# 두 봉투의 역설에 대하여

송하석(아주대)

【요약문】 1989년 네일버프(B. Nalebuff)에 의해서 두 봉투의 역설이 제시된 이래 많은 철학자들이 이 역설에 많은 논의가 있었데, 이 역설에 대한 논의를 다시 촉발시킨 글은 1997년의 잭슨(F. Jackson)과 멘지스(P. Menzies)와 오피(G. Oppy)가 발표한 논문이다. 이 글은 잭슨 등의 논문에서 제시된 두 봉투의 역설에 대한 설명을 토대로 두 봉투의 역설이 생긴다고 생각하는 이유는 무엇이고, 구체적으로 어떤 경우에 역설이 생길 수 있는지를 살펴본다. 그리하여 필자는 다음 두 가지를 주장할 것이다. 첫째 두 봉투의 역설이라고 불리는 것은 봉투에 들어있는 금액이 무한하고 동시에 그 금액의 평균값이 무한한 경우만 발생할 수 있을 뿐이고, 이는 내가 선택한 봉투 안의 기대값을 실제로 계산할 수 없는 경우이다. 둘째로 기대값이 실제로 계산될 수 있는 그 외의 경우, 역설은 발생하지 않으며, 역설이 발생한다고 생각하는 것은 자신의 봉투 안의 금액을 고정적인 것으로, 상대방의 금액을 가변적인 것으로 해석하는 잘못으로부터 기인하는 것이다. 요컨대 기대값이 계산될 수 있는 일반적인 경우 두 봉투의 역설이 생긴다고 생각하는 것은 잘못이고, 그 경우 역설은 발생하지 않는다는 것을 논증할 것이다.

【주제어】 두 봉투의 역설, 카질의 역설, 크라이취크의 역설, 확률

## 1. 들어가는 말

합리적인 의사결정이 확률 계산에 근거해야 한다는 생각은 18세기에 이미 시작되었다. 18세기의 철학자 버틀러(J. Butler)는 "확률은 삶의 안내자"라고 말했고, 여기서 발전하여 아르노(A. Arnauld)는 불확실한 조건 아래에서 이루어지는 결정은 가능한 사건의 확률값과 그 사건에 따르는 손실의 평가라는 두 요소가 고려되어야 한다고 주장함으로써 기대값 개념이 확률 계산에 도입되어야 한다는 통찰을 제공하였다. 이렇게 확률 계산이 합리적인 결정이론에 중요한 만큼, 확률 계산에 역설이 포함된다면 이를 해소하려는 노력 또한 중요한 일일 것이다. 확률 계산과 관련된 역설이 많이 있지

만, 이 글은 확률과 관련된 유명한 역설인 "두 봉투의 역설(Two-envelope paradox)"에 대해서 살펴볼 것이다.

1989년 네일버프(B. Nalebuff)에 의해서 이 역설이 제시된[1] 이래 많은 철학자들이 이 역설에 대한 진단과 처방을 제시해 왔는데, 이 역설에 대한 논의를 다시 촉발시킨 글은 1997년의 잭슨(F. Jackson)과 멘지스(P. Menzies)와 오피(G. Oppy)가 공동으로 발표한 논문[2]이다. 그 이후부터 현재까지 그들의 논문에 대한 많은 찬, 반의 글들이 쏟아져 나오고 있다. 따라서 이 글은 잭슨과 멘지스와 오피(이후 "잭슨 등"으로 표기)가 발표한 논문에서 제시된 두 봉투의 역설에 대한 설명을 정리해 본 후, 이후의 논문들을 통해서 그들의 설명에 대한 평가를 시도해 볼 것이다. 그리고 끝으로 잭슨 등이 제시한 두 봉투의 역설에 대한 진단을 토대로 두 봉투 역설과 관련된 가능한 모든 경우를 살펴보고 필자는 기대값이 계산될 수 있는 일반적인 경우 두 봉투의 역설이 생긴다고 생각하는 것은 잘못임을 보이고, 그 경우 역설은 발생하지 않는다는 것을 논증할 것이다.

먼저 두 봉투의 역설이라고 알려진 역설을 우선 간단히 설명해 보자. 내 앞에 구별이 안 되는 두 개의 봉투(A, B)가 놓여져 있다. 그 봉투 안에는 각각 돈이 들어 있는데, 하나의 봉투에는 다른 봉투에 들어 있는 액수의 두 배가 들어있다. 그리고 나는 그 봉투 중의 하나를 선택하여 그 봉투에 들어 있는 돈을 가질 수 있다. 내가 하나의 봉투를 선택했는데, 사려깊은 판단을 통해서 봉투를 교환할 수 있는 기회가 주어졌다. 교환하는 것이 좋은가, 그냥 원래의 선택을 유지하는 것이 좋은가?

내가 선택한 봉투(A)에 x원이 들어있다고 하자. 그러면 다른 봉투(B)에는 x/2원이나 2x원이 들어 있을 것이고, 그 각각의 확률은 동일하게(equiprobable) 1/2일 것이다. 그렇다면 내가 교환하는 경우 기대값은 다음과 같은 계산에 의해서 1.25x원으로 원래의 x원보다 많다. 따라서 교환하

---

1) B. Nalebuff, 'The other person's envelope is always greener' *Journal of Economic Perspectives* 3 (1983).

2) F. Jackson, & P. Menzies, & G. Oppy, "The Two-envelope 'Paradox'" *Analysis* 54.1 (1994, Jan.).

는 것이 현명하다.

$$E = (2x \times \frac{1}{2}) + (\frac{x}{2} \times \frac{1}{2}) = \frac{5x}{4} \quad (> x)$$

그런데 내가 처음 선택한 봉투가 A가 아니라 B라고 할지라도 위의 추론은 동일하게 적용될 수 있다. 즉 이 사건은 대칭적이기 때문에, 그러한 교환은 처음 내가 어떤 봉투를 선택했다고 할지라도 교환해야 한다는 결론이 나온다. 나아가서 한번 교환한 후 다시금 동일한 추론에 의해서 또 다시 교환해야 한다는 결론이 나온다. 그러나 이 경우는 명백하게 두 봉투 안에 들어있는 금액의 합이 일정하고, 따라서 교환해서 한 사람이 이익을 본다면 다른 사람은 그만큼 손해를 보아야 하는 제로섬(zero-sum) 게임이다.[3] 그런데 처음 어떤 봉투를 선택해도 합리적인 검토 결과 교환해야 한다는 결론을 얻어내고, 그리고 한번 교환 후 또 다시 계속해서 교환해야 한다는 결론이 나온다면 분명히 이것은 역설적이라고 할 수 있을 것이다.[4]

## 2. 잭슨과 멘지스와 오피의 설명

잭슨 등은 E와 같은 계산이 잘못이라고 말하고, 그 이유는 내가 선택한 봉투 A에 x원이 들어 있다는 조건 하에서 다른 봉투 B에 2x원이 들어 있을 확률과 x/2원이 들어 있을 확률이 동일하지 않다고 주장한다. 두 개의 봉투의 금액의 쌍이 (x, 2x)일 확률과 (2x, x)일 확률이 동일한 것이지, 한 봉

3) 제로섬게임에 대한 설명이 요구된다는 익명의 심사위원의 지적이 있었다. 제로섬게임이란 간단히 말해서 하나의 게임에서 손실의 총합과 이득의 총합을 더하면 제로가 되는 게임이다.

4) 익명의 심사위원 중 한 분이 이것이 역설적인 이유에 대해서 의문을 제기했다. 주어진 봉투 A와 B 중에서 처음 A를 선택했는데, 본문에서 설명한 것처럼 계산 결과 B로 교환하는 것이 낫다는 결론을 내렸다. 이는 A를 선택했을 때의 기대값보다 B를 선택했을 때의 기대값이 크다는 것을 의미한다. 그리고 B로 교환한 후, 다시 계산 결과 A로 교환하는 것이 낫다는 결론을 얻게 된다. 그런데 그것은 B를 선택했을 때의 기대값보다 A를 선택했을 때의 기대값이 크다는 것을 의미한다. A를 처음 선택해서 B로 교환하는 것이 유리하고, 또한 B를 선택했다가 A로 교환하는 것이 유리하다는 것은 역설적이다.

투의 금액이 고정된 상태에서 다른 봉투의 금액이 그것의 두 배이거나 절반일 확률이 동일하다고 주장하는 것은 잘못이라는 것이 두 봉투 역설에 관한 그들의 진단에서 가장 핵심적인 것이다.

또한 잭슨 등은 이 세상에 있는 돈이 무한하다고 생각할 수 없으며, 돈의 단위 또한 임의적인 것이 아니기 때문에, 내가 선택한 봉투의 금액이 주어질 때, 우리는 다른 봉투의 금액이 그것의 두 배가 될 것인지, 절반이 될 것인지에 대한 일정한 확률분포(probablity distribution)를 갖는다고 주장한다.[5] 다시 말해서 우리가 상식적이라면, 우리는 두 개의 봉투에 들어 있을 총액에 대하여 하나의 선행하는 확률분포(prior probability distribution)를 갖는데, 이것은 모든 x에 대해서 하나의 봉투의 금액이 x라면 다른 봉투의 금액이 2x거나 x/2일 확률이 동일하지 않을 수 있다는 것을 의미한다. 예컨대 어떤 상황 때문에[6] 우리가 두 봉투의 총액이 100만원을 넘을 가능성이 아주 작다고 생각한다고 하자. 그런데 A에 40만원이 들어 있다면, B에 들어 있을 금액이 20만원과 80만원일 확률이 동일하게 1/2이라고 생각하지 않을 것이다. 그리고 잭슨 등은 우리가 그러한 선행하는 확률 분포를 갖는 것에 대한 정당성을 우리가 일반적으로 공유하고 있다고 생각하는 상식과 합리성에서 찾는다.

우선 잭슨 등의 첫 번째 진단은 옳다. 다시 말해서 한 봉투 안의 임의의 금액 x에 대해서 다른 봉투의 금액이 2x이거나 x/2일 확률이 동일하다는 생각은 잘못이고, 그것이 역설을 낳는다고 잘못 생각하게 만드는 결정적인 이유라는 그들의 생각은 옳다. 잭슨 등은 역설을 낳는다고 주장되는 두 봉투

---

5) 확률분포란 가능한 사건들이 갖는 각각의 확률값의 분포를 말한다. 여기서 우리가 봉투 A를 선택했을 때 B에 들어있는 금액이 A의 금액의 절반일지 두배일지에 대한 확률분포를 갖는다는 의미는, B의 금액이 A의 금액의 두배가 될 확률값을 m, 절반이 될 확률값을 n이라고 생각할 수 있다는 것이다. 물론 여기서 m과 n의 합은 1이어야 정상적인 확률분포가 된다. 우리가 가능한 사건들에 대해서 각각의 확률값을 갖는다고 할지라도 그 가능한 사건에 대한 확률값의 총합은 반드시 1이어야 할 것이다. 그러한 확률 분포를 정상적 확률분포라고 한다. 그러므로 우리가 가능한 사건에 대해서 임의의 확률값을 부여할 때, 그것이 최소한의 합리적 조건을 만족하기 위해서는 그 확률분포가 정상적임이라는 것을 보여야 한다.

6) 예를 들어서 봉투를 제시한 사람의 성격이나 게임의 성격 상, 두 봉투 안의 금액의 총합이 100만원이 넘기 어려운 상황을 생각할 수 있다.

의 경우를 다음과 같은 카질(J. Cargile)이 제시한 경우와 구별하면서 그들의 설명의 정당성을 강조한다. 즉 카질의 경우란 두 개의 봉투의 색이 빨간색과 파란색으로 구별되어 있는데, 빨간색의 봉투에 x원이 들어 있고, 정상적인 동전을 던져서 앞면이 나오면 파란색의 봉투에 2x원을 뒷면이 나오면 x/2원을 넣는 경우이다. 이 경우에는 당연히 파란색 봉투를 선택해야 하는데, 그 이유는 빨간색 봉투를 선택했을 때의 기대값은 x이지만, 파란색의 봉투를 선택했을 때의 기대값은 E와 같이 계산되어 5x/4이기 때문이다. 즉 앞에서 제시된 E와 같은 계산은 문제가 되는 두 봉투의 경우와 구별되는 카질의 경우에 대한 계산법이다. 그 두 경우의 결정적인 차이는 카질의 경우는 두 개의 봉투가 색으로 구별되어 있고, 하나의 금액에 대해서 다른 하나의 금액이 두 배이거나 절반일 확률이 동일한 반면, 역설을 낳는다고 주장되는 두 봉투의 경우는 어느 하나의 봉투에 다른 하나의 봉투에 들어 있는 것의 두 배가 들어 있다는 사실만 주어져 있기 때문에 (x, 2x)와 (2x, x)일 확률이 동일한 것이지, 다른 하나의 봉투에 들어 있는 금액이 고정되었을 경우에 다른 봉투의 금액이 그 고정된 금액의 절반이거나 두 배일 확률이 동일한 것은 아니다.[7)]

그러나 책슨 등이 봉투에 들어 있을 금액이 유한하며, 각 금액에 대한 선행하는 확률분포를 갖기 때문에 역설이 발생하지 않는다는 주장은 몇 가지 점에서 비판이 되어 왔다. 우선 현실 세계의 돈의 액수는 무한하지 않고, 봉투 안의 금액에도 한계가 있다는 현실적이고도 상식적인 지적에 대해서 브룸(J. Broome)이 지적하듯이, 세계의 돈의 액수가 유한하다는 것은 현실적 경우만을 고려한다면 가능하고 옳은 처방일 수 있지만, 얼마든지 무한한 양에 대하여 동일한 종류의 역설이 발생할 수 있음을 보일 수 있다는 문제가 있다. 다시 말해서 두 봉투의 경우에서 봉투 안의 금액이 유한하다고 가정함으로써 역설이 해소된다는 것이 책슨 등의 처방이라면, 첫째 봉투 안에 무한

7) 그런 의미에서 원래의 두 봉투의 역설은 성 페테르스부르그의 게임과 관련하여 발생된다고 주장되는 역설(St. Petersburg paradox)과 구별되어야 한다. 페테르스부르그 역설은 카질의 경우에 발생하는 것이지, 원래의 두 봉투의 경우에서 발생된다고 주장되는 역설과 다르다. 그러므로 두 경우를 비슷하게 생각하는 안쩨니우스와 맥카티의 주장과 브룸의 주장은 옳지 않다. F. Arntzenius & D. McCarthy (1997)와 J. Broome (1995) 참조.

한 금액이 들어 있을 수 있는 사유실험이 가능하고, 둘째 그러한 경우 역설이 발생한다고 주장할 수 있는 선행하는 확률분포를 부여할 수 있다는 비판이 그것이다.

첫 번째에 대한 예를 들어 보자. 우리가 어떤 세계의 행복을 담당하는 천사라고 하자. 하느님이 행복은행에서 인출한 수표들을 봉투에 넣어 두었는데, 우리가 선택한 봉투 안의 금액만큼의 행복이 그 세계의 행복의 양이 된다고 하자. 물론 행복의 양은 무한할 수 있다. 그리고 그 봉투 안의 금액 중의 하나는 다른 하나의 두 배라고 하자. 유한성을 가정함으로써 역설을 막을 수 있다고 한다면 이 경우에 교환과 관련한 역설이 발생할 수 있다는 비판을 피할 수 없을 것이다. 결국 잭슨 등의 설명을 따른다고 할지라도, 한계가 없는 경우에는 여전히 역설이 발생할 수 있다는 것을 인정할 수밖에 없을 것이다. 그리고 봉투 안의 금액이 무한할 수 있다는 것을 인정한다면 여러 철학자들이 보였듯이 역설을 낳는 확률분포를 가정할 수 있기 때문에[8], 잭슨 등이 우리가 봉투 안에 들어 있을 금액에 대하여 선행하는 확률분포를 갖는다는 주장으로 역설이 완전히 해소될 수는 없어 보인다. 물론 잭슨 등은 역설을 낳는다고 제시되는 확률분포는 브룸도 인정하듯이 유한한 평균값(finite mean)을 갖지 않는 경우이다.[9] 그러므로 잭슨 등은 우리가 봉투에 들어 있을 금액에 최저와 최고 한계가 있다는 것을 안다는 사실이 이 역설의 진단에 핵심이라고 주장한다. 잭슨 등도 물론 경계가 없는 경우도 상상할 수 있음을 인정하지만, 그들은 "유한성의 가정은 그러한 추론을 정당화하는 굳건한 토대"[10]라는 제프리(R. Jeffrey)의 말을 인용하면서, 무한영역(infinite domains)을 포함하는 경우에 확률과 기대값 추론의 표준적인 방식이 적용될 수 없다고 주장한다.

요컨대 잭슨 등이 두 봉투의 상황에서 역설이 발생한다고 진단한 원인은 A에 x가 들어 있을 때, B에 2x가 들어 있을 확률과 x/2가 들어 있을 확률

8) 이에 대한 구체적인 확률분포의 예는 J. Broome을 비롯해서 최근에는 M. Clark & N. Shakel 등이 제시한 바 있고, 이에 대해서는 이 글의 후반부에서 자세하게 다룰 것이다.

9) J. Broome, "The Two-envelope Paradox" *Analysis* 55.1 (1995, Jan.), 9쪽.

10) R. Jeffrey, The Logic of Decision (London: University of Chicago Press, 2nd ed. 1983), ch. 10.

이 동일하다고 잘못 생각한 때문이라는 것이고, 이에 대하여 그들은 A와 B에 들어 있을 금액은 유한하고 또한 합리적이고 상식적인 근거에 의해 선행하는 확률분포 값을 가지고 있다고 주장함으로써 역설을 해소하고자 한다.

이제 잭슨 등이 두 봉투의 역설에 대하여 행한 진단과 처방을 구체적으로 평가해 보기 위해서 두 봉투의 경우에 대하여 다음과 같은 몇 가지로 나누어 생각해 보기로 하자. 즉 봉투에 든 금액이 유한한 경우와 무한한 경우, 그리고 무한한 경우는 다시 한 봉투의 금액이 유한한 평균값(finite mean)을 갖는 경우와 그렇지 않는 경우로 나누어 어떤 경우에 역설이 발생하고, 그 이유는 무엇인지 살펴보기로 하자.

## 3. 봉투의 금액이 유한한 경우

잭슨 등이 봉투의 금액에 한계가 있는 경우에는 역설이 발생하지 않고, 그것은 봉투 A에 x가 들어 있을 경우에 봉투 B에 2x가 들어 있을 확률과 x/2가 들어 있을 확률이 동일하다고 잘못 가정하기 때문이라고 주장하는 것은 옳다. 그러나 잭슨 등은 그들의 논문에서 그것을 실제로 증명해 보이지는 않았다. 이제 그것을 증명해 보기로 하자. 논의를 간단하게 하기 위해서 대부분의 철학자들이 가정하듯이 두 개의 가정을 하기로 하자. 첫째, 봉투 안의 금액 중에서 최소 금액을 1원이라고 하고, 봉투 안에 1원이 들어 있을 수 있다는 것을 가정하자. 둘째, 봉투 안에 들어 있을 금액은 모두 2의 제곱수로 표현된다고 하자. 즉 두 봉투에 들어 있을 금액의 쌍은 $(2^0, 2^1)$, $(2^1, 2^2)$, $(2^2, 2^3)$, … $(2^n, 2^{n+1})$이고, 그 각각의 확률값을 $p_0$, $p_1$, $p_2$, … $p_n$이라고 하자.[11]

봉투 A에 $2^1$원이 들어있을 경우, 봉투 B에 $1(=2^0)$원이 들어 있을 확률은 $p_0$, $2^2$원이 들어 있을 확률은 $p_1$이 된다. 내가 A를 선택한 후에 교환했을 때의 기대값은 다음과 같다.

---

11) 이렇게 가정하는 것은 순전히 증명의 단순화를 위한 것이다. 봉투 안의 금액의 쌍을 다음과 같이 a 곱하기 2의 제곱수로 표현할 수도 있다: (20a, 21a), (21a, 22a), (22a, 23a), ... (2na, 2n+1a). 물론 단순화된 가정 하에서 이루어진 증명은 그외의 다른 경우에 대한 증명에도 그대로 적용될 수 있음은 자명하다.

$$E_1 = \frac{p_0 + 4p_1}{p_0 + p_1}$$

$E_1$이 2보다 크기 위해서는 $p_1 > p_0/2$이어야 한다. 즉, $p_1 > p_0/2$일 경우에 그 기대값은 2보다 크게 되고, 그 경우에는 교환하는 것이 합리적이다. 이것을 일반화하여 말하면, 앞에서 가정한 것처럼 봉투 A에 $2^n$원이 들어 있을 경우, 봉투 B에 $2^{n-1}$원이 들어 있을 확률은 $p_{n-1}$, $2^{n+1}$원이 들어 있을 확률은 $p_n$이다. 내가 선택한 봉투가 A인데, B로 교환했을 때의 기대값은 다음과 같다.

$$E_n = \frac{2^{n-1}.p_{n-1} + 2^{n+1}.p_n}{p_{n-1} + p_n}$$

따라서 교환했을 때 기대되는 이익인 $(E_n\text{-}A)$는 $(2^n.p_n - 2^n\text{-}1.p_{n-1})/(p_{n-1} + p_n)$이다. 이제 A에 든 액수의 모든 경우에 대하여 교환했을 때 생길 수 있는 이득을 계산해 보자.

$$\sum_{n=0}^{k}(E_n\text{-}A) = (E_0\text{-}A_0) + \sum_{n=1}^{k}(E_n\text{-}A_n)$$

$$= 2^0.p_0 + \sum_{n=1}^{k}(2^n.p_n - 2^{n-1}.p_{n-1}) - 2^k.p_k$$

$$= 2^0.p_0 + (-2^0.p_0 + 2^1.p_1 - 2^1.p_1 + \ldots.. + 2^k.p_k) - 2^k.p_k$$

$$= (2^0.p_0 - 2^0.p_0) + (2^1.p_1 - 2^1.p_1) + \ldots.. + (2^k.p_k - 2^k.p_k)$$

$$= 0$$

다음 표는 위 식의 계산을 간단하게 보여주고 있다. 다음 표에서 A는 내가 처음 선택한 봉투이다. A에 20가 들어 있고 B에 21이 들어있을 경우

교환할 경우 이득은 +20.p0이다. 그런데 A에 21이 들어있고 B에 20가 들어있을 경우 교환해서 얻게 되는 이득은 -20.p0이다. 이런 식으로 계속해서 계산을 하면, 결국 하나의 봉투를 선택한 다음 교환했을 때의 얻을 수 있는 이득은 0라는 것을 보여준다.

요컨대 위의 식이 말해주는 것은 교환해서 기대되는 이익은 없다는 것, 즉 교환을 계속해야 하는 역설적 상황은 발생하지 않는다는 것이다. 잭슨 등이 주장하는 것처럼 봉투 안의 금액의 최고액에 제한이 있다면 이처럼 각 경우에 어떤 정상적인 확률분포가 주어진다고 할지라도 역설은 발생하지 않는다. 이 점에서 이 경우에 역설이 발생한다는 주장은 바로 A의 봉투에 x원이 있을 때 B의 봉투에 2x원이 들어 있을 확률과 x/2원이 들어 있을 확률이 같다고 잘못 생각하는 데 있다는 잭슨 등의 주장은 옳다.

| B<br>A | $2^0$ | $2^1$ | $2^2$ | $2^3$ | … | $2^{n-1}$ | $2^n$ |
|---|---|---|---|---|---|---|---|
| $2^0$ | | $+2^0.p_0$ | | | | | |
| $2^1$ | $-2^0.p_0$ | | $+2^1.p_1$ | | | | |
| $2^2$ | | $-2^1.p_1$ | | $+2^2.p_2$ | | | |
| $2^3$ | | | $-2^2.p_2$ | | … | | |
| ⋮ | | | | … | | $+2^{n-2}.p_{n-2}$ | |
| $2^{n-1}$ | | | | | $-2^{n-2}.p_{n-2}$ | | $+2^{n-1}.p_{n-1}$ |
| $2^n$ | | | | | | $-2^{n-1}.p_{n-1}$ | |

## 4. 금액에 한계가 없다면?

이제 봉투 안의 최고액에 한계가 없는 경우를 살펴보자. 이 경우는 한 봉투의 금액이 유한한 평균값을 갖는 경우와 그렇지 않는 경우로 나누어진다. 즉 봉투 A에 든 금액의 최대 한계가 없다고 할지라도 A에 대한 평균값이 유한한 경우를 생각할 수 있고, 그러한 경우 우리는 교환해서 얻으리라고 기대할 수 있는 평균 기대이익이 0임을 증명할 수 있다. 먼저 그런 경우를 살펴보기 위해서 다음과 같이 0에 수렴하는 수열을 생각해보자.

$$1/2,\ (1/2)^2,\ (1/2)^3,\ .....\ (1/2)^n,\ ....$$

이렇게 0으로 수렴하는 수열의 총합은 그 수열의 부분의 합의 수열의 극한값으로 나타내진다. 즉 위 무한수열의 총합을 S라고 하면,

$$S = \lim_{n\to\infty} \sum_{k=1}^{n} \left(\frac{1}{2}\right)^k = \lim_{n\to\infty} \{1-\left(\frac{1}{2}\right)^n\} = 1$$

이다. 결국 무한 수열이 그 총합으로 유한한 값을 가진다는 것은 그 수열의 부분의 합들의 수열이 어떤 유한한 값에 수렴한다는 것과 동치이고, 어떤 무한수열의 부분의 합들의 수열이 유한한 값에 수렴하기 위해서는 그 무한수열이 0에 수렴해야 한다.

이제 두 봉투의 경우를 생각해 보자. 봉투 A에 들어 있을 금액에 한계가 없지만, 그 총합이 유한한 값을 갖는다면, 위의 조건에 의해서 A에 들어있는 수열은 0에 수렴하는 수열일 것이다. 그리고 E(A)는 다음과 같다.

$$E(A) = 2^0.p_0 + \sum_{n=1}^{\infty} (2^n.p_{n-1} + 2^n.p_n)$$

그런데 E(A)가 유한한 값을 갖기 때문에 E(A)의 부분들인 $\sum_{n=1}^{\infty} 2^n.p_{n-1}$ 과 $\sum_{n=1}^{\infty} 2^n.p_n$도 유한한 값을 가져야 하며, 위의 조건에 의해서 $\langle 2^n.p_{n-1}\rangle$과 $\langle 2^n.p_n\rangle$도 각각 0에 수렴하는 수열이어야 한다. 그리고 내가 A를 선택한 후에 B와 교환해서 얻을 수 있는 이익은 다음과 같이 계산된다.

$$E(B/A) - E(A) = 2^0.p_0 + \sum_{n=1}^{\infty}(2^n.p_n - 2^{n-1}.p_{n-1}) = \lim_{n\to\infty} 2^n.p_n$$

그런데 E(A)가 유한한 값에 수렴하고 따라서 $\langle 2^n.p_n\rangle$는 0에 수렴하므로, 결국 E(B/A) - E(A) = 0 이다. 이것은 A를 선택한 후 B와 교환해서 얻을 수 있는 이익은 0이라는 뜻이므로 이 경우에 역설은 발생하지 않는다.

이에 대한 구체적인 사례를 살펴보자. 두 봉투에 들어 있는 금액 중에서 작은 액수가 1일 확률을 1/2, 즉 $2p_0=1/2$라고 하고, $2^n$이 작은 액수일 확률을 $(\frac{1}{3})^n$, 즉 $2p_n = (\frac{1}{3})^n$라고 하자. 이러한 확률 분포는 $\frac{1}{2} + \sum_{n=1}^{\infty}(\frac{1}{3})^n = 1$이므로 정상적인 확률분포이다. 이 경우 E(A)를 앞의 방식과 같이 구하면 다음과 같다.

$$\begin{aligned} E(A) &= 2^0.p_0 + \sum_{n=1}^{\infty}(2^n.p_{n-1} + 2^n.p_n) \\ &= p_0 + \sum_{n=1}^{\infty}2^n.p_{n-1} + \sum_{n=1}^{\infty}2^n.p_n \\ &= p_0 + 2^1.p_0 + \sum_{n=1}^{\infty}2^{n+1}.p_n + \sum_{n=1}^{\infty}2^n.p_n \\ &= (1/4) + 2.(1/4) + \sum_{n=1}^{\infty}3.2^n.p_n \end{aligned}$$

$$= 3/4 + \sum_{n=1}^{\infty} 3.2^n . \frac{1}{2} . (\frac{1}{3})^n$$

$$= 3/4 + (3/2) \sum_{n=1}^{\infty} (\frac{2}{3})^n = 3\frac{3}{4}$$

그리고 이 경우 B로 교환했을 때 기대할 수 있는 이익, E(B/A) - E(A)는 $2^n.p_n$이 0에 수렴하는 수열이기 때문에 0이고 따라서 교환해서 기대할 수 있는 이익은 없다. 즉 이 경우에도 역설적인 상황은 발생하지 않는다.

끝으로 봉투의 금액이 무한하면서 한 봉투에 든 금액의 평균값이 유한하지 않는 경우를 살펴보자. 이러한 경우는 또 다음과 같은 두 가지로 나누어 질 수 있다. 즉 A를 선택한 후 B와 교환해서 얻을 수 있는 기대값이 유한한 값을 갖는 경우와 유한한 값을 갖지 않는 경우이다. 다시 말해서 E(B/A) - E(A)가 유한한 값에 수렴하거나 무한대 발산 혹은 진동하는 경우가 있다.

이제 그 값이 유한한 값에 수렴하는 경우를 살펴보기로 하자. 이를 위해서 두 봉투에 들어 있는 금액 중에서 작은 액수를 s라고 하고, s=1에 대해서 1/3의 확률값을, s=2에 대해서는 2/9를, s=4에는 4/27을 각각 부여하는 확률분포를 생각해 보자. 즉 이것은 일반적으로 $s=2^n$에 대해서 $2^n/3^{n+1}$의 확률값을 부여하는 확률분포이다. 당연히 우리는 이 확률분포가 정상적(normalized)이라는 것을 먼저 보여야 할 것이다. 다음과 같이 이 확률 분포가 정상적임을 보일 수 있다.

$$\sum_{n=0}^{\infty} \frac{2^n}{3^{n+1}} = \frac{1}{3} + \sum_{n=1}^{\infty} \frac{2^n}{3^{n+1}}$$

$$= \frac{1}{3} + \lim_{n\to\infty} \sum_{k=1}^{n} \frac{2^k}{3^{k+1}}$$

$$= \frac{1}{3} + \lim_{n\to\infty} \{\frac{2}{3} - (\frac{2}{3})^{n+1}\}$$

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

그러므로 이 확률분포는 정상적인 것이다. 여기서 내가 선택한 봉투의 금액이 x라고 하고, 다른 봉투의 금액을 y라고 하자.

그러면 x=1이라면 y=2이고, 이 경우에는 교환하는 것이 명백하게 이익을 얻을 것이다. 만약 x의 액수가 1이 아니라면, y가 2x일 확률은 얼마인가? y가 2x라는 것은 두 개의 봉투 중 최소 금액인 s가 x라는 것이다. 만약 내가 선택한 금액이 x라면 s=x이거나 s=x/2일 것이다. 따라서 s=x이거나 s=x/2라는 조건 하에서 s=x일 확률을 계산해 보면 다음과 같다.

$$\Pr\{(s=x)/(s=x \vee s=x/2)\} = \frac{\Pr[(s=x)\wedge((s=x)\vee(s=x/2))]}{\Pr((s=x)\vee(s=x/2))}$$

$$= \frac{\Pr(s=x)}{\Pr[(s=x)\vee(s=x/2)]}$$

$$= \frac{\frac{2^n}{3^{n+1}}}{\frac{2^n}{3^{n+1}} + \frac{2^{n-1}}{3^n}} = \frac{2^n}{2^n + 3.2^{n-1}}$$

$$= \frac{2}{5}$$

결국 이와 같은 확률분포의 경우 y가 2x일 확률은 2/5이고 y가 x/2일 확률은 3/5이다. 그러므로 내가 선택한 봉투의 금액이 x일 경우 봉투를 교환해서 얻을 수 있는 기대값은 다음과 같다.

$$E(y/x) = (2x \times \frac{2}{5}) + (\frac{x}{2} \times 3/5) = \frac{11x}{10} \quad (> x)$$

교환해서 얻을 수 있는 기대값이 x보다 크므로 이 경우 교환하는 것이 합리적이라는 결론이 나오고 이것은 분명히 역설적인 상황을 낳는다. 그런데 이 경우는 봉투의 금액의 최고액이 제한되어 있지 않을 뿐만 아니라, 어떤 봉투를 선택했을 때 기대할 수 있는 평균 금액을 계산할 수 없는 경우이다. 즉 E(A)의 값은 유한한 값이 아니기 때문에 계산할 수 없다. 그런 의미에서 이 경우에 발생하는 역설은 결국 찰머스의 말대로 "무한성의 기이한 행태로서 무한성의 익숙한 현상의 한 예"[12]에 불과한 것이라고 할 수 있을 것이다.

끝으로 E(B/A) - E(A)의 값이 무한대 발산하거나 진동하는 경우는 극한값 계산이 되지 않고, 따라서 이 경우는 봉투를 교환해서 얻을 수 있는 손익을 계산할 수 없으므로, 고려의 대상이 되지 않는다.

## 5. 맺음말

두 봉투의 역설에 대한 필자의 결론을 제시하기 전에, 그와 유사한 역설로 알려진 크라이취크(Kraitchick)의 역설에 대해서 생각해 보자.[13] 그것은 두 사람이 각자의 지갑에 든 돈의 액수에 관한 내기와 관련하여 발생하는 역설이다. 즉 두 사람이 다음과 같은 내기를 한다고 하자. 그 두 사람 중에서 자신의 지갑에 든 돈의 액수가 상대방의 지갑에 든 돈의 액수보다 적은 사람이 두 개의 지갑에 든 돈 전부를 갖는 것이다. 이 경우에 이 내기에 참여한 사람은 다음과 같이 추론할 수 있을 것이다. "내가 만약 잃는다면 내가 가진 액수를 잃을 것이고, 내가 만약 번다면 상대방의 액수를 벌 것이다. 그런데 내가 벌 금액은 내가 잃을 금액보다 항상 크고, 내가 잃거나 벌 확률은 동일하다. 따라서 이 내기에서 나의 기대값은 0보다 크고 나는 내기를 해야 한다."

그런데 두 사람 모두 이렇게 추론할 것이고, 그것은 돈의 총액이 고정되

12) D. Chalmers, "The Two-envelope Paradox: A Complete Analysis?" Internet at http://avocado.wustl.edu/~chalmers/chalmers.envelope.html (1996), 4쪽.

13) Christopher Langan, "Paradox Resolved: The Kritchick and Two Envelopes Paradoxes," Internet at http://www.megafoundation.org/Ubiquity/Paradox.html.

어 있는 일종의 제로섬(zero-sum) 게임에서 둘 모두 0보다 큰 기대값을 가질 수 있다면 그것은 분명히 역설이다.

이러한 역설을 낳는 추론에 어떤 문제가 있는가? 그 내기가 어떻게 끝나든지 자신의 지갑에 든 돈의 액수는 고정된(fixed) 것으로 가정하면서, 상대방의 지갑에 든 돈의 액수는 "게임의 결과에 따라 변하는 것으로, 즉 가변적인 것으로 가정하는 것이 문제이다. 예를 들어서 내 지갑 안에 x원, 상대방의 지갑에 y원이 들어 있다고 가정하고 위의 역설을 낳는 추론을 검토해 보자. 위 추론에 따르면, 내가 잃는다면 x원을 잃을 것이고, 내가 번다면 y원을 벌 것이다. 그런데 내가 이겨서 벌 때, y는 내가 가지고 있는 금액 x보다 크다. 그러므로 내기의 기대값은 0보다 크다는 것이다.

그러나 이 추론을 조금만 자세히 검토해보면 그 문제는 분명하게 드러난다. 즉 내가 이겨서 벌 때는 $y>x$이지만, 내가 내기에 져서 잃을 때는 $x>y$이다. 다시 말하면 이 두 사건은 두 개의 별개의 조건에 따라 발생하는 것이고, 다른 조건 하에서 금액들이 비교될 수는 없는 것이다. 요컨대 역설을 낳는 추론은 이길 경우와 질 경우라는 두 조건에서 내가 가지고 있는 금액 x는 고정되어 있고, y는 가변적인 것이라고 가정하고 추론을 하고 있는 것이다. 그러니까 옳은 추론 방식은 다음과 같이 계산하는 것이다. 즉 내가 $x_1$원의 돈을 가지고 있고 상대방이 y원을 가지고 있는데 $x_1$이 y보다 커서, 내가 내기에서 지고 돈을 잃을 경우 내가 잃는 금액은 $x_1$원이다. 반면에 내가 $x_2$원을 가지고 있고 상대방이 y원을 가지고 있는데 $x_2$가 y보다 작아서, 내가 내기에서 이기고 돈을 벌 경우 내가 벌 금액은 y원이다. 그런데 각각의 경우의 확률은 동일하기 때문에 그 기대값은 다음과 같다.

$x_1 > y \Rightarrow -x_1$

$x_2 < y \Rightarrow +y$ (y는 $x_1$보다 클 것이고 그 차이를 n이라고 하자.)

$$\therefore \frac{-x_1}{2} + \frac{x_1 - n}{2} = -\frac{n}{2} < 0$$

그리고 이 경우 상대방의 기대값은 다음과 같이 계산될 것이다.

$$x_1 > y \Rightarrow +x_1$$
$$x_2 < y \Rightarrow -y \quad \text{(위에서와 같이 } y = x_1 + n)$$
$$\therefore \frac{x_1}{2} - \frac{x_1 - n}{2} = \frac{n}{2} > 0$$

그러므로 A와 B의 기대값을 보면, 이 내기는 결국 제로섬 게임임을 알 수 있고, 역설을 낳는다는 추론은 게임의 결과에 따라 변할 수 있는 가변적인 금액을 임의로 고정적인 것으로 가정함으로써 생겨났음을 알 수 있다.

문제의 두 봉투의 경우 역설이 발생한다고 주장하는 추론은 크라이취크의 역설의 경우에서와 같은 잘못을 포함하고 있다.[14] 즉 두 봉투의 역설을 낳는다고 주장하는 추론은 "내가 잃는다면 내가 가진 것(x원)의 절반을 잃을 것이고, 내가 번다면 내가 가진 것의 두배를 벌고, 이 가능성은 각각 1/2로 동일하다"는 것이다. 이 추론의 문제도 크라이취크의 역설에서와 마찬가지로, 내가 선택한 봉투의 금액 x는 고정적이라고 가정하고 상대방의 봉투 안에 있는 금액은 가변적이라고 생각하는 점이다. 두 봉투의 경우를 약간 수정하면, 크라이취크의 상황과 비슷하게 만들 수 있다. 한 독지가가 순이와 철이에게 장학금을 주기로 한다고 하자. 그런데 그 독지가는 겉으로 구별이 불가능한 두 개의 봉투에 장학금을 넣었는데, 하나의 봉투에는 전체 장학금의 1/3을 다른 봉투에는 2/3를 넣었다. 순이와 철이가 두 봉투 중의 하나씩을 선택했다. 그런 다음 순이와 철이에게 교환할 기회를 주었다고 하자. 이 경우가 바로 정확하게 크라이취크의 경우와 같다. 결국 두 봉투의 경우에 역설이 발생한다고 추론하는 것은 크라이취크의 상황에서 역설이 발생한다고 주장하는 것과 같은데, 이는 명백하게 잘못이다.

요컨대 두 봉투의 역설에 대한 옳은 처방은 내가 이길 경우와 질 경우 모두에 대해서 상대방과 대칭적이며, 따라서 이것도 결국은 제로섬 게임이라는

14) 물론 크라이취크의 역설의 경우는 지갑에 무한한 돈이 들어있을 수 없다는 점에서 엄격하게 말하면, 두 봉투의 역설과는 구별된다. 그러므로 여기서 크라이취크의 역설과 비슷하다고 말하는 두 봉투의 역설은 유한한 금액이 들어 있는 경우이다.

점을 인정하면서 두 봉투의 경우에 대한 옳은 추론은 다음과 같이 해야 할 것이다.

$$E_2 = \{(x+d) \times \frac{1}{2}\} + \{(x-d) \times \frac{1}{2}\} = x$$

즉 A와 B 봉투에 들어 있는 금액의 차이를 d라고 하자. A에 x원이 들어 있다면 B에는 (x+d)원이나 (x-d)원이 들어 있다. 그런데 내가 교환을 해서 이익을 볼 확률과 손해를 볼 확률은 각각 1/2이고, 이익을 볼 때도 d원의 이익을 볼 것이고, 손해를 볼 때도 d원의 손해를 볼 것이다. 그러므로 교환할 때의 기대값은 x이고, 따라서 교환할 필요가 없다.

결론적으로 두 봉투의 역설이라고 불리는 것은 봉투에 들어있는 금액이 무한할 뿐만 아니라 각 봉투 안의 금액의 평균값이 무한한 경우만 발생할 수 있을 뿐이고, 이는 내가 선택한 봉투 안의 기대값을 실제로 계산할 수 없는 경우이다. 그것은 무한성의 특이성의 하나라고 여겨질 뿐이다. 그러나 기대값이 실제로 계산될 수 있는 그 외의 경우, 역설은 발생하지 않으며, 역설이 발생한다고 생각하는 것은 자신의 봉투 안의 금액을 고정적인 것으로, 상대방의 금액을 가변적인 것으로 해석하는 잘못으로부터 기인하는 것이다.[15]

---

15) 이러한 설명은 잭슨 등의 주장과 유사한 점이 있다. 그러나 잭슨 등은 두 봉투에 들어있는 금액이 유한한 경우로 논의를 제한했고, 또한 그들은 사람들이 그러한 경우에 역설이 생긴다고 생각하는 이유는 내가 선택한 봉투의 금액에 대해서 다른 봉투의 금액이 절반일 확률과 두 배일 확률이 동일하다고 생각하는 오류 때문이라고 주장하는 것으로 그치고 있다. 필자는 그렇다면 왜 사람들이 일반적으로 자신의 봉투의 금액에 대해서 다른 봉투의 금액이 절반일 확률과 두 배일 확률이 동일하다고 생각하는지를 설명하고자 하고, 그에 대해서 현실적 상황과 관련하여 자신의 봉투 안의 금액은 "고정적"이라고 보고 다른 봉투 안의 금액은 게임의 결과에 따라 변할 수 있는 "가변적"이라고 보는 잘못 때문이라고 대답하는 것이다.

* 이 논문은 2001년 논리학회 여름 정기발표회에서 발표한 것을 수정한 것이다. 발표회에서 많은 비판적 질문과 코멘트를 해주신 여러 선생님들께 감사를 드린다. 특히 이 논문을 심사해주신 익명의 심사위원의 비판은 논문을 보다 명료하게 하는 데 큰 도움이 되었음을 밝히며 감사를 드린다.

## 참고문헌

Arntzenius, F. & McCarthy, D. "The Two-envelope Paradox and Infinite Expectations" *Analysis* 57.1 (1997, Jan.)

Broome J. "The Two-envelope Paradox" *Analysis* 55.1 (1995, Jan.)

Castel, P. & Batens, D. "The Two-envelope Paradox: The Infinite Case" *Analysis* 54.1 (1995, Jan.)

Chalmers, D. "The Two-envelope Paradox: A Complete Analysis?" Internet at http://avocado.wustl.edu/~chalmers/chalmers.envelope.html (1996)

Clark, M. & Shackel N. "The Two-envelope Paradox" *Mind* (2000, June)

Jackson, F. & Menzies, P. & Oppy, G. "The Two-envelope 'Paradox'" *Analysis* 54.1 (1994, Jan.)

Jeffrey, R. *The Logic of Decision* (London: University of Chicago Press, 2nd ed. 1983).

Langan, C. "Paradox Resolved: The Kritchick and Two Envelopes Paradoxes," Internet at http://www.megafoundation.org/Ubiquity/Paradox.html.

Nalebuff, B. 'The other person's envelope is always greener' *Journal of Economic Perspectives* 3 (1983)

Scott Alexander D. & Scott, M. "What's in the Two Envelope Paradox?" *Analysis* 57.1 (1997, Jan.)