

실습 2: 확률과 베이지 정리

<https://github.com/wcjang/hhi>

➔ Day 1 -> 실습자료

2024chapter2_1.ipynb,

실습2_1.PDF 파일 다운로드

loans_income.csv 다운로드

Sp500_data.csv.gz 다운로드.

랜덤이란?

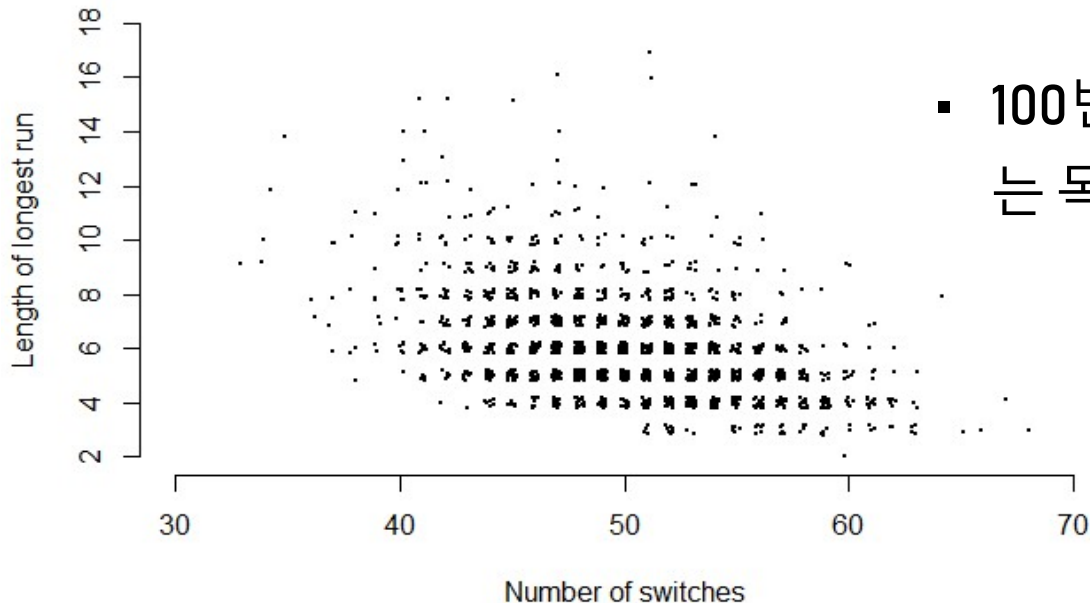
- 동전 던지기 실험을 통해서 랜덤에 대해서 알아보자.
- 3명이 하나의 조를 만들어서 다음 실험을 진행한다.
- 한명은 동전을 100번 던진다. 이때 다른 한명은 앞면 (이순신이 나온 면)이 나오면 0, 뒷면이 나오면 1이라고 표기하고 결과를 10개의 줄로 정리하여라. 각 줄에는 10개의 0과 1로 이루어진다.
- 마지막 한명은 동전던지기를 결과를 마음속으로 상상하여 100개의 0과 1을 생성한 후 같은방식으로 결과를 정리한다.
- 종이 뒷면에 작게 어느 것이 랜덤이고 어느것이 마음속으로 생각한것인지 자그맣게 적고,
- 옆 조와 결과지 를 교환하여, 무엇이 랜덤인지 맞추어보자.

랜덤이란?

- 둘 중 어느 것이 진짜 랜덤인지 맞추어 보자.

동전던지기 패턴 맞추기

- 왼쪽 그림은 100번의 동전 던지기를 2000번 반복한 결과를 정리한 것이다.
- 앞면과 뒷면이 서로 전환되는 횟수가 50-60 정도 일어나는 경우가 제일 많고 또 같은 결과가 연달아 나오는 것도 최대 6-7회 정도 되는 경우가 빈번하다는 것을 알 수 있다.
- 100번 던질 경우 앞면이 연달아 6번 나올 경우는 목격할 경우는 얼마나 될까?



죄수의 딜레마

- 3명의 사형수 A, B, C가 독방에 갇혀있다. 대통령이 이 중 한명을 임의로 선정해서 특별사면하기로 결정했다. 간수는 누가 사면이 될지 알고 있지만 누구인지 누설하면 안된다. A가 간수에게 B와 C중 누가 사형이 되는지 알려달라고 사정했다.
- A는 간수에게 “B가 사면대상자이면 C의 이름을 얘기하고 C가 사면대상자이면 B의 이름을 얘기하고 만약 내가 사면대상자라면 동전을 던져서 앞면이 나오면 B, 뒷면이 나오면 C라고 얘기해달라”라고 애원한다. 간수가 A에게 B가 처형될 것이라고 얘기해주었다.

죄수의 딜레마

- A는 이 사실은 안 후 본인이 사면될 확률이 $1/3$ 에서 $1/2$ 로 증가하였다고 믿는다. 사실인가?
- A는 이 사실을 우연히 만난 C에게 알려준다. 이 사실을 안 후 C가 사면될 확률은 얼마인가?

Monty Hall과 죄수의 딜레마

- Monty Hall와 죄수의 딜레마의 유사점에 대해 생각해보자.
 - 자동차 = 사면
 - 염소 = 처형
 - 몬티 홀 = 간수
 - 염소가 뒤에 있다고 알려준 문 = 처형대상자의 이름
 - 처음 선택한 문 = 죄수 A
 - 문을 바꾸는 경우 = 죄수 C에게 정보를 알려준 경우

죄수의 딜레마

- A는 이 사실은 안 후 본인이 사면될 확률이 $1/3$ 에서 $1/2$ 로 증가하였다고 믿는다. 사실인가?
- 간수가 B가 처형대상자라고 얘기할 수 있는 경우는 (1) C가 사면대상자이거나 (2) A가 사면대상자이지만 동전을 던져서 앞면이 나온 경우이다. 따라서

$$\Pr(\text{A가 사면될 사건} | \text{간수가 B가 처형대상자라고 얘기한 사건}) = \frac{\frac{1}{3} \times \frac{1}{2}}{\frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2}} = \frac{1}{3}$$

- A는 이 사실을 우연히 만난 C에게 알려준다. 이 사실을 안 후 C가 사면될 확률은 얼마인가?

$$\Pr(\text{C가 사면될 사건} | \text{간수가 B가 처형대상자라고 A에게 얘기한 사건}) = \frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2}} = \frac{2}{3}$$

- 베이즈 정리:

$$\Pr(A_k | B) = \frac{\Pr(B | A_k) \cdot \Pr(A_k)}{\sum_{i=1}^n \Pr(B | A_i) \cdot \Pr(A_i)}$$

Pr(A가 사면될 사건 | 간수가 B가 처형대상자라고 얘기한 사건) =

P(A가 사면대상자이고, 간수가 B가 처형대상자라고 얘기한 사건) %
 { P(간수가 B가 처형대상자라고 얘기한 사건 | A가 사면될 사건) X P(A가 사면될 사건) +
 P(간수가 B가 처형대상자라고 얘기한 사건 | B가 사면될 사건) X P(B가 사면될 사건) +
 P(간수가 B가 처형대상자라고 얘기한 사건 | C가 사면될 사건) X P(C가 사면될 사건) }

$$= \frac{\frac{1}{3} \times \frac{1}{2}}{\frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2}} = \frac{1}{3}$$

Pr(C가 사면될 사건 | 간수가 B가 처형대상자라고 얘기한 사건) =

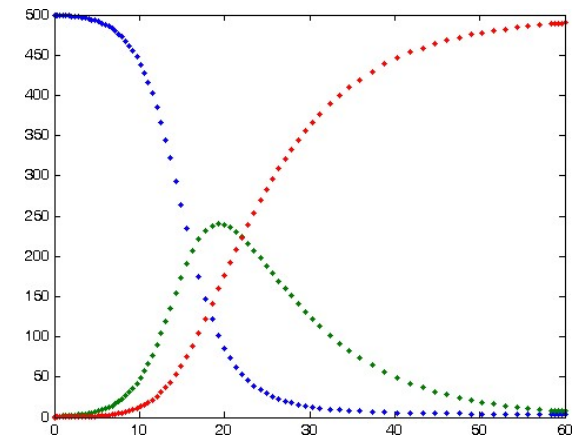
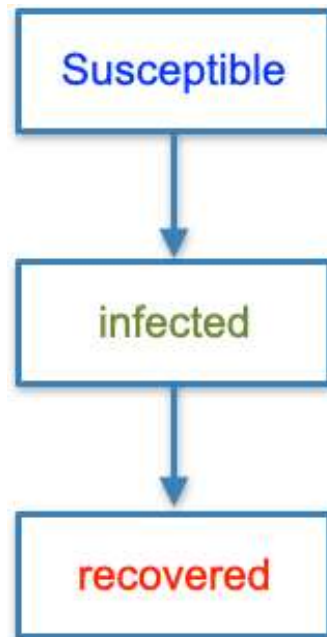
P(C가 사면대상자이고, 간수가 B가 처형대상자라고 얘기한 사건) %

{ P(간수가 B가 처형대상자라고 얘기한 사건 | A가 사면될 사건) X P(A가 사면될 사건)+
P(간수가 B가 처형대상자라고 얘기한 사건 | B가 사면될 사건) X P(B가 사면될 사건)+
P(간수가 B가 처형대상자라고 얘기한 사건 | C가 사면될 사건) X P(C가 사면될 사건) }

$$= \frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2}} = \frac{2}{3}$$

SIR model

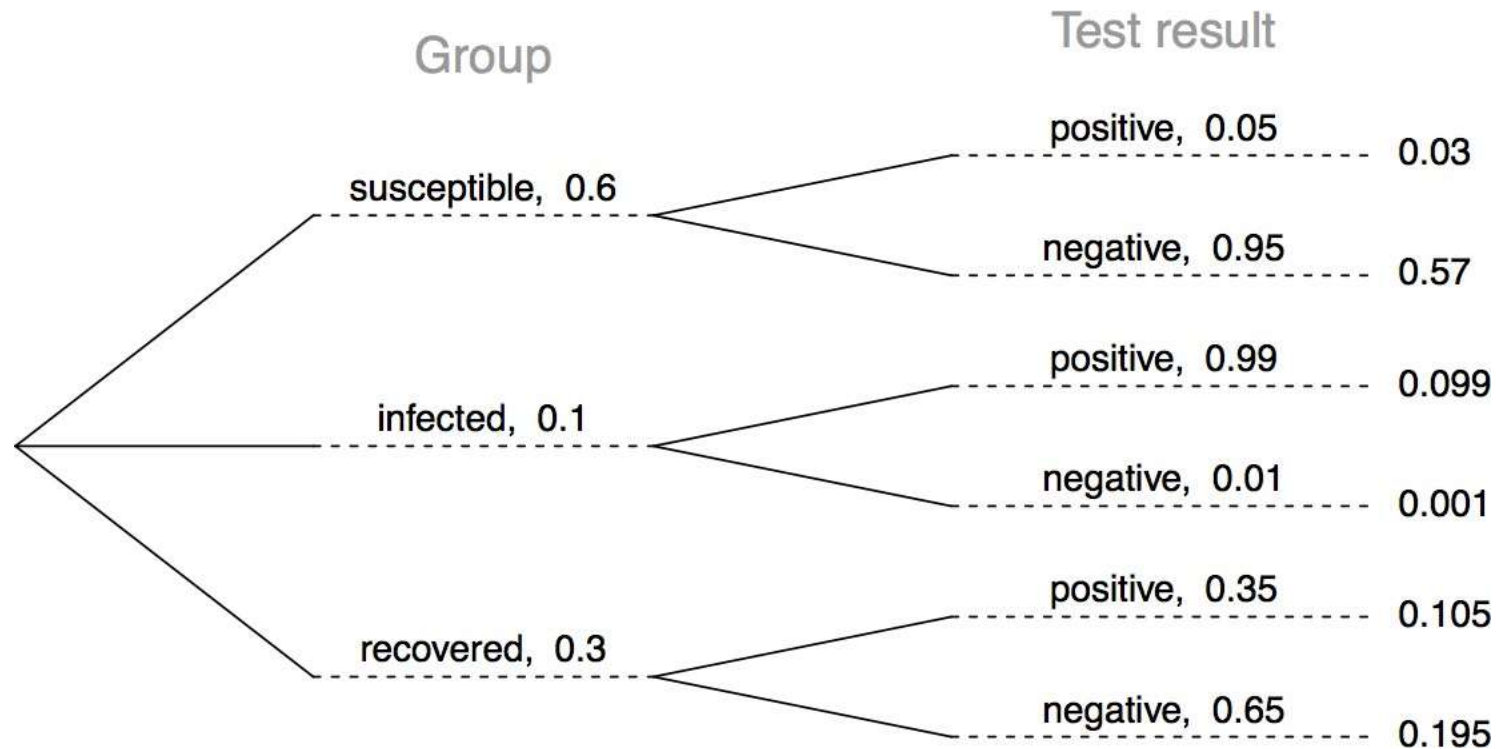
- SIR model은 ordinary differential equation model로서 역학 (epidemiology)에서 질병 전파를 설명하는 가장 보편적인 모형이다.
- SIR은 질병 전파 3단계인 susceptible, infected, recovered의 앞글자를 따온 것이다.



SIR model

- 신종플루 유행시즌중 60%의 인구가 susceptible, 10%는 infected, 30%는 recovered라고 가정하자.
- 신종플루 검사의 정확성은 다음과 같다.
- susceptible: 95% (음성 판정비율)
- infected: 99% (양성판정 비율)
- recovered: 65% (음성 판정비율)
- 만약 검사결과가 양성일 경우 실제로 신종플루에 감염되었을 확률을 구하여 보아라.

연습문제: SIR model



$$\Pr(\text{inf} \mid +) = \frac{\Pr(\text{inf} \cap +)}{\Pr(+)} = \frac{0.099}{0.03 + 0.099 + 0.105} \approx 0.423$$