확률 및 통계학

- 3장 확률변수와 확률분포 -

명 세인(sein@pel.smuc.ac.kr)

상명대학교 프로토콜공학연구실

목 차

- 확률변수의 개념
- 이산형 확률분포
- 연속형 확률분포
- 결합확률분포

확률변수의 개념

- 확률변수(Random Variable)
  - 표본공간의 각 원소를 실수값으로 대응하는 함수
  - 4개의 붉은공(R)과 3개의 검은 공(B)이 들어있는 항아리에서 연속적으로 2개의 공을 비복원 추출하며 **Y를 붉은 공의 개수**로 할 때, 출현가능한 결과와 확률변수 Y의 값 y

표본공간	Y
RR	2
RB	1
BR	1
BB	0

확률변수의 개념

- 확률변수(Random Variable)
  - 무한개의 원소를 가지는 표본공간
    - 주사위를 5의 눈이 나타날 때 까지 시행
    - S = {F, NF, NNF, NNNF, … }

확률변수의 개념

- 확률변수(Random Variable)
  - 가능, 불가능을 나타낼때 1과 0을 사용하는 베르누이 (Bernoulli) 확률변수

$$X = \begin{cases} 1, \text{부품이 불량일 때} \\ 0, \text{부품이 양품일 때} \end{cases}$$

확률변수의 개념

- 이산, 연속형 표본공간
  - 이산표본공간(Discrete Sample Space)
    - 표본공간이 유한개 혹은 셀 수 있는 무한개의 원소
    - 계수자료(Count Data)
      - 사고횟수, 불량품횟수, 발생건수
  - 연속표본공간(Continuous Sample Space)
    - 표본공간이 실선의 어떤 구간 내의 모든 수를 포함
    - 측정자료(Measured Data)
      - 높이, 무게, 온도, 거리, 수명

이산형 확률분포

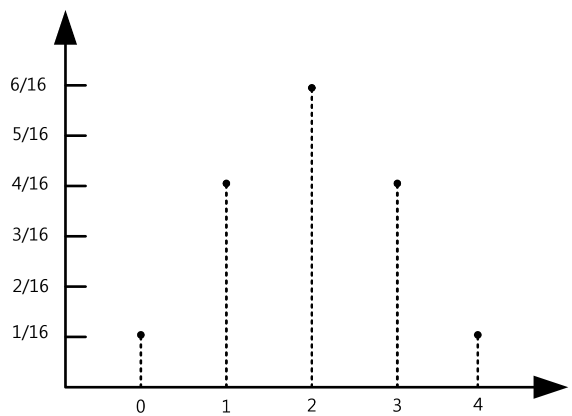
- 이산형 확률분포
  - 확률 분포(Probability Distribution)는 확률변수의 각 값에 확률을 할당
  - 모든 x에 대해 순서쌍(x, f(x))의 집합이 아래 식을 만족하면, 이산형 확률변수 X의 확률함수, 확률질량함수, 확률분포
    - $f(x) \geq 0$
    - $\sum_x f(x) = 0$
    - $P(X = x) = f(x)$

이산형 확률분포

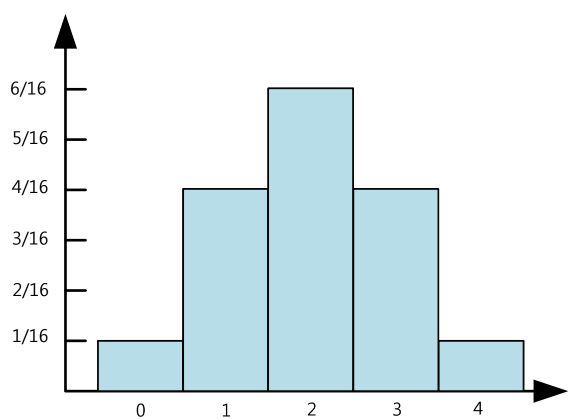
- 이산형 확률분포
  - 예제
    - 판매된 자동차의 50%는 디젤엔진일 때, 다음에 판매될 4대의 자동차중 **디젤엔진이 장착된 차의 수의 확률분포** 식은?
      - 발생확률이 동일한 표본점 2^4개
      - $f(x) = \frac{\binom{4}{x}}{16}, x = 0, 1, 2, 3, 4$

이산형 확률분포

- 확률질량 함수도(Probability Mass Function)



- 확률 히스토그램(Probability Histogram)

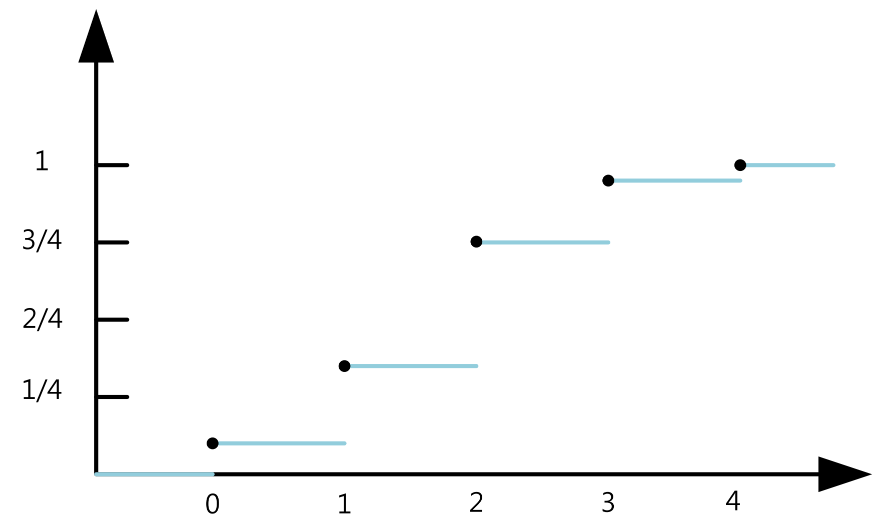


이산형 확률분포

- 누적분포함수 (Cumulative Distribution Function)
  - 특정값 이상, 이하 또는 두값 사이의 확률을 구하기 위함
  - 확률분포 f(x)를 가지는 이산형 확률변수 X의 누적분포함수 F(x)
    - 모든 실수 x에 대한 $f(x) = P(X \le x)$
    - $\mathrm{F}(x) = P(X \le x) = \sum_{t \le x} f(t)\,,\ -\infty < x < \infty$

이산형 확률분포

- 누적분포함수 (Cumulative Distribution Function)
  - 예제
    - 자동차 판매 확률분포의 누적분포
      - $F(x) = \begin{cases} 0, x < 0 \\ \frac{1}{16}, 0 \le x < 1 \\ \frac{5}{16}, 1 \le x < 2 \\ \frac{11}{16}, 2 \le x < 3 \\ \frac{15}{16}, 3 \le x < 4 \\ 1, x \ge 4 \end{cases}$



연속형 확률분포

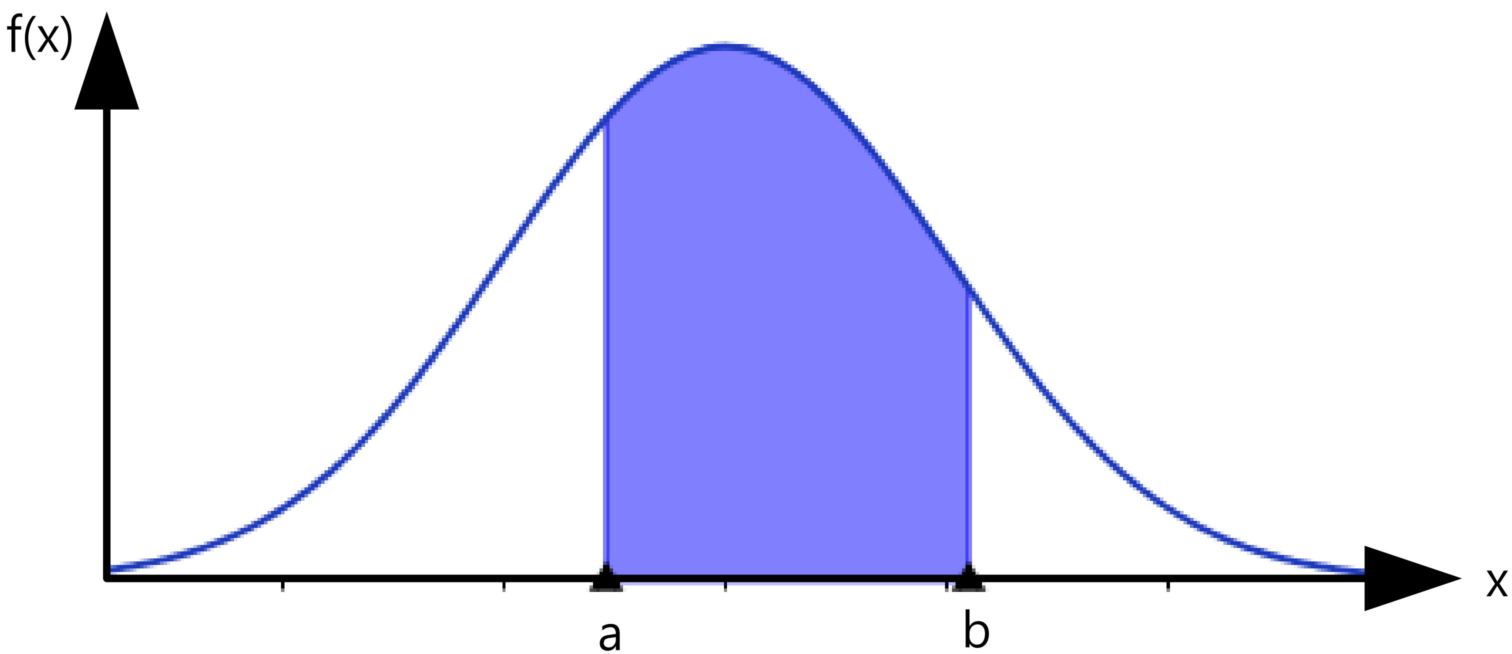
- 연속형 확률분포
  - 확률분포에서 확률변수값의 한 점이아닌, 어떤 구간의 값, 끝 점의 포함 여부는 문제되지 않음
    - $P(a < X \leq b) = P(a < X < b) + P(X = b) = P(a < X < b)$
  - 확률밀도함수(Probability Density Function)
    - 연속형 확률변수를 취급하는 식
    - 밀도함수(Density Function)와 같은말
    - 다음을 만족하는 f(x)는 집합R 에서정의된 연속형확률변수에대한 확률밀도함수
      - 모든$x \in R$에 대하여 $f(x) \geq 0$
      - $\int_{-\infty}^{\infty} f(x)dx = 1$
      - $P(a < X < b) = \int_{a}^{b} f(x)dx$

연속형 확률분포

- 확률밀도함수 (Probability Density Function)
  - 제어실험에서 반응온도의 변화에 따른 오차는 다음과 같은 확률분포를 가지는 연속확률 변수X일 때
    - $f(x) \begin{cases} \frac{x^2}{3}, -1 < x < 2 \\ 0, \text{구간 외} \end{cases}$
    - $f(x)$가 확률밀도 함수임을 증명하라
      - $\int_{-\infty}^{\infty} f(x)\,dx = \int_{-1}^{2} \frac{x^2}{3}\,dx = \left.\frac{x^3}{9}\right|_{-1}^{2} = \frac{8}{9} - \frac{(-1)^3}{9} = 1$
    - P($0 < x \leq 1$)는?
      - P($0 < x \leq 1$)= $\int_0^1 \frac{x^2}{3} = \left.\frac{x^3}{9}\right|_0^1 = \frac{1}{9}$

연속형 확률분포

- 확률밀도함수 (Probability Density Function)
  - $P(a < X < b) = \int_a^b f(x)\,dx$



연속형 확률분포

- 연속형 누적분포함수
  - 확률밀도함수가 f(x)인 연속형 확률변수 X의 누적분포함수 F(x)
    - $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, -\infty < x < \infty$
    - $P(a < X < b) = F(b) - F(a)$
    - 미분이 가능하면(Differentiable), $f(x) = \frac{dF(x)}{dx}$

연속형 확률분포

- 연속형 누적분포함수
  - 제어실험에서 반응온도의 변화에 따른 오차는 다음과 같은 확률분포를 가지는 연속확률 변수X일 때
    - $f(x)\begin{cases} \frac{x^2}{3}, -1 < x < 2 \\ 0, \text{구간 외} \end{cases}$
    - 확률밀도함수에 대하여 F(x)를 구하라

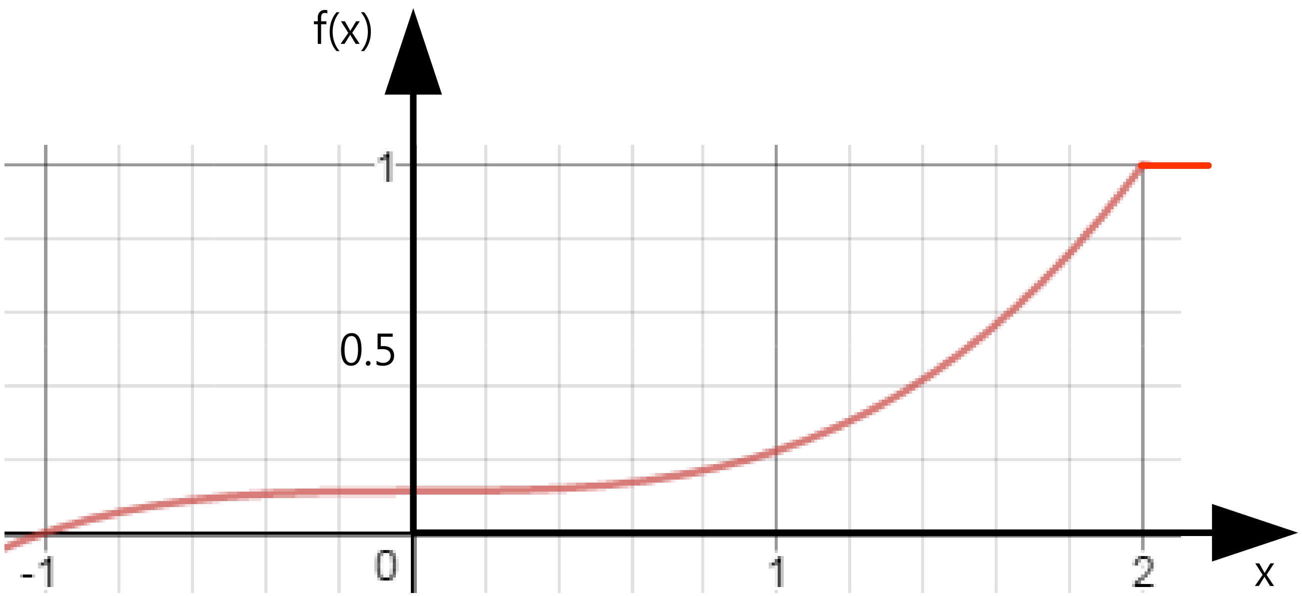
$$F(x)\int_{-\infty}^{x} f(t)\,dt = \int_{-1}^{x} \frac{t^2}{3}\,dt = \frac{x^3}{9}\bigg|_{-1}^{x} = \frac{x^3+1}{9}$$

$$\therefore F(x)\begin{cases} 0, x < -1 \\ \frac{x^3+1}{9}, -1 < x < 2 \\ 1, x \geq 2 \end{cases}$$

연속형 확률분포

- 연속형 누적분포함수

$$F(x)\begin{cases} 0, x < -1 \\ \dfrac{x^3+1}{9}, -1 < x < 2 \\ 1, x \geq 2 \end{cases}$$



결합분포함수

- 결합확률분포 (Joint Probability Distribution)
  - X와 Y가 두 이산형 확률변수일 때, 어떤 (x,y)에 대하여 f(x,y)는 x와 y가 동시에 일어날 확률
  - 이산형 확률변수 X와 Y의 결합확률분포
    - 모든 (x, y)에 대하여 $f(x, y) \geq 0$
    - $\sum_x \sum_y f(x, y) = 1$
    - $P(X = x,\ Y = y) = f(x, y)$
    - $P[(X, Y) \in A] = \sum\sum_A f(x, y)$

결합분포함수

- 결합확률분포 (Joint Probability Distribution)
  - 예제
    - 3개의 청색, 2개의 적색, 3개의 녹색볼펜이 들어있는 상자에서 임의의 2개를 추출
      - 결합확률분포 $f(x,y) = \dfrac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}}$
      - A={(x,y)|x+y≤1}일 때, P[(X, Y)∈A]

$f(x,y)$		x = 0	x = 1	x = 2	행의 합
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$		$\frac{3}{7}$
	2	$\frac{1}{28}$			$\frac{1}{28}$
열의 합		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

결합분포함수

- 결합밀도함수(Joint Density Function)
  - 연속확률변수 X와 Y의 결합밀도함수
    - 모든(x, y)에 대하여 $f(x, y) \geq 0$
    - $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx\, dy = 1$
    - $P[(X, Y) \in A] = \iint_A f(x, y) dx\, dy$ 이때 A는 xy평면상의 임의의 영역

결합분포함수

- 결합밀도함수(Joint Density Function)
  - 예제
    - 과자회사에서 연한초콜릿과 진한초콜릿을 입힌 과자상자를 판매, 임의로 하나의 과자상자를 선택했을 때 X와 Y를 각각 연한초콜릿 진한초콜릿의 비율이라 하면, 결합확률분포는 다음과 같다
    - $f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), 0 \le x \le 1, 0 \le y \le 1 \\ 0, \text{구간 외} \end{cases}$
    - $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dx\ dy = 1$ 증명

결합분포함수

- 결합밀도함수(Joint Density Function)
  - 예제
    - $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dx\ \ dy = 1$ 증명

$$= \int_0^1 \int_0^1 \frac{2}{5}(2x+3y)dx\,dy = \int_0^1 \left(\frac{2x^2}{5} + \frac{6xy}{5}\right)\Bigg|_{x=0}^{x=1} dy$$

$$= \int_0^1 \left(\frac{2}{5} + \frac{6y}{5}\right) dy = \left(\frac{2y}{5} + \frac{3y^2}{5}\right)\Bigg|_0^1 = \frac{2}{5} + \frac{3}{5} - (0+0) = 1$$

결합분포함수

- 주변분포(Marginal Distribution)
  - 결합확률분포에서 하나의 확률변수만의 확률분포
  - 이산형인 경우 합을, 연속형인 경우 적분을 취함

  - 이산형 $g(x) = \sum_y f(x, y),\ h(y) = \sum_x f(x, y)$

  - 연속형 $g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy,\ h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$

결합분포함수

- 주변분포(Marginal Distribution)
  - 이산형 예제
    - 청색, 적색, 녹색상자 예제에서 X와 Y의 주변분포

$f(x,y)$		x = 0	x = 1	x = 2	행의 합
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$		$\frac{3}{7}$
	2	$\frac{1}{28}$			$\frac{1}{28}$
열의 합		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

결합분포함수

- 조건부 분포(Conditional Distribution)
  - X와 Y를 이산형 또는 연속형인 두 확률변수이며, $X = x$로 주어졌을때 확률변수 Y의 조건부분포
    - $f(y|x) = \frac{f(x,y)}{g(x)}, g(x) > 0$
  - 같은 방법으로 $Y = y$로 주어졌을 때 확률변수 X의 조건부 분포
    - $f(x|y) = \frac{f(x,y)}{h(y)}, h(y) > 0$

결합분포함수

- 조건부 분포(Conditional Distribution)
  - 연속형 확률변수 $Y = y$로 주어졌을 때 연속형 확률변수 X가 a와 b사이의 값을 가질 확률은
    - $P(a < X < b | Y = y) = \int_a^b f(x|y)\ dx$

결합분포함수

- 조건부 분포(Conditional Distribution)
  - 예제
    - 3청색, 2적색, 3녹색 폴펜상자 예제에서 X를 청색볼펜의 수, Y를 적색 볼펜의 수이며, Y=1로 주어졌을 때 X의 조건부 분포와 P(X=0 | Y=1)의 확률은?

$f(x,y)$		$x=0$	$x=1$	$x=2$	행의 합
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$		$\frac{3}{7}$
	2	$\frac{1}{28}$			$\frac{1}{28}$
열의 합		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

결합분포함수

- 통계적 독립(Statistically Independent)
  - X와 Y를 결합확률분포 f(x,y)와 주변분포 g(x), h(y)를 가지는 이산형 혹은 연속형 확률변수
    - 모든(x, y)에 대하여 $f(x,y) = g(x)h(y)$가 성립하면 통계적으로 독립
    - 예제
      - $f(0,1) = g(0)h(1)$인가?

감사합니다!