4. 확 률

정보통계학과 : 김 덕 기

http://cafe.daum.net/cb-stat

toby123@cbnu.ac.kr

확률의 기본개념-1

➔ 확률(Probability) : 確(확실할 확), 率(비율 률)

Ex1) Lotto 당첨확률, 한국 야구시리즈에서 특정 팀이 우승할 확률, 폐암에 걸릴 확률

Ex2) 내일 비가 올 확률 60% ➔ 이 정보를 어떻게 유용하게 사용할 것인가 ?

➔ 확률은 불확실성을 이해하는데 중요한 개념 :

표 4.1 의사결정과 예상손실

의사결정	예상손실
우산을 가져간다	m
우산을 가져가지 않는다	$0.6\times r+0.4\times 0=0.6r$

➔ 불확실성을 계량화하는 수단.

➔ 우산을 가져가서 겪게 되는 불편~m원의 비용, 안 가져가서 젖게 되는 손실~r비용

$m < 0.6r$?　　$m > 0.6r$?　　$m=0.6r$?

확률의 기본개념-2

(1) 확률실험(random experiment) :

➔ 모든 가능한 결과는 알고 있지만 어떤 결과가 나올지 예측할 수 없고 동일한 조건에서 반복 시행할 수 있는 실험.

(2) 표본공간(sample space) : 이산표본공간, 연속표본공간

➔ 확률실험에서 나올 수 있는 가능한 모든 결과들의 집합.

[EX1] 공정한 동전을 연속해서 두 번 던지는 실험(확률실험)에서 앞면이 나오면 H, 뒷면이 나오면 T라 할 경우 표본공간 : S={(H,H), (H,T), (T,H), (T,T)} ➔ P.44 예제4.1

[EX2] 전구의 수명을 검사하는 실험에서 전구의 수명을 t라 했을 때 표본공간 : P.45 예제4.2

$$S\text{는}\quad S = \{t, t \geq 0\}$$

(3) 사건(event) : 단순사건, 복수사건

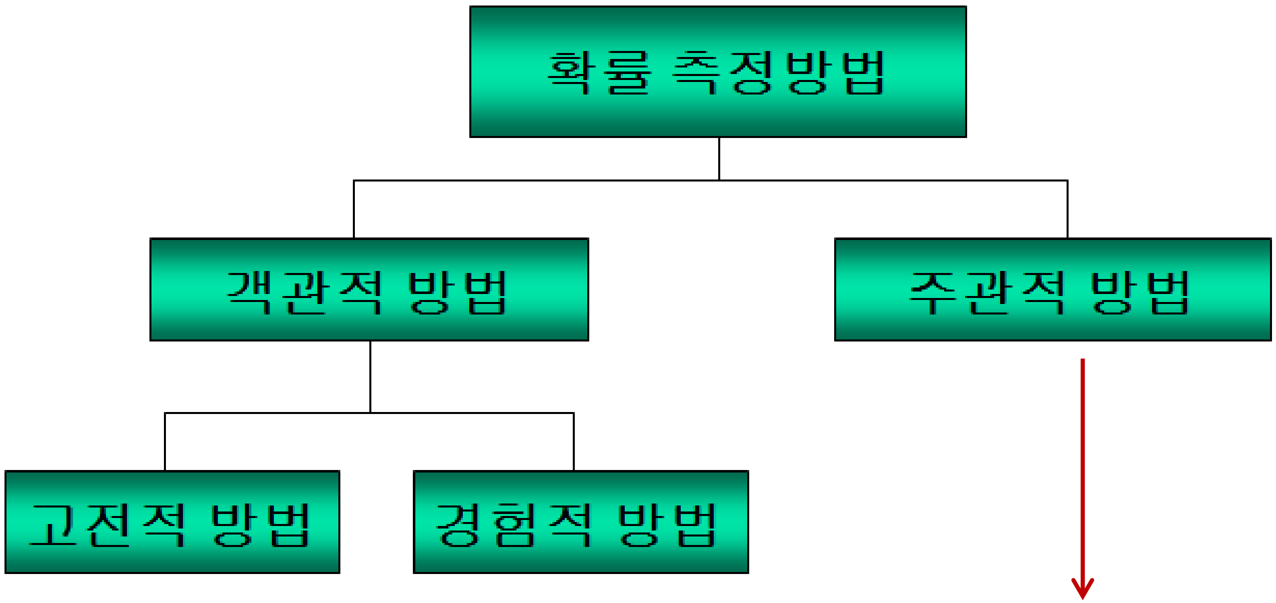
➔ 표본공간에서 특정한 결과에 관한 집합, 표본공간의 부분집합.

[EX3] 공정한 동전을 두 번 던질 때 동일한 면이 나오는 경우를 사건 A라 하면

A={(H,H), (T,T,)}

확률의 기본개념-3

- 확률의 측정방법



주관적 확률

사건 A가 발생할 확률을 개인적인 경험이나 지식에 근거한 추측 또는 간접적인 증거에 의하여 의사결정자가 주관적으로 정의하는 것을 주관적 확률이라고 한다.

확률의 기본개념-4

고전적 확률의 정의

사건 A의 확률의 고전적 정의는 표본공간 안의 n개의 원소들이 일어날 가능성은 모두 같고, 그 가운데 반드시 한 개는 일어날 때, 사건 A에 속한 모든 원소들의 개수를 표본공간 안의 원소의 개수 n으로 나눈 값이 된다.

$$P(A) = \frac{\text{사건 } A\text{에 속하는 모든 원소의 개수}}{\text{표본공간 안의 모든 원소의 개수}}$$

상대도수의 극한적 확률의 정의

사건 A의 확률의 상대도수의 극한적 정의는 어떤 통계적 실험을 무한히 반복하였을 때, 사건 A가 일어난 상대도수를 의미하고 다음과 정의한다.

$$P(A) = \lim_{n \to \infty} \frac{\#(A)}{n}$$

여기서 n은 시행횟수, $\#(A)$는 n번의 시행 중 사건 A가 일어나는 횟수를 나타낸다.

확률의 법칙-1

확률의 공리

(1) 표본공간 S 안의 임의의 사건 A는 $0 \leq P(A) \leq 1$를 만족해야 한다.

(2) 표본공간 S의 확률은 $P(S) = 1$이다.

(3) 서로 배반인 사건 $A_1, A_2, \cdots$에 대하여 다음 식이 성립한다.

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

여사건 A^c의 확률

여사건 A^c는 표본공간 S 안에 있으며, A에 속하지 않는 원소들의 집합을 의미하고, 다음 식이 성립한다.

$$P(A^c) = 1 - P(A)$$

P.47~ex.4.4

확률의 법칙-2

두 사건 A와 B의 덧셈정리

두 사건 A와 B의 합사건은 $A \cup B$로 적고, 표본공간 S 안에서 사건 A에 속하거나 사건 B에 속하거나 사건 A와 B 모두에 속하는 원소들로 구성된 사건을 의미하고, 합사건의 확률에 대해서는 다음 식이 성립한다.

P.47~ex.4.5

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

조건부확률

두 사건 A와 B가 있다면 사건 B가 주어질 때 사건 A가 일어날 조건부확률은 $P(A|B)$ 라고 적고, 다음 식으로 정의한다.

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

P.48~ex.4.6

단, 여기서 $P(B) > 0$이어야 한다.

확률의 법칙-3

확률의 곱셈정리

표본공간 S 안의 두 사건 A와 B가 일어난다면 교사건의 확률 $P(A \cap B)$에 대하여 다음이 성립한다.

$$P(A \cap B) = \begin{cases} P(A)P(B \mid A): \ P(A) > 0 \text{일 때} \\ P(B)P(A \mid B): \ P(B) > 0 \text{일 때} \end{cases}$$

P.49~ex.4.7

곱셈법칙의 일반화 :

$$P(E_1E_2 \cdots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2) \cdots P(E_n|E_1 \cdots E_{n-1})$$

$$=> \ P(E_1)\frac{P(E_1E_2)}{P(E_1)}\frac{P(E_1E_2E_3)}{P(E_1E_2)} \cdots \frac{P(E_1E_2 \cdots E_n)}{P(E_1E_2 \cdots E_{n-1})} = P(E_1E_2 \cdots E_n)$$

독립사건

두 사건의 독립
두 사건 $A, B(\subset S)$에 대해 $$P(A\cap B)=P(A)P(B)$$ 가 성립하면, 두 사건 A, B는 서로 독립(mutually independent)이라 한다.

P.53~ex.4.10

$$P(B|A)=P(B) \text{ 또는 } P(A|B)=P(A)$$

[예제 2−10] 3개의 공정한 동전을 던질 때, 앞면의 수가 1 또는 2일 사건을 A라 하고, 뒷면이 1개일 사건을 B라고 하자. 두 사건 A와 B가 서로 독립인지 조사하여라.

[풀이] 표본공간은 $S=\{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$이며, 사건 A와 B는 각각

$$A=\{HHT, HTH, THH, HTT, THT, TTH\}$$

$$B=\{HHT, HTH, THH\}$$

이다. 따라서 $A\cap B=B$이다. 그런데, $P(A)=\frac{6}{8}$, $P(B)=\frac{3}{8}$이므로, $P(A\cap B)\neq P(A)P(B)$. 즉, 두 사건 A와 B는 서로 종속이다.

분할과 전확률 공식

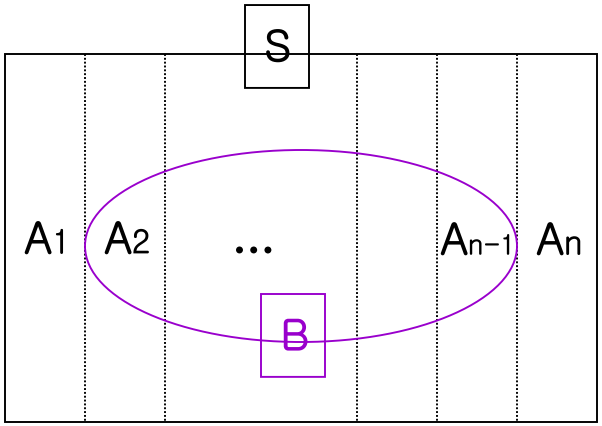
전확률 공식(Total Probability Formula)
집합 $A_1, A_2, \cdots, A_n$이 표본공간 S의 분할이고, 모든 i에 대해 $P(A_i) > 0$이면, 임의의 사건 B에 대해 $$P(B) = P(A_1)P(B \mid A_1) + \cdots + P(A_n)P(B \mid A_n)$$ 이 성립한다.

Proof

$$P(B) = P(S \cap B) = P\left[\left(\cup_{i=1}^{n} A_i\right) \cap B\right]$$

$$= P\left[\cup_{i=1}^{n} (A_i \cap B)\right] = \sum_{i=1}^{n} P(A_i \cap B)$$

$$= \sum_{i=1}^{n} P(A_i) P(B \mid A_i)$$



전확률 공식(예제)

[예제] 어느 중고차 시장은 미국산 자동차가 50%를 점유하며, 이 중 15%는 소형이라고 하고, 유럽산 자동차가 전체 시장의 30%를 점유하며, 이 중 40%는 소형이며, 그리고 일본산 자동차는 전체 시장의 20%를 점유하며, 이 중 60%는 소형으로 구성되어있다고 한다. 전체시장에서 임의로 자동차 한 대를 선택했을 때, 소형일 확률을 구하여라.

[풀이] 선택된 자동차가 미국산일 사건을 A_1, 유럽산일 사건을 A_2, 그리고 일본산일 사건을 A_3라 하고, 선택된 자동차가 소형차일 사건을 B라고 하자. 사건 A_1, A_2, A_3는 서로 배반적인 관계에 있으므로, 전확률 공식에 의하여 구하는 확률은

$$\begin{aligned} P(B) &= P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + P(A_3)P(B \mid A_3) \\ &= 0.50 \times 0.15 + 0.30 \times 0.40 + 0.20 \times 0.60 \\ &= 0.315 \end{aligned}$$

베이즈 정리(Bayes Theorem)

조건부확률의 정의와 곱셈법칙, 전확률 공식으로부터, 임의의 사건 A_k 에 대한 조건부 확률 $P(A_k|B)$이 아래와 같이 성립함을 알 수 있다. (교재 P.51)

베이즈 정리(Bayes Theorem)

사건 $A_1, A_2, \cdots, A_n$이 표본공간 S의 분할이고, 모든 i에 대해 $P(A_i) > 0$이고 $P(B) > 0$이면, 조건부 확률 $P(A_k \mid B)$은 다음과 같이 얻어진다.

$$P(A_k \mid B) = \frac{P(A_k)P(B \mid A_k)}{\sum_{i=1}^{n} P(A_i)P(B \mid A_i)}$$

➔ 곱셈정리: $P(A_k \cap B)$

➔ 전 확률 : $P(B)$

※ 베이즈 정리의 원리

사전적 확률
$P(A_k)$

추가 정보
$P(B|A_k)$

사후적 확률
$P(A_k|B)$

관측 전의 사건에 대한 가능성(사전적 확률)과 관측 후의 사건에 대한 가능성(사후적 확률) 사이의 관계를 규명

베이즈 정리(예제)

[예제] 앞의 전확률 공식 예제에서 임의로 선택한 자동차가 소형이라는 조건 하에서 그 차가 미국산, 유럽산, 일본산일 확률을 계산하여라.

[풀이] 먼저 그 차가 미국산일 확률은

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{P(B)} = \frac{0.50 \times 0.15}{0.315} = \frac{5}{21}(\approx 24\%)$$

P.52~ex.4.9

이다. 비슷한 방법으로 $P(A_2 \mid B) = \frac{8}{21}(\approx 38\%)$, $P(A_3 \mid B) = \frac{8}{21}(\approx 38\%)$ 임을 알 수 있는데, 이 확률을 50%, 30% 그리고 20%와 비교해 보면 약간 흥미로운 결과로, 이는 임의로 선택된 자동차가 소형이었다는 사실이 조건부확률에 영향을 미친 결과로 생각할 수 있다.

제 5 장 확률분포와 기대값

김 덕 기

toby123@cbnu.ac.kr

확률변수와 확률분포

(1) 확률변수(random variable) :

표본공간의 각 원소에 하나의 실수 값을 대응시키는 함수.

(예) 컴퓨터 공장에서 생산한 두 대의 제품에 대한 불량품 여부를 판정하는 실험.

D(불량품), N(양품) ➔ 표본공간 S={DD, DN, ND, NN} ➔ 확률변수 X:{x=0,x=1,x=2}.

X(DD)=2, X(DN)=X(ND)=1, X(NN)=0 ➔ 확률 : P(X=0)=1/4, P(X=1)=1/2, P(X=2)=1/4

(2) 이산형 확률변수(discrete random variable) :

확률변수가 취할 수 있는 실수 값의 수를 셀 수 있는 변수.

(예) 10번 주사위를 던지는 실험에서 1이 나오는 수 : X={0, 1, 2, ..., 10}

(3) 연속형 확률변수(continuous random variable) :

확률변수가 취할 수 있는 실수 값이 어떤 특정구간 전체에 해당하며 그 수를 셀 수 없는 변수.

(예) 우리나라에서 수입하는 금의 양 : X={x|x >0, x는 실수}

(4) 확률분포(probability distribution) :

확률변수가 취할 수 있는 값에 대한 확률적 구조를 합이 1인 양수로 표현한 것.

확률변수와 확률분포(예)

➔ 확률실험(random experiment) :

- 모든 가능한 결과는 알고 있지만 어떤 결과가 나올지 예측할 수 없고 동일한 조건에서 반복 시행할 수 있는 실험.
- 한 아파트 단지 1000세대의 각 가정에 있는 TV의 수를 조사하는 경우 :

➔ X를 선택된 가정의 TV 수라 하면 X=0, 1, 2, 3이 되고, 그 값은 어느 가정이 선택 되는가에 따라 달라지며, 확률실험의 결과에 따라 달라져 확률변수라 함.

각 가정의 TV 수와 상대도수분포

보유 TV의 수(X)	도수	상대도수
0	10	0.010
1	840	0.840
2	145	0.145
3	5	0.005
합계	N = 1000	1.000

각 가정의 TV 수와 확률분포

보유 TV의 수(X)	$P(X=x)$
0	0.010
1	0.840
2	0.145
3	0.005
합계	1.000

이산형 확률변수(확률질량함수:PMF)

이산형 확률변수의 확률 구조 :

$$f(x)=\begin{cases} P(X=x_i) \text{ 또는 } f(x_i),\ x=x_i\text{일 때 } (i=1,2,\cdots) \\ 0, \qquad\qquad\qquad\qquad \text{그 외} \end{cases}$$

이산확률변수의 확률질량함수(Probability Mass Function : pmf)

이산확률변수 X가 취할 수 있는 값 $x_1,x_2,\cdots$에 대하여 $f(x_i)=\Pr\{X=x_i\}$ $(i=1,2,\cdots)$ 라고 하면,

(1) $f(x_i)\ge 0,\ i=1,2,\cdots$

(2) $\sum_{i=1}^{\infty} f(x_i)=1$

(3) $P(\alpha\le X\le\beta)=\sum_{i=\alpha}^{\beta} f(x_i)$ 단, α와 β는 양의 정수

을 만족하는 함수 $f(x)$를 확률변수 X의 확률함수

앞의 예) TV 수가 2대 이상일 확률 : P(X>=2)=P(2<= X <= 3) ➔ alpha=2, beta=3

이산확률분포

[확률실험] 공정한 동전을 연속해서 두 번 던지는 실험.

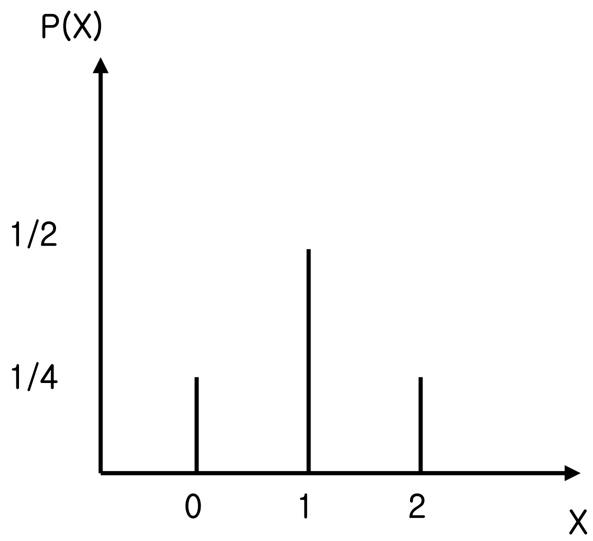
<표본공간> S = {(H,H), (H,T), (T,H), (T,T)}, 확률변수 X : 동전의 앞면이 나온 수.

X의 확률분포

x	0	1	2	합계
$P\{X=x\}$	1/4	1/2	1/4	1

확률밀도함수

$$P\{X=x\} = \begin{cases} \frac{1}{4}, x = 0, 2 \\ \frac{1}{2}, \quad x = 1 \\ 0, \quad \text{그 외} \end{cases}$$



X의 분포그래프

이산확률분포-확률계산 예제

▶ 예제

\- 주당 고장횟수에 대한 확률분포를 그래프로 나타내고 각 확률을 구하라.

(1) 정확히 2회 고장 날 확률 ?

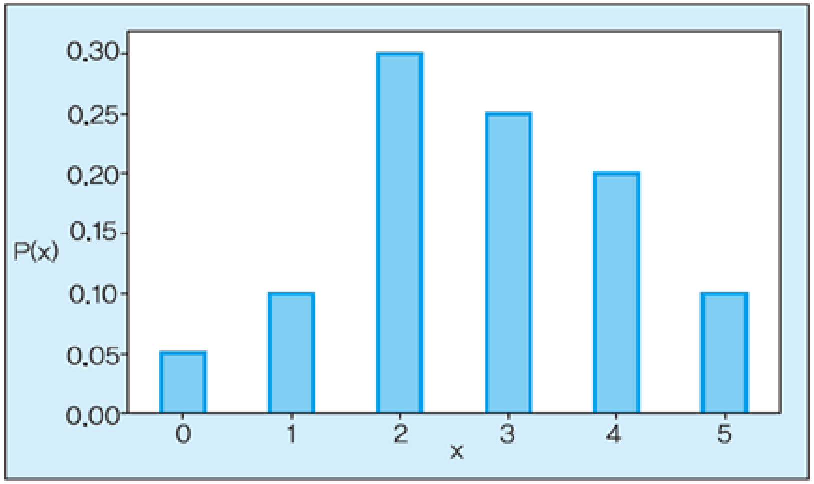
(2) 0회에서 2회까지 고장 날 확률 ?

(3) 2회 이상 고장 날 확률 ?

(4) 적어도 1회 고장 날 확률 ?

표 고장횟수의 확률분포

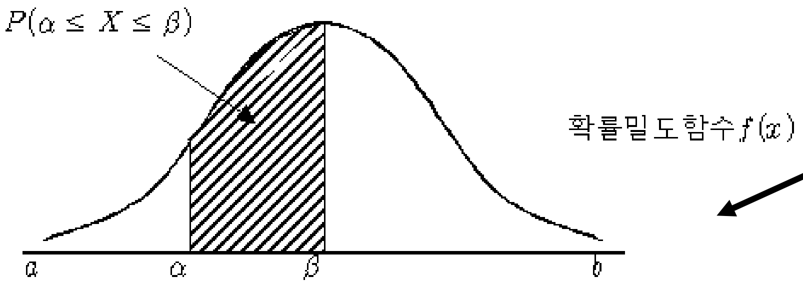
X	$P(X=x)$
0	0.05
1	0.10
2	0.30
3	0.25
4	0.20
5	0.10
합 계	1.00



그래프로 표현된 [표]의 확률분포

연속형 확률변수(확률밀도함수)

연속확률변수의 확률밀도함수(Probability Density Function : pdf)
확률변수 X가 구간 $[a,b]$의 모든 값을 취할 수 있으며, 이 구간에서 정의된 $f(x)$가 (1) $f(x) \geq 0$, x : 실수 (2) $\int_a^b f(x)dx = 1$ (3) $P(\alpha \leq X \leq \beta) = \int_\alpha^\beta f(x)dx$, 단, $a \leq \alpha < \beta \leq b$ 을 만족할 때, 함수 $f(x)$를 연속확률변수 X의 확률밀도함수(probability density function)라고 한다.

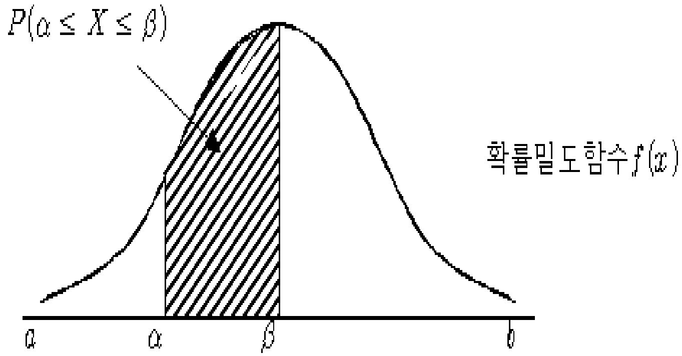


이산확률변수와 달리 확률밀도함수의 값은 확률이 아님.

연속확률분포

[확률분포] 전구의 수명(X)을 검사하는 실험.

확률변수 X : 전구의 수명, X={x|x>0, x∈R}, X의 확률분포



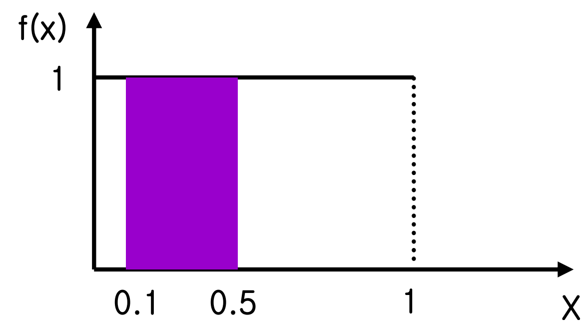
연속적인 구간(a<X<b)에서 정의된 확률밀도함수의 형태

확률분포(연속적인 구간내의 확률의 합=1)

(예) 확률변수 X의 pdf가 다음과 같다.

f(x)=1, 0<x<1, f(x)=0, 그 외에

f(x)의 그래프



$$P(0.1 \leqq X \leqq 0.5) = \int_{0.1}^{0.5} 1\,dx = 0.5 - 0.1 = 0.4$$

확률변수의 기대 값(이산)

확률변수의 기대값(Expected Value of r.v)
확률변수 X의 확률분포함수를 $f(x)$라 할 때, $$E(X)=\begin{cases}\sum_i x_i f(x_i) & X:\ \text{이산확률변수}\\ \int_{-\infty}^{\infty} x f(x)dx & X:\ \text{연속확률변수}\end{cases}$$ 라 한다면, $E(X)$를 확률변수 X의 기대값(expected value)이라고 한다.

[예제] 공정한 주사위를 한 번 던지는 시행에서 윗면에 나타난 눈의 개수를 나타내는 확률변수를 X라고 할 때, X의 확률밀도함수를 구하고 기대 값을 계산하라.

$$X\sim f(x)=P(X=x)=\begin{cases}\frac{1}{6} & ,x=1,2,3,4,5,6\\ 0 & ,\text{그 외}\end{cases}$$

$$E(X)=\sum_x xf(x)=\frac{1}{6}\sum_{x=1}^{6}x=3.5$$

확률변수의 기대 값 : 예제

다음 표는 어떤 기계의 주당 고장횟수에 대한 확률분포를 나타낸다. 이 업체의 주당 고장횟수의 기대값을 구하여라.

주당 고장횟수(X)	0	1	2	3	4	5
P(X=x)	0.05	0.1	0.3	0.25	0.2	0.1

고장횟수의 확률분포에 대한 평균값의 계산

X	$p(x)$	$xp(x)$
0	0.05	$0(0.05) = 0.00$
1	0.10	$1(0.10) = 0.10$
2	0.30	$2(0.30) = 0.60$
3	0.25	$3(0.25) = 0.75$
4	0.20	$4(0.20) = 0.80$
5	0.10	$5(0.10) = 0.50$
		$\sum xp(x) = 2.75$

$$E(X) = \mu_X = \sum_x x \cdot p(x)$$

확률변수의 기대 값(연속)

[예제] 확률변수 X의 확률밀도함수가 다음과 같다.

확률밀도함수 :

$$X \sim f(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{그 외} \end{cases}$$

새로운 확률변수 $Y = \left(X - \frac{1}{2}\right)^2$의 기대값을 계산하여라.

$$E[g(x)] = \begin{cases} \sum_{i=1}^{n} g(x_i)p(x_i), & \text{이산형} \\ \int_a^b g(x)f(x)dx, & \text{연속형} \end{cases}$$

$$E(Y) = E\left[\left(X - \frac{1}{2}\right)^2\right]$$
$$= \int_{-\infty}^{\infty} \left(x - \frac{1}{2}\right)^2 f(x)\,dx$$
$$= \int_0^1 \left(x^2 - x + \frac{1}{4}\right) 2x\,dx$$
$$= \frac{1}{12}$$

확률변수의 분산

확률변수의 분산(Variance)

확률변수 X의 확률분포함수 $f(x)$가 주어져 있을 때, 특히 확률변수 $g(X)=(X-\mu)^2$에 대한 기대값을 확률변수 X의 분산(variance)이라고 한다. 즉,

$$Var(X)=E[(X-\mu)^2], \qquad \text{단}, \mu=E(X)$$

$$=\begin{cases} \sum_x (x-\mu)^2 f(x) & , X\text{: 이산확률변수} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx & , X\text{: 연속확률변수} \end{cases}$$

간편식

$$Var(X)=E[(X-\mu)^2]=E[X^2-2\mu X+\mu^2]$$
$$=E(X^2)-2\mu E(X)+E(\mu^2)$$
$$=E(X^2)-\mu^2$$

$$Var(aX+b)=E[(aX+b-a\mu-b)^2]$$
$$=E[a^2(X-\mu)^2]$$
$$=a^2 Var(X)$$

확률변수의 기대 값, 분산 : 예제

[예 제] 자동차 대리점에서 일주일간 판매되는 자동차 대수를 X라고 할 때, X의 평균과 분산을 계산하여라.

x	0	1	2	3	계
$f(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

풀이 X의 평균은

$$E(X) = \mu = (0)\left(\frac{1}{8}\right) + (1)\left(\frac{3}{8}\right) + (2)\left(\frac{3}{8}\right) + (3)\left(\frac{1}{8}\right) = 1.5$$

이고, 또

$$\begin{aligned} E(X^2) &= \sum x_i^2 f(x_i) \\ &= 0^2\left(\frac{1}{8}\right) + 1^2\left(\frac{3}{8}\right) + 2^2\left(\frac{3}{8}\right) + 3^2\left(\frac{1}{8}\right) \\ &= 3 \end{aligned}$$

따라서 X의 분산은

$$\begin{aligned} \sigma^2 &= E(X^2) - \mu^2 \\ &= 3 - (1.5)^2 = 0.75 \end{aligned}$$

이다.

두 확률변수의 결합확률분포

(1) 이산형 확률변수의 결합 확률분포

실험　　A와 B 두 종류의 곤충에 대한 공존상태를 연구하는 실험

확률변수　X, Y : 각각, 한 나무에 서식하는 곤충 A와 B의 수

① X와 Y의 결합확률분포

Y \ X	1	2	3	4	계
0	0	0.05	0.05	0.10	0.20
1	0.08	0.15	0.10	0.10	0.43
2	0.20	0.12	0.05	0	0.37
계	0.28	0.32	0.20	0.20	1

두 확률변수의 주변확률분포

② 주변 확률분포

x	1	2	3	4	계
$P(X=x)$	0.28	0.32	0.20	0.20	1

[확률변수 X의 주변 확률분포]

y	0	1	2	계
$P(Y=y)$	0.20	0.43	0.37	1

[확률변수 Y의 주변 확률분포]

Question 한 나무에 서식하는 곤충 B의 수가 곤충 A의 수보다 많을 확률은?

이러한 확률은 확률변수 X, Y의 각각의 확률분포로부터 구할 수 없다. 따라서 결합 확률분포의 정보가 각각의 확률분포의 정보보다 더 많은 정보를 제공해준다는 사실을 알 수 있다.

결합확률정보=
각각의 확률변수의 주변확률 정보+두 변수의 Correlation(상관정도) 정보

두 확률변수의 기대 값과 분산계산

확률변수 X의 주변확률분포

x	1	2	3	4	계
$P(X=x)$	0.28	0.32	0.20	0.20	1

$$E(X)=\sum_{all\,x} xf(x)=1\times 0.28+2\times 0.32+3\times 0.20+4\times 0.20=2.32$$

$$E(X^2)=\sum_{all\,x} x^2 f(X)=1^2\times 0.28+2^2\times 0.32+3^2\times 0.20+4\times 0.20=6.56$$

$$Var(X)=E(X^2)-[E(X)^2]=6.56-(2.32)^2=1.1776$$

확률변수 Y의 주변확률분포

y	0	1	2	계
$P(Y=y)$	0.20	0.43	0.37	1

$$E(Y)=\sum_{all\,y} yf(y)=0\times 0.20+1\times 0.43+2\times 0.37=1.17$$

$$E(Y^2)=\sum_{all\,y} x^2 f(y)=0^2\times 0.20+1^2\times 0.43+2\times 0.37=1.91$$

$$Var(Y)=E(Y^2)-[E(Y)]^2=1.91-(1.17)^2=0.5411$$

두 확률변수의 상관계수추정

결합 확률분포로부터 :

$$E(XY) = \sum_x \sum_y xyf(xy)$$

$$= 1 \times 0 \times 0 + 1 \times 1 \times 0.08 + 1 \times 2 \times 0.20 + \cdots + 4 \times 1 \times 0.10 + 4 \times 2 \times 0$$

$$= 2.26$$

$$\therefore \quad Cov(X, Y) = E(XY) - E(X)E(Y)$$

$$= 2.26 - 2.32 \times 1.17 = -0.4544$$

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

$$= \frac{-0.4544}{\sqrt{1.1776}\sqrt{0.5411}} = -0.5692$$

$$\boxed{-1 \le corr(x, y) \le 1}$$

두 확률변수의 독립

두 확률변수 X와 Y의 결합분포에서 모든 실수 x와 y에 대하여

이산형 => $P(X=x, Y=y) = P(X=x)P(Y=y)$

연속형 => $f(x,y) = f(x)f(y)$

$f_{X,Y}(x,y) = f_X(x)f_Y(y)$ ⇒ 두 확률변수 X, Y는 서로 독립(mutually independent)

Tips

X, Y : 확률변수, a, b : 상수

① $E(XY) = E(X)E(Y)$

② $Cov(X,Y) = 0,\ Corr(X,Y) = 0$

③ $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$

교재 P.73~예5.8 참조

두 확률변수의 독립(예제)

(교재 P.68) [표 5-1]의 예제

Y \ X	1	2	3	4	계
0	0	0.05	0.05	0.10	0.20
1	0.08	0.15	0.10	0.10	0.43
2	0.20	0.12	0.05	0	0.37
계	0.28	0.32	0.20	0.20	1

$f_x(1)f_y(0) = 0.28 \times 0.20 \neq 0 = f_{xy}(1,0)$ $\quad$ $f_x(1)f_y(1) = 0.28 \times 0.43 \neq 0.08 = f_{xy}(1,1)$

$f_x(1)f_y(2) = 0.28 \times 0.37 \neq 0.20 = f_{xy}(1,2)$ $\quad$ $f_x(2)f_y(0) = 0.32 \times 0.20 \neq 0.05 = f_{xy}(2,0)$

$f_x(2)f_y(1) = 0.32 \times 0.43 \neq 0.15 = f_{xy}(2,1)$ $\quad$ $f_x(2)f_y(2) = 0.32 \times 0.37 \neq 0.12 = f_{xy}(2,2)$

$f_x(3)f_y(0) = 0.20 \times 0.20 \neq 0.05 = f_{xy}(3,0)$ $\quad$ $f_x(3)f_y(1) = 0.20 \times 0.43 \neq 0.10 = f_{xy}(3,1)$

$f_x(3)f_y(2) = 0.20 \times 0.37 \neq 0.05 = f_{xy}(3,2)$ $\quad$ $f_x(4)f_y(0) = 0.20 \times 0.20 \neq 0.10 = f_{xy}(4,0)$

$f_x(4)f_y(1) = 0.20 \times 0.43 \neq 0.10 = f_{xy}(4,1)$ $\quad$ $f_x(4)f_y(2) = 0.20 \times 0.37 \neq 0 = f_{xy}(4,2)$

$\therefore$ 두 확률변수 X와 Y는 서로 종속이다.

(교재 P.73) 예제5-8의 경우 독립.