통계의 이해와 활용 2차 과제

소프트웨어학과 2021041017 김규현

[1] 베이즈정리에 대한 아래 3문제를 풀이 과정을 상세히 하여 풀어볼 것.

예1) (사전확률 정보) 발병률이 0.05인 A질병이 있다고 하자. 실제로 A병에 걸린 사람이 검사하면 0.90의 비율로 양성판정, 병에 걸리지 않은 사람이 검사하면 0.10의 비율로 양성판정으로 나온다고 한다. 그렇다면 이 질병 검사에서 양성판정이 나왔을 때, (사후확률 추정) 검사 대상자가 진짜 질병에 걸렸을 확률은?

Sol)

A : A질병에 걸린 사건, B : 양성판정이 나온 사건

P(A) = 0.05, P(A^c) = 0.95, P(B|A) = 0.90, P(B|A^c) = 0.10

양성판정이 나왔을 때, 진짜 질병에 걸렸을 확률: P(A|B)

P(B) = P(B|A) * P(A) + P(B|A^c) * P(A^c) = (0.90 * 0.05) + (0.10 * 0.95) = 0.045 + 0.095 = 0.14

P(A|B) = P(B|A) * P(A) / P(B) = (0.90 * 0.05) / (0.14) = 0.32142857...

∴ 약 32.14%

예2) 실험자 학생이 거짓말을 하는지 알아보기 위해 거짓말 탐지기를 사용하기로 했다. (사전확률 정보) 학생이 거짓말을 할 확률이 0.25이고, 진실을 말할 확률은 0.75이다. 그리고 경험적으로 거짓말 탐지기의 작동이 맞을 확률은 0.85이고 틀릴 확률은 0.15이다. (사후확률 추정) 거짓말 탐지기가 거짓말이라고 응답했을 때, 실제로 이 학생이 진실을 말했을 확률은 얼마인가?

Sol)

A: 학생이 진실을 말한 사건, B: 거짓말 탐지기가 거짓말이라고 응답

P(A) = 0.75, P(A^c) = 0.25, P(B|A) = 0.15, P(B|A^c) = 0.85

거짓말 탐지기가 거짓말이라고 응답했을 때, 이 학생이 진실을 말할 확률: P(A|B)

P(B) = P(B|A) * P(A) + P(B|A^c) * P(A^c) = (0.15 * 0.75) + (0.85 * 0.25) = 0.1125 + 0.2125 = 0.325

P(A|B) = P(B|A) * P(A) / P(B) = (0.15 * 0.75) / 0.325 = 0.34621538...

∴ 약 34.62%

예3) 한 대학교 재학생의 35%는 여학생이다. (사전정보 확률) 이 학교의 남학생이 70%, 여학생의 20%가 흡연을 한다고 조사되었다. (사후정보 추정) 이 학교의 학생 한 명을 임으로 선발하였을 때, 선출된 학생이 흡연자라면, 이 학생이 여학생일 확률은 얼마인가?

Sol)

A: 선출된 학생이 여학생일 사건, B: 선출된 학생이 흡연자일 사건

P(A) = 0.35, P(A^c) = 0.65, P(B|A) = 0.20, P(B|A^c) = 0.70

선출된 학생이 흡연자라면, 여학생일 확률: P(A|B)

P(B) = P(B|A) * P(A) + P(B|A^c) * P(A^c) = (0.20 * 0.35) + (0.70 * 0.65) = 0.07 + 0.455 = 0.525

P(A|B) = P(B|A) * P(A) / P(B) = (0.20 * 0.35) / 0.525 = 0.13333333...

∴ 약 13.33%

[2] 가구당 보유하고 있는 TV 수 X와 자동차 수 Y를 조사하기 위하여 서울시에서는 100가구를 랜덤으로 뽑아 다음과 같은 자료를 얻었다.

y / x	1	2	3	4
0	20	15	10	0
1	10	20	7	3
2	4	5	4	2

1) X, Y의 결합확률분포를 구하여라.

Sol)

P(X,Y)	P(X=1)	P(X=2)	P(X=3)	P(X=4)
P(Y=0)	0.20	0.15	0.10	0.00
P(Y=1)	0.10	0.20	0.07	0.03
P(Y=2)	0.04	0.05	0.04	0.02

2) X, Y의 주변확률분포를 구하여라.

Sol)

X의 주변확률분포: $P(X = x) = \sum_y P(X = x,\ Y = y)$

P(X=1) = 0.20 + 0.10 + 0.04 = 0.34

P(X=2) = 0.15 + 0.20 + 0.05 = 0.40

P(X=3) = 0.10 + 0.07 + 0.04 = 0.21

P(X=4) = 0.00 + 0.03 + 0.02 = 0.05

Y의 주변확률분포: $P(Y=y) = \sum_x P(X=x,\ Y=y)$

P(Y=0) = 0.20 + 0.15 + 0.10 + 0.00 = 0.45

P(Y=1) = 0.10 + 0.20 + 0.07 + 0.03 = 0.40

P(Y=2) = 0.04 + 0.05 + 0.04 + 0.02 = 0.15

3) E(XY), E(X), E(Y)를 각각 구하여라.

Sol)

E(X) = (1 * 0.34) + (2 * 0.40) + (3 * 0.21) + (4 * 0.05) = 1.97

E(Y) = (0 * 0.45) + (1 * 0.40) + (2 * 0.15) = 0.70

E(XY) = (1 * 0 * 0.20) + (1 * 1 * 0.010) + (1 * 2 * 0.04) + (2 * 0 * 0.15) + (2 * 1 * 0.20) + (2 * 2 * 0.05) + (3 * 0 * 0.10) + (3 * 1 * 0.07) + (3 * 2 * 0.04) + (4 * 0 * 0.00) + (4 * 1 * 0.03) + (4 * 2 * 0.02)

= 0 + 0.010 + 0.08 + 0 + 0.40 + 0.20 + 0 + 0.21 + 0.24 + 0 + 0.12 + 0.16 = 1.51

4) Cov(X,Y)를 구하여라.

Sol)

Cov(X,Y) (공분산) = E(XY) – E(X)E(Y) = 1.51 – (1.97 * 0.70) = 1.51 – 1.379 = 0.131

~~[3] 어떤 교차로에서 발생하는 교통사고 건수 X는 일주일에 평균 3.5건인 포아송분포를 따른다고 한다. 다음 확률을 구하여라.~~

~~1) 일주일에 사고가 한 건도 없을 확률?~~

~~2) 일주일에 4건 이상의 사고가 발생할 확률?~~

~~3) 오늘 한 건의 사고가 발생할 확률?~~

※ 포아송분포 문제는 진도 문제로 생략.

[4] 어느 동아리 회원 수가 20명이라 하자. 이 중 8명은 안경을 쓰고 있다. 회원들 중 4명을 임의로 추출한다고 했을 때, 4명 중 안경을 쓴 학생의 수를 확률변수 X라 하자.

1) 확률변수 X의 확률질량함수를 정의하고, 확률분포를 구하여라.

Sol)

모집단(population) N = 20, 안경을 쓴 학생 수 K = 8, 표본 크기(sample size) n = 4

확률변수 X: 임의로 추출한 4명 중 안경을 쓴 학생의 수

확률질량함수(pmf) = $P(X = x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$

P(X=0) = $\frac{\binom{8}{0}\binom{12}{4}}{\binom{20}{4}} = \frac{1 * 495}{4845}$ ≈ 0.1022

P(X=1) = $\frac{\binom{8}{1}\binom{12}{3}}{\binom{20}{4}} = \frac{8 * 220}{4845}$ ≈ 0.3634

P(X=2) = $\frac{\binom{8}{2}\binom{12}{2}}{\binom{20}{4}} = \frac{28 * 66}{4845}$ ≈ 0.3817

P(X=3) = $\frac{\binom{8}{3}\binom{12}{1}}{\binom{20}{4}} = \frac{56 * 12}{4845}$ ≈ 0.1387

P(X=4) = $\frac{\binom{8}{4}\binom{12}{0}}{\binom{20}{4}} = \frac{70 * 1}{4845}$ ≈ 0.0145

2) 확률변수 X의 평균과 분산을 각각 구하여라.

Sol)

평균: E(X) = n * $\frac{K}{N}$= 4 * 8 / 20 = 1.6

분산: Var(X) = E(X) * $\frac{N-K}{N}$ * $\frac{N-n}{N-1}$ = 1.6 * (12 / 20) * (16 / 19) ≈ 0.808

[5] 원통 안에 1에서 10까지 숫자가 쓰여 있는 공이 있고 3개를 추첨하여 모두 맞으면 1등, 2개를 맞추면 2등, 1개를 맞추면 3등이 정해진다. 1과 10사이에서 임의의 숫자 3개를 고른 후 추첨에서 3개의 공을 임의로 뽑을 때 2개 이상 맞출 확률을 구하여라.

1) 추첨방식이 복원방법일 경우 확률을 구하라. => 독립

Sol)

당첨 확률 p = $\frac{3}{10}$, 시행 횟수 n = 3

P(X=3) = $(\frac{3}{10})^3$ = 0.027, P(X=2) = ${}_3C_2$ * $(\frac{3}{10})^2$ * $(\frac{7}{10})^1$ = 3 * 0.09 * 0.7 = 0.189

P(X ≥ 2) = P(X=2) + P(X=3) = 0.189 + 0.027 = 0.216

∴ 21.6%

2) 추첨방식이 비복원방법일 경우 확률을 구하라. => 종속

Sol)

첫 번째 당첨 확률: $\frac{3}{10}$, 두 번째 당첨 확률: $\frac{2}{9}$, 세 번째 당첨 확률: $\frac{1}{8}$

P(X=3) = $\frac{3}{10} * \frac{2}{9} * \frac{1}{8}$ = $\frac{6}{720}$ = 0.00833

P(X=2) (2개는 당첨, 1개는 불당첨) = ${}_3C_2$ * $\frac{3}{10} * \frac{2}{9} * \frac{7}{8}$ = $3 * \frac{42}{720}$ = 0.175

P(X ≥ 2) = 0.00833 + 0.175 = 0.18333

∴ 18.333%

3) 1등은 10만원, 2등은 5만원, 3등은 1만원의 상금이 주어진다면, 기대수익은 얼마이겠는가?

Sol)

복원방식에서의 기댓값(E_1):

P(X=1) = 1 – P(X ≥ 2) = 0.784

E_1 = 0.027 * 100000 + 0.189 * 50000 + 0.784 * 10000 = 2700 + 9450 + 7840 = 19900원

∴ 복원방식에서의 기대수익: 19900원

비복원방식에서의 기댓값(E_2)

P(X=1) = 1 – P(X ≥ 2) = 0.81667

E_2 = 0.00833 * 100000 + 0.175 * 50000 + 0.81667 * 10000 = 17750원

∴ 비복원방식에서의 기대수익: 17750원