평 균 법 칙

“자네는 허리케인이 이곳을 덮칠 확률이 얼마인가를 알 수 있는 식을 만들어 낼 수 있겠나?” 하고 메신저가 말했다.

“못합니다. 허리케인이란 놈은 마치 동전과도 같아서 기억력이 없거든요. 동전을 50번 던저서 다 앞면이 나왔다고 해도 그 다음 던졌을 때 앞면이 나올 확률은 50 대 50이니까요. 하지만 10,000번 정도 던지면 5,000번은 앞면이 나올 겁니다. 오차가 조금 있을 수는 있겠지만…….”

존 맥도널드(John D. MacDonald), 『콘도미니엄*(Condominium)*』

그들이 결혼한 직후, 학교 동료 한 사람이 코드에게 축하 인사를 건넨 후 이렇게 말했다. “자네 혹시 확률이론에서 나온 오래된 문구를 기억하나. 백만 마리의 원숭이들이 백만 년 동안 타자기 위에서 이리 뛰고 저리 뛰면 그 중 하나는 『실락원』을 타자로 쳐낼지도 모른다는 이야기 말일세. 어쨌든 자넨 여자 문제에서 만큼은 그런 것 같아. 이리 뛰고 저리 뛰고 하다가 엄청난 횡재를 했으니까 말일세.”

솔 벨로(Saul Bellow), 『학장의 12월*(The Dean's December)*』

여러분에게 간단한 질문을 한 가지 하겠다. 앞뒤가 반듯한 동전을 하나 가지고 있다고 하자. 세심하게 동전을 살펴본 다음, 여러분은 이 동전을 던지면 앞면이 나올 확률과 뒷면이 나올 확률이 같다는 결론을 내렸다. 그런데 여러분이 자기만큼이나 할 일이 없다고 생각하는 교수가 여러분에게 동전을 100번 던져보라고 했다. 99번 던지고 나서 보니까 56번은 앞면이 나왔고 43번은 뒷면이 나왔다. 마지막 던질 때에

(1) 뒷면이 나올 확률이 큰가?

(2) 앞면이 나올 확률이 큰가?

(3) 앞면이 나올 확률과 뒷면이 나올 확률이 모드 1/2인가?

놀라울 정도로 많은 사람들이 (1)번이 맞다고 대답한다! 이들은 주로 평균법칙을 근거로 그런 답을 했다고 말한다. 그 동전이 **공정하다면** 평균적으로 앞면이 나오는 횟수와 뒷면이 나오는 횟수가 같아야 한다는 것이다. 그런데 99번 던졌을 때까지는 앞면이 뒷면보다 많이 나왔으므로 균형이 맞으려면 이번에는 뒷면이 나와야 한다는 것이다. 이들의 주장은 평균법칙에서 이론적으로 요구되는 대로, 뒷면이 나오는 비율이 1/2에 가까워져야 한다는 것이다. 반대로 101번째 던졌을 때 앞면이 나오면 이들은 뒷면이 나오는 비율이 이론적 요구값인 1/2에서 멀어지게 된다고 말한다.

대부분의 사람들에게는 다음과 같은 사실이 놀랍게 느껴질 것이다.

A. 바로 위에서 설명한 방식의 평균법칙은 완전히 잘못된 것이다.

B. 동전이 정말 공정하다면 100번째도 앞면이 나올 확률과 뒷면이 나올 확률은 같아야 한다.

C. 동전에 어떤 물리적인 결함이 있다면 마지막으로 던질 때 뒷면이 나올 확률은 앞면이 나올 확률보다 **작다.**

어떤 독자들에게는 이단적인 말로까지 들릴지 모르지만 이런 주장들이 정당하다는 사실을 지금부터 증명해보도록 하겠다.

A, B. 방금 설명한 평균법칙은 계속해서 동전을 던질 때 어떤 일이 일어날 것인가를 **예측하는** 데 쓰일 수 있도록 수정되어야 한다. 동전을 99번 던져서 56번은 앞면이 나오고 43번은 뒷면이 나왔다는 사실은, 동전을 100번째 던질 때 어느 면이 나올 것인가 하는 문제에는 아무런 영향도 미칠 수 없다. 어쨌든 동전에는 기억력이 없으니까 말이다. 동전은 이전에 99번 던졌을 때 어느 면이 나왔는가를 기억하지 못한다. 정말로 공정한 동전이라면 앞면이 나올 확률과 뒷면이 나올 확률은 던질 때마다 모두 1/2이다. 이러한 사실은 100번째 동전을 던질 때에도 그대로 적용될 수 있다. 과거에 어떤 결과가 나왔는가와는 아무 상관없이 말이다.

(통계학적 변동을 고려하면서도)연속적으로 동전을 던질 때 어느 면이 나올 것인가를 예측하게 해주는 통계학 법칙들이 있다. 앞에서 설명한 평균법칙보다는 좀더 정확한 설명을 해보기로 하자. 공정한 동전이라면 던지는 횟수가 많아질수록 앞면이 나오는 비율이 0.5라는 이론상의 수치에서 벗어날 가능성은 점점 적어진다.

이것은 앞에서 말한 평균법칙보다는 훨씬 세련되고 복잡한 말이다.

이것은 앞으로 동전을 던지게 되는 경우에 대해 말하고 있다. 과거를 이용해서 미래를 예측하려는 것이 아니다. 이것은 앞에서 평균법칙에 대해 언급할 때와 마찬가지로 동전이 공정하다는 가정에 근거하고 있다. 또 이 말은 앞으로 뒷면이 나올 횟수에 대해서도 고려하고 있다. 게다가 모호한 말을 즐겨 쓰는 통계학자들에게는 다반사이지만, 이 말은 '오차'의 여지도 남겨두고 있다. 실험 횟수가 증가할수록 이러한 오차는 줄어들게 된다. 마지막으로 이 말은 어떤 결과들이 나올 확률에 대해 언급한 것이다. 이 말은 어떤 특정한 결과들이 분명히 일어나리라고 말하지는 않는다. (이 모든 사실로부터 통계학자라는 사람들은 뭔가 분명한 구석이 없고 모호한 말을 즐기는 사람들이라는 인상을 받을지도 모르겠다. 더 정확하게 말하면 확률에 좌우되는 사건들을 논할 때 통계학자들은 다양한 결과가 나올 확률에 대해서만 정확한 말을 한다. 통계학자들은 예언자가 아니니까 말이다!)

C. 동전에 물리적으로 미세한 결함이 있다고 가정하자. 어쨌든 **완전하게** 공정한 동전이라는 것은 이상적인 생각이며, 실제 생활에서 접하기보다는 통계학 문헌에서 관념적으로 정의하기에 쉬운 것이다. 따라서 실제로 앞면이 나올 확률은 1/2이 안 될 수 있고 그보다 조금 큰 0.501이 될 수도 있다. 이런 경우에 우리는 동전을 계속 던지다 보면 앞면이 나오는 비율이 1/2보다 약간 크리라고 예상할 수 있다. 따라서 99번 던졌을 때 뒷면보다 앞면이 많이 나왔다는 사실을 알게 되었다면(이 경우에는 앞면이 56번 나왔고 뒷면은 43번 나왔다) 마지막으로 동전을 던질 때 앞면이 나올 확률이 뒷면이 나올 확률보다 많다는 쪽에 돈을 거는 것이 현명하다. 그 이유는 간단하다. 동전이 완전히 반듯하지 않고 약간 기울어져 **있다**

면 지금까지 나타난 증거로 볼 때 앞면이 나올 가능성이 뒷면이 나올 가능성보다 훨씬 크다는 사실을 보여주고 있기 때문이다. 게다가 동전이 기억력을 가지고 있다는 식의 조금은 우스꽝스러운 생각이라든가, 통계학 이론에 대한 미심쩍은 지식, 평균법칙이라고 불리는 것을 따르고 싶어하는 어쩔 수 없는 충동을 버린다면, 우리는 앞면이 나온다는 데 거는 것이 훨씬 그럴듯하다는 사실을 더욱 강하게 믿을 수 있을 것이다.

이 정도면 평균법칙에 대해서는 충분히 말한 것 같다!

주의 : 평균법칙의 일반화는 다양하게 존재한다. 한 가지만 예를 들면, 정글에서 원숭이들이 이리 뛰고 저리 뛰면서 타자기 위를 뛰어다니다 보면 그 중에 한 마리는 연대 순서도 틀리지 않고 단 하나의 오자도 없이 셰익스피어의 작품을 완전히 쳐낼 것이라는 환상적인 비유가 있다. 또 다른 원숭이는 프루스트의 작품을 완전히 쳐낼 것이고, 또 다른 한 마리는 밀턴의 『실락원』을 쳐낼 것이다. 이런 식의 이야기는 확률론 학자들의 흥미는 유발하고 문필가들의 미간은 찌푸리게 한다.

요 약

1. 공정한 동전을 여러 번 던질 때 앞면이 나오는 비율이 0.5에 가까울 확률은 매우 높다. 동전을 여러 번 던질수록 앞면이 나오는 비율이 이론적인 값 0.5에서 어느 정도의 오차 이상으로 벗어날 확률은 점점 작아진다. 대략적으로 말하면 동전을 던지는 횟수가 많아질수록 앞면이 나오는 비율은 이론적인 수치인 0.5에 가까워진다.
2. 1번에서 말한 것은 미래의 결과에 대한 확률이지 어떤 특정한 결과(예를 들면 다음 번 던졌을 때 앞면이 나오는 것)나 일련의 결과들이 분명히 일어날 것이라고 주장하는 것은 아니다.
3. 동전이나 카드, 또는 주사위같이 생명이 없는 물체들은 기억력이 없다. 컴퓨터의 메모리를 제외하고는 말이다.

<통계 마인드 길들이기, 마일즈 홀랜더 · 프랭크 프로쉬안 지음, 김동훈 옮김, 새날, 1995년 9월>