

Probability and Statistics / 확률과 통계

강의노트 05

**확률 3**

47. 베이즈의 정리, 잘못된 양성반응 패러독스

Question :

인구 천명당 한명꼴로 걸리는 희귀병이 있다. 이 병에 걸렸는지를 판단하는 진단법이 있다. 병을 가진 사람이 이 진단법을 시행했을 때는 99%로 양성반응을 나타낸다. 반면 병을 가지지 않은 사람이 이 진단법에 따를 경우 2%만이 양성반응을 보인다. 만약 당신이 이 진단법에 따라 검사를 했고 그 결과로 양성반응이 나왔다면 여러분이 병에 걸려있을 확률은 얼마인가?

Answer :

사건A : 피검자가 병에 걸려 있다.

사건B : 피검자가 양성반응을 보인다.

$$P[A] = .001$$

$$P[B | A] = .99$$

$$P[B | \text{not } A] = .02$$

$$P[A | B] = ?$$

이 질병을 치료하는데는 심각한 부작용이 나타날 수 있다. 그래서 의사는 변호사와 함께 확률학자인 조 베이즈를 방문했다. 이 문제는 토머스 베이즈(조 베이즈의 선조, 1744~1809)가 최초로 증명했던 정리를 사용하여 풀어내었다.

주의 방법

표본 공간을 4개의 배반사건으로 나눈다.

	A	NOT A
B	A <b>AND</b> B	NOT A <b>AND</b> B
NOT B	A <b>AND</b> NOT B	NOT A <b>AND</b> NOT B

표에 있는 각 사건의 확률을 찾는다.

	A	NOT A	합계
B	P[A <b>AND</b> B]	P[NOT A <b>AND</b> B]	P[B]
NOT B	P[A <b>AND</b> NOT B]	P[NOT A <b>AND</b> NOT B]	P[NOT B]
합계	P[A]	P[NOT A]	1

계산하면,

$$\begin{aligned} P[A \text{ AND } B] &= P[B|A]P[A] \\ &= .99 \times .001 \\ &= .00099 \end{aligned}$$

$$\begin{aligned} P[\text{NOT } A \text{ AND } B] &= P[B|\text{NOT } A]P[\text{NOT } A] \\ &= .02 \times .999 \\ &= .01998 \end{aligned}$$

계산결과를 정리하면,

	A	NOT A	합계
B	.00099	.01998	.02098
NOT B	P[A <b>AND</b> NOT B]	P[NOT A <b>AND</b> NOT B]	P[NOT B]
합계	.001	.999	1

남은 부분을 구한다.

$$P[A \text{ AND } \text{NOT } B] = .001 - .00099 = .00001$$

$$P[\text{NOT } A \text{ AND } \text{NOT } B] = .999 - .01998 = .97902$$

최종적인 표

	A	NOT A	합계	
B	.00099	.01998	.02098	P[B]
NOT B	.00001	.97902	.97903	P[NOT B]
합계	.001	.999	1	
	P[A]	P[NOT A]		

위 표를 이용하면,

$$P[A|B] = \frac{P[A \text{ AND } B]}{P[B]} = .0472$$

.0472 의 의미는?

-> 양성반응이 나온 사람 중 4.72%가 병에 걸려 있다. 즉, 1000명이 검사를 받았을 때 그중 21명이 양성반응이 나오고, 그 21명중 한명만이 병을 가지고 있다. - 잘못된 양성반응 패러독스

	감염자	비감염자	
양성반응	1	20	21
음성반응	0	979	979
	1	999	1000

결과:치료는 잘못된 부작용이 심각하므로 양성반응을 보인 사람들에 대해서 정밀 검사를 시행(의사)

베이즈의 정리

$$P[A|B] = \frac{P[B|A] \cdot P[A]}{P[B|A] \cdot P[A] + P[B|NOT A] \cdot P[NOT A]}$$

정리 :

1. 알고자 하는 것 : P[A|B]
2. 알고 있는 것 : P[A], P[B|A], P[B|NOT A]

48. 베이즈의 정리 일반식

$$P[A_j|B] = \frac{P[B|A_j] \cdot P[A_j]}{\sum_{i=1}^n P[B|A_i] \cdot P[A_i]}$$

49. 베이즈의 정리 예제

태아의 기형여부를 검사하려는 목적으로 이루어지는 트리플테스트는 산모의 혈액속에 있는 3가지(triple)의 표지 물질(알파태아단백 (alpha-fetoprotein), 비포합성 에스트리올 (unconjugated estriol), 융모성 성선자극호르몬 (human chorionic gonadotropin))의 농도를 통해 다운증후군을 비롯한 태아 기형들을 간접적으로 테스트하는 방법이다.

통계적으로 1만명의 산모에서 태어난 태아 중 다운증후군인 태아는 13명이다. 태아가 다운증후군일 때 산모가 트리플테스트에 양성반응이 나올 확률은 70% 이고, 태아가 다운증후군이 아닐 때 양성반응이 나올 확률은 4.92% 이다. 만약 어떤 산모가 트리플테스트 결과 다운증후군 양성진단을 받았다면 그때 태아가 다운증후군일 확률은?

Answer :

- P[A] : 태아가 다운증후군일 확률
- P[B] : 트리플테스트에 양성일 확률
- P[A|B] = ?

$$P[A] = .0013 \quad (P[\text{not}A]=1-P[A]=.9987)$$

$$P[B|A] = .7$$

$$P[B|\text{not} A] = .0492$$

$$P[A|B] = \frac{P[B|A]P[A]}{P[B|A]P[A] + P[B|\text{not}A]P[\text{not}A]} = \frac{.7 * .0013}{.7 * .0013 + .0492 * .9987} = .00185$$

※ 1.8% 의 태아가 다운증후군을 가진다.