확률변수와 확률분포

1. 확률변수 (Random Variables)

A. 정의

표본공간의 모든 결과들을 새로운 표본공간인 수직선상의 수치값으로 대응시켜 주는 함수

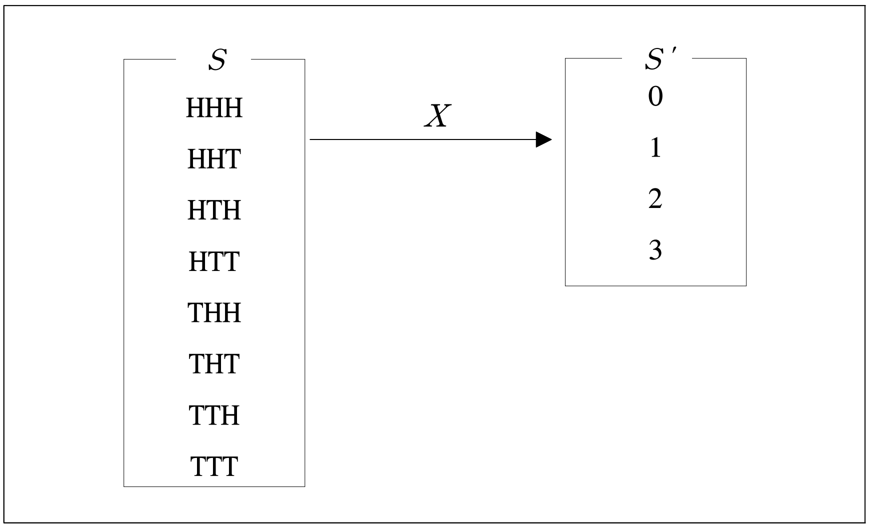
동전을 세 번 반복해서 던지는 실험

표본공간 S={HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

확률변수 X를 앞면이 나타날 횟수라고 정의.

새로운 표본공간 S'={0, 1, 2, 3}

동전던지기에서의 확률변수



B. 확률분포 (Probability Distribution)

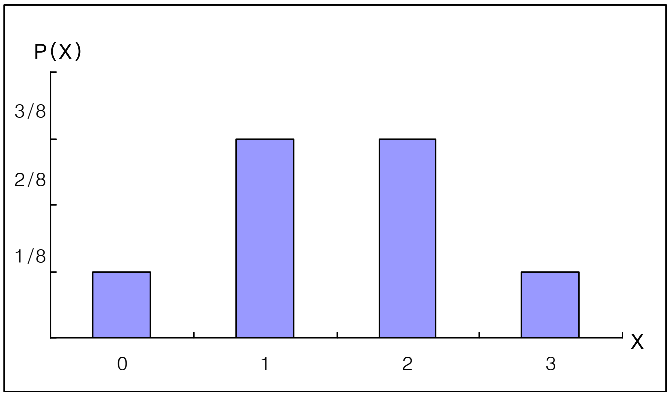
동전던지기의 확률분포

가능한 결과	확률
HHH	1/8
HHT	1/8
HTH	1/8
HTT	1/8
THH	1/8
THT	1/8
TTH	1/8
TTT	1/8

⇔

X	$P(X)$
0	1/8
1	3/8
2	3/8
3	1/8
Total	1

확률분포



C. 이산형 확률분포

1) 확률변수가 이산형

$$0 \leqq P(X=x_i) \leqq 1\ , \qquad i=1,\cdots,n$$

$$\sum_{i=1}^{n} P(X=x_i)=1$$

2) $P(X=x_i)$는 확률변수 X가 x_i의 값($i=1,\cdots,n$)을 갖는 확률.

3) 일반적으로 $P(X=x_i)\ =\ P(x_i)$

D. 연속형 확률분포

1) 확률변수가 연속형

2) 확률밀도함수 ($f(x)$)의 특징

① $f(x) \geqq 0, \qquad -\infty < x < \infty$

② $P(X = x) = 0$

③ $P(\alpha \leqq X \leqq \beta) = \int_{\alpha}^{\beta} f(x)\, dx$

④ $P(-\infty \leqq X \leqq \infty) = \int_{-\infty}^{\infty} f(x)\, dx = 1$

2. 기대값과 분산

A. 개념 : 얼마만큼을 기대할 수 있는가 ?

동전을 세 번 던져서 앞면이 세 번 나타나면 1,000원을 받고 두 번 나오면 500원을 받으며 한번 나타나거나 한번도 나타나지 않으면 800원을 잃게 되는 게임이 있다고 하자. 이러한 게임에서 확률변수 X는 손해 또는 이득을 나타내는 금액이므로 확률분포표를 작성하면 다음과 같다.

동전던지기 게임의 확률변수와 확률분포

앞면의 수	0번	1번	2번	3번
확률변수(X)	-800	-800	500	1,000
확률: $P(X)$	1/8	3/8	3/8	1/8

여기서 기대이익을 계산해 보면 다음과 같다.

기대이익=(-800원)×1/8 + (-800원)×3/8 + 500원×3/8 + 1,000원×1/8
=-87.5원

B. 기대값 : 확률변수에 확률변수가 나올 확률을 곱하여 합한 것.

1) 확률변수 X의 기대값을 $E(X)$로 표기
2) X의 위치를 나타내는 모수로서 평균(mean)이라 부르기도 함.
3) 평균 : 데이터가 얻어진 후의 무게 중심
4) 기대값 : 아직 데이터를 얻지 않은 상황에서 미래의 결과를 확률로 예측하여 얻은 평균

C. 분산

1) X의 분산은 $Var(X)$로 표기.
2) $(X-\mu)^2$의 기대값으로 표시될 수 있음.
3) 확률변수 X의 평균을 μ라고 하면,
   확률변수 X의 분포가 μ로부터 떨어진 정도인 $E[(X-\mu)^2]$를 확률변수 X의 분산(variance)이라고 할 수 있음.

D. 이산형 확률변수의 기대값과 분산

$$\mu = E(X) = \sum_{i=1}^{n} x_i \, P(x_i)$$

$$\sigma^2 = E[(X-\mu)^2] = \sum_{i=1}^{n} (x_i - \mu)^2 \, P(x_i)$$

$$= \sum_{i=1}^{n} x_i^2 \, P(x_i) - 2\mu \sum_{i=1}^{n} x_i \, P(x_i) + \mu^2 \sum_{i=1}^{n} P(x_i)$$

$$= E(X^2) - \mu^2$$

E. 연속형 확률변수의 기대값과 분산

$$\mu = E(X) = \int_{-\infty}^{\infty} x \, f(x) \, dx$$

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 \, f(x) \, dx$$

$$= \int_{-\infty}^{\infty} x^2 \, f(x) \, dx - 2\mu \int_{-\infty}^{\infty} x \, f(x) dx + \mu^2 \int_{-\infty}^{\infty} f(x) dx$$

$$= E(X^2) - \mu^2$$

F. 기대값과 분산의 성질

X와 Y를 확률변수 그리고 a와 b는 상수라 할 때,

1) 기대값의 성질

① $E(a) = a$

② $E(aX) = aE(X)$

③ $E(aX \pm bY) = aE(X) \pm bE(Y)$

④ $E(XY) = E(X)E(Y)$

단, X와 Y가 독립인 경우

2) 분산의 성질

① $Var(a) = 0$

② $Var(X) > 0$

③ $Var(X \pm a) = Var(X)$

$Var(aX) = a^2 Var(X)$

④ $Var(aX \pm b) = a^2 Var(X)$

3. 연습문제

A. 한 개의 동전을 던져 앞면이 나오면 1000원을 받고, 뒷면이 나오면 500원을 준다고 할 때,

1) 확률변수의 분포를 구하시오.

2) 기대금액을 구하시오

B. 어느 학생의 부모가 통계학 과목에서 A+를 받으면 5천원, A 학점을 받으면 3천원, B+ 학점을 받으면 2천원, B학점을 받으면 1천원을 준다는 약속을 했다. 각 경우의 확률이 다음과 같을 때, 기대값과 분산을 구하시오.

	A^+	A	B^+	B
확률	0.12	0.18	0.33	0.37

C. 확률변수 X의 분포가 다음과 같다.

X	0	1	2	3	4	5
$f(x)$	0.05	0.3	0.3	0.2	0.1	0.05

1) $E(X)$와 $Var(X)$를 구하시오.

2) $E(2X-1)$와 $Var(2X-1)$를 구하시오.