The Korean Journal of Applied Statistics (2014) DOI: http://dx.doi.org/10.5351/KJAS.2014.27.1.115
27(1), 115–122

A Test for Randomness of the Binary Random Sequence

In-Kwon Yeo[a,1]

[a]Department of Statistics, Sookmyung Women's University

(Received November 18, 2013; Revised December 16, 2013; Accepted January 9, 2014)

Abstract

A test for randomness of the binary random sequence is proposed in this paper. The proposed test statistic is based on the mean length of runs distributed with truncated geometric distribution and asymptotically χ^2_2-distributed when the size of the sequences is large. A small Monte Carlo simulation compared the size of the test with a significant level as well as evaluated the test power. We applied the proposed method to the sequence of yes or no numbers in Lotto 6/45 and concluded that the randomness of Lotto is retained.

Keywords: Length of runs, Lotto 6/45, truncated geometric distribution.

1. 서론

로또 6/45는 사회적으로 사행성, 한탕주의와 같은 많은 부작용도 낳았지만 저소득층이나 소외 계층을 위한 각종 복지사업이나 임대주택 건설 등 다양한 사업의 재원으로 사용되고 있다. 2002년 12월 7일 처음 추첨을 실시된 이후 현재까지 각종 사건과 이슈를 만들어 왔고 많은 사람들에게 인생역전이라는 환상을 주며 운영되고 있다. 엄청난 금액의 누적당첨금으로 로또광풍이 불기도 했으며 당첨번호를 예측하여 회원에게 제공해 주는 회사도 우우죽순처럼 생겨나고 있다. 또한 당첨번호의 조작이나 추첨 후에도 해당 회차의 복권을 발행 할 수 있다는 의혹들이 인터넷이나 언론을 통해 제기되기도 했다.

당첨번호에 대한 균일성(uniformity)이나 무작위성(randomness)에 대한 연구는 Genest 등 (2002), Helman (2005) 등에 의해 연구되었다. 특히 Johnson과 Klotz (1993)는 Lotto America Megabucks Lottery의 200회의 당첨번호를 사용하여 54개 각 번호가 당첨번호에 속할 확률을 최대가능도추정법을 통해 추정하고 이를 이용하여 균일성 검정을 실시하였다. 그 결과 p-값이 0.084로 확실히 균일하다고 보기 어려우며 낮은 번호의 공이 조금 많이 선택되는 것을 보아 이는 매 회 공을 넣을 때 크기순서대로 넣었기 때문일 것으로 추측했다. 우리나라의 경우 Kim (2004)는 71회까지의 당첨번호에서 가장 많이 나온 번호의 표집분포를 몬테칼로 모의실험을 통해 유도하여 균일성에 위배되지 않은 것을 보였으며 Lim과 Baek (2009)는 331회까지의 당첨번호에 대해 순서통계량 평균을 이용한 Coronel-Brizio 등 (2008)의 방법을 이용하여 로또 추첨이 무작위적인 것을 확인했다.

이 논문에서는 임의의 어떤 번호가 매 회에서 전 회에서의 추첨여부와 관계없이 독립적으로 추첨되는지를 검정하는 방법에 대해 알아본다. 이 번호가 어떤 회에서 추첨되었으면 1, 아니면 0이라고 표시하

This Research was supported by the Sookmyung Women's University Research Grants 2012.
[1]Professor, Department of Statistics, Sookmyung Women's University, Chongpa-dong 2ga, Yongsan-gu, Seoul 140-742, Korea. E-mail: inkwon@sm.ac.kr

면 자료는 이진확률수열(binary random sequence) 구조를 가진다. 이러한 수열에서의 독립성은 연검정(run test)을 통해 확인할 수 있으나 연검정의 경우 1이 발생할 확률 π가 0.5일 때 사용된다. 이 논문에서는 π가 임의의 값을 가질 때에도 이진확률수열의 무작위성을 검정하기 위한 검정통계량을 제안하고 이 통계량의 점근적 성질에 대해 알아본다. 이 방법은 로또뿐만 아니라 각종 컴퓨터 프로그램에서 제공하는 난수의 무작위성 여부를 확인하는데 사용될 수 있다. Rukhin 등 (2010)에는 적합도검정, 연검정, 최대연길이 등 15가지의 난수검정방법이 제시되어 있으나 우리가 알아보고자 하는 형태를 가진 이진확률수열의 무작위성을 검정하는 방법을 제시하지 않고 있다.

2. 이진확률수열의 무작위성 검정통계량

$\{X_i\}_{i=1}^n$는 0과 1로 이루어진 이진확률수열이라고 하자. 이 수열이 베르누이 시행에 의한 수열이라고 하면 각각의 X_i는 독립이고 1일 확률이 π가 된다. 로또에서 어떤 번호가 i번째 회에서 당첨번호로 뽑히면 1 아니면 0이라고 하자. 이 때 1이 추출될 확률이 6/45인 것처럼 이 논문에서는 π의 값을 알고 있다고 가정한다. 확률 π가 1/2이라고 하면 연(run)의 개수를 이용하여 독립성을 검정하는 연검정을 실시할 수 있으나 π가 1/2이 아닌 경우에는 연검정으로 독립성을 확인하는데 무리가 있다. 이 논문에서는 연의 개수 대신 연의 길이를 이용하여 π가 1/2이 아닌 경우에도 적용할 수 있는 검정통계량을 소개하고자 한다.

통계량 Y_j^+는 j번째 1로 이루어진 연의 길이, Y_j^-는 j번째 0으로 이루어진 연의 길이라고 하자. 예를 들어, 이진확률수열이 다음과 같이 이루어졌다고 하면

$$111\ 00\ 1\ 00\ 11\ 0\ 1\ 000\ 1$$

$Y_1^+=3$, $Y_2^+=1$, $Y_3^+=2$, $Y_4^+=1$, $Y_5^+=1$이 되고 $Y_1^-=2$, $Y_2^-=2$, $Y_3^-=1$, $Y_4^-=3$이 된다. 이때 n^+와 n^-를 각각 1과 0으로 이루어진 런의 개수라고 하면 위의 예제의 경우 $n^+=5$, $n^-=4$가 된다. 연은 순환구조를 가지기 때문에 $\left|n^+-n^-\right|\le 1$이 된다. 만약 $\{X_i\}_{i=1}^n$가 베르누이 확률변수라고 하면 각각의 Y_j^+는 0이 나올 때까지 베르누이를 반복할 때의 1의 개수이고 Y_j^-는 1이 나올 때까지의 0의 개수로 n이 무한이 커진다면 기하분포를 이용하여 모형화 할 수 있다. 즉 Y_j^+는 절사된 기하분포로

$$P\left(Y_j^+=y|Y_j^+\ge 1\right)=\frac{P\left(Y_j^+=y\right)}{1-P\left(Y_j^+=0\right)}=\frac{\pi^y(1-\pi)}{1-(1-\pi)}=\pi^{y-1}(1-\pi),\quad y=1,2,\ldots$$

가 된다. 마찬가지로 Y_j^-의 분포는 다음과 같다.

$$P\left(Y_j^-=y|Y_j^-\ge 1\right)=\pi(1-\pi)^{y-1},\quad y=1,2,\ldots,$$

여기서 Y_j^+와 Y_j^-의 평균은 각각 $1/(1-\pi)$과 $1/\pi$이고 분산은 $\pi/(1-\pi)^2$와 $(1-\pi)/\pi^2$가 된다. 만약 n^+와 n^-가 커지면 중심극한정리에 의해

$$Z_+=\frac{\bar{Y}^+-1/(1-\pi)}{\sqrt{\pi}/(\sqrt{n^+}(1-\pi))}=\frac{(1-\pi)\bar{Y}^+-1}{\sqrt{\pi/n^+}},\qquad Z_-=\frac{\bar{Y}^--1/\pi}{\sqrt{1-\pi}/(\sqrt{n^-}\pi)}=\frac{\pi\bar{Y}^+-1}{\sqrt{(1-\pi)/n^-}}$$

는 점근적으로 표준정규분포를 따른다. 여기서 $\bar{Y}^+=\sum_j Y_j^+/n^+$이고 $\bar{Y}^-=\sum_j Y_j^-/n^-$이다.

표본크기 n이 유한하더라도 매우 크다면 근사적으로 위의 결과들을 활용할 수 있다. 여기서 한 가지 주의해야 할 것은 실제분석에 있어서는 n이 유한하기 때문에 마지막 연의 길이는 더 길어질 수 있으나 n에 의해 강제로 잘릴 수 있다. 그러므로 위의 예제에서 Y_5^+의 성질은 앞의 Y_j^+와는 통계적 성질이 다르다

고 볼 수 있다. 즉, n^+와 n^-가 2보다 작은 경우를 제외하고 마지막 연의 길이는 해당 값을 가지는 앞의 연 길이와 통계적 성질이 조금 차이가 난다. 그러므로 이 논문에서는 마지막에 해당되는 연의 경우 분석에서 제외하기로 한다.

결론적으로 수열의 크기 n이 큰 경우 Y_j^+와 Y_j^-는 점근적으로 독립이기 때문에, 이 논문에서 제안하는 검정통계량은

$$Z^2 = Z_+^2 + Z_-^2$$

이며 이는 n^+와 n^-가 클 때 점근적으로 자유도가 2인 카이제곱분포를 따른다.

3. 모의실험과 실증분석

제안 방법에 대한 타당성을 보이기 위해 제안 검정통계량의 검정크기(test size)와 검정력에 대한 모의실험을 실시하였다. 이진확률수열을 생성하기 위해 아래와 같은 결합분포를 따른다고 가정하였다.

$$\begin{aligned}
&P(X_i = 1, X_{i+1} = 1) = \pi^2 + \pi(1-\pi)\rho, \\
&P(X_i = 0, X_{i+1} = 1) = P(X_i = 1, X_{i+1} = 0) = \pi(1-\pi)(1-\rho), \\
&P(X_i = 0, X_{i+1} = 0) = (1-\pi)^2 + \pi(1-\pi)\rho.
\end{aligned}$$

이 분포 하에서 X_i와 X_{i+1}의 상관계수는 ρ가 되며 $\rho = 0$이면 1일 확률이 π인 베르누이 확률변수가 된다. 참고로 위의 결합확률이 0에서 1사이에 있어야 하기 때문에 $-\min(\pi/(1-\pi), (1-\pi)/\pi) \le \rho \le 1$를 만족해야 한다. 조건부 확률의 정의에 의해

$$\begin{aligned}
P(X_{i+1} = 1 | X_i = 1) &= \frac{\pi^2 + \pi(1-\pi)\rho}{\pi} = \pi + (1-\pi)\rho, \\
P(X_{i+1} = 1 | X_i = 0) &= \frac{\pi(1-\pi)(1-\rho)}{1-\pi} = \pi(1-\rho)
\end{aligned}$$

이므로 다음과 같은 단계를 거쳐 이진난수를 발생시켰다.

- 단계 1: 초기값으로 1일 확률이 π 이진난수를 생성.
- 단계 2: 앞의 난수가 1이면 1일 확률이 $\pi + (1-\pi)\rho$, 0이면 1일 확률이 $\pi(1-\rho)$인 이진난수를 생성.
- 단계 3: 단계 2를 n개의 난수를 얻을 때까지 반복.

모의실험은 R의 rbinom함수를 이용했으며 아래의 결과에 의하면 무작위성에 문제가 없는 것으로 나타났다. Table 3.1, 3.2, 3.3은 n = 100, 300, 500, 1000이고 $\rho = -0.3, -0.2, \ldots, 0.3$일 때 위의 ρ의 구간이 만족하는 π = 0.1, 0.2, 0.3, 0.4, 0.5에 대해 각각 10000번씩 반복하여 얻어진 모의실험한 결과이다. $n = 100$인 경우 n^+와 n^-가 크지 않아 안정적인 결과라고 할 수 없으나 $n = 300$ 이상인 경우 앞 절의 이론을 적용하는데 문제가 없는 것으로 나타났다. Table 3.1과 Table 3.2에서 $\bar{n}^+$와 $\bar{n}^-$는 n^+와 n^-들의 평균, $\bar{\bar{y}}^+$와 $\bar{\bar{y}}^-$는 $\bar{y}^+$와 $\bar{y}^-$들의 평균, s^+와 s^-는 $\bar{y}^+$와 $\bar{y}^-$들의 표준편차를 표시한 것으로 각각의 ρ에서 π가 작아질수록 $\bar{n}^+$와 $\bar{n}^-$가 작아지고 $\bar{\bar{y}}^+$가 $\bar{\bar{y}}^-$보다 작아지는 것을 볼 수 있으며 s^+는 작아지나 s^-는 오히려 커지는 경향이 있다. 이것은 π이 작아진다는 것은 반대로 $1-\pi$가 커지기 때문에 0의 연 길이가 길어지는 경향이 있어 위의 결과가 발생하는 것이다. 또한 동일한 π일지라도 ρ 값이 작아질수록 $\bar{\bar{y}}^+$, $\bar{\bar{y}}^-$, s^+, s^- 모두 작아지는 경향이 있다.

Table 3.1. Simulation results(the average number of runs, average and standard deviation of run's length when $n = 100$ and 300)

ρ	π	100						300					
		$\bar{n}^+$	$\bar{y}^+$	s^+	$\bar{n}^-$	$\bar{y}^-$	s^-	$\bar{n}^+$	$\bar{y}^+$	s^+	$\bar{n}^-$	$\bar{y}^-$	s^-
−0.3	0.3	26.6	1.1	0.06	27.0	2.6	0.40	81.3	1.1	0.04	81.7	2.56	0.23
	0.4	30.6	1.3	0.11	30.8	1.9	0.24	93.0	1.3	0.06	93.2	1.92	0.14
	0.5	32.0	1.5	0.16	32.0	1.5	0.16	97.0	1.5	0.09	97.0	1.54	0.09
−0.2	0.2	18.4	1.0	0.05	19.0	4.2	0.88	56.8	1.0	0.03	57.4	4.17	0.49
	0.3	24.5	1.2	0.10	24.9	2.8	0.47	74.9	1.2	0.06	75.3	2.78	0.26
	0.4	28.2	1.4	0.14	28.4	2.1	0.29	85.8	1.4	0.08	86.0	2.08	0.16
	0.5	29.5	1.7	0.20	29.5	1.7	0.20	89.4	1.7	0.11	89.4	1.67	0.11
−0.1	0.1	9.0	1.0	0.03	9.8	9.1	3.18	28.7	1.0	0.02	29.5	9.12	1.68
	0.2	16.8	1.1	0.10	17.4	4.5	1.02	52.0	1.1	0.05	52.6	4.55	0.56
	0.3	22.4	1.3	0.13	22.8	3.0	0.54	68.6	1.3	0.08	69.0	3.03	0.30
	0.4	25.8	1.5	0.18	26.0	2.3	0.34	78.6	1.5	0.10	78.8	2.27	0.19
	0.5	27.0	1.8	0.24	27.0	1.8	0.23	81.9	1.8	0.14	81.9	1.82	0.14
0	0.1	8.1	1.1	0.13	8.9	9.9	3.62	26.1	1.1	0.07	26.9	10.00	1.95
	0.2	15.2	1.2	0.15	15.8	5.0	1.21	47.2	1.2	0.08	47.8	5.00	0.66
	0.3	20.3	1.4	0.18	20.7	3.3	0.63	62.3	1.4	0.10	62.7	3.33	0.36
	0.4	23.4	1.7	0.22	23.6	2.5	0.40	71.4	1.7	0.12	71.6	2.50	0.23
	0.5	24.5	2.0	0.29	24.5	2.0	0.29	74.5	2.0	0.16	74.5	2.00	0.16
0.1	0.1	7.2	1.2	0.22	8.0	11.0	4.49	23.4	1.2	0.11	24.2	11.11	2.24
	0.2	13.6	1.4	0.20	14.2	5.5	1.45	42.4	1.4	0.11	43.0	5.56	0.79
	0.3	18.2	1.6	0.23	18.6	3.7	0.77	56.0	1.6	0.13	56.4	3.71	0.43
	0.4	21.1	1.9	0.28	21.3	2.8	0.50	64.3	1.9	0.16	64.5	2.77	0.28
	0.5	22.0	2.2	0.36	22.0	2.2	0.37	66.9	2.2	0.20	67.0	2.22	0.20
0.2	0.1	6.3	1.4	0.33	7.1	12.4	5.38	20.7	1.4	0.17	21.5	12.52	2.80
	0.2	12.0	1.6	0.28	12.6	6.3	1.78	37.6	1.6	0.16	38.2	6.25	0.97
	0.3	16.1	1.8	0.30	16.5	4.2	0.93	49.7	1.8	0.17	50.1	4.17	0.53
	0.4	18.6	2.1	0.36	18.8	3.1	0.61	57.0	2.1	0.20	57.2	3.13	0.34
	0.5	19.5	2.5	0.45	19.5	2.5	0.46	59.6	2.5	0.25	59.6	2.50	0.25
0.3	0.1	5.4	1.6	0.47	6.2	14.0	6.34	18.0	1.6	0.23	18.8	14.33	3.43
	0.2	10.4	1.8	0.39	11.0	7.1	2.22	32.9	1.8	0.21	33.5	7.12	1.19
	0.3	14.0	2.0	0.40	14.4	4.8	1.23	43.4	2.0	0.22	43.8	4.76	0.65
	0.4	16.2	2.4	0.46	16.4	3.6	0.79	49.9	2.4	0.26	50.1	3.57	0.43
	0.5	17.0	2.9	0.58	17.0	2.9	0.57	51.9	2.9	0.32	51.9	2.86	0.32

Table 3.3은 유의수준 5%와 1%에서의 검정의 크기(size of test)와 검정력이 얼마나 되는지를 알아보기 위한 모의실험 결과를 정리한 것이다. 검정의 크기는 ρ가 0인 경우로 유의수준 0.05나 0.01와 큰 차이가 나지 않은 것을 볼 수 있다. 여기서 $n = 300$ 이상인 경우에는 동일한 π에서 ρ가 0에서 멀어질수록 검정력이 커지는 것으로 나타났다. 앞으로 개발되는 검정법의 검정력은 이 Table 3.3과 비교되어야 할 것이다.

Table 3.4는 2002년 12월 7일 1회부터 2013년 10월 26일 569회까지의 로또6/45 자료를 이용하여 각각의 번호에 대하여 무작위성을 검정한 것이다. 총 45개의 n^+와 n^-의 평균은 각각 4.13, 5.02로 n^-의 평균이 큰 것은 나타났는데 이것은 첫 번째 자료가 1인 것이 6개인 반면 0인 것은 39개로 0으로 시작되는

Table 3.2. Simulation results(the average number of runs, average and standard deviation of run's length when $n = 500$ and 1000)

ρ	π	500						1000					
		$\bar{n}^+$	$\bar{\bar{y}}^+$	s^+	$\bar{n}^-$	$\bar{\bar{y}}^-$	s^-	$\bar{n}^+$	$\bar{\bar{y}}^+$	s^+	$\bar{n}^-$	$\bar{\bar{y}}^-$	s^-
-0.3	0.3	135.7	1.10	0.03	136.1	2.57	0.17	272.4	1.10	0.02	272.8	2.56	0.12
	0.4	155.4	1.28	0.05	155.5	1.92	0.11	311.3	1.28	0.03	311.5	1.92	0.08
	0.5	162.0	1.54	0.07	162.0	1.54	0.07	324.4	1.54	0.05	324.4	1.54	0.05
-0.2	0.2	95.3	1.04	0.02	95.9	4.16	0.38	191.4	1.04	0.02	192.0	4.16	0.27
	0.3	125.3	1.19	0.04	125.7	2.78	0.20	251.3	1.19	0.03	251.7	2.78	0.14
	0.4	143.4	1.39	0.06	143.6	2.08	0.13	287.3	1.39	0.04	287.5	2.08	0.09
	0.5	149.5	1.67	0.09	149.5	1.67	0.09	299.5	1.67	0.06	299.5	1.67	0.06
-0.1	0.1	48.6	1.01	0.01	49.4	9.09	1.25	98.2	1.01	0.01	99.0	9.08	0.87
	0.2	87.2	1.14	0.04	87.8	4.55	0.43	175.2	1.14	0.03	175.8	4.55	0.31
	0.3	114.8	1.30	0.06	115.2	3.03	0.23	230.4	1.30	0.04	230.8	3.03	0.16
	0.4	131.4	1.52	0.08	131.6	2.27	0.15	263.4	1.52	0.05	263.7	2.27	0.10
	0.5	137.0	1.82	0.11	137.0	1.82	0.10	274.6	1.82	0.07	274.6	1.82	0.07
0	0.1	44.1	1.11	0.05	44.9	10.01	1.44	89.2	1.11	0.04	90.0	9.98	1.01
	0.2	79.2	1.25	0.06	79.8	5.00	0.49	159.2	1.25	0.04	159.8	5.00	0.35
	0.3	104.4	1.43	0.08	104.8	3.33	0.28	209.2	1.43	0.05	209.6	3.34	0.19
	0.4	119.3	1.67	0.10	119.5	2.50	0.18	239.4	1.67	0.07	239.6	2.50	0.12
	0.5	124.5	2.00	0.13	124.5	2.00	0.13	249.5	2.00	0.09	249.5	2.00	0.09
0.1	0.1	39.6	1.23	0.09	40.4	11.09	1.70	80.1	1.24	0.06	80.9	11.11	1.20
	0.2	71.2	1.39	0.09	71.8	5.55	0.60	143.2	1.39	0.06	143.8	5.56	0.42
	0.3	93.9	1.59	0.10	94.3	3.70	0.33	188.3	1.59	0.07	188.6	3.70	0.23
	0.4	107.3	1.85	0.12	107.5	2.78	0.22	215.4	1.85	0.09	215.6	2.78	0.15
	0.5	112.1	2.22	0.16	112.1	2.22	0.16	224.4	2.22	0.11	224.5	2.22	0.11
0.2	0.1	35.1	1.39	0.13	35.9	12.52	2.08	71.1	1.39	0.09	71.9	12.50	1.45
	0.2	63.2	1.56	0.12	63.7	6.26	0.72	127.1	1.56	0.08	127.7	6.25	0.51
	0.3	83.2	1.78	0.13	83.7	4.17	0.40	167.4	1.79	0.09	167.8	4.16	0.28
	0.4	95.4	2.08	0.15	95.6	3.12	0.27	191.5	2.08	0.11	191.7	3.12	0.19
	0.5	99.6	2.50	0.19	99.6	2.50	0.19	199.3	2.51	0.14	199.3	2.50	0.14
0.3	0.1	30.5	1.59	0.18	31.3	14.31	2.60	62.0	1.59	0.12	62.8	14.30	1.75
	0.2	55.2	1.79	0.16	55.8	7.15	0.91	111.1	1.78	0.11	111.7	7.15	0.64
	0.3	72.9	2.04	0.17	73.2	4.76	0.50	146.3	2.04	0.12	146.7	4.76	0.35
	0.4	83.5	2.38	0.20	83.7	3.56	0.33	167.4	2.38	0.14	167.6	3.57	0.24
	0.5	87.1	2.85	0.25	87.1	2.85	0.25	174.4	2.86	0.17	174.4	2.86	0.17

것이 많기 때문이다. $\bar{y}^+$와 $\bar{y}^-$의 평균은 1.206, 6.844이고 표준편차는 0.237, 1.577이였으며 z^2의 평균은 1.393, 표준편차는 1.46인 것으로 나타났다. 여기서 유의수준 5%에서 기준이 되는 임계값은 5.99인데 가장 큰 z^2값을 가지는 6번과 13번의 경우 4.63으로 모든 번호가 무작위적으로 추출되었다고 볼 수 있다.

4. 결론

이 논문에서는 이진확률수열의 무작위성을 확인하기 위한 검정통계량과 점근적 성질에 대해 알아보았

Table 3.3. Estimation of the size of test and power($\alpha = 0.05, 0.01$)

ρ	π	100		300		500		1000	
		0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
−0.3	0.3	0.677	0.297	0.999	0.991	1.000	1.000	1.000	1.000
	0.4	0.672	0.329	0.999	0.986	1.000	1.000	1.000	1.000
	0.5	0.664	0.336	0.997	0.981	1.000	1.000	1.000	1.000
−0.2	0.2	0.182	0.032	0.923	0.647	0.999	0.973	1.000	1.000
	0.3	0.258	0.062	0.867	0.623	0.987	0.935	1.000	1.000
	0.4	0.285	0.073	0.853	0.620	0.983	0.923	1.000	0.999
	0.5	0.281	0.075	0.841	0.609	0.981	0.918	1.000	1.000
−0.1	0.1	0.022	0.003	0.112	0.018	0.327	0.078	0.934	0.549
	0.2	0.045	0.005	0.225	0.053	0.443	0.169	0.820	0.565
	0.3	0.059	0.008	0.253	0.072	0.444	0.192	0.805	0.566
	0.4	0.071	0.010	0.254	0.076	0.447	0.204	0.795	0.554
	0.5	0.072	0.011	0.251	0.079	0.451	0.201	0.799	0.565
0	0.1	0.050	0.019	0.051	0.017	0.048	0.014	0.050	0.012
	0.2	0.050	0.017	0.051	0.014	0.047	0.011	0.051	0.010
	0.3	0.053	0.013	0.055	0.013	0.047	0.010	0.053	0.010
	0.4	0.047	0.013	0.049	0.011	0.054	0.013	0.051	0.010
	0.5	0.049	0.013	0.049	0.010	0.051	0.012	0.054	0.012
0.1	0.1	0.193	0.117	0.365	0.236	0.514	0.346	0.771	0.612
	0.2	0.186	0.099	0.364	0.208	0.532	0.336	0.818	0.648
	0.3	0.182	0.085	0.375	0.204	0.540	0.330	0.831	0.658
	0.4	0.177	0.087	0.366	0.196	0.549	0.337	0.829	0.661
	0.5	0.183	0.083	0.375	0.200	0.544	0.332	0.835	0.662
0.2	0.1	0.420	0.309	0.790	0.671	0.937	0.867	0.999	0.994
	0.2	0.458	0.318	0.860	0.734	0.978	0.930	1.000	0.999
	0.3	0.481	0.306	0.893	0.766	0.985	0.951	1.000	1.000
	0.4	0.487	0.318	0.898	0.784	0.989	0.962	1.000	1.000
	0.5	0.492	0.315	0.901	0.785	0.989	0.961	1.000	1.000
0.3	0.1	0.611	0.500	0.963	0.921	0.997	0.993	1.000	1.000
	0.2	0.736	0.603	0.993	0.978	1.000	1.000	1.000	1.000
	0.3	0.797	0.650	0.998	0.990	1.000	1.000	1.000	1.000
	0.4	0.816	0.674	0.999	0.993	1.000	1.000	1.000	1.000
	0.5	0.824	0.671	0.998	0.994	1.000	1.000	1.000	1.000

다. 또한 모의실험을 통해 유의수준 5%와 1%에서의 검정의 크기와 검정력을 추정해 보았으며 1회부터 569회까지의 자료를 이용하여 로또 6/45의 각 번호에 대해 당첨번호에 포함여부의 수열을 유도하여 무작위적인지 아닌지를 검정해 보았다.

이 방법은 각종 난수발생기에서 발생되는 난수들 간의 독립성 여부를 확인하는데 사용될 수 있으며 이진뿐만 아니라 항이 여러개인 경우에도 확장시킬 수 있다. 이 논문에서는 제시된 통계량의 근사적 독립성을 확보하기 위해 수열의 크기가 큰 경우를 중심으로 이론을 유도하였나 향후 n이 크지 않은 경우에도 적용할 수 있는 방법에 대해 알아볼 예정이다.

Table 3.4. Randomness test of Lotto 6/45

번호	n^+	$\bar{y}^+$	n^-	$\bar{y}^-$	z^2	번호	n^+	$\bar{y}^+$	n^-	$\bar{y}^-$	z^2
1	5	1.00	6	5.67	1.08	24	4	1.25	5	7.40	0.21
2	5	1.00	6	6.00	0.94	25	4	1.00	5	7.60	0.53
3	5	1.00	6	4.17	2.03	26	4	1.25	5	3.80	1.61
4	4	1.25	5	7.20	0.22	27	5	1.00	5	7.40	0.67
5	3	1.00	4	9.00	0.58	28	5	1.00	6	5.17	1.34
6	3	1.67	4	9.00	4.63	29	4	1.00	4	8.50	0.62
7	4	1.25	5	6.80	0.26	30	4	1.25	5	7.60	0.21
8	5	1.00	6	5.50	1.16	31	5	1.00	6	4.83	1.54
9	4	1.25	5	6.60	0.29	32	5	1.00	6	6.33	0.83
10	5	1.00	6	6.33	0.83	33	3	1.33	4	8.75	0.67
11	3	1.33	3	12.00	1.79	34	5	1.00	6	6.00	0.94
12	5	1.00	6	6.50	0.79	35	4	1.25	5	7.40	0.21
13	3	1.67	4	9.00	4.63	36	5	1.00	5	6.80	0.72
14	5	1.00	6	5.67	1.08	37	5	1.00	6	5.17	1.34
15	3	1.67	4	7.75	4.45	38	4	1.25	5	7.40	0.21
16	3	1.67	4	8.50	4.53	39	4	1.25	5	7.40	0.21
17	4	1.25	5	6.20	0.38	40	3	1.33	4	5.00	1.06
18	4	1.25	5	5.80	0.50	41	5	1.00	6	6.17	0.89
19	3	1.67	4	9.25	4.70	42	3	1.67	4	6.50	4.53
20	5	1.00	6	4.00	2.17	43	4	1.25	5	7.60	0.21
21	3	1.33	4	6.50	0.63	44	5	1.00	6	6.50	0.79
22	3	1.67	4	8.25	4.49	45	5	1.00	5	5.80	0.96
23	4	1.25	5	7.20	0.22						

References

Coronel-Brizio, H. G., Hern´andez-Montoya, A. R., Rapallo, F. and Scalas, E. (2008). Statistical auditing and randomness test of lotto k/N-type games, *Phisica A: Statistical Mechanics and its Applications*, **387**, 6385–6390.

Genest, C., Lockhart, R. A. and Stephens, M. A. (2002). χ^2 and the lottery, *Journal of the Royal Statistical Society: Series D*, **51**, 243–257.

Helman, D. (2005). Combinatorial interdependence in lottery, *Teaching Mathematics and its Applications*, **24**, 203–207.

Johnson, R. A. and Klotz, J. (1993). Estimating hot numbers and testing uniformity for the lottery, *Journal of the American Statistical Association*, **88**, 662–668.

Kim, J. H. (2004). Are there hot numbers in the Lotto Korean lottery?, *The Korean Communications in Statistics*, **11**, 413–418.

Lim, S. and Baek, J. (2009). Statistical randomness test for Korean lotto game, *Journal of the Korean Data & Information Science Society*, **20**, 779–786.

Rukhin, A., Soto, J., Nechvatal, J., Smid, M., Barker, E., Leigh, S., Levenson, M., Vangel, M., Banks, D., Heckert, A., Dray, J. and Vo, S. (2010). A Statistical test suite for random and pseudorandom number generators for cryptographic applications, *NIST Special Publication 800-22 Revision 1a*, NIST, Gaithersburg.

이진확률수열의 무작위성 검정

여인권[a,1]

[a]숙명여대 통계학과

(2013년 11월 18일 접수, 2013년 12월 16일 수정, 2014년 1월 9일 채택)

요 약

이 논문에서는 이진확률수열의 무작위성을 검정하는 방법을 제안한다. 연의 길이는 절사된 기하분포를 따르는데 제안하고자 하는 검정통계량은 연의 평균길이를 기초로 하고 있으며 표본크기가 커지면 점근적으로 χ^2_2-분포를 따른다. 검정크기와 검정력을 비교하기 위해 몬테칼로모의실험을 실시했다. 로또 6/45에서의 추첨여부에 대한 수열에 적용해 보았으며 로또는 무작위성을 만족하는 것으로 나타났다.

주요용어: 로또 6/45, 연의 길이, 절사된 기하분포.

이 논문은 2012년도 숙명여자대학교의 교내연구비에 의하여 수행되었음.

[1](140-742) 서울시 용산구 청파로47길 100, 숙명여대 통계학과, 교수. E-mail: inkwon@sm.ac.kr