이론 통계학

2021년 봄학기

전북대학교 통계학과

Outline

소개

확률밀도함수 및 확률분포함수

결합분포

기댓값

소개

소개

- 송성주, 전명식(2015) 수리통계학의 2장을 기본으로하여 만든 강의 노트입니다.
- 강기훈, 박진호 (2015) 수리통계학의 2장, 3장, 4장에 해당합니다.

예

- 동전 3회 던지는 실험
- 어떤 기계의 수명시간 측정

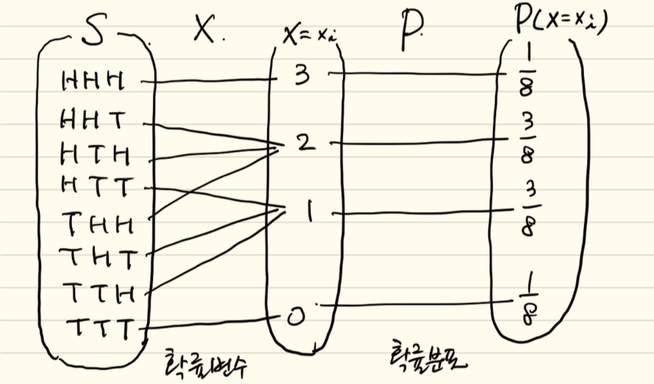
2장에서는

- 표본공간 S에 정의된 사건들을 실수를 사용하여 나타내고
- 확률을 이용하여 수치적인 양에 대한 수리적 모형 고려

확률변수 - 표본공간 *S*에 정의된 실수값을 가지는 함수 (real-valued function), *X* 영문대문자

이산형, **연속형**

예제: 동전 3회 던지는 실험



확률밀도함수 및 확률분포함수

확률밀도함수 및 확률분포함수

- 확률밀도함수(probability density function)- $f(x)$

- 확률분포함수 (probability distribution function) - $F(x)$

- 확률변수의 분포형태를 나타내는 데 사용

이산형 확률변수의 경우

다음 조건 만족

- 모든 실수 x에 대하여 $f(x) \geq 0$
- 확률변수가 가질 수 있는 값 $x_1, x_2, \ldots$에 대하여 $f(x_i) > 0, \quad \sum_{\text{all}_{x_i}} f(x_i) = 1,$

- 확률질량함수(probability mass function)으로 부르기도 함.
- $f(x) = P(X = x)$

연속형 확률변수의 경우

- 셀 수 없이 무한히 많은 가능한 값 하나하나에 확률을 부여하지 않고 구간에 확률 부여,

  즉 $P(X = x) = 0$,

  다음 조건 만족

  - 모든 실수 x에 대하여 $f(x) \geq 0$
  - $\int_{-\infty}^{\infty} f(x)dx = 1$

- $P(a \leq X \leq b) = \int_a^b f(x)dx$

$$P(X \in A) = \begin{cases} \sum_{x_i \in A} f(x_i) & \text{discrete} \quad X \\ \int_A f(x)dx & \text{continuous} \quad X \end{cases}$$

예제 2.1
구간 (0,3)에서 정의된 함수

$$f(x) = \frac{x^2}{9}$$

확률밀도함수인가?

(누적)분포함수(cumulative distribution function)
확률변수 X가 주어진 점 x이하인 값을 가질 확률

$$F(x) = P(X \leq x)$$

참고

- $X \sim f(x)$: 확률변수 X가 확률밀도함수 $f(x)$를 가짐
- $X \sim F(x)$: 확률변수 X가 확률분포함수 $F(x)$를 가짐

$$F(x) = \begin{cases} \sum_{x_i \leq x} f(x_i) & \text{discrete} \quad X \\ \int_{-\infty}^{x} f(t)dt & \text{continuous} \quad X \end{cases}$$

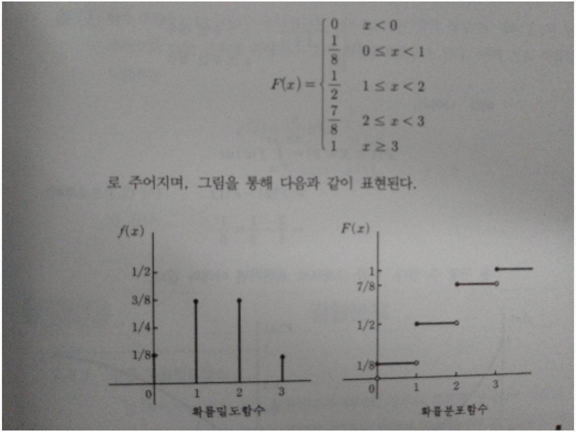
정리 2.1
함수 $F(x)$가 어떤 확률변수 X의 누적분포함수가 되는 필요충분조건

- $\lim_{x\to-\infty} F(x) = 0$
- $\lim_{x\to\infty} F(x) = 1$
- $\lim_{h\to 0+} F(x+h) = F(x)$
- $a < b$ 이면 $F(a) \leq F(b)$

$$P(a < X \leq b) = F(b) - F(a)$$

예제 2.2

앞면 나올 확률이 1/2인 동전 3회 던지는 실험에서 관심있는 변수 X=앞면의 수 일 때 $f(x)$와 $F(x)$는?



$$F(x)=\begin{cases} 0 & x<0 \\ \frac{1}{8} & 0\le x<1 \\ \frac{1}{2} & 1\le x<2 \\ \frac{7}{8} & 2\le x<3 \\ 1 & x\ge 3 \end{cases}$$

로 주어지며, 그림을 통해 다음과 같이 표현된다.

예제 2.3
연속형 확률 변수 X의 확률밀도함수

$$f(x) = \begin{cases} c(1+x)^{-2} & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (= c(1+x)^{-2} I(x > 0))$$

$c, \; P(1 < X < 2)$?

정리 2.2
연속형 확률 변수 X의 확률밀도함수 $f(x)$와 확률분포함수 $F(x)$는

$$f(x) = \frac{d}{dx} F(x)$$

결합분포

예

환자의 건강기록: 나이, 혈압, 몸무게 등

아버지의 키와 아들의 키

종종 여러 개의 확률변수들을 동시에 고려

$\Rightarrow$ 결합분포(joint distribution) 이론 필요

확률벡터 (random vector)
$\mathbf{X} = (X_1, X_2, \ldots, X_k)$

결합확률밀도함수

결합 확률밀도함수 (joint probability density function)
두 확률변수 X와 Y의 결합 확률밀도함수 $f_{X,Y}(x,y)$

- 이산형인 경우

$$f_{X,Y}(x,y) = P(X = x, Y = y)$$

- 연속형인 경우

  임의의 영역 A에 대하여

$$P[(X,Y) \in A] = \int\int_A f_{X,Y}(x,y)dxdy$$

결합확률분포함수

결합 확률밀도함수
확률변수 X와 Y의 결합 확률밀도함수 $f_{X,Y}(x,y)$는 모든 실수 x, y에 대하여 다음을 만족

$$f_{X,Y}(x,y) \geq 0$$

$$\sum_{(x,y)} f_{X,Y}(x,y) = 1 \quad \text{이산형인 경우}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1 \quad \text{연속형인 경우}$$

k개의 확률변수 분포로 확장
$f(x_1, x_2, \ldots, x_k)$

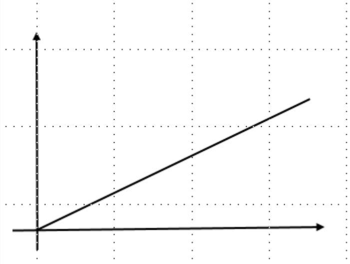
예제 2.4
4개의 빨간 공과 3개의 하얀 공과 2개의 검은 공이 들어 있는 주머니에서 3개의 공을 임의로 꺼낼 때, X=하얀공의 수, Y=검은 공의 수라고 하자. 두 변수 (X, Y)의 결합 확률밀도함수는?

$$f_{X,Y}(x,y) = \frac{\binom{3}{x}\binom{2}{y}\binom{4}{3-x-y}}{\binom{9}{3}} \quad (x = 0,1,2,3; y = 0,1,2; 0 \le x+y \le 3)$$

예제 2.5
두 확률변수 X(=TV 시청시간), Y(=숙제하는 데에 보내는 시간)의 결합 확률밀도함수가

$$f_{X,Y}(x,y) = xye^{-(x+y)} I(x>0, y>0)$$

라고 주어졌다고 할 때, $X \geq 2Y$의 확률은?



결합확률분포함수

확률변수 $X_1, X_2, \ldots, X_k$의 결합 확률분포함수

$$F(x_1, x_2, \ldots, x_k) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_k \leq x_k)$$

연속형인 경우

$$f(x_1, x_2, \ldots, x_k) = \frac{\partial^k}{\partial x_1 \cdots \partial x_k} F(x_1, x_2, \ldots, x_k)$$

예제 2.6
두 확률변수 X, Y의 결합 확률분포함수가

$$F_{X,Y}(x,y) = xy \quad (0 < x < 1, 0 < y < 1)$$

라고 주어졌을 때, X, Y의 결합 확률밀도함수를 구하고 $P(X^2 < Y)$를 구하여 보자.

주변 확률밀도함수

결합분포로부터 각 변수만의 분포를 구할 필요가 있을 때

주변 확률밀도함수(marginal probability density function)

- 이산형인 경우

$$f_X(x) = \sum_y f_{X,Y}(x, y), \ \ f_Y(y) = \sum_x f_{X,Y}(x, y)$$

- 연속형인 경우

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy, \ \ f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$$

예제 2.7
두 확률변수 X, Y의 결합 확률밀도함수가

$$f_{X,Y}(x,y) = xye^{-(x+y)} I(x > 0, y > 0)$$

라고 주어졌다고 할 때, 이 때 X, Y의 주변 확률밀도함수?

조건부 확률밀도함수

- 여러 개의 확률변수들 중 몇 개의 변수값이 주어져 있을 때 나머지 변수들의 분포에 대한 이론
- 회귀분석 등에서 필수적으로 사용
- 변수들의 독립개념과 관련

조건부 확률밀도함수(conditional probability density function)
$X = x$가 주어졌을 때, $Y|x$의 조건부 확률밀도함수

$$f_{Y|x}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

예제2.8
두 확률변수 X와 Y의 결합 확률밀도함수 $f_{X,Y}(x,y)$가 다음과 같이 주어져 있다.

	x=0	x=1	x=2	x=3
y=0	4/84	18/84	12/84	1/84
y=1	12/84	24/84	6/84	0
y=2	4/84	3/84	0	0

$f_{Y|X}(y|x=0)$을 구하시오.

예제 2.9
두 확률변수 X와 Y의 결합 확률밀도함수 $f_{X,Y}(x,y)$가 다음과 같이 주어져 있다.

$$f_{X,Y}(x,y) = x^2 e^{-x(y+1)} I(x > 0, y > 0)$$

$X = x$가 주어지면, $Y|x$의 조건부 확률밀도함수 $f_{Y|x}(y|x)$을 구하시오.

독립확률변수

두 확률변수의 독립(independent)
모든 실수 x, y에 대하여

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

여러 확률변수의 독립(independent)
모든 실수 $x_1, \dots, x_k$에 대하여

$$f(x_1, \dots, x_k) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_k}(x_k)$$

랜덤표본(random sample)
확률밀도함수 $f(x)$를 갖는 모집단으로부터 크기가 n인 랜덤표본 $X_1, X_2, \dots, X_n$을 얻었을 때 이들의 결합 확률밀도함수는

$$f(x_1, \dots, x_n) = f(x_1) f(x_2) \cdots f(x_n)$$

기댓값

기댓값

- 확률밀도함수나 확률분포함수는 확률변수의 전체적인 성격 설명
- 몇 개의 수치들 (평균, 분산,...)로 확률분포의 성질 요약

확률변수 X의 기댓값 또는 평균
확률변수 X (분포)의 중심위치 척도

$$E(X) = \begin{cases} \sum_{\text{all } x_i} x_i f_X(x_i) & X : \text{discrete} \\ \int_{-\infty}^{\infty} x f_X(x) dx & X : \text{continuous} \end{cases}$$

(단, $E(|X|) < \infty$)

기댓값

$X \sim f_X(x), \quad g(X)$: a function of X

$g(X)$의 기댓값

$$E[g(X)] = \begin{cases} \sum_{\text{all } x_i} g(x_i) f_X(x_i) & X : \text{discrete} \\ \int_{-\infty}^{\infty} g(x) f_X(x) dx & X : \text{continuous} \end{cases}$$

- $X_1, X_2, \ldots, X_n \sim f(x_1, x_2, \ldots, x_n)$
- $g(X_1, X_2, \ldots, X_n)$: a function of $X_1, X_2, \ldots, X_n$

$$E[g(X_1, X_2, \ldots, X_n)] =$$

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \cdots dx_n$$

Week 2

- random variable
- probability density function (pdf)
- cumulative distribution function (cdf)
- joint pdf
- joint cdf
- marginal pdf
- conditional pdf
- independent random variables
- random sample