### Reviewer A

자세한 리뷰에 대해서 감사드립니다. 지적사항에 대한 답변은 다음과 같습니다.

Major Comments

1. 첫번째 모의 실험에서는 복권 구매자의 복권 번호 선호 현상이 1등 당첨자수의 분포에 끼치는 영향을 알아보기 위해 1등 당첨자 수의 분포를 알아내기 위해 *D=*500회 반복을 실시했습니다. 하지만 *K=*25일때 선호하는 번호의 조합에서 당첨이 나올 확률은 으로 500회 반복 중에서 선호하는 번호의 조합에서 당첨이 나오는 경우의 수의 기댓값이 매우 낮습니다. 이를 보완하기 위해 좀 더 많은 횟수의 반복이 필요해보입니다. 참고로 *K=*30일때는 이고 *K=*35일때는 이어서 500회 반복에서도 선호하는 번호의 조합에서 나오는 경우의 수에 대한 기댓값이 충분히 큰 값을 가집니다.

복권 구매자를 1억명으로 가정하였기에 D를 너무 크게 할 경우 컴퓨팅 타임이 많이 소요된다는 점과 11월 16일 기준으로 D=1146인점을 고려해서 D=1000번으로 늘였습니다.

1. 첫번째 모의실험에서 가정한 구매패턴이 현실을 충분히 반영하는지에 대한 설명이 부족합니다. Table 1에 나와있는 가장 많이 구매된 번호조합의 빈도를 보면 모의실험에서 가정한 구매패턴으로는 나올 수 없는 결과라고 생각됩니다. 왜 그러한 구매패턴을 가정했는지 에 대한 설명을 추가해주시기 바랍니다.

본 연구의 주요 목적은 실제 구매패턴을 파악하는 것이 아니라 랜덤하지 않은 구매패턴이 다수의 1등 당첨자가 나오는 원인이 된다는 것을 설명하는데 목표가 있습니다. 실제 서울대 통계연구소에서 로또의 공정성에 관한 용역연구를 진행했는데 동행복권 위원회에서 구매패턴 자료도 제공한 것으로 알고 있습니다. 본 연구는 용역과제와는 무관하게 진행되었고 구매패턴에 관한 연구는 Haigh (1997)과 Lien and Yuan (2015)에 자세히 나와 있는데 구매패턴은 향후 연구주제로 남겨두는 것으로 하고 아래 내용을 결론 마지막 부분에 추가하였습니다.

“한편, 동행복권 위원회에서는 구매패턴에 관한 자료가 있지만 복권 구매패턴에 영향을 주지 않기 위해서 공표하지 않고 있다. 추가적인 정보가 주어진다면, Haigh (1997)과 Lien and Yuan (2015)에서와 같이 구매자들의 구매패턴을 알아보고 이를 통한 보다 현실에 부합하는 시뮬레이션을 하는 것을 고려할 수 있다. 본 논문에서는 랜덤하지 않은 구매패턴이 다수의 1등 당첨자가 나오는 원인이 된다는 것을 설명하는데 목표가

있으므로 구매패턴에 관한 연구는 향후 연구주제로 남겨두고자 한다.”

1. 두번째 모의실험에서는 선호하는 번호의 집합에 따른 1등 당첨자 수의 분포 변화를 분석했습니다. 하지만 모의실험 세팅에서 이미 45개 번호들 간의 대칭성이 만족되므로 2번 시뮬례이션은 단순히 1번 시뮬레이션의 반복에 지나지 않으며 굳이 반복하여 확인할 필요가 없는 명제라고 생각됩니다.

이 논문은 통계학 전공자 뿐만 아니라 일반인들도 이해할 수 있도록 하는 것을 목표로 하였기 때문에 두번째 모의실험이 필요하다고 생각됩니다. 즉 첫번째 모의실험에서 관측되는 현상이 특정번호에 국한되어 발생하는 것이 아니라는 사실을 일반인들이 보다 쉽게 이해하기 위해서 두번째 모의실험이 추가되었다고 생각하시면 됩니다. 사실 방송국과 신문사 기자들이 로또 당첨자가 다수 나오는 것을 보다 쉽게 설명해 달라는 다수의 요청을 받은 것이 이 논문 작성의 동기중 하나이기도 합니다.

1. 세번째 모의실험에서는 복권이 공정하지 않을 경우 1등 당첨자 수의 분포를 분석하였습니다. 하지만 복권이 공정하지 않은 경우 복권 번호의 추출이 어떻게 이뤄지는지에 대한 설명이 없으며 이로 인해 모의 실험 결과를 해석할 수가 없습니다. 구체적으로 9페이지에서 “즉 다음과 같은 비복원 추철을 가정하고” 부분에서 어떤식으로 추출이 실행됐는지 묘사되어 있지 않습니다.

해당문장은 다음과 같이 수정되었습니다.

“ 마지막으로 로또가 공정하지 않다는 가정을 하고 일부 구매자들이 사전정보를 통하여 어떤 번호가 많이 나오는지 알았다고 가정해보자. 즉 이전의 시뮬레이션 절차의 각 단계가 동일하지만, 복권 당첨번호가 $p$의 확률로 사람들이 선호하는 번호 집단 $F$에서 추첨되는 새로운 상황을 고려하자. ”

1. 추가적으로 현재는 톡정 번호 집단 에 속하는 숫자들만으로 이루어진 조합을 선택하는 복권구매자 집단을 가정했는데 집단을 소형화 다양화하여 모의실험을 진행해보는 것도 괜찮을 것 같습니다. 즉 여러 번호 집단을 고려하고 () 각 번 집단에 속하는 숫자들만으로 이루어진 조합을 선택하는 복권구매자들의 그룹을 가정한다면 현실의 복권구매 패턴에 조금 더 가까운 구매패턴을 구현할 수 있을 것으로 생각됩니다.

2번과 3번의 답변에서 이미 언급했듯이 본 논문은 통계학전공자 뿐만 아니라 일반인들도 대상으로 하였고 구매패턴에 관한 연구는 별도의 향후 연구주제로 적절하다고 생각되어서 이 논문의 범위를 벗어나는 것으로 생각됩니다.

Minor Comments

1. Page 3 (line 3): ‘기존에’ → ‘기존의 ’

제안하신 대로 수정하였습니다.

1. Page 5: ‘숫자쌍 {i, j}’ → ‘숫자쌍 {i, j}’

제안하신 대로 수정하였습니다.

1. Page 5 (식 (2.3)):

제안하신 대로 수정하였습니다.

1. Page 9 (line 4) ‘선호하는 번호과 관련이 없다는 것을’ → ‘선호하는 번호와 관련이 없다는 것을’

제안하신 대로 수정하였습니다.

### Reviewer B

이 논문은 사회적/실용적 문제에 대해 엄밀한 통계적 평가를 통해 설득력 있는 결론에 도달하였습니다. 이 연구는 통계적 검정, 시뮬레이션, 모델링 등 통계 방법론의 활용 사례를 명확히 보여주고 있어 응용통계 연구 및 교육에 기여할 수 있는 잠재력이 있어 보입니다. 아래 minor point들만 수정하면 응용통계연구 출판이 적절하다고 생각합니다.

감사합니다. 지적사항에 대한 답변은 다음과 같습니다.

Suggestion

1. 5쪽, 2.1절 제목 바로 아랫줄 첫 문장: $i=1, ldots N$ --> $i=1, ldots , N$ 수정이 필요하고, 문장의 비문여부 확인도 필요합니다.

제안하신 대로 수정을 하였으며 해당 문장은 다음과 같이 바꾸었습니다.

“총 D회차의 k/N 로또 추첨에서 각 번호가 뽑힐 확률이 동일한지 여부에 관한 적합도 검정의 피어슨 카이제곱 검정통계량은 다음과 같이 정의된다.

여기서 Oi는 i번째 번호가 D회차 동안 관측된 횟수이며, Ei은 k/N 로또에서 i번째 번호가 D회차 동안 나타날 기대횟수를 나타낸다.

1. 수식 번호 중에서 본문에서 언급되지 않는 번호는 삭제가 필요합니다 (예를 들어 2.5, 2.6).

본문에서 언급되지 않은 수식은 삭제하였습니다.

1. 7쪽, Table 3아래: 선호되는 번호집단 F의 원소의 개수가 K로 정의되어 있는데, 독자 입장에서는 본문 앞의 small k ("총 D회차의 k/N 로또")와 혼동의 가능성이 커 보입니다. small k나 large K 중 하나의 알파벳을 변경하면 좋을 것 같습니다.

*F*의 원소의 개수 *K*를 모두 *M*으로 변경하였습니다.

1. 9쪽, 3번째 문단, "비율 p", "p=0.3" --> p가 수식으로 처리되어야 할까요?

수식으로 변경하였습니다.

1. 10쪽, 밑에서 두번째 줄: 촤근 복권1등 --> 최근 복권 1등?

제안하신 대로 수정하였습니다.

1. 10쪽, 밑에서 두번째 줄: "복권1등 당첨자 수가 다수 발생하는 현상은 복권 구매자의 특정번호 선택으로 인한다는 것을 보였다": 복권 구매자의 특정번호 선택은 가능한 이유 중의 하나이므로, 주장의 tone down이 필요합니다 (e.g.... 다수 발생하는 현상의 이유 중 하나로, 복권 구매자의 특정번호 선택이 1등 당첨자 수를 증폭시킬 수 있음을 예증하였다.)

제안하신 대로 수정하였습니다.

1. 국문초록: 구체적으로 본 연구 당첨자가 .. -> 구체적으로 본 연구는 당첨자가 .. ?

제안하신 대로 수정하였습니다.