Explorations into and beyond the Game of Bingo

Ashutosh Srivastava

April 21, 2018

1 Specifying the model

1.1 Defining the parameters

k players are given a ticket of digits between 1 to n. Each ticket consists of r distinct integers drawn one at a time from a pool of n integers without replacement. As there is no constraint on two tickets being same or different therefore $k \geq 1$.

1.2 Playing the Game

A host draws one digit at a time from the pool of integers at hand and each player checks for the presence of that digit in her/his own ticket. If present, it is striked off the player's ticket otherwise he waits for the next call without taking any action. The player whose ticket gets exhausted first i.e. all of whose numbers are striked off first wins the round and takes home the jackpot.

1.3 Investigation

As the game will run as long as no ticket gets exhausted, we are interested in the behavior of the time it takes for the game to be over.

1.4 Notation

- T_q - time it takes for the ticket of q^{th} player to get exhausted
- T_{min} - $\inf(T_q)$ where $r \leq q \leq n$

2 Analysis

2.1 Computing Probabilities

Clearly, as per the above conditions $T_r, T_{r+1}, \ldots\ldots, T_n$ are i.i.d random variables. Without much complication we can say that

$P[T_j=i] = \binom{i-1}{r-1}/\binom{n}{r}$ for $r \leq i, j \leq n$

$P[T_j \leq i] = \sum_{m=r}^{i} \binom{m-1}{r-1}/\binom{n}{r}$

$P[T_j \leq i] = \binom{i}{r}/\binom{n}{r}$

Consider probability of the event $T_{min} \geq i$, this is equivalent to each of the individual T_js being greater than i,as $T_r, T_{r+1}, \ldots\ldots, T_n$ are i.i.d random variables

$P[T_{min} \geq i] = P[T_1 \geq i]^k$

$P[T_{min} > i] = (1 - \binom{i}{r}/\binom{n}{r})^k$ (**1**)

$P[T_{min}=i] = P[T_{min} \geq i] - P[T_{min} \geq i+1]$

$P[T_{min}=i] = P[T_1 \geq i]^k - P[T_1 \geq i+1]^k$ (**2**)

$P[T_{min}=i+1] = (1 - P[T_1 \leq i])^k - (1 - P[T_1 \leq i+1])^k$

$P[T_{min} = i+1] = (1 - \binom{i}{r}/\binom{n}{r})^k - (1 - \binom{i+1}{r}/\binom{n}{r})^k$ (**3**)

Computing the terminal probabilities using above equations i.e. for the cases i=r and i=n, we get For i=r, which is the event where a ticket gets exhausted in first r turns we have
$P[T_{min}=r] = 1 - (1 - P[T_1 \leq r])^k$

$= 1 - (1 - \binom{r}{r}/\binom{n}{r})^k$

$P[T_{min}=r] = 1 - (1 - 1/\binom{n}{r})^k$

Now considering the case where the game lasts till the n^{th} draw

$P[T_{min}=n] = (1 - \binom{n-1}{r}/\binom{n}{r})^k$

= (1 - (n-r)/n)^k

P[T$_{min}$=n] = $(r/n)^k$

2.2 Expression for Expected value of T$_{min}$

Invoking the tail sum formula and applying it to calculate the expectation of T$_{min}$

E[T$_{min}$] = $\sum_{q=1}^{\infty}$P[T$_{min}$ $\geq$q]

= $\sum_{q=1}^{r}$P[T$_{min}$ $\geq$q] + $\sum_{q=r+1}^{n}$P[T$_{min}$ $\geq$q] + $\sum_{q=n+1}^{\infty}$P[T$_{min}$ $\geq$q]

Now, as we know from previous section P[T$_{min}$ $\geq$q] can be defined piecewise as follows

$$\begin{cases} 1 & q \leq r \\ P[T_{min} \geq q] & r+1 \leq q \leq n \\ 0 & n+1 \leq q \end{cases}$$

Using result from eqn 3, we can write the final form as follows

E[T$_{min}$] = r + $\sum_{j=r}^{n-1}(1 - \binom{j}{r}/\binom{n}{r})^k$ (**4**)

This makes intuitive sense, because the average time it takes for the game to get over should be at least the size of the ticket for each of the players.

2.3 Bounds for Expected value of T$_{min}$

As it is extremely difficult to compute the sum of series in eqn 4 analytically, we shall try to put a handle onto the upper and lower bound for the average time it takes for the game to finish. For that lets consider the product of terms involved in X = $\binom{j}{r}/\binom{n}{r}$ we get

X = $\prod_{i=0}^{r-1} \frac{j-i}{n-i}$ (**5**)

Also, we know that for a,b> 0, a<b and $\forall 0 < z < a$ we have $\frac{a}{b} > \frac{a-z}{b-z}$. Using this fact and substituting it into eqn 5 we get the following results

- X < $(\frac{j}{n})^r$ (**6**)
- X > $(\frac{j-r+1}{n-r+1})^r$ (**7**)

For the ratio $\binom{j}{r}/\binom{n}{r}$ we substitute the results from eqns 6 and 7 to get

$= (\frac{j-r+1}{n-r+1})^r < \binom{j}{r}/\binom{n}{r} < (\frac{j}{n})^r$

$= (1-(\frac{j}{n})^r) < (1-\binom{j}{r}/\binom{n}{r}) < (1-(\frac{j-r+1}{n-r+1})^r)$

$= (1-(\frac{j}{n})^r)^k < (1-\binom{j}{r}/\binom{n}{r})^k < (1-(\frac{j-r+1}{n-r+1})^r)^k$

$= \sum_{j=r}^{n-1}(1-(\frac{j}{n})^r)^k < \sum_{j=r}^{n-1}(1-\binom{j}{r}/\binom{n}{r})^k < \sum_{j=r}^{n-1}(1-(\frac{j-r+1}{n-r+1})^r)^k$

$= r+\sum_{j=r}^{n-1}(1-(\frac{j}{n})^r)^k < r+\sum_{j=r}^{n-1}(1-\binom{j}{r}/\binom{n}{r})^k < r+\sum_{j=r}^{n-1}(1-(\frac{j-r+1}{n-r+1})^r)^k$

$= r+\sum_{j=r}^{n-1}(1-(\frac{j}{n})^r)^k < E[T_{min}] < r+\sum_{j=r}^{n-1}(1-(\frac{j-r+1}{n-r+1})^r)^k$

$$\boxed{\sum_{j=r}^{n-1}(1-(\frac{j}{n})^r)^k < E[T_{min}-r] < \sum_{j=r}^{n-1}(1-(\frac{j-r+1}{n-r+1})^r)^k}$$

The upper and lower limits can be condensed down to a single expression using Binomial expansion and Faulhaber's formula for summing series involving powers of natural numbers(which can be found here) but that would not give us any further insight into the probabilistic structure of the problem. Also, if we take the usual case of bingo where the ticket size $r = 15$, the pool size $n = 90$ and the number of players, say $k = 8$. Using the above formulas we observe the following:

- On an average we can expect the game to end in 77 turns as exact computation gives $\mathrm{E}[\mathrm{T}_{min}] = 77$
- Upper bound = 78 and Lower bound = 75, these are quite precise for the normal game of Bingo

3 Questions of Interest

The whole situation of Bingo can be translated in statistical terms as the behavior of set of finite simple random samples taken from a finite population without replacement. This implies that it is indeed a fertile field for studying the same. The following questions are asked keeping the same in mind, not from any academic necessity but purely out of curiosity.

3.1 Involving the dependence structure

The version considered here resembles the scenario where the players are blinded in the sense that they cannot see tickets of other players, which means that the Jackpot prize has a chance of being shared. Now if we were to say that every player is given the right to check every other player's ticket and declare that the game is legitimate iff at least z digits are different, where, $1 \leq z \leq r$.
This slight modification subtely encodes the dependence structure of the tickets or samples in our terminology. I think it would be interesting to analyze the behavior of T_{min} under different dependent structures i.e parameters of game now being *n,r,k,z*.

3.2 Continuous extension of Bingo

Another change can be to make the tickets "continuous" in some sense and then draw numbers from a distribution wihch would enable a "continuous" extension of Bingo an add on to the discrete version we have right now. Also, I am not sure how to go about this but finding out the constraints on the game so that it ends in linear time, polynomial time etc. might give us something much more in return.

3.3 Framework for Testing Randomness among set of samples

I intend to use this as a measure of randomness in any given set of samples. As in its purest form, the behavior of $T_m in$ is a consequence of simple random samples and independence among tickets i.e. if a given set of samples are truly random then their exhaustion time or T_{min} must follow the results derived above. Additionally, if we combine this with the above two questions, this can be used to test samples from continuous domains as well and assess the extent of dependency amongst them. As the assumption of independent random samples is the key to not only classic statistics but even critical to Machine Learning algorithms, it would be worthy to test and quantify the impact of this framework with respect to the performance of various algorithms.