김은영

제3장 확률과 확률분포

3.1 서론

(1) 확률 (probability)
: 통계학의 주된 목표는 표본에서 얻어진 정보를 기초로 하여 관심 있는 모집단의 특성을 추론하는 것이다. 이런 추론이 얼마나 믿을만한 것인지를 제시해 줄 수 있는 이론적 근거가 '확률'이다.

(2) 확률모형에 대한 법칙
① 표본공간(sample space Ω): 확률실험에서 발생할 수 있는 모든 결과들의 집합
② 근원사상(elementary events ω)[표본공간에 있는 각각의 결과]에 확률을 부여
❑ 근원사상의 예: 동전던지기 실험
- $\Omega = \{HH, HT, TH, TT\}$ - $\{HH\}$는 근원사상 ω의 한 예

3.2 사상의 확률

(1) 사상 (event): 어떤 특성을 갖는 근원사상들의 모임

(2) 표기: $P(A)$

(3) 확률의 성질
① $0 \le P(A) \le 1$ for all $A \subset \Omega$
② $P(A) = \sum_{\omega \in A} P(\omega)$
③ $P(\Omega) = \sum_{\omega \in \Omega} P(\omega) = 1$

3.3 확률의 계산

3.3.1 균일 확률모형(uniform probability model) [예제 3.2~4]

$\Omega=\{\omega_1, \cdots, \omega_k\}$, $A=\{\omega_1, \cdots, \omega_m\}$, $A\subset\Omega$

$$\boxed{P(\omega_1)= \cdots = P(w_k)= \frac{1}{k}}$$

$$P(A)= \frac{m}{k}\left(=\frac{A\text{에 있는 근원사상의 개수}}{\Omega\text{에 있는 근원사상의 개수}}\right)$$

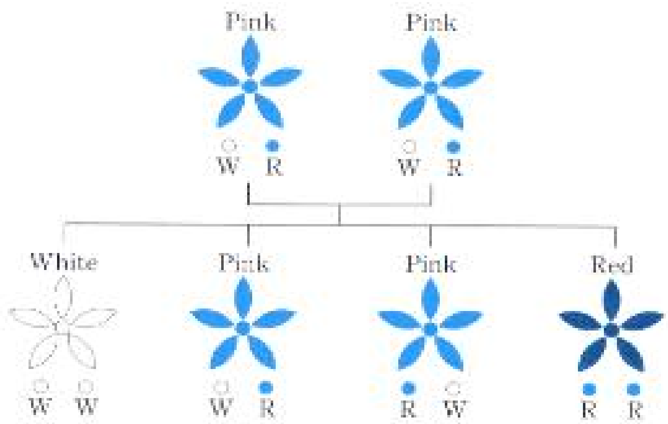
[예제 3.2] 하나의 공정한 동전을 세 번 던져 정확히 앞면이 한 번 나올 확률을 구하라.

【풀이】

표본공간 $\Omega=\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

공정한 동전 → 표본공간 Ω에 있는 8개의 근원사상이 발생할 확률이 동일

$A=\{\text{하나의 앞면}\}=\{HTT, THT, TTH\}$ → $P(A)=\frac{3}{8}$



[예제 3.3] 유전학의 선구자인 멘델은 완두콩의 특성에 어떤 형태가 있다는 확신을 갖고, 그것을 설명할 유전학 이론을 연구하였다. 그에 따르면, 유전된 특성은 유전자들에 의하여 한 세대에서 다른 세대로 전해지고, 유전자들은 쌍으로 나타나며 부모로부터 하나씩 받은 유전자들로 이루어져 있다. 멘델의 유전이론 실험은 다음과 같다. 순수 혈통인 빨간 꽃과 하얀 꽃을 교배하면 각각의 유전자를 가진 분홍 꽃 잡종이 만들어진다. 이 잡종들을

다시 교배하면 가능한 네 가지 유전자쌍 중 하나가 나타난다. 멘델 법칙에 의하면 이 네 가지 경우들은 균일하게 나타난다. 결론적으로

$$P(\text{분홍 꽃})=\frac{1}{2},\ P(\text{흰 꽃})=P(\text{빨간 꽃})=\frac{1}{4}$$

[예제 3.4] 어느 학급 50명 중, 42명이 오른손잡이이고 나머지 8명이 왼손잡이다. 그 학급에서 한 학생을 임의 선택할 때 선택된 학생이 왼손잡이일 확률을 구하라.

【풀이】

임의선택: 각 학생이 선택될 가능성이 같다

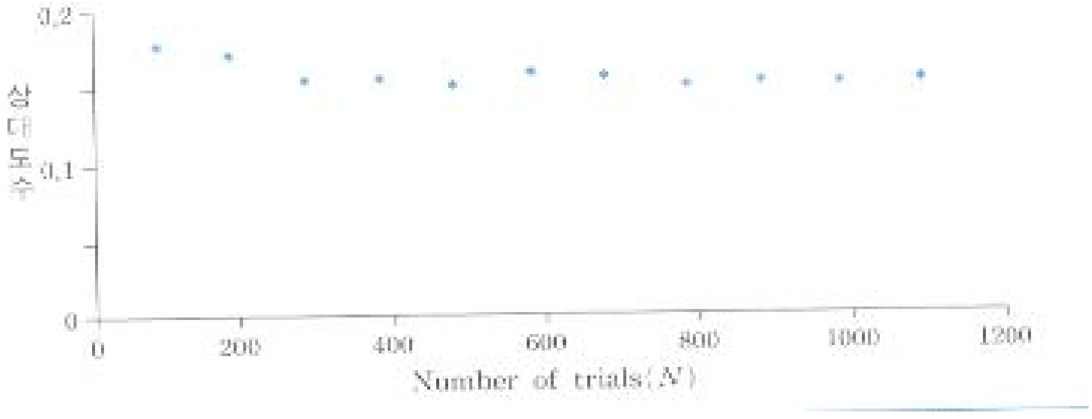
→ 표본공간은 50개의 근원사상으로 구성

$P(\text{왼손잡이일 사상})=\frac{8}{50}=0.16$

3.3.2 상대도수 수렴치를 이용한 확률

(1) 근원 사상들에 대한 균일성(uniformity)의 가정이 유지될 수 없을 때, 한 사상의 확률[예컨대, $P(A)$]을 어떻게 정의해야 하는가?

→ 실험을 여러 번 반복한 후 그 사건이 일어날 비율을 관찰한다.



(2) 상대도수 수렴치를 이용한 확률

사상 A의 확률 $P(A)$는 시행 수가 증가되어질 때 상대도수가 안정되는 값으로 정의한다. 비록 $P(A)$의 정확한 값은 모르지만, 실험을 수 없이 많이 반복함으로써 거의 비슷한 값으로 추정할 수 있다.

3.3.3 사상 간의 관계와 두 가지 확률의 법칙

$P(\overline{A}) = 1 - P(A)$	$\overline{A}$: 사상 A의 여집합
$P(A \cup B) = P(A) + P(B) - P(AB)$	A와 B는 배반사상이 아니다.
$P(A \cup B) = P(A) + P(B)$	A와 B는 배반사상이다. $AB = \varnothing, P(AB) = 0$

[예제 3.5] 동전을 두 번 던지는 실험의 벤다이어그램을 그리고 다음 사상을 규정지어라.

(1) 두 번째 던졌을 때 뒷면이 나옴 (A)

【풀이】

표본공간 $\Omega = \{HH, HT, TH, TT\}$

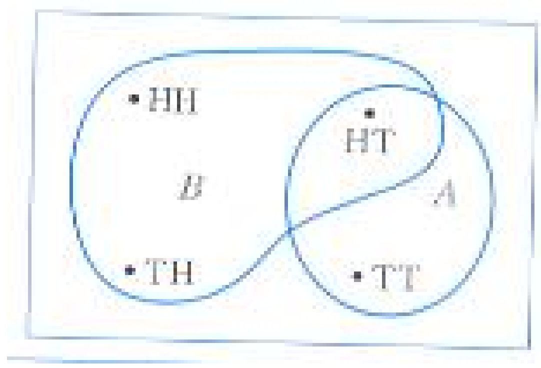
사상 $A = \{HT, TT\}$

(2) 적어도 한 번의 앞면 (B)

【풀이】

표본공간 $\Omega = \{HH, HT, TH, TT\}$

사상 $B = \{HH, HT, TH\}$



[예제 3.6] 다음 표는 실험실에서 약물실험에 쓰이는 원숭이 네 마리의 종류와 나이다.

원숭이	종류	나이
1	Baboon	6
2	Baboon	8
3	Spider	6
4	Spider	6

두 마리의 원숭이가 임의로 선택되고 실험할 약물을 투여한다. 두 마리의 원숭이를 선택하는 모든 가능한 경우를 고려하여 벤다이어그램을 만들고 다음 사건들을 나타내어라.

A: 선택된 원숭이들은 같은 종류다.

B: 선택된 원숭이들은 같은 나이다.

【풀이】

<table>
<tr><td>e_1</td><td>$\{1,2\}$</td><td>같은 종류 / 다른 나이</td><td rowspan="6">$A=\{e_1,\ e_6\}$
$B=\{e_2,\ e_3,\ e_6\}$</td></tr>
<tr><td>e_2</td><td>$\{1,3\}$</td><td>다른 종류 / 같은 나이</td></tr>
<tr><td>e_3</td><td>$\{1,4\}$</td><td>다른 종류 / 같은 나이</td></tr>
<tr><td>e_4</td><td>$\{2,3\}$</td><td>다른 종류 / 다른 나이</td></tr>
<tr><td>e_5</td><td>$\{2,4\}$</td><td>다른 종류 / 다른 나이</td></tr>
<tr><td>e_6</td><td>$\{3,4\}$</td><td>같은 종류 / 같은 나이</td></tr>
</table>

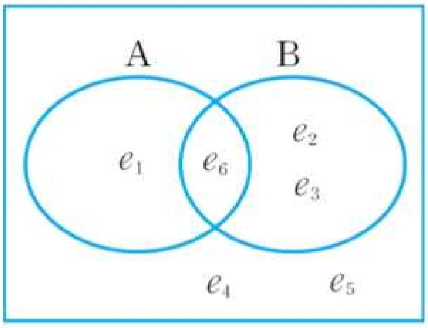
[예제 3.7] 네 마리 원숭이 중 두 마리를 뽑는 [예제 3.6]의 실험을 다시 생각해 보자.

$$A=\{\text{같은 종류}\},\ B=\{\text{같은 나이}\},\ C=\{\text{다른 종류}\}$$

다음 사상들을 정의해 보라.

$$C,\ \overline{A},\ A\cup B,\ AB,\ BC$$

<table>
<tr><td>e_1</td><td>$\{1,2\}$</td><td>같은 종류 / 다른 나이</td><td rowspan="6">$A=\{e_1,\ e_6\}$
$B=\{e_2,\ e_3,\ e_6\}$
$C=\{e_2,\ e_3,\ e_4,\ e_5\}$</td></tr>
<tr><td>e_2</td><td>$\{1,3\}$</td><td>다른 종류 / 같은 나이</td></tr>
<tr><td>e_3</td><td>$\{1,4\}$</td><td>다른 종류 / 같은 나이</td></tr>
<tr><td>e_4</td><td>$\{2,3\}$</td><td>다른 종류 / 다른 나이</td></tr>
<tr><td>e_5</td><td>$\{2,4\}$</td><td>다른 종류 / 다른 나이</td></tr>
<tr><td>e_6</td><td>$\{3,4\}$</td><td>같은 종류 / 같은 나이</td></tr>
</table>



【풀이】

$C=\{e_2, e_3, e_4, e_5\}$

$\overline{A}=\{e_2, e_3, e_4, e_5\} \leftarrow A=\{e_1,\ e_6\}$

$A\cup B=\{e_1, e_2, e_3, e_6\} \leftarrow A=\{e_1,\ e_6\}$, $B=\{e_2,\ e_3,\ e_6\}$

$AB=\{e_6\} \leftarrow A=\{e_1,\ e_6\}$, $B=\{e_2,\ e_3,\ e_6\}$

$BC=\{e_2, e_3\} \leftarrow B=\{e_2,\ e_3,\ e_6\}$, $C=\{e_2, e_3, e_4, e_5\}$

[예제 3.8] 임의의 4마리의 원숭이로부터 2마리를 뽑는 [예제 3.6]을 다시 보자. 뽑혀진 원숭이가 같은 종류이거나 같은 나이일 확률은 얼마인가?

【풀이】

$A=\{\text{같은 종류}\} \rightarrow A=\{e_1, e_6\} \rightarrow P(A)=\dfrac{2}{6}$

$B=\{\text{같은 나이}\} \rightarrow B=\{e_2, e_3, e_6\} \rightarrow P(B)=\dfrac{3}{6}$

$AB=\{e_6\} \leftarrow A=\{e_1, e_6\},\ B=\{e_2, e_3, e_6\} \Rightarrow P(AB)=\dfrac{1}{6}$

$\therefore\ P(A\cup B)=P(A)+P(B)-P(AB)=\dfrac{2}{6}+\dfrac{3}{6}-\dfrac{1}{6}=\dfrac{4}{6}=\dfrac{2}{3}$

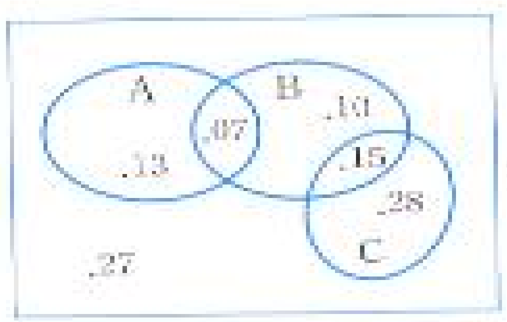
$cf)\ A\cup B=\{e_1, e_2, e_3, e_6\} \leftarrow A=\{e_1, e_6\},\ B=\{e_2, e_3, e_6\} \Rightarrow P(A\cup B)=\dfrac{4}{6}=\dfrac{2}{3}$

[예제 3.9] 벤다이어그램을 이용하면 세 사상 A, B, C 뿐 아니라 다양한 교집합의 확률값도 쉽게 찾을 수 있다(예를 들어 $P(AB)=0.07$, $P(A\overline{B})=0.13$). 다음을 구하라.

(1) $P(A)$

(2) $P(B\overline{C})$

(3) $P(A\cup B)$



【풀이】

(1) $P(A)=0.13+0.07=0.2$

(2) $P(B\overline{C})=0.1+0.07=0.17$

(3) $P(A\cup B)=0.13+0.07+0.1+0.15=0.45$

3.4 조건부 확률과 독립

(1) $P(A|B)$

: 사상 A의 확률이 사상 B가 일어난 것과 관계가 있으면 그 값이 달라질 수 있다. 사상 B가 일어난 것을 알 때 사상 A가 일어날 확률을 B가 주어졌을 때 **A의 조건부 확률 (conditional probability)**라고 한다.

[예제 3.10] 어떤 사무실 직원들 몸무게와 고혈압의 상태에 따라 분류하였다. 그 비율은 다양한 범주로 나타난다.

	비만	정상	체중미달	계
고혈압	0.10	0.08	0.02	0.20
비고혈압	0.15	0.45	0.20	0.80
계	0.25	0.53	0.22	1.00

① 임의로 한 사람을 뽑았을 때 고혈압일 확률은 얼마인가?

【풀이】

A: 고혈압일 사상

B: 비만일 사상

$P(A)=0.2 \rightarrow A$의 비조건부 확률

② 임의로 뽑은 한 사람이 비만이었다. 그 사람이 또한 고혈압일 확률은 얼마인가?

【풀이】

$$P(A|B)=\frac{0.10}{0.25}=0.4 \rightarrow P(A|B)=\frac{P(AB)}{P(B)}$$

(2) 확률의 승법 법칙 (multiplication law of probability)

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

$$P(AB)=P(B)P(A|B)$$

[예제 3.11] 주요 고객의 명단에 25명의 이름이 있다. 그들 중 20명은 우수한 상태의 거래를 지속하지만 5명은 체납한다. 이 명단에서 2명을 뽑고 그들의 계좌상황을 조사해 보기로 하자. 다음의 확률을 구하라.

① 두 명의 계좌 모두 체납했을 확률

【풀이】

D: 체납한 사상

D_1: 첫 번째 계좌가 체납일 사상, D_2: 두 번째 계좌가 체납일 사상

G: 우수한 상태의 거래인 사상

G_1: 첫 번째 계좌가 우수거래계좌일 사상

G_2: 두 번째 계좌가 우수거래계좌일 사상

$$P(\text{둘 다 체납계좌}) = P(D_1 D_2) = P(D_1)P(D_2|D_1) = \frac{5}{25} \times \frac{4}{24} = \frac{1}{30}$$

② 한 명의 계좌는 체납했고 다른 한 명의 계좌는 우수한 상태일 확률

【풀이】

$$P(G_1 D_2) + P(D_1 G_2) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

$$P(G_1 D_2) = P(G_1)P(D_2|G_1) = \frac{20}{25} \times \frac{5}{24} = \frac{1}{6}$$

$$P(D_1 G_2) = P(D_1)P(G_2|D_1) = \frac{5}{25} \times \frac{20}{24} = \frac{1}{6}$$

(3) 사상들 간의 독립 (Independence between events)

$P(A|B) = P(A)$, $P(B|A) = P(B) \Leftrightarrow P(AB) = P(A)P(B)$

$\Rightarrow$ 두 사상 A와 B는 독립이다.

[예제 3.12] [예제 3.10]의 모집단에서 두 사상 $A = \{\text{고혈압}\}$과 $B = \{\text{비만}\}$은 독립인가?

【풀이】

$$P(A) = 0.2,\ P(A|B) = \frac{0.10}{0.25} = 0.4 \rightarrow P(A) \neq P(A|B) \Rightarrow A\text{와 } B\text{는 독립이 아니다.}$$

3.5 유한모집단으로부터의 확률표본

Q: 모집단과 표본의 크기가 모두 큰 경우, 어떻게 모든 가능한 근원사상을 열거할 수 있을까?
A: ‘열거법(counting rule)’은 이런 문제를 해결하는데 도움이 된다.

(1) 조합법

① N개의 다른 객체를 가지고 있는 한 집단으로부터 r개의 객체를 선택할 때, 가능한 선택의 수를 $\binom{N}{r}$라 표기한다.

② 공식: $\binom{N}{r} = \dfrac{N\times(N-1)\times\cdots\times(N-r+1)}{r\times(r-1)\times\cdots\times 2\times 1} \leftarrow {}_nC_r = \dfrac{{}_nP_r}{r!}$

③ 대칭의 성질: $\binom{N}{r} = \binom{N}{N-r}$

[예제 3.16] $\binom{5}{2}$, $\binom{15}{4}$와 $\binom{15}{11}$의 값을 계산하라.

【풀이】

$$\binom{5}{2} = \frac{5\times 4}{2\times 1} = 10$$

$$\binom{15}{4} = \frac{15\times 14\times 13\times 12}{4\times 3\times 2\times 1} = 1365$$

$$\binom{15}{11} = \binom{15}{4} = 1365$$

[예제 3.17] 대학본부는 대학자치위원회에서 일할 4명의 학생대표를 뽑으려고 한다. 후보자는 12명이 추천되었다. 이들 중 5명은 공학부 소속이고 나머지 7명은 자연과학부 소속이다.

① 12명의 후보자 집단으로부터 4명의 학생이 선택되는 경우는 모두 몇 가지인가?
【풀이】
$\binom{12}{4} = \frac{12 \times 11 \times 10 \times 9}{4 \times 3 \times 2 \times 1} = 495$

② 1명의 자연과학부 학생과 3명의 공학부 학생이 선택되는 경우는 모두 몇 가지인가?
【풀이】
7명의 자연과학부 학생 중 한 명이 선택될 경우의 수: $\binom{7}{1} = 7$
5명의 공학부 학생 중 3명이 뽑힐 경우의 수: $\binom{5}{3} = \frac{5 \times 4 \times 3}{3 \times 2 \times 1} = 10$
가능한 모든 경우의 수: $\binom{7}{1} \times \binom{5}{3} = 7 \times 10 = 70$

③ 전 과정을 임의선택으로 한다면, 1명의 자연과학부 학생과 3명의 공학부 학생이 선택될 확률은 얼마인가?
【풀이】
임의추출 → 495가지 모든 가능한 표본이 똑같은 확률을 가진다.
1명의 자연과학부 학생과 3명의 공학부 학생이 선택되는 사건 A의 경우의 수: 70가지

$$P(A) = \frac{70}{495} = 0.141$$

(2) 확률표본

: 크기가 n인 모든 묶음들이 똑같은 확률 $1/\binom{N}{r}$을 갖는다면, N개의 서로 다른 객체를 가지고 있는 한 모집단으로부터 뽑혀진 크기 n의 표본을 '확률표본'이라 한다.

3.6 확률변수 (random variable)

(1) 정의: 확률변수 X란 실험의 결과들에 수치를 대응시키는 것이다.

$[X=a]=\{\omega\in\Omega|X(\omega)=a\}$ for any real number a

[예제 3.18] 동전을 세 번 던지는 실험에서 앞면이 나오는 횟수를 X라고 하자. X가 가지는 값들과 이에 대응하는 결과들을 나열하라.

X: 근원사상	X값
HHH	3
HHT	2
HTH	2
HTT	1
THH	2
THT	1
TTH	1
TTT	0

X가 가질 수 있는 각각의 값에 따라서 대응되는 사상

X값	각 X값에 대응하는 사상
$[X=0]$ =	$\{TTT\}$
$[X=1]$ =	$\{HTT, THT, TTH\}$
$[X=2]$ =	$\{HHT, HTH, THH\}$
$[X=3]$ =	$\{HHH\}$

(2) 성질

① 서로 다른 X값들에 대응하는 사상들은 배반적이다.

② 이런 사상들의 합사상은 전체 표본공간이 된다.

[예제 3.19] 12명의 품질 검사단이 유명 제품의 옥수수 튀김과 새로 개발된 제품을 비교하고자 한다. 이때 X를 새로운 제품이 유명제품만큼의 맛을 갖는다고 느끼는 사람의 수라고 하자. 그러면 X가 취할 수 있는 값은 $0, 1, 2, \cdots, 12$이다.

[예제 3.20] 교차로에서 특정한 차종을 발견할 때까지 지나가는 자동차의 대수를 X라고 할 때, X가 취할 수 있는 값들은 $0, 1, 2, 3, \cdots$과 같이 무한히 많은 값을 가질 수 있다.

(3) 유형

① 이산확률변수(discrete random variables)

: 확률변수가 유한값을 혹은 자연수와 일대일대응이 되는 무한히 많은 값을 가질 때

② 연속확률변수(continuous random variables)

: 확률변수가 어떤 연속적인 양의 측도로 표현이 되고 모든 값들이 어떤 구간에 있을 때

3.7 이산확률변수의 확률분포

[예제 3.21] 공정한 동전을 3번 던지는 실험에서 앞면이 나오는 횟수를 X라 할 때, X의 확률분포를 구하라.

【풀이】

X: 근원사상	X값
HHH	3
HHT	2
HTH	2
HTT	1
THH	2
THT	1
TTH	1
TTT	0

X값	확률
0	1/8
1	3/8
2	3/8
3	1/8
합계	1

< 이산확률분포 >

X값	확률
x_1	$f(x_1)$
x_2	$f(x_2)$
$\vdots$	$\vdots$
x_k	$f(x_k)$
합계	1

(1) 표기: $f(x_i) = P[X = x_i]$

(2) 성질

① $0 \leq f(x_i) \leq 1 \quad \forall x_i$

② $\sum_{i=1}^{k} f(x_i) = 1$

3.8 확률분포의 기댓값과 표준편차

(1) 평균 (mean): $E(X) = \mu = \sum(\text{값} \times \text{확률}) = \sum x_i f(x_i)$

[예제 3.24] 공정한 동전을 세 번 던져서 앞면이 나오는 횟수를 X라고 할 때, X의 평균을 구하라.

【풀이】

x	$f(x)$	$xf(x)$
0	1/8	0
1	3/8	3/8
2	3/8	6/8
3	1/8	3/8
합계	1	$E(X) = \mu = \sum x_i f(x_i) = 1.5$

(2) 분산 (variance): $\sigma^2 = Var(X) = \sum(x_i - \mu)^2 f(x_i) = \sum x_i^2 f(x_i) - \mu^2$

(3) 표준편차 (standard deviation): $\sigma = sd(X) = \sqrt{Var(X)}$

[예제 3.26] [예제 3.25]의 확률분포를 하에서 σ^2의 간편식을 이용하여 분산과 표준편차를 구하라.

【풀이】

x	$f(x)$	$xf(x)$	$x^2f(x)$
0	0.1	0.0	0.0
1	0.2	0.2	0.2
2	0.4	0.8	1.6
3	0.2	0.6	1.8
4	0.1	0.4	1.6
합계	1	2.0	5.2

$\sigma^2 = Var(X) = \sum(x_i - \mu)^2 f(x_i) = \sum x_i^2 f(x_i) - \mu^2 = 5.2 - (2.0)^2 = 1.2$

$\sigma = sd(X) = \sqrt{Var(X)} = \sqrt{1.2} = 1.095$