9.9 생일 문제

학습지

문제

STEP II 2007, 13번 문제

더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:

step.maths.org/questions/07-s2-q13

힌트

먼저 혼자 힘으로 문제를 풀어 보자. 막히면 힌트를 **하나씩** 읽고, 힌트를 하나 읽을 때마다 다시 문제로 돌아가자. **잠깐, 직접 해 보기**라고 표시된 힌트가 핵심 단계이다. 각 힌트 뒤에는 풀이를 적을 공간이 있다.

> ? **생각해 보기** 왜 어떤 두 손님도 생일이 같지 **않을** 확률을 계산하는 편이 더 쉬울까? 손님들을 한 줄로 세워 보자. 두 번째 손님이 첫 번째 손님의 생일을 피할 확률은 얼마인가? 세 번째 손님이 두 사람의 생일을 모두 피할 확률은?

1단계 유용한 근사(approximation)

다루는 내용: 근사

> ? **생각해 보기** e^{-x}의 급수(series) 전개를 쓰고 $x = r/n$을 대입한다. r이 n보다 훨씬 작을 때 무시해도 되는 항은 어느 것인가?

2단계 생일이 같을 확률(probability)

다루는 내용: 적어도 두 손님의 생일이 같을 확률

잠깐, 직접 해 보기 k개의 생일이 모두 다를 확률을 분수 $\frac{365-r}{365}$들의 곱으로 나타낸다. 각 인수를 지수함수(exponential)로 근사하고, 지수들을 모두 더한다.

3단계 스물셋

다루는 내용: 확률이 반 이상이 되는 가장 작은 모임

생각해 보기 구한 확률이 $\frac{1}{2}$ 이상이 되도록 놓고 로그(logarithm)를 취한다. $\ln 2 \approx 0.693$을 이용하고, 답 근처의 자연수들을 대입해 확인한다.

4단계 주최자와 생일이 같은 경우

다루는 내용: 주최자와 생일이 같을 확률

? **생각해 보기** 각 손님에 대해, 주최자와 생일이 같지 않을 확률은 얼마인가? 이 경우 손님들은 서로 독립(independent)인가? 앞에서와 같이 근사하여 푼다.

숨은 수학

제곱근 법칙

365일 대신 똑같이 가능한 날이 n일 있다면, 같은 논증으로

$$\mathrm{P}(\text{일치가 있을 확률, 인원}k) \approx 1 - e^{-k^2/(2n)},$$

을 얻는다. 따라서 k가 n이 아니라 $\sqrt{n}$ 정도만 되어도 일치가 일어날 가능성이 커진다. 더 정확히 말하면, 확률은 $k \approx 1.18\sqrt{n}$에서 $\frac{1}{2}$에 이르고, 처음으로 겹치는 사람이 나올 때까지 만나야 하는 사람 수의 기댓값은 약 $\sqrt{\pi n/2}$이다($n = 365$이면 약 24.6). 이 **제곱근 법칙**(square-root law)이 우연의 일치가 그토록 흔한 이유이다. 사건들을 크게 모아 놓으면, 특정한 일치 하나하나는 일어날 것 같지 않아도 어떤 쌍은 어떤 식으로든 일치할 가능성이 크다.

생일 공격

컴퓨터는 파일이 변조되지 않았는지 확인하기 위해 **해시**(hash)를 계산한다. 해시는 예컨대 $n = 2^{128}$가지의 값 중 하나를 갖는 지문과 같은 것이다. 해시가 같은 두 문서(무해한 계약서와 사기성 계약서)를 원하는 위조자는 2^{128}개의 문서를 시도할 필요가 없다. 생일 논증에 따르면 각 문서의 변형을 약 $\sqrt{2^{128}} = 2^{64}$개씩 만들어 내면 아마도 일치가 생길 것이다. 이 **생일 공격**(birthday attack) 때문에 현대의 해시 함수는 256비트 이상을 사용한다. 제곱근 법칙 때문에 비트의 절반을 잃기 때문이다. 비슷한 계산으로, 두 컴퓨터가 실수로 같은 식별자를 고르는 일이 결코 없도록 무작위 식별자의 길이를 얼마로 해야 하는지도 정한다.

처음 물은 사람은 누구인가?

생일 문제는 보통 1939년에 이 문제를 다룬 리하르트 폰 미제스(Richard von Mises)의 공으로 돌려지지만, 그 이전부터 퍼져 있었다. 해럴드 대븐포트(Harold Davenport)도 그보다 먼저 이 문제를 알고 있었으나 너무 단순해 보여서 발표하지 않았다고 전해진다. 워런 위버(Warren Weaver)가 전하는 이야기에 따르면, 그는 제2차 세계 대전 중 육군 장교들의 만찬에서 이 문제를 보여 주기를 좋아했다. 한번은 그 자리에 있던 스물두 명 중 두 사람의 생일이 같다는 데 내기를 걸었다가 졌는데, 그때 한 웨이트리스가 자신이 장교 중 한 명과 생일이 같다고 밝혔다.

더 도전하기

1 **첫 번째 반복.** STEP II 2014, 13번 문제 (step.maths.org/questions/14-s2-q13)에 도전해 보자. 이 문제는 난수 생성기에서 처음으로 반복되는 값을 다루는데, 이는 모습을 바꾼 생일 문제이며, 확률들을 놀라운 두 항등식으로 바꾸어 놓는다.

2 **세 사람의 생일이 같은 경우.** 어떤 세 사람의 생일이 같을 확률이 절반보다 커지려면 대략 몇 명이 필요한가? (**힌트:** 생일이 같은 세 사람 묶음의 개수의 기댓값은 $\binom{k}{3}/365^2$이다. 이 값을 약 $\ln 2$로 놓아라.) 실제 답은 88이다.

3 **아슬아슬한 경우.** 두 사람의 생일이 같은 날이거나 이웃한 날일 확률이 절반이 되려면 몇 명이 필요한가?

4 **고르지 않은 생일.** 실제 생일은 균등하게 분포하지 않는다. 각 날의 확률이 $p_1, \ldots, p_{365}$일 때 두 사람의 생일이 같을 확률은 $\sum p_i^2$이고, 이 값은 모든 p_i가 같을 때 가장 작음을 보여라. (따라서 분포가 고르지 않으면 일치가 오히려 더 일어나기 쉬워진다.)

더 읽을거리

- **William Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1** (Wiley), 제II장. 생일 문제와 그 근사.
- **Persi Diaconis & Frederick Mosteller, "Methods for studying coincidences"**, *Journal of the American Statistical Association* 84 (1989). 우연의 일치가 흔한 이유를 다루며, 매우 읽기 쉽다.
- **David Hand, *The Improbability Principle*** (Bantam). 제곱근 법칙을 비롯해, 일어날 것 같지 않은 일이 늘 일어나는 이유들을 다룬다.

안내에 따른 전체 풀이는 제 9.9절에 있다:

https://beyondstep.keithmcnulty.org/ko/ch09/birthday.html

매주 STEP 문제 하나를 끝까지 풀어 본다. The STEP Up은 이메일로 무료로 받아 볼 수 있다: keithmcnulty.substack.com