Lottery Numbers and Ordered Statistics

Kung-Kuen Tse

Department of Mathematics, Kean University, Union, NJ, USA
Email: ktse@kean.edu

Abstract

The lottery has long captivated the imagination of players worldwide, offering the tantalizing possibility of life-changing wins. While winning the lottery is largely a matter of chance, as lottery drawings are typically random and unpredictable. Some people use the lottery terminal randomly generates numbers for them, some players choose numbers that hold personal significance to them, such as birthdays, anniversaries, or other important dates, some enthusiasts have turned to statistical analysis as a means to analyze past winning numbers identify patterns or frequencies. In this paper, we use order statistics to estimate the probability of specific order of numbers or number combinations being drawn in future drawings.

Keywords

Lottery, Order Statistics, Hypergeometric Distribution, Expectation, Uniform

1. Introduction

Various sophisticated statistical methods have been employed to analyze and predict lottery numbers: frequency analysis [1], regression analysis [2], machine learning [3], artificial intelligence [4], computer simulation [5], clustering and pattern recognition [6]. The mathematics behind the theory is based on previous draws and patterns which arise from them—previous draws dictate the future probability of certain number being drawn. In this work, we analyze what numbers are likely to be drawn (independent of past draws) by using elementary probability.

2. Ordered Statistics

We choose K balls among N numbered balls and order them in ascending order. Let X_k be the kth largest. For example, X_1 is the smallest and X_K is the largest among the K chosen balls. For $k = 1, \cdots, K$, X_k has the following probability

mass function:

Theorem 1

$$p(X_k = x) = \frac{\binom{x-1}{k-1}\binom{N-x}{K-k}}{\binom{N}{K}} \text{ for } x = k, \cdots, N-K+k.$$

Proof The event $X_k = x$ means that we need to choose $k-1$ numbers among $1, \cdots, x-1$ and we need to choose $K-k$ numbers among $x+1, \cdots, N$. Hence,

$$p(X_k = x) = \frac{\binom{x-1}{k-1}\binom{N-x}{K-k}}{\binom{N}{K}} \text{ for } x = k, \cdots, N-K+k. \quad \blacksquare$$

Remark This is not the same as hypergeometric distribution discussed in [7].

Example For the Mega Millions in the U.S. [8], players pick six numbers from two separate pools of numbers—five different numbers from 1 to 70 and one bonus number (Mega Ball) from 1 to 25. Here, we ignore the bonus number because it does not affect the distribution of the order statistics. Using the Theorem, the table on the next page displays the numbers of X_1, X_2, X_3, X_4 and X_5 with the top five highest probability.

	X_1	X_2	X_3	X_4	X_5
1	0.0714286				
2	0.0672878				
3	0.0633297				
4	0.0595488				
5	0.0559398				
⋮					
16		0.0307411			
17		0.0309688			
18		0.0310419			
19		0.0309716			
20		0.0307692			
⋮					
31			0.0266326		
32			0.0270094		
33			0.0272937		
34			0.0274841		
35			0.0275795		
36			0.0275795		

Continued

37	0.0274841	
38	0.0272937	
39	0.0270094	
40	0.0266326	
⋮		
51		0.0307692
52		0.0309716
53		0.0310419
54		0.0309688
55		0.0307411
⋮		
66		0.0559398
67		0.0595488
68		0.0633297
69		0.0672878
70		0.0714286

Remark At the time this work is carried out, according to Lotto America [9] and USAMega [10], the most frequent Mega Millions numbers are 3, 10, 14, 17, 31, 46, 64, … Some of the numbers are not showing up in our calculation because these are the statistics for the sixth/current version of Mega Millions (October 31, 2017 to present: first 5 numbers are chosen from 1 to 70 and the Mega Ball is chosen from 1 to 25). Statistical analysis is typically based on a sufficient sample size to draw meaningful conclusions. In the context of lotteries, the number of past draws available for analysis is often limited. With a small sample size, it becomes challenging to identify statistically significant patterns or trends.

We next describe the long-term behavior of X_k.

Corollary 2 *The expectation of X_k is*

$$E[X_k] = \sum_{x=k}^{N-K+k} x \frac{\binom{x-1}{k-1}\binom{N-x}{K-k}}{\binom{N}{K}}$$

for $1 \le k \le K$.

We now simplify of $E[X_k]$ by using a different approach.

Theorem 3 $E[X_k] = k \cdot \dfrac{N+1}{K+1}$ *for* $1 \le k \le K$.

Proof If K numbers are randomly selected in the interval $(0, N+1)$ and each number is equally likely to be picked, then $X_k = (N+1)Y_k$, where $Y_1 < Y_2 < \cdots < Y_K$ are the order statistics over the unit interval $(0,1)$. Y_k satisfies [7] [11] [12].

$$f_{Y_k}(y) = \frac{K!}{(k-1)!(K-k+1)!} y^{k-1}(1-y)^{K-k} \text{ for } 0 < y < 1.$$

and

$$E[Y_k] = \frac{k}{K+1} \text{ for } k = 1, \cdots K.$$

Hence,

$$E[X_k] = E[(N+1)Y_k] = (N+1)E[Y_k] = k \cdot \frac{N+1}{K+1}.$$

Corollary 4

$$\sum_{x=k}^{N-K+k} x \frac{\binom{x-1}{k-1}\binom{N-x}{K-k}}{\binom{N}{K}} = k\frac{N+1}{K+1}$$

for $k = 1, \cdots, K$.

3. Conclusion

It's important to note that while statistical analysis can provide insights into patterns and frequencies, lottery drawings are still random, and there is no guaranteed method to predict future winning numbers. These methods should be used for informational purposes and to assist in making informed choices, but the element of chance always remains dominant in lottery games. Moreover, lottery systems are complex, involving various factors such as ball machines, condition of the balls, number selection methods, and multiple games within a lottery. It can be challenging to capture all the intricacies and variables accurately in a statistical model. Finally, lottery games are games of chance, and the odds of winning are typically very low. It's essential to approach playing the lottery with the understanding that it is purely for entertainment purposes.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Finkelstein, M. (1995) Estimating the Frequency Distribution of the Numbers Bet on the California Lottery. *Applied Mathematics and Computation*, **69**, 195-207. https://doi.org/10.1016/0096-3003(94)00126-O

[2] Combs, K., Kim, J. and Spry, J. (2008) The Relative Regressivity of Seven Lottery Games. *Applied Economics*, **40**, 35-39. https://doi.org/10.1080/13504850701439327

[3] Chen, T., Cheng, Y., Gan, Z., Liu, J. and Wang, Z. (2021) Data-Efficient GAN Training Beyond (Just) Augmentations: A Lottery Ticket Perspective. *Advances in Neural Information Processing Systems*, **34**, 20941-20955.

[4] Kalibhat, N.M., Balaji, Y. and Feizi, S. (2021) Winning Lottery Tickets in Deep Generative Models. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 8038-8046. https://doi.org/10.1609/aaai.v35i9.16980

[5] Nishizaki, I. and Hayashida, T. (2013) Simulation Analysis for Choice of Binary

Lotteries. *Computational Economics*, **41**, 195-211.
https://doi.org/10.1007/s10614-012-9348-5

[6] Li, H., Mao, L., Zhang, J. and Xu, J. (2015) Classifying and Profiling Sports Lottery Gamblers: A Cluster Analysis Approach. *Social Behavior and Personality: An International Journal*, **43**, 1299-1317. https://doi.org/10.2224/sbp.2015.43.8.1299

[7] Ross, S. (2023) A First Course in Probability. 10th Edition, Pearson, Harlow. https://doi.org/10.1017/9781009179928

[8] https://www.megamillions.com/How-to-Play

[9] https://www.lottoamerica.com/mega-millions/statistics

[10] https://www.usamega.com/mega-millions/statistics

[11] David, H.A. and Nagaraja, N.A. (2003) Order Statistics. John Wiley & Sons, Hoboken.

[12] Casella, G. and Berger, R. (2001) Statistical Inference. Cengage Learning, Boston.