경영통계

원광대학교 경영학부

담당교수: 정호일

제 5 장 확률변수와 확률분포

- 확률변수
- 확률분포
- 확률함수
- 기대값과 분산
- 결합확률분포
- 공분산과 상관계수

1. 확률변수

확률변수 (random variable)

확률실험의 결과로 결정되는 수치를 취할 가능성을 확률로 표시할 수 있는 변수를 말한다.

이산확률변수와 연속확률변수

➢이산확률변수(discrete random variable)

셀 수 있는 정수값을 취하는 변수를 말한다. (예: 가족중 남자의 수,한시간 동안 어떤은행에 도착하는 고객의 수 등)

➢연속확률변수(continuous random variable)

일정한 실수구간 내에서 연속적인 값을 취할 수 있는 확률변수를 말한다.
(예:자족의 연간소득,증권의 수익률,체중,키,온도)

2. 확률분포

확률분포 (probability distribution)

실험의 가능한 모든 결과를 수치로 나타내고 각 결과에 대응하는 확률을 도수분포표나 그래프로 나타낸 것을 말한다.

이산확률분포와 연속확률분포

- 이산확률분포(discrete probability distribution)란 이산확률변수의 확률분포를 말한다. 예) 이항분포, 포아송분포, 초기하분포 등

- 연속확률분포(continuous probability distribution)란 연속확률변수의 확률분포를 말한다.
예) 균등분포,정규분포,지수분포,t분포,F분포,카이제곱(X^2)분포 등

3. 확률함수

확률함수 (probability function)

➢확률변수 X가 어떤 특정한 실수값 x를 취할 확률을 일일이 나열하지 않고 x의 함수로 간편하게 나타낸 것을 말한다.

➢ 확률함수는 대상이 되는 변수가 이산이냐 또는 연속이냐에 따라 확률질량함수(probability mass function : pmf)와 확률밀도함수(probability density function : pdf)로 나뉜다.

확률질량함수

➢확률함수는 이산점 x에서 0보다 큰 값을 취할 수 있기 때문에 이는 확률질량함수라고 부른다.

➢**확률질량함수**란 이산확률변수 X가 취할 수 있는 각 실수값 x에 확률을 대응시키는 함수를 말한다.

➢확률질량함수의 조건

$P(X=x) \geq 0$

$$\sum_{\text{모든}x} P(X=x)=1$$

확률밀도함수

연속확률변수 X가 취할 수 있는 어떤 실수구간 속의 실수값 x 에 확률을 대응시키는 함수를 말한다.

4. 기대값과 분산

기대값

이산확률변수의 확률분포가 구해지면 확률변수의 기대값 (expected value)과 분산(variance)을 계산할 수 있다.

확률변수 X의 기대값 E(x) 또는 μ 는 확률변수 X의 가능한 모든 값들의 가중평균(weighted mean)인데 가중치는 각 값들의 확률이다.

기대값의 특성

$$E(a) = a$$

$$E(bX) = bE(X)$$

$$E(a + bX) = a + bE(X)$$

$$E(X + Y) = E(X) + E(Y)$$

$$E(aX + bY) = aE(X) + bE(Y)$$

분산

기대값은 확률분포의 집중경향을 나타내는데 반하여 분산 (variance) 또는 표준편차(standard deviation)는 확률변수들이 기대값을 중심으로 얼마나 흩어져 있는가를 나타내는 산포도(degree of dispersion)의 측정치이다

분산의 특징

$$Var(a) = 0$$

$$Var(a + x) = Var(x)$$

$$Var(bx) = b^2 Var(x)$$

$$Var(x + y) = Var(x) + Var(y) \text{(} X \text{와 } Y \text{는 독립적인 확률변수)}$$

5. 결합확률분포

결합확률함수

두 이산변수 X와 Y의 결합확률함수란 변수 X가 실수값 x, 변수 Y가 실수값 y를 취할때 대응하는 확률을 발생시키는 함수를 말한다.

$P(X,Y) = P(X = x, Y = y)$

결합확률함수의 조건

모든 실수값 x, y에 대하여 $P(X,Y) \geq 0$이다.

두 변수 X와 Y가 취할 수 있는 모든 값에 대한 확률의 합은 1이다.

$$\sum_x \sum_y P(X,Y) = 1$$

주변확률함수

두 이산확률변수 X와 Y의 주변확률함수는 다음과 같다.

$$P(X) = \sum_{\text{모든 } y} P(X, Y)$$

$$P(Y) = \sum_{\text{모든 } x} P(X, Y)$$

조건확률함수

확률변수 Y가 실수값 y를 취하였다는 조건하에서 변수 X가 실수값 x를 취할 조건확률함수는 다음과 같이 정의한다.

$$P(X \mid Y) = \mathrm{P}(\mathrm{X} = x \mid \mathrm{Y} = y) = \frac{P[(X = x) \cap (Y = y)]}{P(Y = y)}$$

두 변수의 독립성

두 변수간의 독립성 조건

두 개의 확률변수 X와 Y가 독립적이면

$$P(X = x \mid Y = y) = P(X = x)$$

$$P(Y = y \mid X = x) = P(Y = y)$$

6. 공분산과 상관계수

공분산

$$모집단 : \mathrm{Cov}(X,Y) = \sigma_{XY} = \sum_{i=1}^{N}[X_i - E(X)][Y_i - E(Y)]P(X_i Y_i)$$

$$= E[(X - E(X))(Y - E(Y))]$$

$$= E(XY) - E(X)E(Y)$$

$P(X_i Y_i) : X$의 i번째 결과와 Y의 j번째 결과의 발생확률

상관계수

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

σ_X : 변수 X의 표준편차

σ_Y : 변수 Y의 표준편차