California Mega-Millions: an Abstract Version[1]

The setting. The top prize is for correctly guessing n numbers out of $1, \dots, N$ AND one extra (Mega) number from $1, \dots, M$.

Basic Analysis. The total number L of possible outcomes of this lottery game is

$$L = \binom{N}{n} \cdot M$$

and therefore the probability to win the top prize is

$$p_n^* = \frac{1}{L}.$$

Full Mathematical Analysis. For a given lottery ticket, there are $2(n+1)$ possible outcomes of the game, corresponding to correctly guessing $k = 0, \dots, n$ numbers, with or without the Mega one. These outcomes are mutually exclusive but not equally likely. Denote by p_k^* and p_k the corresponding probabilities. For example, p_1 is the probability to correctly guess one number out of n while getting a wrong Mega number; p_0^* is the probability to get correctly only the Mega number. Then

$$p_k^* = \frac{\binom{n}{k}\binom{N-n}{n-k}}{L}, \quad p_k = (M-1)p_k^*, \quad p_k + p_k^* = \frac{\binom{n}{k}\binom{N-n}{n-k}}{\binom{N}{n}},$$

and

$$\sum_{k=0}^{n}(p_k + p_k^*) = \frac{\sum_{k=0}^{n}\binom{n}{k}\binom{N-n}{n-k}}{\binom{N}{n}} = 1;$$

the equality

$$\sum_{k=0}^{n}\binom{n}{k}\binom{N-n}{n-k} = \binom{N}{n}$$

is a particular case of `Vandermonde's identity` and admits a simple combinatorial interpretation. Note also that all p_k and p_k^* are multiples of $p_n^* = 1/L$.

The rules of the game are such that not all outcomes with at least one correct guess lead to the prize. In other words, winning nothing is not just getting all numbers wrong:

$$\mathbb{P}(\text{winning nothing}) > p_0 = \mathbb{P}(\text{all numbers are wrong}).$$

Alternatively,

$$\mathbb{P}(\text{winning something}) < 1 - p_0;$$

the actual design of the game is such that $1 - p_0$ is *much larger* than the probability to win something.

A concrete numerical analysis. Take $N = 75$, $n = 5$, $M = 15$ and assume that there are "nine ways to win" corresponding to

$$p_0^*, p_1^*, p_2^*, p_3, p_3^*, p_4, p_4^*, p_5, p_5^*.$$

Then

$$L = \binom{75}{5} \cdot 15 = 258,890,850$$

and the probability to win something is

$$p_0^* + p_1^* + p_2^* + p_3 + p_3^* + p_4 + p_4^* + p_5 + p_5^* \approx \frac{1}{14.7} \approx 0.068.$$

The probability to get something correctly is much higher:

$$1 - p_0 = 1 - \frac{\binom{70}{5} \cdot 14}{\binom{75}{5} \cdot 15} \approx 0.3455.$$

This is how MATLAB evaluates this number:

```
>> 1-(nchoosek(70,5)*14)/(nchoosek(75,5)*15)

ans =

0.3455
```

[1]Sergey Lototsky, USC, Version of August 31, 2023