

수DA쟁이

헤드퍼스트 DA - chap 6  
베이지안 통계  
(첫걸음을 내딛다.)

Department of Mathematics  
Gyeongsang National University  
Youngmin Shin

## 1. 의사가 충격적인 소식을 전해주었습니다.

- 도마뱀 독감에 걸려버렸다! (양성 판정, 엉덩이에 상당한 통증  $\pi\pi$ )
- 이 검사가 얼마나 정확한지 확인 하고 싶다.

## 1. 의사가 충격적인 소식을 전해주었습니다.

### 통계 결과

- 도마뱀 독감에 걸렸을 경우 이 검사에서 양성 판정이 나올 확률: 90%
- 도마뱀 독감에 걸리지 않았을 경우 이 검사에서 양성 판정이 나올 확률: 9%

## 2. 정확도 분석을 한번에 하나씩 해봅시다.

### 시나리오 1.

- 100명 중 90명이 이 병에 걸렸을 경우, 걸리지 않은 사람이 이 검사에서 양성 판정을 받는 경우

→ 10명이 이 병에 걸리지 않음. 이 중에 양성 판정을 받았지만, 실제로 걸리지 않는 사람은 10명의 9%인 1명

- 도마뱀 독감에 걸리지 않았을 경우 이 검사에서 양성 판정이 나올 확률: 9%

## 2. 정확도 분석을 한번에 하나씩 해봅시다.

### 시나리오 2.

- 100명 중 10명이 이 병에 걸렸을 경우, 걸리지 않은 사람이 이 검사에서 양성 판정을 받는 경우

→ 90명이 이병에 걸리지 않았음. 양성 판정을 받았지만 실제로 걸리지 않은 사람은 90명 중의 9%인 8명

- 도마뱀 독감에 걸리지 않았을 경우 이 검사에서 양성 판정이 나올 확률: 9%

### 3. 여러분은 거짓 양성을 세고 있습니다.

- 진짜 양성: 질병에 걸린 사람이 양성으로 결과가 나옴
- 거짓 양성: 질병에 걸리지 않은 사람이 양성으로 결과가 나옴
- 진짜 음성: 질병에 걸리지 않은 사람이 음성으로 결과가 나옴
- 가짜 음성: 질병에 걸린 사람이 음성으로 결과가 나옴

### 3. 여러분은 거짓 양성을 세고 있습니다.

- 진짜 양성: 질병에 걸린 사람이 양성으로 결과가 나옴
- 거짓 양성: 질병에 걸리지 않은 사람이 양성으로 결과가 나옴
- 진짜 음성: 질병에 걸리지 않은 사람이 음성으로 결과가 나옴
- 가짜 음성: 질병에 걸린 사람이 음성으로 결과가 나옴

거짓 양성 ↔ 진짜 음성  
(거짓 양성: 9%)  
(진짜 음성: 91%)

진짜 양성 ↔ 가짜 음성  
(진짜 양성: 90%)  
(가짜 음성: 10%)

#### 4. 여기 나오는 모든 용어는 조건부 확률을 나타냅니다.

- 조건부 확률: 다른 어떤 사건이 발생했을 경우에 사건이 일어날 확률

$P(\text{확률}|\text{조건})$



#### 4. 여기 나오는 모든 용어는 조건부 확률을 나타냅니다.

$+$ : 양성 판정

$-$ : 음성 판정

$L$ : 독감에 걸림

$\sim L$ : 독감에 걸리지 않음

- $P(+|\sim L)$ : 도마뱀 독감에 걸리지 않았을 경우, 검사에서 양성판정이 나올 확률.
- $P(+|L)$ : 도마뱀 독감에 걸린 경우, 검사에서 양성판정이 나올 확률.
- $P(-|L)$ : 도마뱀 독감에 걸린 경우, 검사에서 음성판정이 나올 확률.
- $P(-|\sim L)$ : 도마뱀 독감에 걸리지 않았을 경우, 검사에서 음성판정이 나올 확률.

## 5. 1%의 사람이 도마뱀 독감에 걸려있습니다.

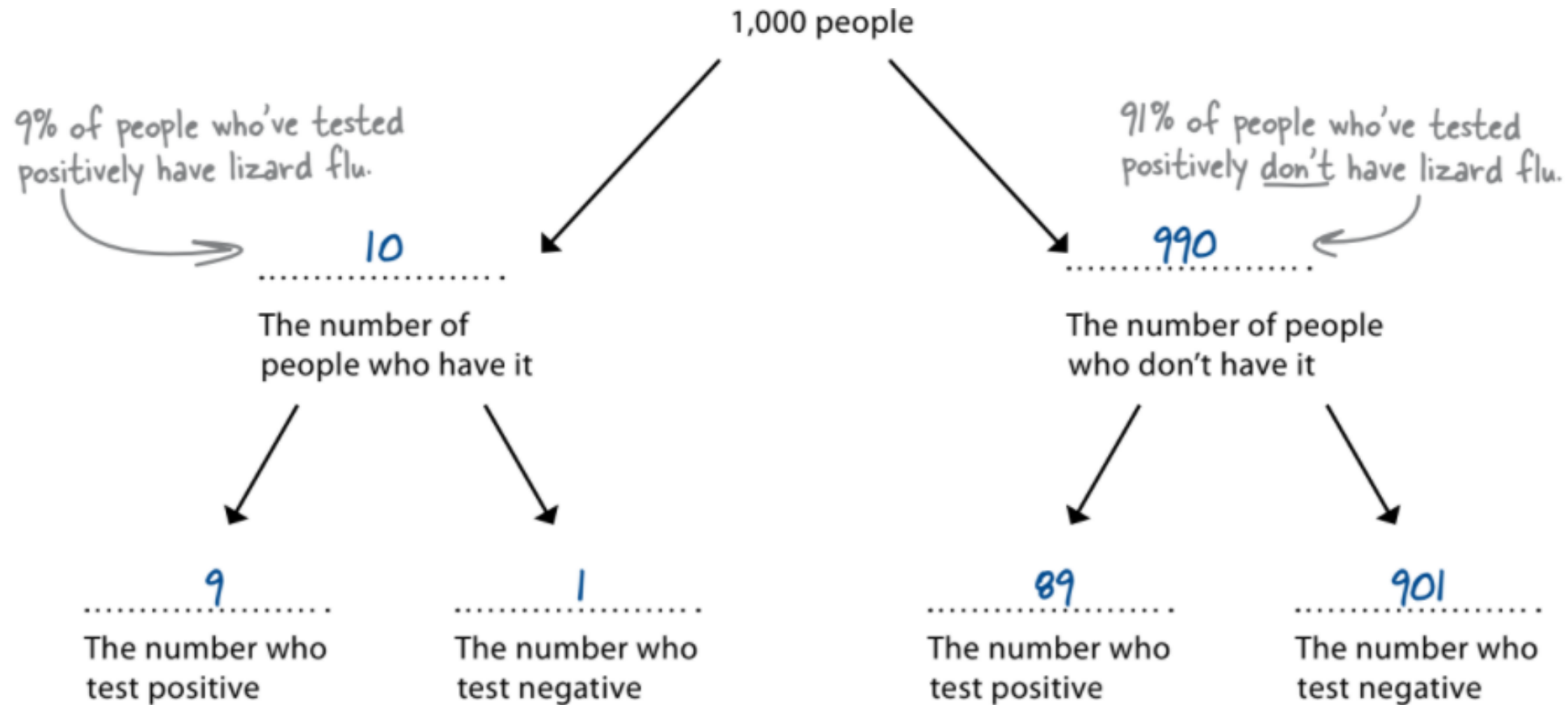
- 조사를 통해 국내 인구의 1%가 도마뱀 독감에 걸려있음을 알아냄.
- ✓ 1%는 기준 비율
- ✓ 검사를 위해 개개인에 대한 새로운 사실을 알기 전에 모집단의 1%만이 도마뱀 독감에 걸려 있음을 알고 있습니다. 따라서 기준 비율을 **사전 비율**이라고 합니다.

## 5. 1%의 사람이 도마뱀 독감에 걸려있습니다.

- **기준 비율 오류를 조심하세요**

- ✓ 진짜 양성 비율이 90%이기 때문에 독감에 걸렸을 가능성이 꽤 높다는 것을 의미 한다고 착각 할 수 있지만, 기준 비율이 1%라는 것을 알면 진짜 양성 비율이 90%나 되는 것은 검사 결과에 별로 설득력이 없이 진다.

## 5. 1%의 사람이 도마뱀 독감에 걸려있습니다.



검사에서 양성 판정을 받았을 경우 독감에 걸렸을 확률=

독감에 걸렸고 결과가 양성인 사람 수

독감에 걸렸고 검사결과가 양성인 사람의 수 + 독감에 걸리지 않았지만 검사 결과가 양성인 사람 수

$$= 9 / (9 + 89)$$

$$= 0.09$$

9

The number who  
test positive

1

The number who  
test negative

89

The number who  
test positive

901

The number who  
test negative

기준 비율: 1%

내가 독감에 걸렸을 확률: 9% (일반인에 비해 도마뱀에 걸렸을 가능성이 9배!)

## 6. 간단한 정수를 사용해 복잡한 확률에 대해 생각해 봅시다.

- 우리 두뇌는 원래 소수의 확률을 처리하는데 별로 적합하지 않기 때문에 확률은 정수로 반환하고 생각하면 실수를 줄일 수 있어서 효과적입니다.

## 7. 베이즈 정리는 새로운 데이터가 들어왔을 때 기준 비율을 계산합니다.

- 베이즈 정리는 기준비율과 조건부 확률을 사용하여 새로운 조건부 확률을 예측할 수 있게 해주는 강력한 통계학적 공식

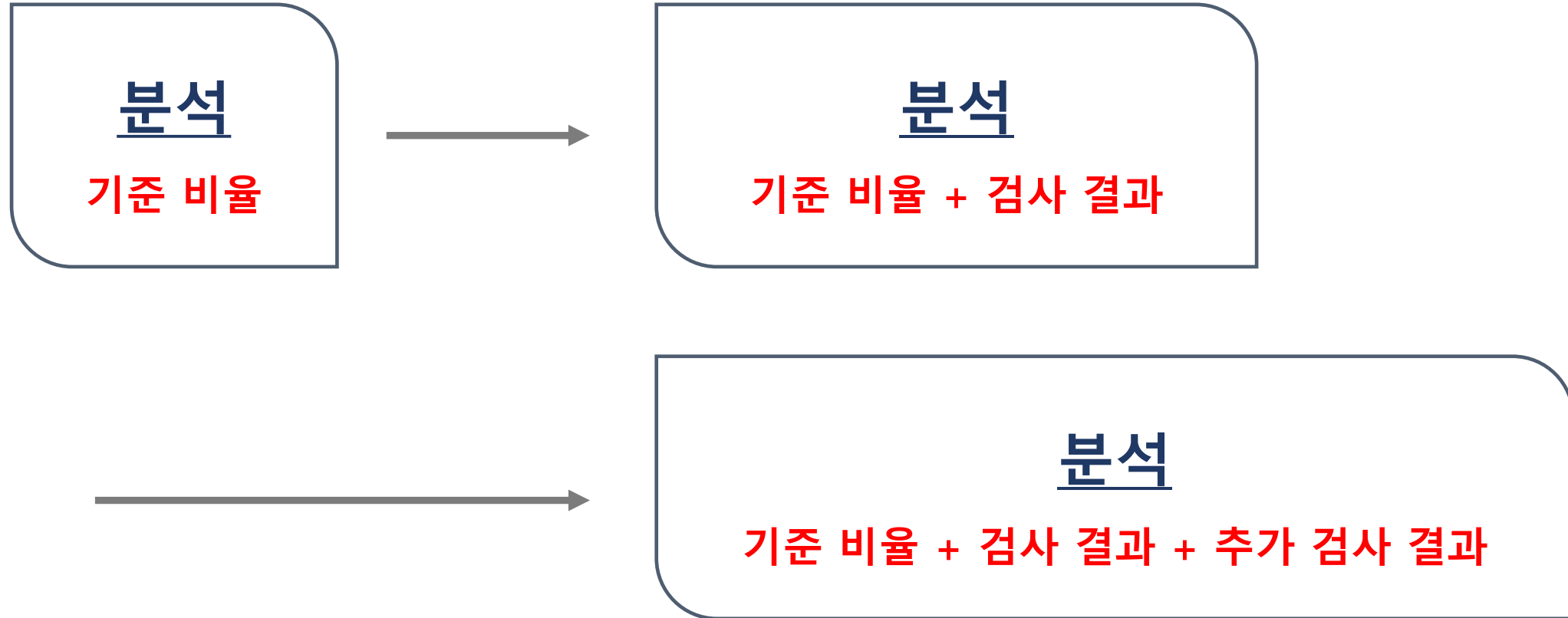
$$P(L|+) = \frac{P(L)P(+|L)}{P(L)P(+|L) + P(\sim L)P(+|\sim L)}$$

## 8. 베이지 정리는 반복해서 사용할 수 있습니다.

- 베이지 정리는 새로운 정보를 분석에 포함 시키는 방법을 제공하기 때문에 분석할 때 매우 중요하게 사용됩니다.



## 8. 베이즈 정리는 반복해서 사용할 수 있습니다.



9. 두번째 검사 결과는 음성입니다.

음성!

## 10. 새로운 검사에서는 다른 정확도 통계가 나왔습니다.

- 도마뱀 독감 진단 검사 → 고급 도마뱀 독감 진단 검사
- 이전과 동일한 기준비율을 사용하면 안된다.

## 11. 새로운 정보로 기준 비율이 바뀔 수 있습니다.

- 첫 번째 검사 결과는 모집단의 발생률을 기준 비율로 사용.
- 검사에서 도마뱀에 걸렸을 확률을 새로운 기준비율로 사용한다. 왜냐하면, 양성 판정을 받은 사람이 그룹이 되기 때문이다.

**새로운 기준비율: 9%**

Thank you!