확률 및 통계학

- 4장 수학적 기대값 -

임연주 (yeonjoo@pel.smuc.ac.kr)

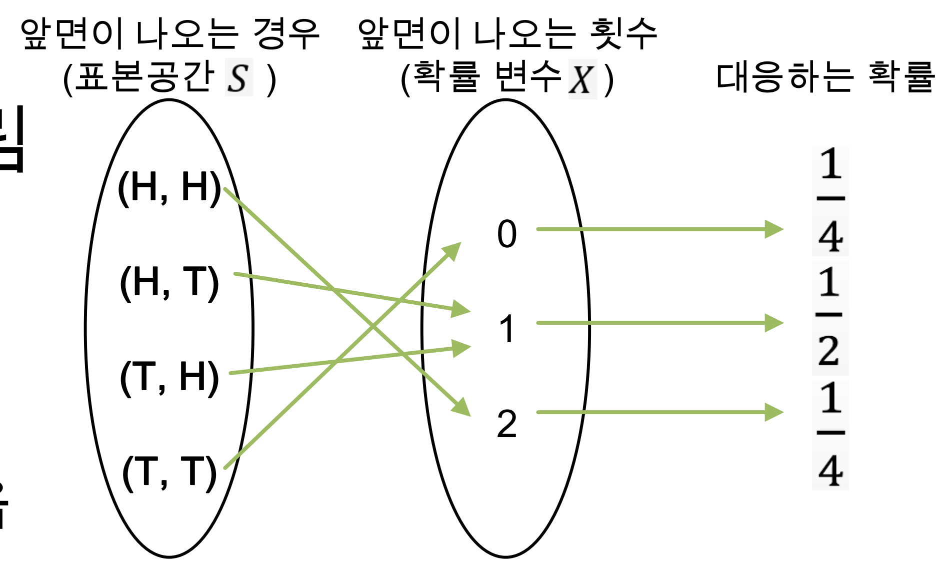
상명대학교 프로토콜공학연구실

목 차

- 확률변수의 평균
- 분산과 공분산
- 선형결합된 확률변수의 평균과 분산
- 체비세프 정리

확률변수의 평균

- 확률변수
  - 정의
    - 표본공간의 각 원소에 하나의 실수를 대응시킬 때 실수를 의미함
  - 표본공간과 확률변수 관계 그림



  - 종류
    - 이산 확률변수
      - 확률변수 X의 집합을 셀 수 있음
    - 연속 확률변수
      - 확률변수 X가 연속적인 구간 내의 값을 취함

확률변수의 평균

- 기대값(Expectation)
  - 각 사건이 발생할 때의 결과와 그 사건이 벌어질 확률을 곱한 것들
  - 전체 사건에 대해 합한 값
    - $E(X) = \sum xf(x)$
  - 확률변수 X가 확률분포 $f(x)$를 가질 때, X의 기대값(μ)
    - 이산형: $E(X) = \sum xf(x)$
    - 연속형: $E(X) = \sum_{-\infty}^{\infty} xf(x)\, dx$

확률변수의 평균

- 기대값 Practice (1/3)
  - 품질검사원이 7개의 부품으로 구성된 로트를 검사
  - 로트에 4개의 양품과 3개의 불량품이 들어있음
  - 검사원이 3개의 부품을 추출할 때 나타나는 양품의 평균개수는?
    - 추출된 표본의 양품의 확률변수 $X = \{0,1,2,3\}$
    - X의 확률분포 $f(x) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}, \mathrm{x} = \{0,1,2,3\}$
    - $\mu = E(X) = (0)\left(\frac{1}{35}\right) + (1)\left(\frac{12}{35}\right) + (2)\left(\frac{18}{35}\right) + (3)\left(\frac{4}{35}\right) = \mathbf{\frac{12}{7}}$

확률변수의 평균

- 기대값 Practice (2/3)
  - 의료기기 외판원이 두 고객에게 거래를 함. 첫 번째 고객과의 거래 성사율은 70%이고 이 경우 1000원을, 두 번째 고객과의 거래 성사율은 40%이고 이 경우 1500원을 벌게 됨
  - 각 고객과의 거래 결과는 서로 독립적이라고 할 때 외판원이 기대할 수 있는 보수는 얼마인가?
    - 추출된 외판원의 성공 보수 $X = \{0,1000,1500, 2500\}$
    - 확률변수 X의 확률분포 $f(x)$
      - $f(0) = (1-0.7)(1-0.4) = 0.18$, $f(1000) = (0.7)(1-0.4) = 0.28$, $f(1500) = (1-0.7)(0.4) = 0.42$, $f(2500) = (0.7)(0.4) = 0.12$
    - $\mu = E(X) = (0)(0.18) + (1000)(0.28) + (1500)(0.42) + (2500)(0.12) = \mathbf{1300}$

확률변수의 평균

- 기대값 Practice (3/3)
  - 전자장치의 수명(단위: 시간)을 확률변수 X, 보기와 같은 확률밀도함수 $f(x)$로 주어졌을 때, 이 장치의 기대수명을 구하라
  - 확률밀도함수 $f(x) = \begin{cases} \frac{20000}{x^3}, x > 100 \\ 0,\ \text{다른구간} \end{cases}$
    - $\mu = E(X) = \int_{100}^{\infty} x \, \frac{20000}{x^3} dx = \int_{100}^{\infty} \frac{20000}{x^2} = \mathbf{200}$

확률변수의 평균

- 확률변수 $g(X)$의 기대값
  - 확률변수 X에 종속(관계)되는 확률변수 $g(X)$
  - 정의
    - 확률변수 X가 확률분포 $f(x)$를 갖는 확률변수 일 때, $g(X)$의 기대값
      - 이산형: $\mu_{g(X)} = E[g(X)] = \sum g(x) f(x)$
      - 연속형: $\mu_{g(X)} = E[g(X)] = \int_{-\infty}^{\infty} \sum g(x) f(x)\ dx$

확률변수의 평균

- 확률변수 $g(X)$의 기대값 Practice (1/2)
  - 금요일 오후 4시에서 5시 사이에 세차장에서 서비스를 받는 차의 수를 X라고 할 때, X의 확률분포가 다음과 같음

x	4	5	6	7	8	9
$P(X=x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

  $g(X) = 2X - 1$을 종업원이 받는 수당(단위: 원)이라고 할 때, 이 시간대의 종업원의 기대수익은?

  - $\mu = E[g(X)] = E(2X-1) = \sum_4^9 (2x-1)f(x) = (7)\left(\frac{1}{12}\right) + (9)\left(\frac{1}{12}\right) + (11)\left(\frac{1}{4}\right) + (13)\left(\frac{1}{4}\right) + (16)\left(\frac{1}{6}\right) + (17)\left(\frac{1}{6}\right) = \mathbf{12.67}$원

확률변수의 평균

- 확률변수 $g(X)$의 기대값 Practice (2/2)
  - 확률변수 X의 밀도함수 $f(x)$가 보기와 같을 때, $g(X) = 4X + 3$의 기대값을 구하라
  - 확률밀도함수 $f(x) = \begin{cases} \frac{x^2}{3} & , -1 < x \\ 0, & \text{다른구간} \end{cases}$
    - $\mu = E(4X+3) = \int_{-1}^{2} (4x+3)\,\frac{x^2}{3}dx = \frac{1}{3}\int_{-1}^{2}(4x^3+3x^2)dx = \mathbf{8}$

확률변수의 평균

- 결합확률분포의 기대값
  - 정의
    - X와 Y는 결합확률분포 $f(x, y)$를 갖는 확률변수일 때, 확률변수 $\mathrm{g}(X, Y)$의 기대값 $\mu_{\mathrm{g}(X,Y)}$
    - 이산형: $\mu_{\mathrm{g}(X,Y)} = \mathrm{E}[\mathrm{g}(X, Y)] = \sum\sum g(x, y) f(x, y)$
    - 연속형: $\mu_{\mathrm{g}(X,Y)} = \mathrm{E}[\mathrm{g}(X, Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f(x, y)\, dx\, dy$

확률변수의 평균

- 결합확률분포의 기대값 Practice (1/2)
  - X와 Y는 아래 표와 같은 결합확률분포를 따르는 확률변수일 때, $\mathrm{g}(X,Y)=XY$의 기대값을 구하라
    - $\mathrm{E}[\mathrm{g}(X,Y)]=\sum_{x=0}^{2}\sum_{y=0}^{2}xyf(x,y)=(0)(0)f(0,0)+(0)(1)f(0,1)+(1)(0)f(1,0)+(1)(1)f(1,1)+(2)(0)f(2,0)+(0)(2)f(0,2)=\boldsymbol{f(1,1)}=\frac{\mathbf{3}}{\mathbf{14}}$

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

확률변수의 평균

- 결합확률분포의 기대값 Practice (2/2)
  - 아래와 같은 결합밀도함수에 대해$\mathrm{E}(Y|X)$의 값을 구하라
  - $f(x,y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2,\ 0 < y < 1 \\ 0, & \text{다른구간} \end{cases}$
    - $E\left(\frac{Y}{X}\right) = \int_0^1 \int_0^2 \left(\frac{y}{x} \times \frac{x(1+3y^2)}{4}\right) dx\, dy = \int_0^1 \frac{y+3y^3}{2}\, dy = \frac{5}{8}$

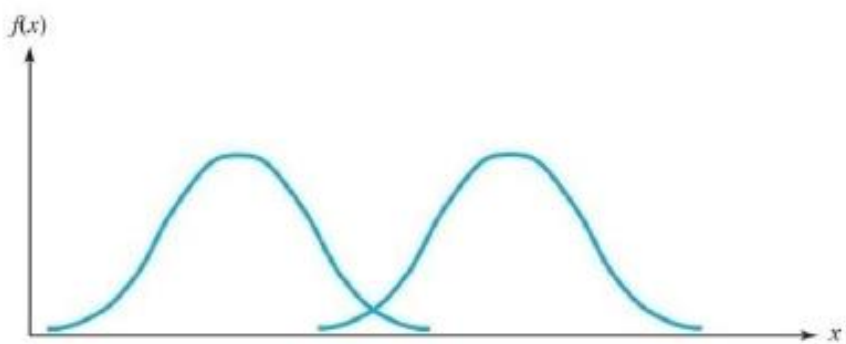
분산과 공분산

- 분산(Variance)
  - 확률변수가 기대값으로부터 얼마나 떨어진 곳에 분포하는지를 가늠하는 숫자
  - 분산은 분포의 형태로 나타남

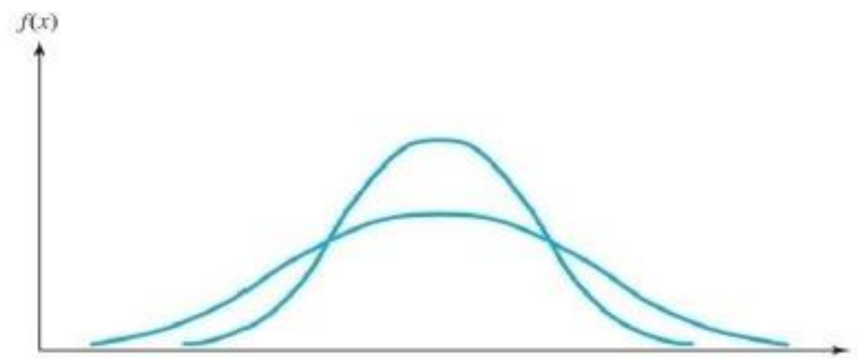
  - 표현
    - 확률변수의 기대값: μ
    - 확률변수 X의 분산: σ^2
    - 확률변수 X 의 표준편차: σ

분산과 공분산

- 평균과 분산의 관계 설명 그림
  - 평균: 중심 위치, 분산: 산포 정도



분산이 같고 평균이 다른 두 분포



평균이 같고 분산이 다른 두 분포

분산과 공분산

- 분산의 정의
  - X를 확률분포 $f(x)$와 평균 를 가지는 확률변수라고 할 때, X의 분산
    - $x - \mu$: 측정된 평균으로부터의 편차(Deviation)이며, 위 식은 편차 제곱의 평균을 의미
    - μ(평균)에 가까울수록 분산의 값이 작음

    - 이산형: $\sigma^2 = E\left[(X - \mu)^{\,2}\right] = \sum (x - \mu\ )^2 f(x)$
    - 연속형: $\sigma^2 = E\left[(X - \mu)^{\,2}\right] = \int_{-\infty}^{\infty} (x - \mu\ )^2 f(x)\, dx$

분산과 공분산

- 분산의 정의
  - 분산의 양의 제곱근 μ를 X의 표준편차(Standard deviation) 이라고 함
  - 좀 더 간단한 공식
    - $\boldsymbol{\sigma^2 = \mathrm{E}[X^2] - \mu^2}$
    - 증명

$$\sigma^2 = \sum_x (x-\mu)^2 f(x) = \sum_x (x^2 - 2\mu x + \mu^2) f(x)$$
$$= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x).$$

Since $\mu = \sum_x x f(x)$ by definition, and $\sum_x f(x) = 1$ for any discrete probability distribution, it follows that

$$\sigma^2 = \sum x^2 f(x) - \mu^2 = E(X^2) - \mu^2.$$

분산과 공분산

- 확률변수 $g(X)$의 분산
  - 정의
    - X를 확률분포 $f(x)$를 갖는 확률변수일 때, 확률변수 $g(x)$의 분산
      - 이산형: $\sigma^2 = E\{[g(X) - \mu_{g(X)}]^2\} = \sum [g(X) - \mu_{g(X)}]^2 f(x)$
      - 연속형: $\sigma^2 = E\{[g(X) - \mu_{g(X)}]^2\} = \int_{-\infty}^{\infty} [g(X) - \mu_{g(X)}]^2 f(x)\, dx$

분산과 공분산

- 확률변수 $g(X)$의 분산 Practice
  - 생산라인으로부터 3개의 부품을 추출하여 검사, 확률변수 X는 결함이 있는 부품의 수이며, X의 확률분포가 아래 표를 따를 때, 분산(σ^2)은?

x	0	1	2	3
$f(x)$	0.51	0.38	0.10	0.01

- 평균(μ)
  - $(0)(0.51) + (1)(0.38) + (2)(0.10) + (3)(0.01) = 0.61$
- $\mathrm{E}(X^2)$
  - $(0)(0.51) + (1)(0.38) + (4)(0.10) + (9)(0.01) = 0.87$
- σ^2
  - $\sigma^2 = \mathrm{E}[X^2] - \mu^2 = 0.87 - (0.61)^2 = \mathbf{0.4979}$

선형결합된 확률변수의 평균과 분산

체비셰프 정리

- 개요
  - 러시아 수학자 체비셰브(Chebyshev)가 발견
    - 평균을 중심으로 대칭인 두 값 사이의 면적이 표준편차와 관계되어있음을 발견
  - 개념
    - 분산(σ^2)은 평균(μ)을 중심으로 관측 값이 얼마나 변동하는지를 나타냄
    - 분산 또는 표준편차(σ)가 작으면 대부분의 값이 평균 근처에 밀집됨
  - 확률변수가 평균을 중심으로 특정 구간내의 값을 나타낼 확률은 표준편차가 작을수록 큼

감사합니다!