(1, 4, figsize=(12, 4))
5.
6. # (a) 비대칭 단봉
7. x = np.linspace(0, 6, 500)
8. y = x * np.exp(-x)
9. axs[0].plot(x, y)
10.
11. # (b) 직선
12. x = np.linspace(0, 5, 500)
13. y = 0.5 * x + 0.2
14. axs[1].plot(x, y)
15.
16. # (c) 종 모양
17. x = np.linspace(-3, 3, 500)
18. y = np.exp(-x**2)
19. axs[2].plot(x, y)
```

```
20. # (d) U자형
21. x = np.linspace(-3, 3, 500)
22. y = x**2
23. axs[3].plot(x, y)
24.
25. for ax in axs:
26.     ax.set_xticks([])
27.     ax.set_yticks([])
28.
29. plt.tight_layout()
30. plt.show()
```

부록#2

• <그림 2> 예제 3.8의 확률질량함수도 코드

```
1. import numpy as np
2. import matplotlib.pyplot as plt
3.
4. # 표본 데이터 생성
5. x = np.array([0, 1, 2, 3, 4])
6. p = np.array([1, 4, 6, 4, 1]) / 16
7.
8. plt.figure(figsize=(7, 4.5))
9.
10. # 수직 점선
11. plt.vlines(x, 0, p, linestyle="dashed",
            linewidth=2)
12. plt.scatter(x, p, color="black", zorder=3, s=60)
13.
14. # 축 설정
15. plt.xlabel("x", fontsize=12)
16. plt.ylabel("f(x)", fontsize=12)
17. plt.xlim(-0.5, 4.5)
18. plt.ylim(0, 7/16)
```

```
19. # y축 눈금을 분수로 표시
20. yticks = np.arange(1, 7) / 16
21. plt.yticks(yticks, [f"{i}/16" for i in
                    range(1, 7)])
22.
23. plt.xticks(x)
24. plt.gca().spines["top"].set_visible(False)
25. plt.gca().spines["right"].set_visible(False)
26.
27. plt.show()
```

부록#3

• <그림 3> 예제 3.8의 확률히스토그램 코드

```
1. import numpy as np
2. import matplotlib.pyplot as plt
3.
4. # 표본 데이터 생성
5. data = np.array(
6.     [0]*1 +
7.     [1]*4 +
8.     [2]*6 +
9.     [3]*4 +
10.    [4]*1
11. )
12.
13. plt.figure(figsize=(7, 4.5))
14.
15. # 히스토그램
16. plt.hist(
17.     data,
18.     bins=[-0.5, 0.5, 1.5, 2.5, 3.5, 4.5],
19.     density=True,
20.     edgecolor="tab:blue",
21.     linewidth=2,
22.     alpha=0.25
23. )
```

```
24. # 축 라벨
25. plt.xlabel("x", fontsize=12)
26. plt.ylabel("f(x)", fontsize=12)
27.
28. # x축 눈금
29. plt.xticks([0, 1, 2, 3, 4])
30.
31. # y축 눈금을 분수로 (1/16 ~ 6/16)
32. yticks = np.arange(1, 7) / 16
33. plt.yticks(yticks, [f"{i}/16" for i in
34.    range(1, 7)])
35.
36.
37. ax = plt.gca()
38. ax.spines["top"].set_visible(False)
39. ax.spines["right"].set_visible(False)
40.
41. plt.show()
```

부록#4

• <그림 4> 정규분포에 대한 확률밀도함수도 코드

```
1. import numpy as np
2. import matplotlib.pyplot as plt
3.
4. # 함수 정의
5. def f(x):
6.     return (1 / np.sqrt(2 * np.pi)) *
7.     np.exp(-x**2 / 3)
8.
9. # 구간
10. a, b = -0.5, 1.0
11.
12. # x 범위
13. x = np.linspace(-4, 4, 2000)
14. y = f(x)
15.
16. plt.figure(figsize=(8, 4.5))
17.
18. # 곡선
19. plt.plot(x, y, linewidth=2)
20.
21. # 음영 구간
22. xs = np.linspace(a, b, 500)
23. plt.fill_between(xs, f(xs), alpha=0.3)
```

```
23. # 수직선
24. plt.vlines([a, b], ymin=0, ymax=f(np.array(
25.     [a, b])), linewidth=2)
26.
27. # 축 스타일
28. plt.xlabel("x", fontsize=12)
29. plt.ylabel("f(x)", fontsize=12)
30. plt.xlim(-4, 4)
31. plt.ylim(0, y.max() * 1.1)
32.
33. # 눈금 제거
34. plt.xticks([a, b], ["a", "b"])
35. plt.yticks([])
36.
37. ax = plt.gca()
38. ax.spines["top"].set_visible(False)
39. ax.spines["right"].set_visible(False)
40. plt.show()
```

부록#5

• <그림 5> 예제 3.10의 이산형 누적분포도 코드

```
1. import numpy as np
2. import matplotlib.pyplot as plt
3.
4. # 표본 데이터 생성
5. x = np.array([0, 1, 2, 3, 4])
6. p = np.array([1, 4, 6, 4, 1]) / 16
7. F = np.cumsum(p)
8.
9. plt.figure(figsize=(7.5, 4.8))
10.
11. # 수평 구간 (실선)
12. x_left = np.r_[-0.5, x]
13. x_right = np.r_[x, 4.5]
14. y_vals = np.r_[0, F]
15.
16. for xl, xr, y in zip(x_left, x_right, y_vals):
17.     plt.hlines(
18.         y, xl, xr,
19.         linewidth=2
20.     )
```

```
21. # 수직 점프 (점선)
22. for xi, pi, Fi in zip(x, p, F):
23.     plt.vlines(
24.         xi,
25.         Fi - pi,
26.         Fi,
27.         linestyle=(0, (4, 4)),
28.         linewidth=2,
29.         alpha=0.9
30.     )
31.
32. # 축 설정
33. plt.xlabel("x", fontsize=12)
34. plt.ylabel("F(x)", fontsize=12)
35. plt.xlim(-0.5, 4.5)
36. plt.ylim(0, 1.05)
37. plt.xticks(x)
38. plt.yticks([1/4, 1/2, 3/4, 1], ["1/4", "1/2",
39.                                     "3/4", "1"])
40.
41. ax = plt.gca()
42. ax.spines["top"].set_visible(False)
43. ax.spines["right"].set_visible(False)
44. plt.show()
```

부록#6

• <그림 6> 예제 3.11의 연속형 누적분포도 코드

```
1. import numpy as np
2. import matplotlib.pyplot as plt
3.
4. def F(x): #연속형 누적분포함수
5.     x = np.asarray(x, dtype=float)
6.     y = np.zeros_like(x)
7.
8.     mid = (x >= -1) & (x < 2)
9.     right = (x >= 2)
10.
11.     y[mid] = (x[mid]**3 + 1.0) / 9.0
12.     y[right] = 1.0
13.     return y
14.
15. xs = np.linspace(-2.0, 3.0, 2000)
16. ys = F(xs)
17.
18. plt.figure(figsize=(8, 4.5))
19. plt.plot(xs, ys, linewidth=2)
```

```
20. plt.xlim(-2, 3)
21. plt.ylim(-0.05, 1.05)
22. plt.xlabel("x")
23. plt.ylabel("F(x)")
24. plt.grid(True, alpha=0.3)
25.
26. ax = plt.gca()
27. ax.spines["top"].set_visible(False)
28. ax.spines["right"].set_visible(False)
29.
30. plt.show()
```