응용통계연구 (2010)
23(6), 1209–1216
DOI: 10.5351/KJAS.2010.23.6.1209

화투 섞기의 과학

허명회[1] · 이용구[2]

[1]고려대학교 통계학과, [2]중앙대학교 통계학과

(2010년 8월 접수, 2010년 9월 채택)

요 약

48장의 화투패가 섞어지는 과정을 수학적으로 기술하고 그것이 완전히 섞여지기 위해서 몇 번을 반복해 쳐야 하는가에 대한 답을 제시한다. 또한 그보다 작은 횟수로 치는 경우 예상되는 불완전한 임의성의 형태를 밝힌다.

주요용어: 카드 섞기, 임의순열, 포아송 분포.

1. 연구 배경과 목적

물리적 임의화(physical randomization)는 결코 쉬운 일이 아니다. 미국의 1970년 징병 추출(drat lottery)이 대표적인 예이다.

임의추출(random selection)을 실현하는 체계적 방법으로서 허명회와 이용구 (2010)는 추첨상자(lottery box) 모형을 제안한 바 있다. 임의추출 외에 카드 게임에 적용되는 섞기(shuffling), 즉 임의순열화도 매우 흥미로운 수학적 토픽이다.

게임 카드(playing cards) 52장의 리플셔플(riffle shuffle)에 대한 Bayer와 Diaconis (1992)의 연구로 상업적 카지노에서 게임 카드를 섞는 횟수가 완전임의화를 달성하기에 상당히 부족하다는 사실이 알려진 바 있다.

우리나라의 게임 카드인 화투(花鬪, flower cards 또는 hanafuda)의 임의순열화에 관한 확률적 연구는 아직 없었던 것으로 보인다. 본 연구의 주제는 확률적 관점에서 본 화투 섞기이다.

48장의 화투 섞기(HwaTu shuffling)는 다음과 같은 확률과정으로 기술될 수 있다.

1) 임의분할(random partitioning): 48장 묶음이 s개 소묶음으로 나뉜다. 이때 각 소묶음의 크기 $n_1, \ldots, n_s$는 等확률의 다항분포 multinomial$(48, \mu)$로부터 생성된다($\mu = 1/s, \ldots, 1/s$). 즉,

$$P(n_1, \ldots, n_s) = \frac{48!}{n_1! \cdots n_s!} \left(\frac{1}{s}\right)^{48}, \quad n_1 \geq 0, \ldots, n_s \geq 0 \ (n_1 + \cdots + n_s = 48).$$

2) 거꾸로 바꾸기(reverse substitution): 소묶음의 순서를 $1, \ldots, s$에서 $s, \ldots, 1$로 바꾼다.

3) 앞의 두 단계를 m회 반복한다.

수치 예를 보기로 하자. 화투 묶음을 위에서 아래로 일련번호화하여 $1, \ldots, 48$로 지칭한다.

제 1 반복: 48장의 화투가 크기가 14장, 19장, 15장인 3개($= s$) 소묶음으로 임의분할됐다고 하자. 즉,

[1]교신저자: (136-701) 서울시 성북구 안암동 5가 1, 고려대학교 통계학과, 교수. E-mail: stat420@korea.ac.kr

소묶음 1: 01 02 03 04 05 06 07 08 09 10 11 12 13 14,

소묶음 2: 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33,

소묶음 3: 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48.

이어서 묶음 순서가 거꾸로 바뀐다. 즉, 화투 배열이 위에서 아래로

34 35 36 37 38 39 40 41 42 43 44 45 46 47 48,

15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33,

01 02 03 04 05 06 07 08 09 10 11 12 13 14

가 된다. 이 배열에서 첫 장 34를 제외한 47장의 카드 중 모두 45장의 카드가 앞 장의 다음 번호임을 관찰할 수 있다. 또한 큰 번호들이 앞에 나타나고 작은 번호들이 뒤에 나타나는 경향이 발견된다(출발배열과의 상관계수 = –0.77).

제 2 반복: 이번에는 48장의 화투가 18장, 18장, 12장으로 나뉘었다면 소묶음은 다음과 같이 구성된다.

소묶음 1: 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 15 16 17,

소묶음 2: 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 01 02,

소묶음 3: 03 04 05 06 07 08 09 10 11 12 13 14.

묶음 순서를 거꾸로 하면 전체 배열은 다음과 같이 나타난다.

03 04 05 06 07 08 09 10 11 12 13 14,

18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 01 02,

34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 15 16 17.

이 배열에서 첫 장 03을 제외한 47장의 카드 중 모두 43장의 카드가 앞 장의 다음 번호임을 관찰할 수 있다. 또한 작은 번호들이 앞에 나타나고 큰 번호들이 뒤에 나타나는 경향이 발견된다(출발배열과의 상관계수 = 0.73).

이와 같이 화투 섞기에서는 연이은 카드 번호가 나타나는 횟수는 우연한 중복이 없다면 반복 1회당 $s-1$개씩 감소한다. 또한 출발배열과의 상관계수도 점차 0에 가까워진다.

이 연구의 목적은 48장 화투를 완전히 섞기 위해서 몇 번을 반복하여 쳐야 하는가에 대한 수학적 답을 제시하고 그보다 작은 횟수로 치는 경우 예상되는 불완전한 임의성의 형태를 파악하는 데 있다.

2절에서 출발배열과의 상관 계수와 연이은 번호의 출현 횟수 등 임의순열(random permutations)의 통계적 성질을 탐구한다. 3절에서 완전한 화투 섞기의 조건을 규명하고 4절에서는 불완전한 화투 섞기에서 예상되는 패턴을 살펴볼 것이다. 마지막 절에서는 게임의 관점에서 연구 결과에 의미를 부여한다.

2. 임의순열

출발배열 $\mathbf{a}\,(=1,\ldots,n)$와 임의순열(random permutation) $\boldsymbol{y}$간 상관 r이 취하는 확률분포를 몬테칼로 방법으로 구하였다(모의시행 수 100,000). 그 결과로 얻은 히스토그램이 그림 2.1이고 분위수로 다음을 얻었다.

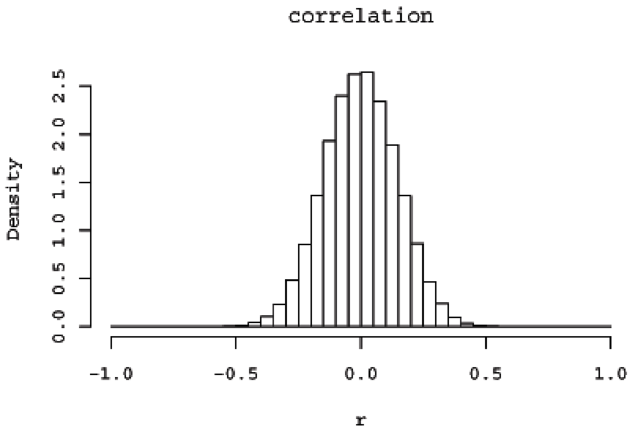

그림 **2.1.** 상관 r의 몬테칼로 분포(모의시행 수 100,000): 임의순열의 경우

1%	5%	10%	25%	50%	75%	90%	95%	99%
−0.337	−0.241	−0.188	−0.101	−0.001	0.099	0.187	0.240	0.336

참고로, 상관 r의 분포에 대한 점근이론은

$$\frac{1}{2}\log_e\frac{1+r}{1-r} \sim N\left(0, \frac{1}{n-3}\right)$$

이며 (Kenney와 Keeping, 1962), 이에 따른 분위수는 다음과 같다.

1%	5%	10%	25%	50%	75%	90%	95%	99%
−0.334	−0.240	−0.189	−0.100	0.000	0.100	0.189	0.240	0.334

순열 $(X_1, \ldots, X_n)$에서 연이은 번호의 출현횟수 Z를 다음과 같이 정의하기로 한다.

$$Z = \sum_{j=1}^{n-1} I(X_{j+1} = (X_j \mod n) + 1).$$

즉, n 다음에 1이 나오는 경우 연이은 번호의 출현으로 간주한다. 여기서 "$a \mod b$"는 a를 b로 나누었을 때의 나머지이다.

수치 예를 보자. 48매$(= n)$의 카드가 다음과 같이 카드가 섞였다고 하자.

19 20 02 08 16 38 34 28 07 44 10 04 22 17 18 31
05 32 03 36 40 48 01 41 15 14 45 37 26 30 35 13
42 29 25 06 46 39 23 27 09 47 21 33 11 12 24 43.

이 예에서 연이은 번호의 출현횟수 Z는 4이다(19 다음 20, 17 다음 18, 48 다음 01, 11 다음 12 등 총 4회).

일반적으로, Z는 다음과 같은 확률적 성질을 갖는다.

1) $E(Z) = 1$이다. 왜냐하면 $P\{X_{j+1} = (X_j \mod n) + 1\} = 1/(n-1)$이기 때문이다.
2) Z는 근사적으로 Poisson(1) 분포를 따른다. 그 이유는 다음과 같다.

$$P\{X_{j+1} = (X_j \mod n) + 1,\ X_{k+1} = (X_k \mod n) + 1\} = \frac{1}{n-1}\frac{1}{n-2}, \quad j \neq k$$

이므로, n이 큰 경우 $I(X_{j+1} = (X_j \mod n) + 1)$과 $I(X_{k+1} = (X_k \mod n) + 1)$이 독립에 가깝다($j \neq k$). 따라서 $n\,(= 48)$장의 카드 게임인 화투에서 Z의 확률분포는 Poisson(1) 분포와 비슷할 것으로 기대할 수 있다.

다음은 100,000번의 모의시행으로 얻은 Z의 몬테칼로 분포를 포아송(1) 분포에 대비시킨 표이다.

Z	0	1	2	3	4	5	6	7	8
모의시행 결과	0.366	0.368	0.185	0.061	0.016	0.003	0.001	0.000	0.000
포아송(1) 분포	0.368	0.368	0.184	0.061	0.015	0.003	0.001	0.000	0.000

3. 이상적인 섞기

48장 화투 뭉치를 s개 소뭉음화하고 거꾸로 바꾸는 섞기를 m회 반복하여 얻은 배열에서 출발배열과의 상관 r의 중간값과 연이은 번호의 출현 횟수 Z의 평균을 몬테칼로 모의시행으로 구해 보았다(모의시행 수 1,000).

다음은 $s = 4$, $m = 6, 12, 24, 48, 96, 144, 192$의 경우이다.

	상관계수 중간값	출현횟수의 평균
$s = 4, m = 6$	0.509	32.7
$s = 4, m = 12$	0.259	23.2
$s = 4, m = 24$	0.069	11.7
$s = 4, m = 48$	−0.006	3.6
$s = 4, m = 96$	−0.002	1.2
$s = 4, m = 144$	0.007	1.0
$s = 4, m = 192$	−0.007	1.0

반복수 m이 커짐에 따라 상관계수 r의 중간값이 0에 가까워지고 연이은 번호의 출현 횟수 Z의 평균이 1에 가까워짐을 볼 수 있다. 상관계수 r을 기준으로 하면 48회 반복이 필요하고 Z의 평균을 기준으로 하면 96회의 반복이 필요하다.

다음은 $s = 8$, $m = 6, 12, 24, 48, 96, 144, 192$의 경우이다.

	상관계수 중간값	출현횟수의 평균
$s = 8, m = 6$	0.788	19.0
$s = 8, m = 12$	0.686	8.5
$s = 8, m = 24$	0.404	2.4
$s = 8, m = 48$	0.158	1.0
$s = 8, m = 96$	0.027	1.0
$s = 8, m = 144$	0.005	1.0
$s = 8, m = 192$	−0.008	1.0

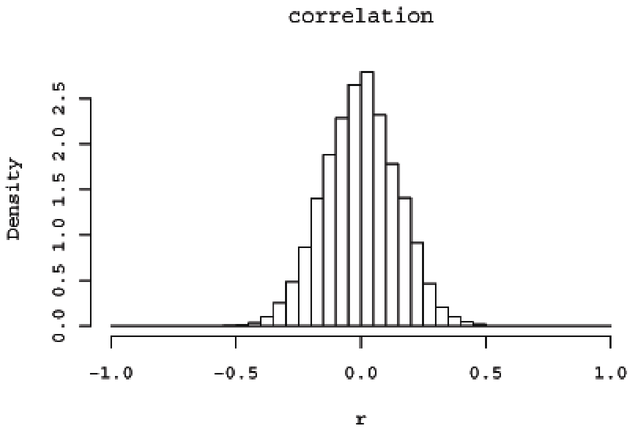

그림 **3.1.** 상관 r의 몬테칼로 분포(모의시행 수 10,000): $s = 4$, $m = 96$

이 경우에도 반복수 m이 커짐에 따라 상관계수 r의 중간값이 0에 가까워지고 연이은 번호의 출현 횟수 Z의 평균이 1에 가까워짐을 볼 수 있다. 상관계수 r을 기준으로 하면 96회 반복이 필요하고 Z의 평균을 기준으로 하면 48회의 반복이 필요하다.

이상의 모의시행 결과로부터 2개의 소결론이 얻어진다. 첫째, 소묶음 수 s가 클수록 연이은 번호의 출현 횟수 Z가 평균적으로 급속히 작아지게 되지만 출발배열과의 상관 r은 매우 느리게 0에 가까워지게 된다. 따라서 s를 크게 하는 것이 임의순열화를 위해 별 도움이 되지 않는다. 둘째, 소묶음 수 s의 현실적 범위에서 완전한 임의순열화를 위해 필요한 최소 반복수 m은 96회 정도이다. 그림 3.1은 $s = 4, m = 96$의 경우에서 상관계수 r의 분포이다(모의시행 수 10,000). 이 모의시행에서 얻은 상관 r의 주요 분위수는 다음과 같다.

1%	5%	10%	25%	50%	75%	90%	95%	99%
−0.333	−0.241	−0.189	–0.101	0.001	0.098	0.189	0.237	0.335

그림 3.1이 임의순열에서의 상관계수 r의 분포인 그림 2.1과 거의 일치함을 볼 수 있다.

모의시행에서 얻은 Z의 출현값 분포는 다음과 같다(평균 1.1, 분산 1.1). 대비를 위해 포아송(1) 분포를 제시하였다.

Z	0	1	2	3	4	5	6	7	8
모의시행 결과	0.318	0.363	0.209	0.081	0.024	0.004	0.001	0.000	0.000
포아송(1) 분포	0.368	0.368	0.184	0.061	0.015	0.003	0.001	0.000	0.000

$Z = 0$의 빈도에서 다소의 차이가 있기는 하지만 대체로 유사함을 볼 수 있다(이 차이는 반복수 m을 2배인 192회로 늘림으로써 해소된다).

소묶음의 수 s의 선택과 관련하여 임의순열화를 기대할 수 없는 3가지 상황을 특기한다.

1) $s = 2$인 경우에는 반복 수 m에 관계없이 연이은 번호의 출현 횟수 Z는 47로 고정된다. 출현 가능한 순열은 48개뿐이다. 따라서 소묶음의 수 s는 3 이상이여야 한다.

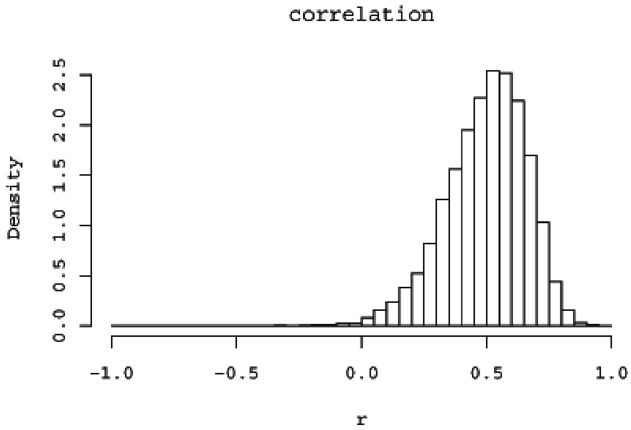

그림 **4.1.** 상관 r의 몬테칼로 분포(모의시행 수 100,000): $s = 4$, $m = 6$

2) s를 크게 하면 반복 수 m이 커지더라고 출발배열과의 상관 r이 ± 1 사이에서 진동하는 모습을 보인다.

3) 임의분할 대신 고정된 분할을 하는 경우, 예컨대 매번 12장, 12장, 12장, 12장으로 소뮦음화가 되는 경우 짝수번의 치기 후 카드 배열은 출발배열에 되돌아오므로 고정분할로써는 반복을 많이 하더라도 카드 배열이 임의순열화되지 않는다.

4. 섞기의 실제

실제 화투 섞는 과정을 관찰해보면 소뭉음의 수 s의 크기는 4 내외, 반복수 m은 6회 이하이다. $s = 4$, $m = 6$의 섞기에서 예상되는 화투 배열의 모습을 r과 Z로 살펴보기로 한다(모의시행 수 10,000). 상관 r의 히스토그램으로 그림 4.1을 얻었으며 주요 분위수는 다음과 같이 나타났다.

1%	5%	10%	25%	50%	75%	90%	95%	99%
0.069	0.206	0.283	0.397	0.514	0.612	0.689	0.729	0.800

모의시행에서 얻은 Z의 출현값 분포는 다음과 같다(평균 32.7, 분산 3.9).

Z	29−	30	31	32	33	34	35	36	37+
모의시행 결과	0.020	0.097	0.178	0.199	0.189	0.145	0.089	0.044	0.038

상관 r의 몬테칼로 분포와 Z의 출현값 분포에서 실제 섞기는 이상적 섞기와는 상당한 거리가 있음을 볼 수 있다.

5. 맺음말

화투를 4–8개 소뭉음 분할로 임의순열화하기 위해서는 96회 이상의 반복이 필요하다. 실제로는 4회 소뭉음 분할로 6회 정도 반복하여 치므로 화투 배열은 매우 불충분하게 임의화된다. 예컨대 연이은 번호의 출현횟수는 이상적 임의순열배열에서는 평균 1회이지만 실제로는 이 값을 훨씬 상회한다. 따라서 초

약, 풍약, 비약 등의 발생 빈도가 커지게 되는 등 게임 리스크가 증대되는 결과가 초래된다. 이 점이 화투를 게임으로서 더 흥미롭게 만드는 요소가 아닌가 싶다.

참고문헌

허명회, 이용구 (2010). 유한 순서열의 임의화, <응용통계연구>, **23**, 189–196.

Bayer, D. and Diaconis, P. (1992). Trailing the dovetail shuffle to its lair, *Annals of Applied Probability*, **2**, 294–313.

Kenney, J. F. and Keeping, E. S. (1962). *Mathematics of Statistics*, Part 1, 3rd Edition, Princeton, New Jersey Van Nostrand, p.266.

The Science of HwaTu Card Shuffling

Myung-Hoe Huh[1] · Yonggoo Lee[2]

[1]Department of Statistics, Korea University; [2]Department of Statistics, ChungAng University

(Received August 2010; accepted September 2010)

Abstract

We describe shuffling process of 48 HwaTu(Flower) cards from a mathematical aspect and give the number of shuffles that are needed to arrive at a state of randomness. We observe that the cards are shuffled much less times in usual plays, that results in much riskier outcomes.

Keywords: Card shuffling, random permutations, Poisson distribution, HwaTu Cards.

[1]Corresponding author: Professor, Department of Statistics, Korea University. 5-1 Anam-dong, Sungbuk-Gu, Seoul 136-701, Korea. E-mail: stat420@korea.ac.kr