Statistical Auditing of Lotto k/N-type Games

Kyle Oddson

Under the direction of Dr. Naghmeh Daneshi

with second reader Dr. Daniel Taylor-Rodriguez

Mathematical literature project

in partial fulfillment of requirements for the

Master of Science in Statistics

Fariborz Maseeh Department of Mathematics and Statistics

Portland State University

Spring 2022

Abstract

In a lottery k/N-type game, k numbers are drawn from a pool of N. Players win prize money based on how many of their chosen k numbers match the k numbers drawn without replacement. Coronel-Brizio et al. [1] utilize the distributions of the order statistics of the drawn numbers to statistically analyze historical results of lotteries vis-à-vis a theoretical distribution of lottery results.

We explicate the findings of Coronel-Brizio et al. [1] and apply them to Oregon's lottery game Megabucks, with additional analysis on the multi-state games Powerball and Mega Millions. Whereas Coronel-Brizio et al. [1] state that their findings "...can be adapted to test pseudorandom number generators", we apply the theory to one lottery game using a number generator as well as to multiple games using ball machines as their selection method.

Analysis is done using two test methods, an asymptotic test and Monte Carlo simulation. Our results show that Oregon's Megabucks ball machine (no longer in use) did not fit the expected distribution, but its number generator (currently in use) does not deviate from the expected distribution. Our results on Powerball and Mega Millions show no evidence that they deviate from their expected distributions.

1 Introduction

According to Coronel-Brizio et al. [1],

> ... although, strictly speaking, it is known that there is no way to "prove" the randomness of a sequence of numbers, it is always possible to statistically test whether or not historical results exhibit the quantitative properties derived from the probabilistic model assumed to explain the selection mechanism. In this respect, the statistical procedure presented here could be easily used as a test of pseudorandom number generators.

The lottery games we will examine (Megabucks, Powerball, and Mega Millions) claim to be random-draw games. Here that means that in any given draw all numbers not already drawn are equally likely to be drawn. Coronel-Brizio et al. [1] call this "fairness".

Our running example will be Oregon's game Megabucks, a 6/48 draw game [6] running since 30 November 1985 [3]. This game began using a ball machine as its selection method. The draw on 21 May 2001 was the first draw using a pseudorandom number generator [3] (hereafter referred to as a PRNG). The multi-state lottery games Powerball and Mega Millions continue to use ball machines as their selection method [3], though rule changes at various times affected the k/N parameters for the games ([8], [9], [2]).

We will use as our basic random variable the number of matches attained by a player of a lottery game; this is a hypergeometric random variable. In order to analyze the fairness of the games, we will determine the distributions of the order statistics of the drawn numbers. We will use two approaches in our hypothesis testing: an asymptotic approach, based on the multivariate central limit theorem; and Monte Carlo sampling to simulate the distribution of our test statistic. Should any ambiguity occur, we will use a default confidence level of $\alpha = 0.05$. Where appropriate, we round to three decimal places.

2 Probabilities

We pick k distinct numbers out of N; the lottery draws k numbers out of N (supposedly randomly and fairly) without replacement. There are

$$\binom{N}{k} = \frac{N!}{k!\,(N-k)!}$$

possible ways to choose k from N. (Recall that $\binom{N}{k}$ is greater than 0 only for $N \geq k > 0$ and is 0 otherwise.) In the case of Megabucks, we pick 6 out of 48; there are $\binom{48}{6} = 12,271,512$ possible outcomes.

Let random variable X be the number of matches between our selection and the drawn numbers. We will find $P(X = i)$, the probability of matching exactly i of our k numbers. First, select which of the k numbers are matches: there are $\binom{k}{i}$ options. The rest of our numbers are not matches. Second, select which of the $N - k$ non-match options our $k - i$ non-matches are: there are $\binom{N-k}{k-i}$ options. By the product principle, multiply these to count all ways we can match and not match our k chosen from N to the game's k drawn from N. To find the probability, divide this product by the total ways we could have chosen our k: $\binom{N}{k}$. Hence

$$P(X = i) = \frac{\binom{k}{i}\binom{N-k}{k-i}}{\binom{N}{k}}$$

for $i = 1, \ldots, k$ and $k = 1, \ldots, N$ (0 otherwise). We see that X follows the hypergeometric distribution.

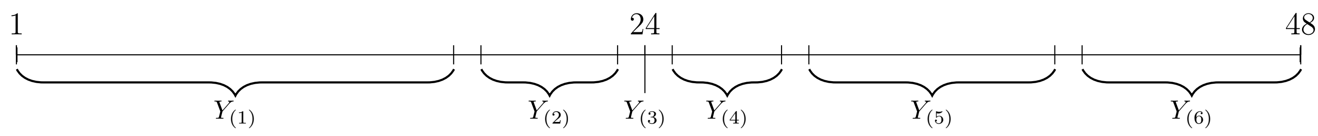
In the Megabucks case,

$$P(X = i) = \frac{\binom{6}{i}\binom{48-6}{6-i}}{\binom{48}{6}} = \frac{\binom{6}{i}\binom{42}{6-i}}{\binom{48}{6}}.$$

(A note on notation: moving forward, we will use negative exponents in place of division, for clarity of display.)

Since we will be testing historical results, which are samples of size k and are usually reported in sorted order, we will perform our hypothesis tests using the order statistics. Let $Y_1, Y_2, \ldots, Y_k$ be the randomly drawn numbers and let $Y_{(i)}$, $i = 1, \ldots, k$ be the sorted randomly drawn numbers, $Y_{(1)} < Y_{(2)} < \cdots < Y_{(k)}$. We will find $P(Y_{(i)} = r)$. Suppose $Y_{(i)} = r$. Then $Y_{(1)}, \ldots, Y_{(i-1)}$ must be less than r; i.e., exactly $i - 1$ numbers must be chosen from $\{1, \ldots, r - 1\}$. Additionally, $Y_{(i+1)}, \ldots, Y_{(k)}$ must be greater than r; i.e., exactly $k - i$ numbers must be chosen from $\{r + 1, \ldots, N\}$. Note that this entails that r is bounded by $i \leq r \leq N - k + i$.

For our Megabucks example, suppose we wish to find $P(Y_{(3)} = 24)$. We have $i = 3$, $r = 24$, $k = 6$, and $N = 48$. We must have $Y_{(1)}, Y_{(2)}$ chosen from $\{1, \ldots, 23\}$; and $Y_{(4)}, Y_{(5)}, Y_{(6)}$ chosen from $\{25, \ldots, 48\}$.



To find $P(Y_{(i)} = r)$, we observe the $i-1$ numbers chosen from $\{1, \ldots, r-1\}$: there are $\binom{r-1}{i-1}$ options. We also observe the $k-i$ numbers chosen from $\{r+1, \ldots, N\}$: there are $\binom{N-r}{k-i}$ options. We last divide by the total number of ways the Y's could be chosen: as ever, there are $\binom{N}{k}$ options.

$$P(Y_{(i)} = r) = \binom{r-1}{i-1}\binom{N-r}{k-i}\binom{N}{k}^{-1}$$

for $i = 1, \ldots, k$, $r = i, \ldots, N-k+i$, and $k = 1, \ldots, N$ (0 otherwise). In our analysis we will use this to find the moments of $Y_{(i)}$.

In our above Megabucks example, we find

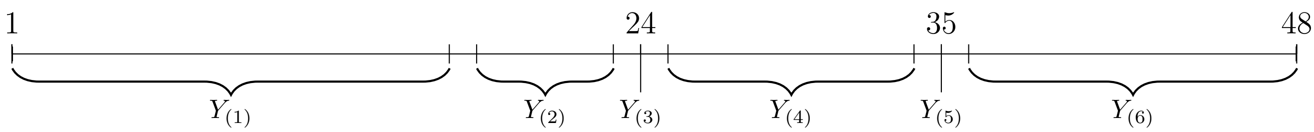
$$P(Y_{(3)} = 24) = \binom{24-1}{3-1}\binom{48-24}{6-3}\binom{48}{6}^{-1} = \binom{23}{2}\binom{24}{3}\binom{48}{6}^{-1} = 253/6,063 \approx 0.042$$

We next compute the joint probability $P(Y_{(i)} = r, Y_{(j)} = s)$. Let $i < j$, $r < s$ (so that $i \in \{1, \ldots, k-1\}$, $j \in \{i+1, \ldots, k\}$, $r \in \{1, \ldots, N-1\}$, and $s \in \{r+1, \ldots, N\}$.) If $Y_{(i)} = r$, then $Y_{(1)}, \ldots, Y_{(i-1)}$ must be less than r; i.e., exactly $i-1$ numbers must be chosen from $\{1, \ldots, r-1\}$: there are $\binom{r-1}{i-1}$ options. We must also have $Y_{(i+1)}, \ldots, Y_{(j-1)}$ greater than r and less than s; i.e., exactly $j-i-1$ numbers must be chosen from $\{r+1, \ldots, s-1\}$: there are $\binom{s-r-1}{j-i-1}$ options. Last, we must have $Y_{(j+1)}, \ldots, Y_{(k)}$ greater than s; i.e., exactly $k-j$ numbers must be chosen from $\{s+1, \ldots, N\}$: there are $\binom{N-s}{k-j}$ options. And finally we divide by the $\binom{N}{k}$ ways to have chosen the Y's.

$$P(Y_{(i)} = r, Y_{(j)} = s) = \binom{r-1}{i-1}\binom{s-r-1}{j-i-1}\binom{N-s}{k-j}\binom{N}{k}^{-1}$$

for $1 \le i < j \le k$ and $1 \le r < s \le N-k+j$ (0 otherwise).

For our Megabucks example, suppose we wish to find $P(Y_{(3)} = 24, Y_{(5)} = 35)$. We have $i = 3$, $r = 24$, $j = 5$, $s = 35$, $k = 6$, and $N = 48$.



We find the probability to be

$$\begin{aligned}P(Y_{(3)} = 24, Y_{(5)} = 35) &= \binom{24-1}{3-1}\binom{35-24-1}{5-3-1}\binom{48-35}{6-5}\binom{48}{6}^{-1} \\ &= \binom{23}{2}\binom{10}{1}\binom{13}{1}\binom{48}{6}^{-1} = 65/24,252 \approx 0.003\end{aligned}$$

3 Moments

We will now use the above found probabilities to calculate the mean and variance of $Y_{(i)}$ and the covariance of $Y_{(i)}, Y_{(j)}$. Note that

$$\mathbb{E}[Y_{(i)}] = \sum_r r \cdot P(Y_{(i)} = r) = \sum_r r \binom{r-1}{i-1}\binom{N-r}{k-i}\binom{N}{k}^{-1}$$

Naively, $r \in \{1, \ldots, N\}$. However, recall that $\binom{r-1}{i-1} > 0$ only when $r - 1 \geq i - 1 \;\Rightarrow r \geq i$. Similarly, $\binom{N-r}{k-i} > 0$ only when $n - r \geq k - i \;\Rightarrow N - k + i \geq r$. That is, the only non-zero terms occur when $i \leq r \leq N - k + i$. Hence

$$\mathbb{E}[Y_{(i)}] = \sum_{r=i}^{N-k+i} r \binom{r-1}{i-1}\binom{N-r}{k-i}\binom{N}{k}^{-1} \tag{1}$$

This expression actually simplifies rather nicely. Recall that $P(Y_{(i)} = r) = \binom{r-1}{i-1}\binom{N-r}{k-i}\binom{N}{k}^{-1}$; and as shown above is only positive for $i \leq r \leq N - k + i$. Therefore,

$$\sum_{r=i}^{N-k+i} \binom{r-1}{i-1}\binom{N-r}{k-i}\binom{N}{k}^{-1} = 1 \quad \Rightarrow \quad \sum_{r=i}^{N-k+i} \binom{r-1}{i-1}\binom{N-r}{k-i} = \binom{N}{k} \tag{2}$$

We express a more general form for later use:

$$\sum_{X=x}^{A-b+x} \binom{X-1}{x-1}\binom{A-X}{b-x} = \binom{A}{b} \tag{3}$$

Recall the identity $x\binom{x-1}{m-1} = m\binom{x}{m}$. Then

$$E[Y_{(i)}] = \sum_{r=i}^{N-k+i} i\binom{r}{i}\binom{N-r}{k-i}\binom{N}{k}^{-1} = i\binom{N}{k}^{-1} \cdot \sum_{r=i}^{N-k+i} \binom{r}{i}\binom{N-r}{k-i} \tag{4}$$

We will make two successive substitutions. First, let $m = r + 1$ (so $r = m - 1$). Then

$$\sum_{r=i}^{N-k+i} \binom{r}{i}\binom{N-r}{k-i} = \sum_{m=i+1}^{N-k+i+1} \binom{m-1}{i}\binom{N-(m-1)}{k-i} \tag{5}$$

Now let $p = i + 1$ (so $i = p - 1$). Then, also adding $+1 - 1$ to our index,

$$\text{eqn. (5)} = \sum_{m=p}^{(N+1)-(k+1)+p} \binom{m-1}{p-1}\binom{(N+1)-m}{(k+1)-p} = \binom{N+1}{k+1} \tag{6}$$

by eqn. (3) with $A = N+1$, $b = k+1$. Hence we find that

$$\mathbb{E}[Y_{(i)}] = i\binom{N}{k}^{-1} \cdot \sum_{r=i}^{N-k+i} \binom{r}{i}\binom{N-r}{k-i} \tag{7}$$

$$= i\binom{N}{k}^{-1}\binom{N+1}{k+1} \tag{8}$$

$$= i \cdot \frac{k!\,(N-k)!}{N!} \cdot \frac{(N+1)!}{(k+1)!\,[(N+1)-(k+1)]!}$$

$$= i \cdot \frac{k!\,(N-k)!}{N!} \cdot \frac{(N+1)N!}{(k+1)k!\,(N-k)!}$$

$$= \frac{N+1}{k+1} i \quad \text{for } i = 1, \ldots, k. \tag{9}$$

In our Megabucks example with $N = 48$, $k = 6$, we see that

$$\mathbb{E}[Y_{(i)}] = \frac{48+1}{6+1} i = \frac{49}{7} i = 7i \tag{10}$$

We now calculate $\text{Var}[Y_{(i)}]$. Observe the following for any random variable X:

$$\begin{aligned}\text{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \mathbb{E}[X^2] - \mathbb{E}[X] + \mathbb{E}[X] - \mathbb{E}[X]^2 \\ &= \mathbb{E}[X^2 - X] + \mathbb{E}[X] - \mathbb{E}[X]^2 \\ &= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2\end{aligned}$$

Hence

$$\text{Var}[Y_{(i)}] = \left[\sum_{r=i}^{N-k+i} r(r-1)\binom{r-1}{i-1}\binom{N-r}{k-i}\binom{N}{k}^{-1}\right] + \frac{N+1}{k+1} i - \left(\frac{N+1}{k+1} i\right)^2 \tag{11}$$

for $i = 1, \ldots, k$.

Consider only the first term in the above and recall the identity $x\binom{x-1}{m-1} = m\binom{x}{m}$:

$$\sum_{r=i}^{N-k+i} r(r-1)\binom{r-1}{i-1}\binom{N-r}{k-i}\binom{N}{k}^{-1}$$

$$= \sum_{r=i}^{N-k+i} (r-1)\cdot i \cdot \binom{r}{i}\binom{N-r}{k-i}\binom{N}{k}^{-1}$$

$$= i\binom{N}{k}^{-1} \cdot \sum_{r=i}^{N-k+i} (r-1)\binom{r}{i}\binom{N-r}{k-i}$$

$$= i\binom{N}{k}^{-1}\left[\sum_{r=i}^{N-k+i} r\binom{r}{i}\binom{N-r}{k-i} - \sum_{r=i}^{N-k+i}\binom{r}{i}\binom{N-r}{k-i}\right] \tag{12}$$

We consider the first summation in (12) above. Let $m = r+1$. Then we have

$$\sum_{r=i}^{N-k+i} r\binom{r}{i}\binom{N-r}{k-i} = \sum_{m=i+1}^{N-k+i+1} (m-1)\binom{m-1}{i}\binom{N-m+1}{k-i} \tag{13}$$

Now let $p = i+1$. Then

$$\text{eqn. (13)} = \sum_{m=p}^{N-k+p} (m-1)\binom{m-1}{p-1}\binom{N-m+1}{k-p+1}$$

$$= \sum_{m=p}^{(N+1)-(k+1)+p} m\binom{m-1}{p-1}\binom{(N+1)-m}{(k+1)-p} - \sum_{m=p}^{(N+1)-(k+1)+p} \binom{m-1}{p-1}\binom{(N+1)-m}{(k+1)-p}$$

$$= \left[\sum_{m=p}^{(N+1)-(k+1)+p} p\binom{m}{p}\binom{(N+1)-m}{(k+1)-p}\right] - \binom{N+1}{k+1} \tag{14}$$

by eqn. (3) with $A = N+1$, $b = k+1$. For the first term of eqn. (14), substitute first $t = m+1$

then $q = p + 1$. Then

$$\begin{aligned}
\text{eqn. (14)} &= \sum_{t=p+1}^{(N+1)-(k+1)+p+1} p\binom{t-1}{p}\binom{(N+1)-m+1}{(k+1)-p} \\
&= \sum t = q^{(N+1)-(k+1)+q}(q-1)\binom{t-1}{q-1}\binom{(N+1)-m+1}{(k+1)-p+1} \\
&= \sum t = 1^{(N+2)-(k+2)+q}(q-1)\binom{t-1}{q-1}\binom{(N+2)-m}{(k+2)-p} \\
&= (q-1)\binom{N+2}{k+2} \qquad (15) \\
&= p\binom{N+2}{k+2} \\
&= (i+1)\binom{N+2}{k+2}
\end{aligned}$$

where eqn. (15) results from eqn. (3) with $A = N + 2$, $b = k + 2$. Then

$$\text{eqn. (14)} = (i+1)\binom{N+2}{k+2} - \binom{N+1}{k+1}$$

We now consider the second summation in (12): by our process of repeated substitution, we see that $\sum_{r=i}^{N-k+i} \binom{r}{i}\binom{N-r}{k-i} = \binom{N+1}{k+1}$. Putting these together, we have

$$\begin{aligned}
\text{eqn. (12)} &= i\binom{N}{k}^{-1}\left[(i+1)\binom{N+2}{k+2} - \binom{N+1}{k+1} - \binom{N+1}{k+1}\right] \\
&= i(i+1)\binom{N+2}{k+2}\binom{N}{k}^{-1} - 2i\binom{N+1}{k+1}\binom{N}{k}^{-1} \\
&= i(i+1)\frac{(N+2)(N+1)}{(k+2)(k+1)} - 2i\frac{N+1}{k+1}
\end{aligned}$$

Returning to eqn. (11), we find

$$\begin{aligned}
\mathrm{Var}[Y_{(i)}] &= \frac{i(i+1)(N+2)(N+1)}{(k+2)(k+1)} - \frac{2i(N+1)}{k+1} + \frac{i(N+1)}{k+1} - \frac{i^2(N+1)^2}{(k+1)^2} \\
&= \frac{i(i+1)(N+2)(N+1)(k+1)}{(k+2)(k+1)^2} - \frac{i(N+1)(k+2)(k+1)}{(k+2)(k+1)^2} - \frac{i^2(N+1)^2(k+2)}{(k+2)(k+1)^2} \\
&= \frac{i(N+1)[(i+1)(N+2)(k+1) - (k+2)(k+1) - i(N+1)(k+2)]}{(k+2)(k+1)^2} \\
&= \frac{i(N+1)}{(k+2)(k+1)^2}\left[i\Big((N+2)(k+1) - (N+1)(k+2)\Big) + (N+2)(k+1) - (k+2)(k+1)\right] \\
&= \frac{i(N+1)}{(k+2)(k+1)^2}\Big[i(-N+k) + (Nk+N+2k+2) - (k^2+3k+2)\Big] \\
&= \frac{i(N+1)}{(k+2)(k+1)^2}\Big[-i(N-k) + Nk - k^2 + N - k\Big] \\
&= \frac{i(N+1)}{(k+2)(k+1)^2}\Big[k(N-k) - i(N-k) + N - k\Big] \\
&= \frac{i(N+1)(k-i+1)(N-k)}{(k+2)(k+1)^2} \quad \text{for } 1 \le i \le k \le N. \qquad (16)
\end{aligned}$$

In our Megabucks example with $N = 48$, $k = 6$, we see that

$$\mathrm{Var}[Y_{(i)}] = \frac{i(48+1)(6-i+1)(48-6)}{(6+2)(6+1)^2} = \frac{i(49)(7-i)(42)}{(8)(49)} = \frac{21i(7-i)}{4} \qquad (17)$$

For example,

$$\mathrm{Var}[Y_{(3)}] = \frac{21(3)(7-3)}{4} = 63$$

We will now calculate $\mathrm{Cov}[Y_{(i)}, Y_{(j)}] = \mathbb{E}[Y_{(i)}Y_{(j)}] - \mathbb{E}[Y_{(i)}]\mathbb{E}[Y_{(j)}]$. Observe the following for any random variables X, Y:

$$\begin{aligned}
\mathbb{E}[XY] &= \mathbb{E}[XY - X^2] + \mathbb{E}[X^2] \\
&= \mathbb{E}[X(Y-X)] + \mathbb{E}[X^2] \\
\Rightarrow \mathrm{Cov}[X,Y] &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[X(Y-X)] + \mathbb{E}[X^2] - \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

Hence

$$\begin{aligned}
\mathrm{Cov}[Y_{(i)}, Y_{(j)}] &= \mathbb{E}[Y_{(i)}(Y_{(j)} - Y_{(i)})] + \mathbb{E}[Y_{(i)}^2] - \mathbb{E}[Y_{(i)}]\mathbb{E}[Y_{(j)}] \\
&= \mathbb{E}[Y_{(i)}(Y_{(j)} - Y_{(i)})] + \mathrm{Var}[Y_{(i)}] + \mathbb{E}[Y_{(i)}]^2 - \mathbb{E}[Y_{(i)}]\mathbb{E}[Y_{(j)}]
\end{aligned}$$

Recall our earlier-found joint probability:

$$P(Y_{(i)} = r, Y_{(j)} = s) = \binom{r-1}{i-1}\binom{s-r-1}{j-i-1}\binom{N-s}{k-j}\binom{N}{k}^{-1}$$

Observe the following:

$$\begin{aligned}
\mathbb{E}[Y_{(i)}(Y_{(j)} - Y_{(i)})] &= \sum_{r=i}^{(N-k+i)} \sum_{s=r+1}^{(N-k+j)} r(s-r)\binom{r-1}{i-1}\binom{s-r-1}{j-i-1}\binom{N-s}{k-j}\binom{N}{k}^{-1} \\
&= \binom{N}{k}^{-1} \sum_{r=i}^{(N-k+i)} \sum_{s=r+1}^{(N-k+j)} r\binom{r-1}{i-1}(s-r)\binom{(s-r)-1}{(j-i)-1}\binom{N-s}{k-j} \\
&= \binom{N}{k}^{-1} \sum_{r=i}^{(N-k+i)} \sum_{s=r+1}^{(N-k+j)} i\binom{r}{i}(j-i)\binom{s-r}{j-i}\binom{N-s}{k-j} \\
&= i(j-i)\binom{N}{k}^{-1} \sum_{r=i}^{N-k+i} \binom{r}{i}\left[\sum_{s=r+1}^{N-k+j} \binom{s-r}{j-i}\binom{N-s}{k-j}\right] \qquad (18)
\end{aligned}$$

where the equality in line three results from the identity $x\binom{x-1}{m-1} = m\binom{x}{m}$. Consider

$$\sum_{s=r+1}^{N-k+j} \binom{s-r}{j-i}\binom{N-s}{k-j} \qquad (19)$$

We use the substitution $m = s - r + 1$. Then

$$\text{eqn. (19)} = \sum_{m=2}^{N-k+j-r+1} \binom{m-1}{j-i}\binom{N-(m+r-1)}{k-j} \qquad (20)$$

Now use the substitution $p = j - i + 1$. Then

$$\text{eqn. (20)} = \sum_{m=2}^{N-k-r+p+i} \binom{m-1}{p-1}\binom{N-m-r+1}{k-(p+i-1)} \qquad (21)$$

Notice that $\binom{m-1}{p-1} = 0$ if $m < p$. Then, adding $+1-1$ to our index,

$$\text{eqn. (21)} = \sum_{m=p}^{(N+1-r)-(k+1-i)+p} \binom{m-1}{p-1}\binom{(N+1-r)-m}{(k+1-i)-p} \qquad (22)$$

By eqn. (3) with $A = N + 1 - r$, $b = k + 1 - p$,

$$\text{eqn. (22)} = \binom{N+1-r}{k+1-i}$$

So

$$\text{eqn. (18)} = \mathbb{E}[Y_{(i)}(Y_{(j)} - Y_{(i)})] = i(j-i)\binom{N}{k}^{-1} \sum_{r=i}^{(N-k+i)} \binom{r}{i}\binom{N+1-r}{k+1-i} \qquad (23)$$

Following the forms of eqn. (7) and eqn. (8), we see that

$$\sum_{r=i}^{N-k+i} \binom{r}{i} \binom{N+1-r}{k+1-i} = \binom{N+2}{k+2}$$

Hence

$$\text{eqn. (23)} = i(j-i) \binom{N}{k}^{-1} \binom{N+2}{k+2} = \frac{i(j-i)(N+2)(N+1)}{(k+2)(k+1)}$$

Finally,

$$\begin{aligned}
\text{Cov}[Y_{(i)}, Y_{(j)}] &= \mathbb{E}[Y_{(i)}(Y_{(j)} - Y_{(i)})] + \text{Var}[Y_{(i)}] + \mathbb{E}[Y_{(i)}]^2 - \mathbb{E}[Y_{(i)}]\mathbb{E}[Y_{(j)}] \\
&= \frac{(i(j-i)(N+2)(N+1)}{(k+2)(k+1)} + \frac{i(k-i+1)(N+1)(N-k)}{(k+2)(k+1)^2} + \frac{i^2(N+1)^2}{(k+1)^2} - \frac{ij(N+1)^2}{(k+1)^2} \\
&= \frac{i(j-i)(N+1)(N+1)(k+1)}{(k+2)(k+1)^2} + \frac{i(k-i+1)(N+1)(N-k)}{(k+2)(k+1)^2} + \frac{i(i-j)(N+1)^2(k+1)^2}{(k+2)(k+1)^2} \\
&= \frac{i(N+1)[(j-i)(N+2)(k+1) + (k-i+1)(N-k) - (j-i)(N+1)(k+2)]}{(k+2)(k+1)^2} \\
&= \frac{i(k-j+1)(N+1)(N-k)}{(k+2)(k+1)^2} \quad \text{for } 1 \leq i \leq j \leq k \leq N. \qquad (24)
\end{aligned}$$

using eqn. (16) and eqn. (9).

In our Megabucks example with $N = 48$, $k = 6$, we find

$$\text{Cov}[Y_{(i)}, Y_{(j)}] = \frac{i(6-j+1)(48+1)(48-6)}{(6+1)^2(6+2)} = \frac{i(7-j)(49)(42)}{(49)(8)} = \frac{21i(7-j)}{4} \qquad (25)$$

For example,

$$\text{Cov}[Y_{(3)}, Y_{(5)}] = \frac{21(3)(7-5)}{4} = 63/2.$$

3.1 Moments: Summary of Equations

$$\mathbb{E}[Y_{(i)}] = \frac{N+1}{k+1} i \qquad i = 1, \ldots, k \qquad (26)$$

$$\text{Var}[Y_{(i)}] = \frac{i(k-i+1)(N+1)(N-k)}{(k+2)(k+1)^2} \qquad 1 \leq i \leq k \qquad (27)$$

$$\text{Cov}[Y_{(i)}, Y_{(j)}] = \frac{i(k-j+1)(N+1)(N-k)}{(k+2)(k+1)^2} \qquad 1 \leq i \leq j \leq k \qquad (28)$$

4 Hypothesis Testing

Let $\boldsymbol{\mu} = [\mathbb{E}[Y_{(1)}, \mathbb{E}[Y_{(2)}], \ldots, \mathbb{E}[Y_{(k)}]]$, the vector of expected order statistics, and let $\mathbf{V}$ be the $k \times k$ covariance matrix of the order statistics (so $\mathbf{V}_{ij} = \mathrm{Cov}[Y_{(i)}, Y_{(j)}]$ for $i < j$; $\mathbf{V}_{ii} = \mathrm{Var}[Y_{(i)}]$; and $\mathbf{V}_{ji} := \mathbf{V}_{ij}$ for $j > i$). Our null hypothesis is that the lotteries under consideration operate under the practice of "fairness" as previously described: any number not already drawn in the current game is equally likely to be drawn. In terms of our described vectors, our hypotheses are thus $H_0: \ \mathbb{E}[\mathbf{Y}] = \boldsymbol{\mu}$; $H_1: \ \mathbb{E}[\mathbf{Y}] \neq \boldsymbol{\mu}$.

Recall the univariate central limit theorem: $\dfrac{\bar{X} - \mu}{\sigma/\sqrt{m}} \rightarrow N(0,1)$. Then $m(\bar{X} - \mu)^2/\sigma^2 \rightarrow \chi^2_1$. Let m be the number of observed game draws, let $\mathbf{y}_1, \ldots, \mathbf{y}_m$ be the ordered outcome vectors of those observed game draws, and let $\bar{\mathbf{y}}$ be the average vector of those ordered outcome vectors. That is, $\bar{\mathbf{y}} = [\overline{y_{(1)}}, \overline{y_{(2)}}, \ldots, \overline{y_{(k)}}]$ across m game draws. Applying the quadratic form to the chi-square application of the univariate form of the CLT and noting the result as a being a sum of k squared standard normal RV's, we see that the random variable $Q = m(\bar{\mathbf{y}} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\bar{\mathbf{y}} - \boldsymbol{\mu})$ converges in distribution to a chi-square distribution with k degrees of freedom. We will use this asymptotic distribution to test the null hypothesis, where the p-value is $P(\chi^2_k > Q)$.

We will also test the null hypothesis using Monte Carlo simulations of lottery drawings, performing 5,000 calculations to simulate the distribution of Q (following the example of [1]), finding as our p-value what proportion of the simulated Q values are greater than our observed Q.

5 Numerical Results: Megabucks

In our Megabucks example with $N = 48$, $k = 6$, using eqn. (10): $\mathbb{E}[Y_{(i)}] = 7i$, we see that $\boldsymbol{\mu} = [7, 14, 21, 28, 35, 42]$. We use eqn. (17): $\mathrm{Var}[Y_{(i)}] = 21i(7-i)/4$ and eqn. (25): $\mathrm{Cov}[Y_{(i)}, Y_{(j)}] = 21i(7-j)/4$ to find

$$\mathbf{V} = \begin{bmatrix} 63/2 & 105/4 & 21 & 63/4 & 21/2 & 21/4 \\ 105/4 & 105/2 & 42 & 63/2 & 21 & 21/2 \\ 21 & 42 & 63 & 189/4 & 63/2 & 63/4 \\ 63/4 & 63/2 & 189/4 & 63 & 42 & 21 \\ 21/2 & 21 & 63/2 & 42 & 105/2 & 105/4 \\ 21/4 & 21/2 & 63/4 & 21 & 105/4 & 63/2 \end{bmatrix}$$

and hence

$$\mathbf{V}^{-1} = \begin{bmatrix} 8/147 & -4/147 & 0 & 0 & 0 & 0 \\ -4/147 & 8/147 & -4/147 & 0 & 0 & 0 \\ 0 & -4/147 & 8/147 & -4/147 & 0 & 0 \\ 0 & 0 & -4/147 & 8/147 & -4/147 & 0 \\ 0 & 0 & 0 & -4/147 & 8/147 & -4/147 \\ 0 & 0 & 0 & 0 & -4/147 & 8/147 \end{bmatrix}$$

As mentioned in the introduction, Megabucks has been running since 1985. The number selection method changed from a ball machine to a PRNG beginning with the drawing on 21 May 2001. As such, we divide the historical results into two periods – Period 1: 30 November 1985 through 19 May 2001 and Period 2: 21 May 2001 through 13 April 2022. During both of these periods Megabucks has been a 6/48 game [3]. We calculate $\bar{\mathbf{y}}$ (the average vectors of the ordered outcome vectors) for each of the two periods:

Period	m	$\bar{\mathbf{y}}$	$\boldsymbol{\mu}$
1	1537	(6.211, 12.635, 18.986, 25.364, 31.690, 38.005)	(7, 14, 21, 28, 35, 42)
2	3272	(6.938, 13.860, 20.894, 28.044, 34.906, 41.974)	

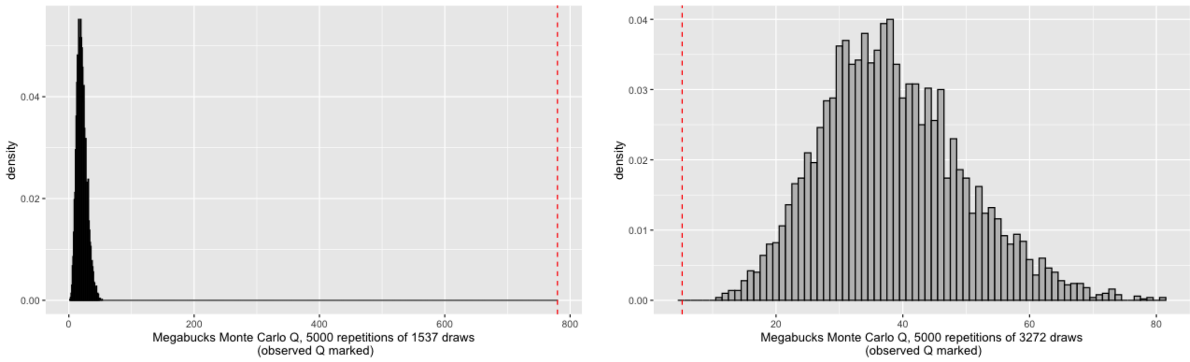
Under the asymptotic distribution, $Q = m(\bar{\mathbf{y}} - \boldsymbol{\mu})^T\mathbf{V}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu})$ approaches a χ^2_6 distribution; the p-value is thus $P(\chi^6_2 > Q)$.

In addition to asymptotic testing of our hypotheses, we employ a Monte Carlo analysis: we generate values for multivariate random variables with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{V}$. Specifically, we generate the same number m of results as in our historical periods and calculate the resulting $\bar{\mathbf{y}}$, and hence the resulting value of Q. We repeat this process 5,000 times to simulate the distribution of Q. To find our Monte Carlo p-value, we calculate what proportion of these generated Q values are greater than our observed Q value. We find the following:

Period	Observed Q	CLT p-value	MC p-value
1	779.754	0	0
2	5.194	0.519	1

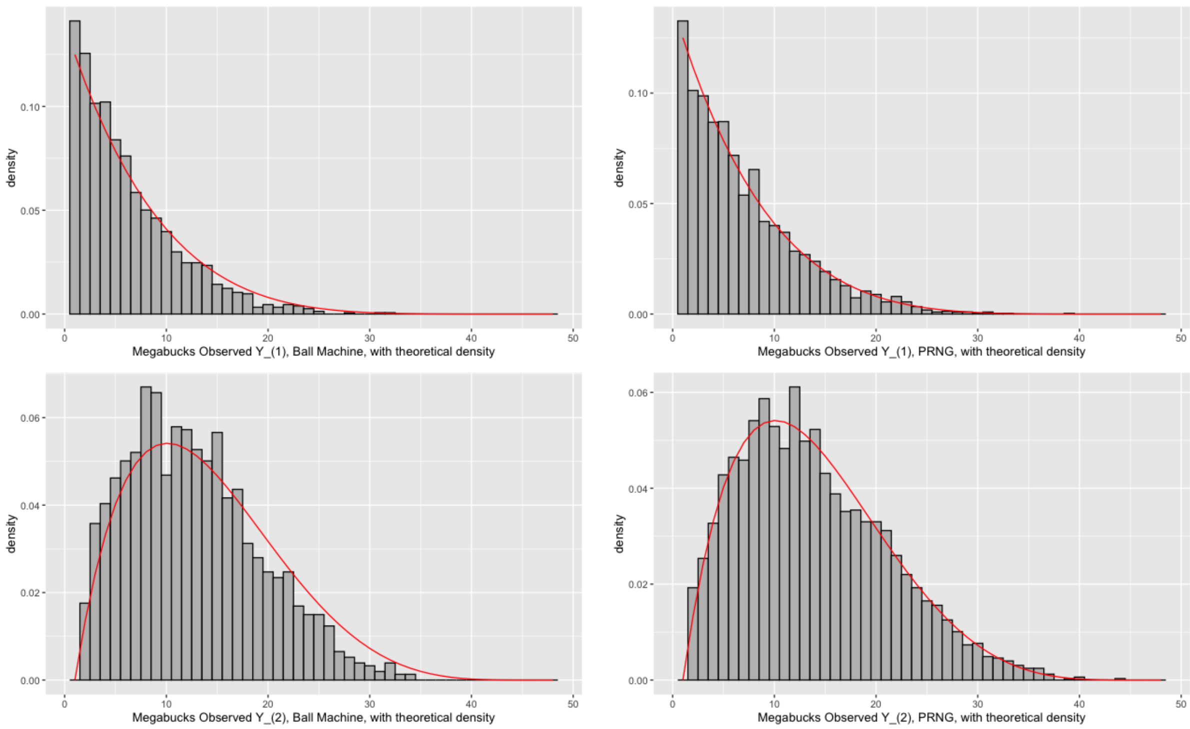
We do see some difference between the two p-values for the PRNG period, but the result is the same for any reasonable confidence level. We find that there is extremely strong evidence that the historical Megabucks results drawn using a physical ball machine do *not* fit the fairness distribution using either testing procedure; whereas there is *no* significant evidence using either testing procedure that the results drawn using a pseudorandom number generator do not fit the fairness distribution.

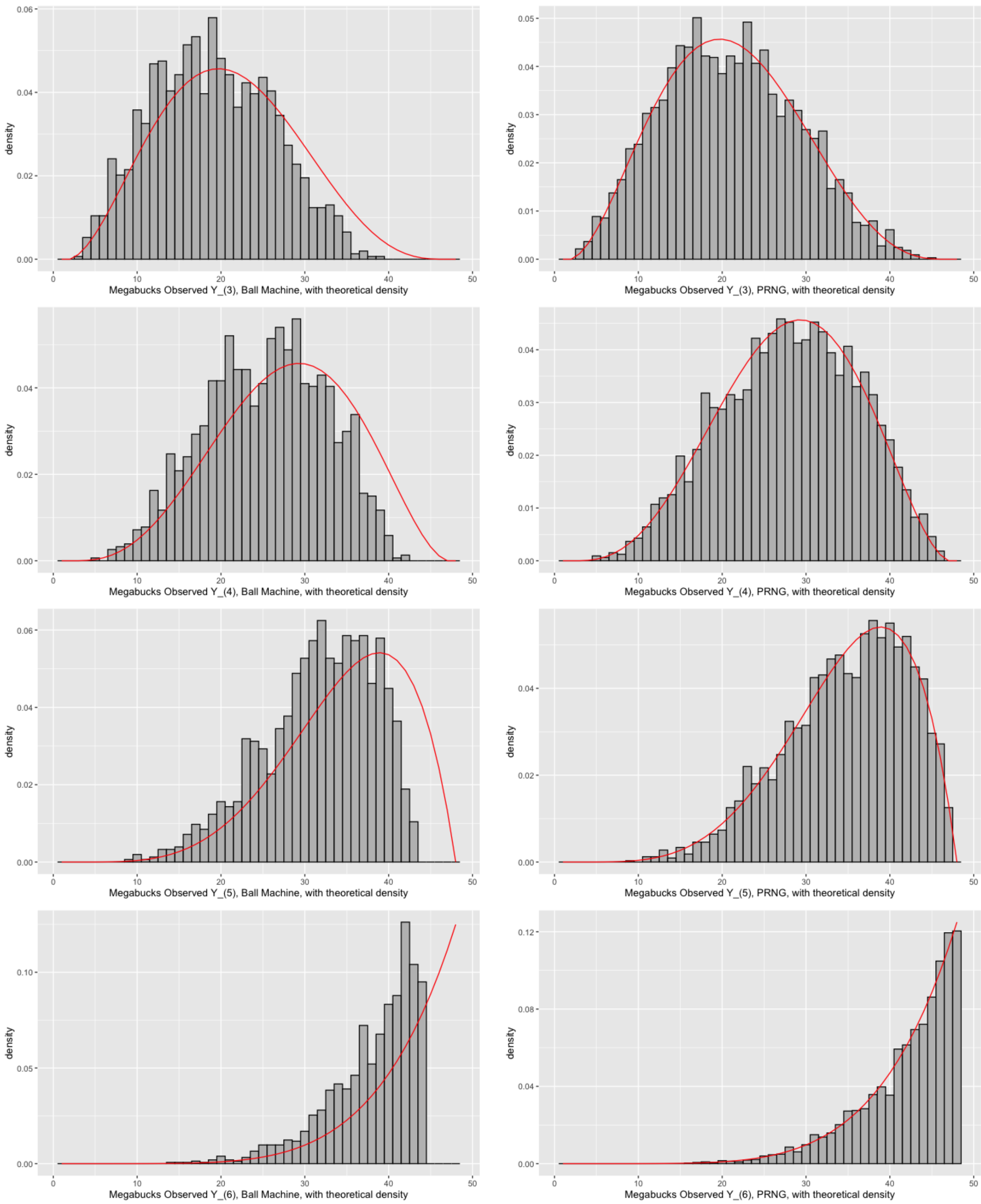
We picture here histograms of the simulated Q distributions; left is ball machine, right is PRNG.



The vertical dash marks the observed Q value for the draw period.

While it is perhaps surprising that the physical selection mechanism that is the ball machine provides less fairness than does the PRNG, we do note that this result matches the histograms of the historical results of the order statistics. We see that the ball machine results (pictured left below) depart from the theoretical distribution more than do the PRNG results (right below):





Our original theory to explain the departure of the physical randomization device from the theoretical distribution was that the $m = 1537$ drawings were insufficient for the Law of Large Numbers to manifest. However, that is not borne out by to-come results of other lotteries with similar or smaller sample sizes also using ball machine draw mechanisms. Currently, we have no hypothesis to explain these results beyond "luck of the draw." However, we note that the PRNG in use by Oregon Lottery for Megabucks seems sufficiently well-programmed.

We continue this two-pronged analysis (asymptotic and Monte Carlo) on two other lottery games, Powerball and Mega Millions.

6 Numerical Results: Powerball

Powerball is a multi-state draw game running since 1992. It has used a ball machine as its draw mechanism for its entire history. However, it has undergone multiple rule changes. While k has consistently been 5, N has changed [9]. This game also features a "bonus ball" (called the Powerball) – this is a secondary drawing from a separate pool of balls. Matching the Powerball increases the prize payout for non-jackpot prizes [7]. For our analysis we will ignore the bonus ball, the rules of which have also changed over time. In that case we divide Powerball into six periods:

Period	First Draw	Last Draw	k/N
1	22-Apr-1992	1-Nov-1997	5/45
2	5-Nov-1997	5-Oct-2002	5/49
3	9-Oct-2002	27-Aug-2005	5/53
4	31-Aug-2005	3-Jan-2009	5/55
5	7-Jan-2009	3-Oct-2015	5/59
6	7-Oct-2015	13-Apr-2022	5/69

For each of these periods we use eqn. (26), eqn. (27), and eqn. (28):

$$\begin{aligned}
\mathbb{E}[Y_{(i)}] &= \frac{N+1}{k+1}i, \\
\mathrm{Var}[Y_{(i)}] &= \frac{i(N+1)(k-i+1)(N-k)}{(k+2)(k+1)^2}, \\
\mathrm{Cov}[Y_{(i)}, Y_{(j)}] &= \frac{i(k-j+1)(N+1)(N-k)}{(k+2)(k+1)^2}
\end{aligned}$$

to calculate $\boldsymbol{\mu}$, $\mathbf{V}$, and $\mathbf{V}^{-1}$ (see the appendix for $\mathbf{V}$ and $\mathbf{V}^{-1}$ for each period); and the historical results to calculate $\bar{\mathbf{y}}$:

Period	m	$\bar{\mathbf{y}}$	$\boldsymbol{\mu}$
1	577	(7.685, 15.224, 22.738, 30.459, 38.139)	(7.667, 15.333, 23, 30.667, 38.333)
2	514	(8.280, 17.019, 24.724, 33.397, 41.728)	(8.333, 16.667, 25, 33.333, 41.667)
3	302	(9.026, 18.142, 27.636, 35.781, 45.093)	(9, 18, 27, 36, 45)
4	804	(9.405, 18.952, 27.712, 37.724, 46.946)	(9.333, 18.667, 28, 37.333, 46.667)
5	704	(10.134, 19.761, 30.121, 40.337, 49.541)	(10, 20, 30, 40, 50)
6	715	(11.945, 23.477, 35.283, 47.180, 58.418)	(11.667, 23.333, 35, 46.667, 58.333)

Under the null hypothesis of fairness, each $Q = m(\bar{\mathbf{y}} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\bar{\mathbf{y}} - \boldsymbol{\mu})$ approaches a χ^2_5 distribution. For our Monte Carlo analysis, we again generate m lottery results 5,000 times and compute Q each time, generating a distribution of Q for each period. We find the following:

Period	Observed Q	CLT p-value	MC p-value
1	1.060	0.958	0.995
2	6.690	0.245	0.628
3	5.348	0.375	0.588
4	4.889	0.430	0.679
5	10.726	0.057	0.455
6	2.405	0.791	0.984

Using our default confidence level of $\alpha = 0.05$, we see that despite the disparity of the results between our testing methods, our conclusions are consistent: there is not sufficient evidence that the ball machine mechanism used for Powerball deviates from its theoretical "fair" distribution. See the appendix for histograms of ordered historical results as compared to their theoretical distributions.

7 Numerical Results: Mega Millions

Similarly to Powerball, Mega Millions is also a multi-state draw game with a bonus ball (the "mega ball") aspect [5]. We will again ignore the bonus ball in our analysis. Again like Powerball, Mega Millions has always used a ball machine as its selection method. Although Mega Millions has been running since 1996 [8], our data runs from 30 March 2010 through 12 April 2022 [3]. Over this period there have been two rule changes to the game [8], and so we divide Mega Millions into three periods:

Period	First Draw	Last Draw	k/N
1	30-Mar-2010	18-Oct-2013	5/56
2	22-Oct-2013	27-Oct-2017	5/75
3	31-Oct-2017	12-Apr-2022	5/70

For each of these periods we use

$$\text{eqn. (26)}, \mathbb{E}[Y_{(i)}] = \frac{N+1}{k+1}i,$$

$$\text{eqn. (27)}, \text{Var}[Y_{(i)}] = \frac{i(N+1)(k-i+1)(N-k)}{(k+2)(k+1)^2},$$

$$\text{eqn. (28)}, \text{Cov}[Y_{(i)}, Y_{(j)}] = \frac{i(k-j+1)(N+1)(N-k)}{(k+2)(k+1)^2}$$

to calculate $\boldsymbol{\mu}$, $\mathbf{V}$, and $\mathbf{V}^{-1}$ (see the appendix for $\mathbf{V}$ and $\mathbf{V}^{-1}$ for each period); and the historical results to calculate $\bar{\mathbf{y}}$:

Period	m	$\bar{\mathbf{y}}$	$\boldsymbol{\mu}$
1	372	(8.715, 18.132, 27.323, 36.954, 46.524)	(9.5, 19, 28.5, 38, 47.5)
2	420	(12.386, 24.833, 36.862, 49.460, 62.450)	(12.667, 25.333, 38, 50.667, 63.333)
3	465	(11.333, 22.989, 34.430, 46.247, 58.232)	(11.833, 23.667, 35.5, 47.333, 59.167)

Under the null hypothesis of fairness, each $Q = m(\bar{\mathbf{y}} - \boldsymbol{\mu})^T\mathbf{V}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu})$ approaches a χ^2_5 distribution. For our Monte Carlo analysis, we again generate m lottery results 5,000 times and compute Q each time, generating a distribution of Q for each period. We find the following:

Period	Observed Q	CLT p-value	MC p-value
1	9.100	0.105	0.303
2	4.722	0.451	0.768
3	5.637	0.343	0.705

As with Powerball, we see some significant disparity between the p-values for our selection methods, but our conclusions are consistent at $\alpha = 0.05$: there is not sufficient evidence to doubt the fairness of the ball machine selection method in use for Mega Millions. See the appendix for histograms of ordered historical results as compared to their theoretical distributions.

8 Conclusion

We observe that for all three lottery games, our conclusions were consistent at confidence levels of $\alpha \leq 0.05$ (only Powerball Period 5 was borderline, with a CLT p-value of 0.057). However, we consistently had quite different p-values between our asymptotic test and our Monte Carlo test, in contradiction with Coronel-Fabrizio et al. [1], who consistently found p-values within 0.01 across methods. Were we to surmise regarding our disparate results, we would question as affecting factors the random seed and the random sampling variable for the Monte Carlo simulations. It is also entirely possible that the lottery results themselves are a significant factor.

Given how different our results were across the two tests, we are less confident in this method overall than Coronel-Fabrizio et al. [1], but we do find it to be an interesting starting point for lottery auditing. We suggest further analysis of sample size for the asymptotic test, simulation size analysis for the Monte Carlo, and power analysis for both methods. We would also suggest that other non-parametric tests might also be worthy of research.

Based on our observed results, it does not seem likely for an organized lottery's draw results to significantly deviate from their theoretical distribution. Therefore, as opposed to the more general question put forth by Coronel-Fabrizio et al. [1] and this paper, "Does this lottery fit its theoretical distribution?", we would suggest as a research question, "Can we determine how closely this lottery fits its theoretical distribution?" Answering this question is beyond the scope of the hypothesis tests proposed by Coronel-Fabrizio et al. [1] and this paper, but is, in this author's opinion, an interesting potential topic of future research.

References

[1] H.F. Coronel-Brizio, A.R. Hernández-Montoya, F. Rapallo, E. Scalas. Statistical auditing and randomness test of lotto k/N-type games. *Physica A: Statistical Mechanics and its Applications*, Volume 387, Issue 25: 6385–6390, 2008. https://arxiv.org/abs/0806.4595

[2] LottoAmerica. *History of Powerball.* Retrieved 10 May 2022, from https://www.lottoamerica.com/powerball/history

[3] J. Nelson, Public Records Specialist, Oregon Lottery. Historical results of Megabucks, Mega Millions, and Powerball draws; and clarification of game rules. Public records request and email correspondence, 3 April 2002 – 11 May 2022. Available on https://github.com/koddson/LotteryStats

[4] K. Oddson. *GitHub Repository of data, R code, and references*, 2022. Retrieve at https://github.com/koddson/LotteryStats

[5] Oregon Lottery. *How to Play Mega Millions.* Retrieved 22 April 2022, from https://www.oregonlottery.org/wp-content/uploads/2020/06/how-to-play-mega-millions.pdf

[6] Oregon Lottery. *How to Play Oregon's Game Megabucks.* Retrieved 23 April 2022, from https://www.oregonlottery.org/wp-content/uploads/2020/06/how-to-play-megabucks.pdf

[7] Oregon Lottery. *How to Play Powerball.* Retrieved 22 April 2022, from https://www.oregonlottery.org/wp-content/uploads/2022/01/HTP_Powerball_Web_EN_01-18-22.pdf

[8] Wikipedia. *Mega Millions.* Retrieved 10 May 2022, from https://en.wikipedia.org/wiki/Mega_Millions

[9] Wikipedia. *Powerball.* Retrieved 13 May 2022, from https://en.wikipedia.org/wiki/Powerball

9 Appendix

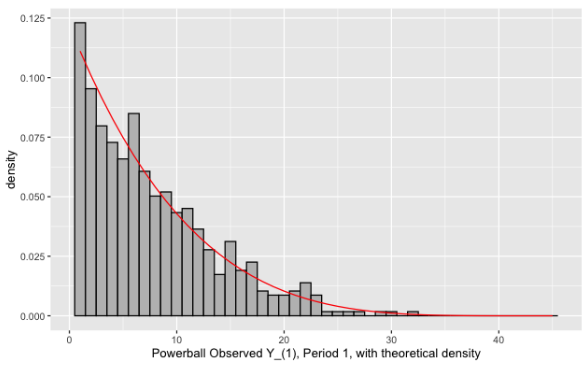
9.1 Powerball Period 1, 22-Apr-1992 - 1-Nov-1997, 5/45 ($m = 577$):

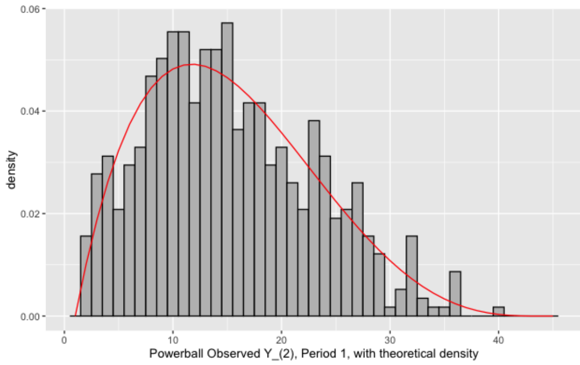
$$\boldsymbol{\mu} = [7.667,\ 15.333,\ 23,\ 30.667,\ 38.333]$$

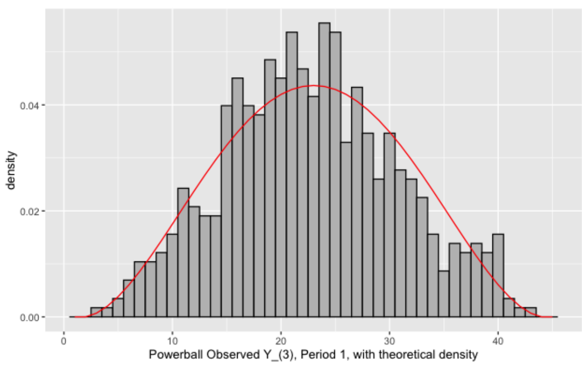
$$\mathbf{V} = \begin{bmatrix} 2300/63 & 1840/63 & 460/21 & 920/63 & 460/63 \\ 1840/63 & 3680/63 & 920/21 & 1840/63 & 920/63 \\ 460/21 & 920/21 & 460/7 & 920/21 & 460/21 \\ 920/63 & 1840/63 & 920/21 & 3680/63 & 1840/63 \\ 460/63 & 920/63 & 460/21 & 1840/63 & 2300/63 \end{bmatrix}$$

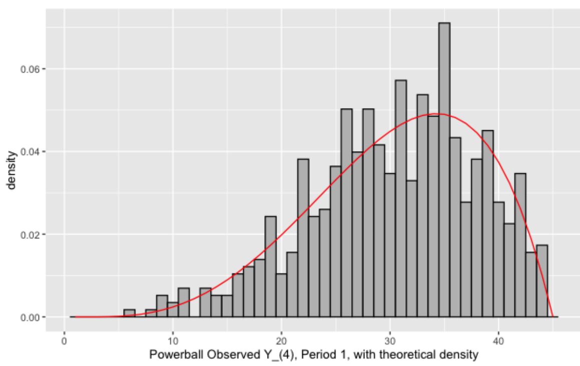
$$\mathbf{V}^{-1} = \begin{bmatrix} 21/460 & -21/920 & 0 & 0 & 0 \\ -21/920 & 21/460 & -21/920 & 0 & 0 \\ 0 & -21/920 & 21/460 & -21/920 & 0 \\ 0 & 0 & -21/920 & 21/460 & -21/920 \\ 0 & 0 & 0 & -21/920 & 21/460 \end{bmatrix}$$

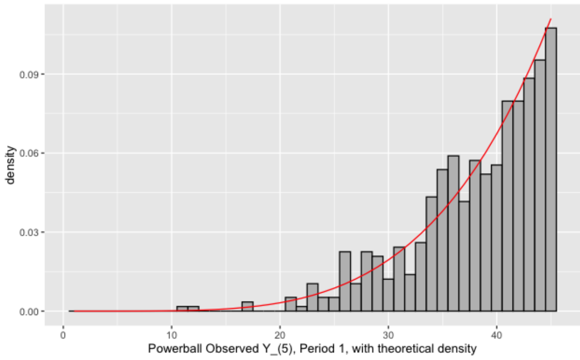
Histograms of observed ordered results with theoretical densities overlaid:



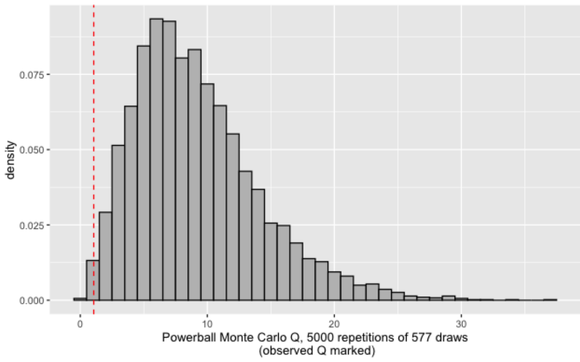








Histogram of simulated Q distribution with observed Q marked:



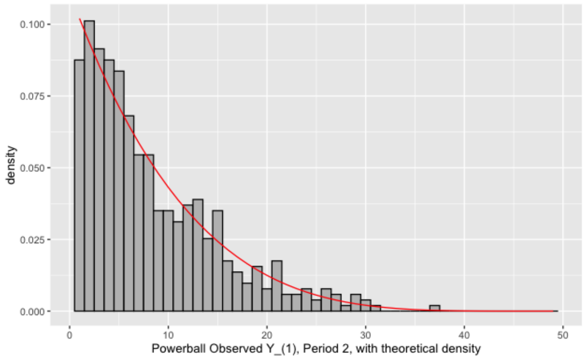
9.2 Powerball Period 2, 5-Nov-1997 - 5-Oct-2002, 5/49 ($m = 514$):

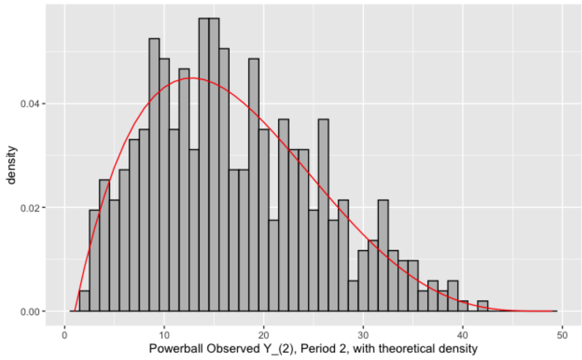
$$\boldsymbol{\mu} = [8.333,\ 16.667,\ 25,\ 33.333,\ 41.667]$$

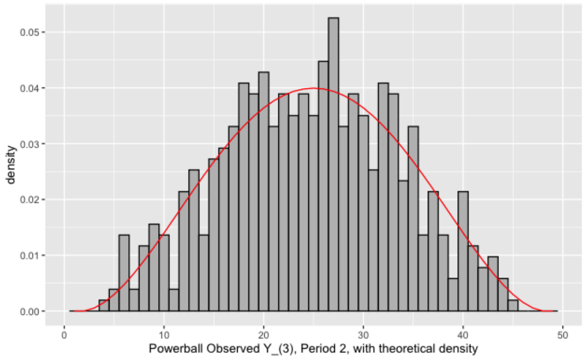
$$\mathbf{V} = \begin{bmatrix} 2750/63 & 2200/63 & 550/21 & 1100/63 & 550/63 \\ 2200/63 & 4400/63 & 1100/21 & 2200/63 & 1100/63 \\ 550/21 & 1100/21 & 550/7 & 1100/21 & 550/21 \\ 1100/63 & 2200/63 & 1100/21 & 4400/63 & 2200/63 \\ 550/63 & 1100/63 & 550/21 & 2200/63 & 2750/63 \end{bmatrix}$$

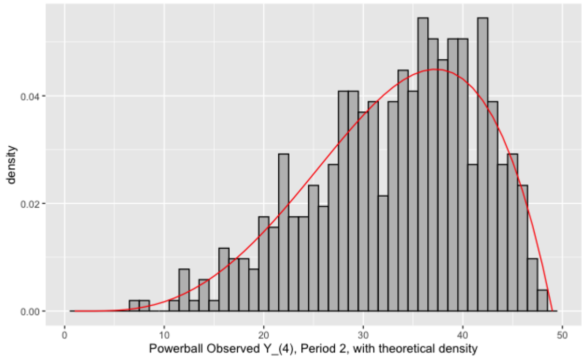
$$\mathbf{V}^{-1} = \begin{bmatrix} 21/550 & -21/1100 & 0 & 0 & 0 \\ -21/1100 & 21/550 & -21/1100 & 0 & 0 \\ 0 & -21/1100 & 21/550 & -21/1100 & 0 \\ 0 & 0 & -21/1100 & 21/550 & -21/1100 \\ 0 & 0 & 0 & -21/1100 & 21/550 \end{bmatrix}$$

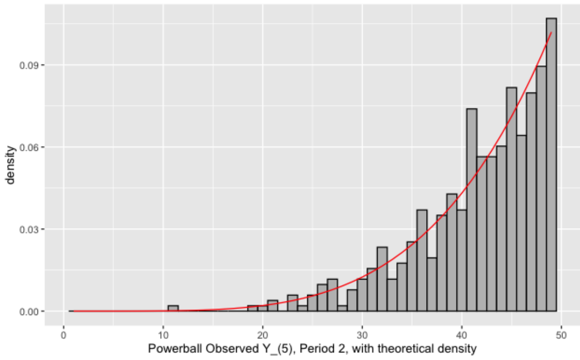
Histograms of observed ordered results with theoretical densities overlaid:



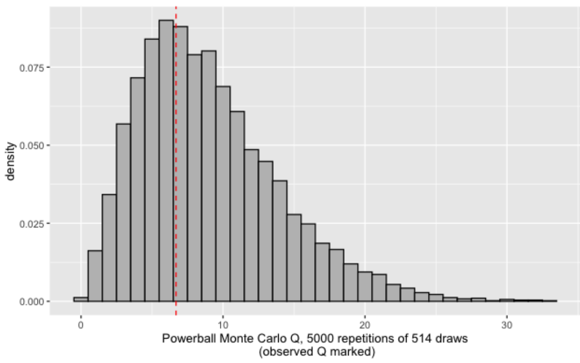








Histogram of simulated Q distribution with observed Q marked:



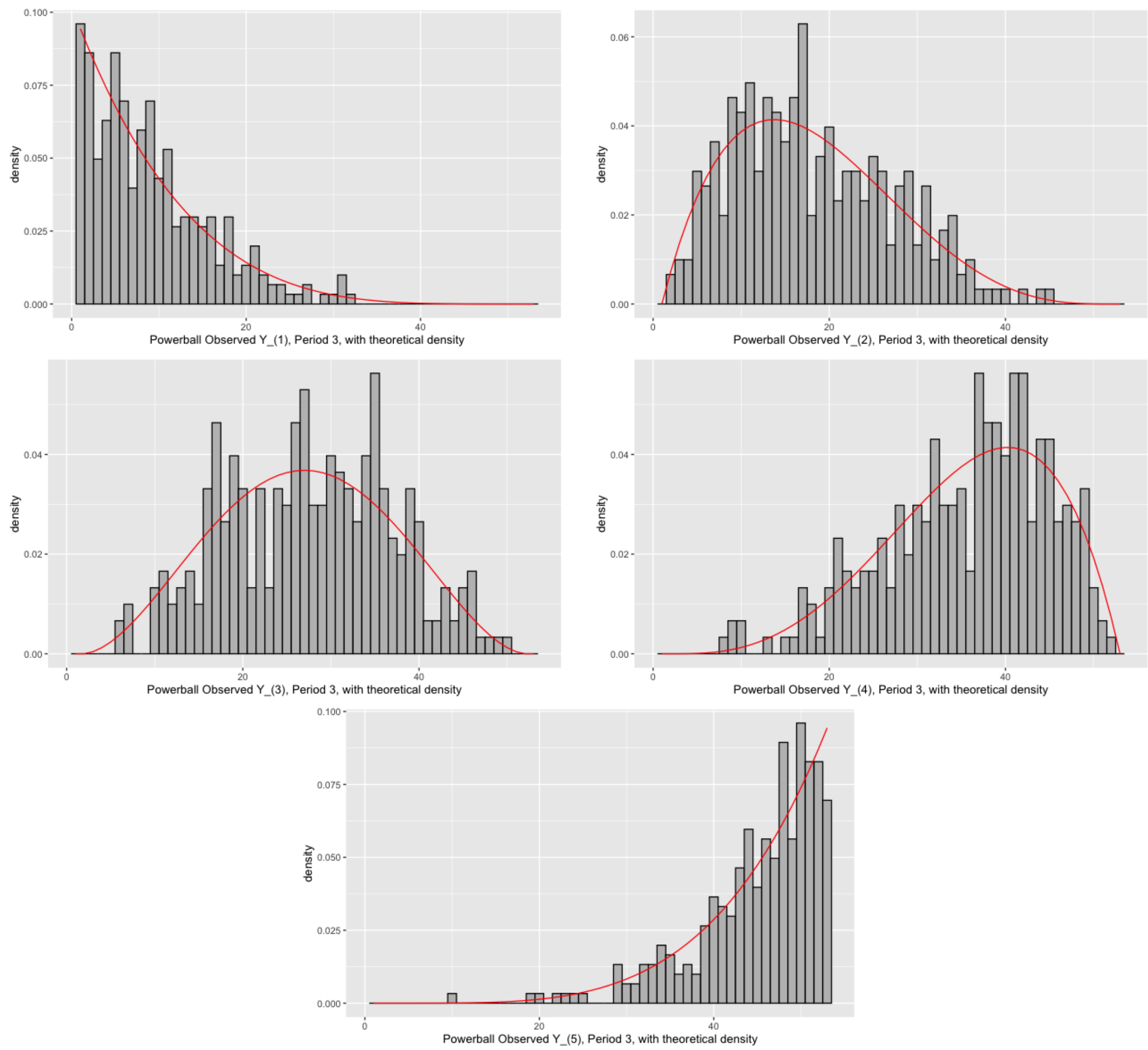
9.3 Powerball Period 3, 9-Oct-2002 - 27-Aug-2005, 5/53 ($m = 302$):

$$\boldsymbol{\mu} = [9,\ 18,\ 27,\ 36,\ 45]$$

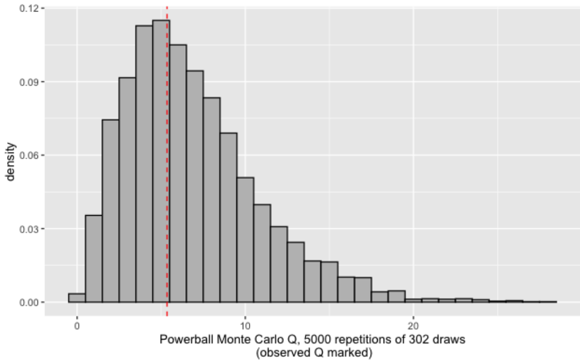
$$\mathbf{V} = \begin{bmatrix} 360/7 & 288/7 & 216/7 & 144/7 & 72/7 \\ 288/7 & 576/7 & 432/7 & 288/7 & 144/7 \\ 216/7 & 432/7 & 648/7 & 432/7 & 216/7 \\ 144/7 & 288/7 & 432/7 & 576/7 & 288/7 \\ 72/7 & 144/7 & 216/7 & 288/7 & 360/7 \end{bmatrix}$$

$$\mathbf{V}^{-1} = \begin{bmatrix} 7/216 & -7/432 & 0 & 0 & \\ -7/432 & 7/216 & -7/432 & 0 & 0 \\ 0 & -7/432 & 7/216 & -7/432 & 0 \\ 0 & 0 & -7/432 & 7/216 & -7/432 \\ 0 & 0 & 0 & -7/432 & 7/216 \end{bmatrix}$$

Histograms of observed ordered results with theoretical densities overlaid:



Histogram of simulated Q distribution with observed Q marked:



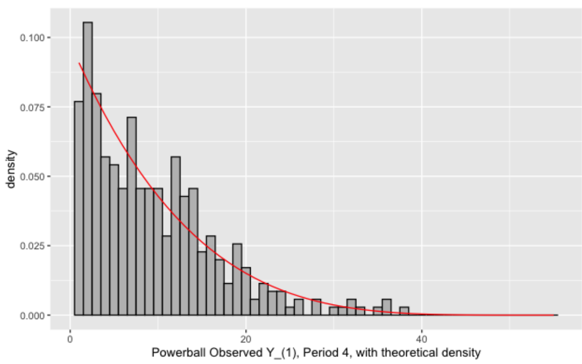
9.4 Powerball Period 4, 31-Aug-2005 - 3-Jan-2009, 5/55 ($m = 804$):

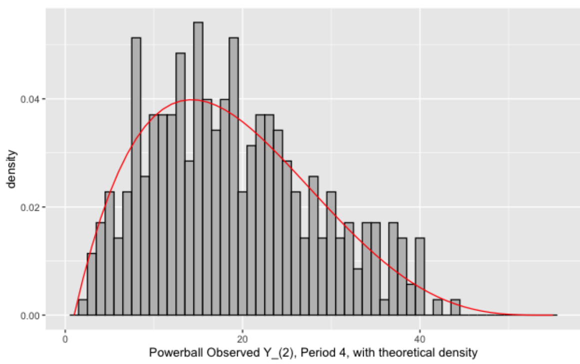
$$\boldsymbol{\mu} = [9.333,\ 18.667,\ 28,\ 37.333,\ 46.667]$$

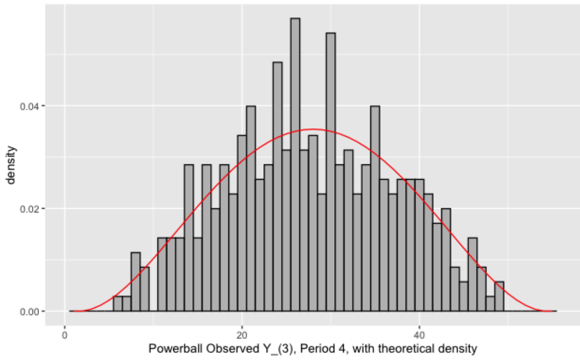
$$\mathbf{V} = \begin{bmatrix} 500/9 & 400/9 & 100/3 & 200/9 & 100/9 \\ 400/9 & 800/9 & 200/3 & 400/9 & 200/9 \\ 100/3 & 200/3 & 100 & 200/3 & 100/3 \\ 200/9 & 400/9 & 200/3 & 800/9 & 400/9 \\ 100/9 & 200/9 & 100/3 & 400/9 & 500/9 \end{bmatrix}$$

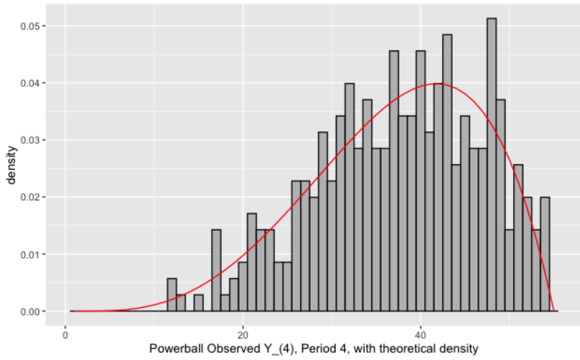
$$\mathbf{V}^{-1} = \begin{bmatrix} 3/100 & -3/200 & 0 & 0 & \\ -3/200 & 3/100 & -3/200 & 0 & 0 \\ 0 & -3/200 & 3/100 & -3/200 & 0 \\ 0 & 0 & -3/200 & 3/100 & -3/200 \\ 0 & 0 & 0 & -3/200 & 3/100 \end{bmatrix}$$

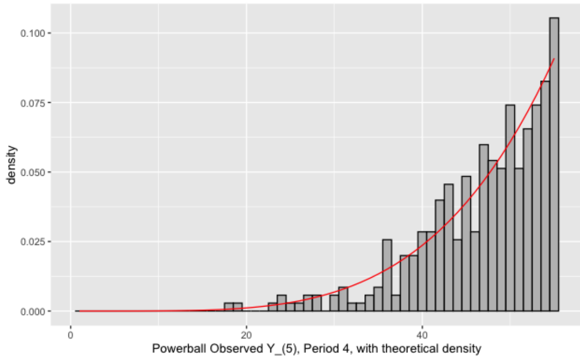
Histograms of observed ordered results with theoretical densities overlaid:



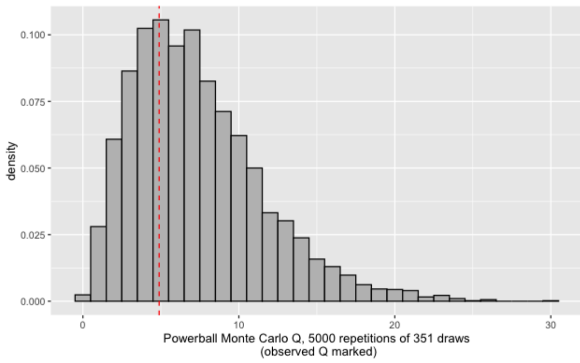








Histogram of simulated Q distribution with observed Q marked:



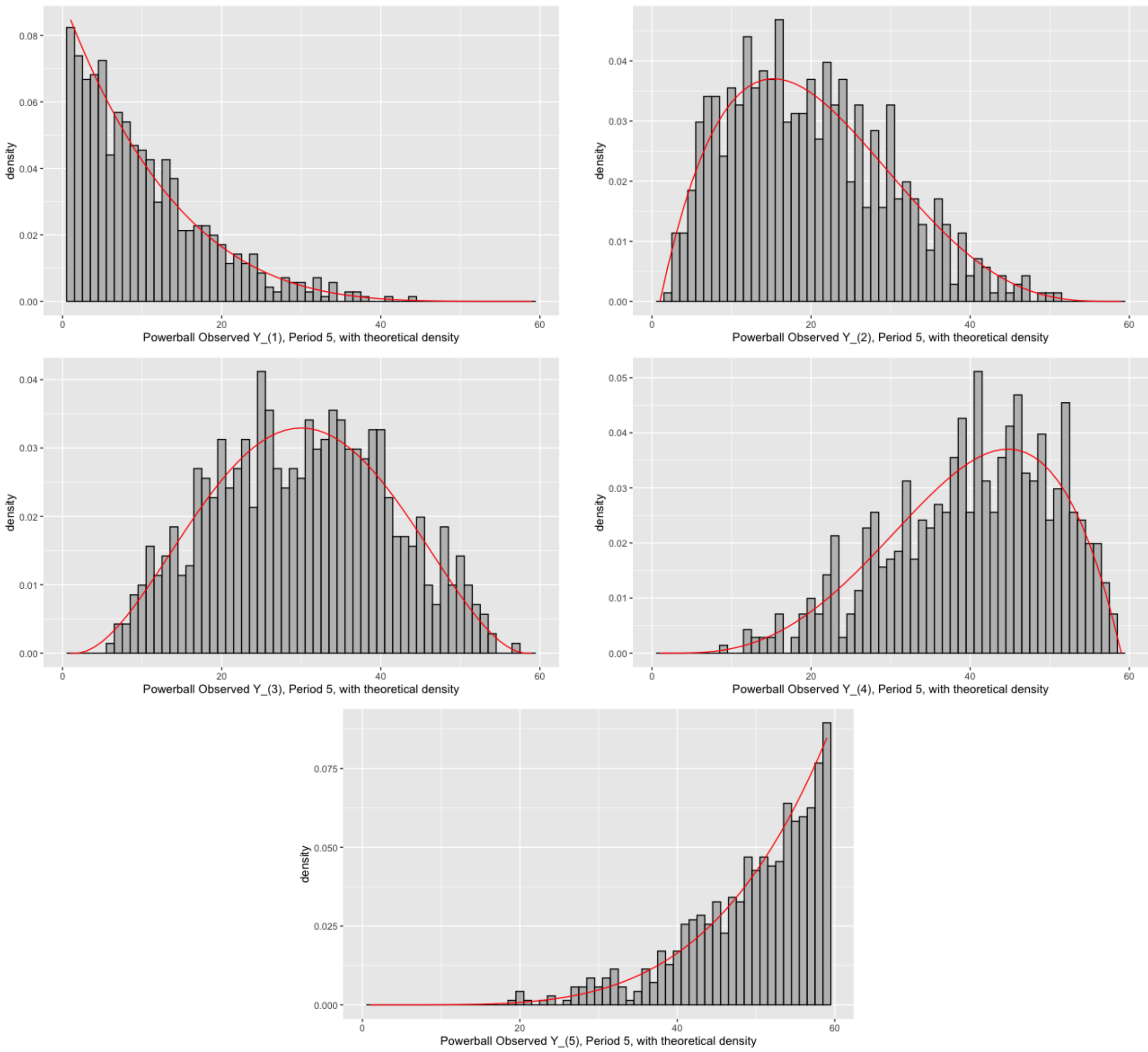
9.5 Powerball Period 5, 7-Jan-2009 - 3-Oct-2015, 5/59 ($m = 704$):

$$\boldsymbol{\mu} = [10,\ 20,\ 30,\ 40,\ 50]$$

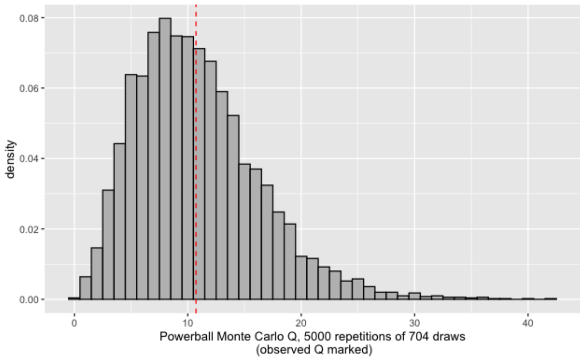
$$\mathbf{V} = \begin{bmatrix} 450/7 & 360/7 & 270/7 & 180/7 & 90/7 \\ 360/7 & 720/7 & 540/7 & 360/7 & 180/7 \\ 270/7 & 540/7 & 810/7 & 540/7 & 270/7 \\ 180/7 & 360/7 & 540/7 & 720/7 & 360/7 \\ 90/7 & 180/7 & 270/7 & 360/7 & 450/7 \end{bmatrix}$$

$$\mathbf{V}^{-1} = \begin{bmatrix} 7/270 & -7/540 & 0 & 0 & \\ -7/540 & 7/270 & -7/540 & 0 & 0 \\ 0 & -7/540 & 7/270 & -7/540 & 0 \\ 0 & 0 & -7/205400 & 7/270 & -7/540 \\ 0 & 0 & 0 & -7/540 & 7/270 \end{bmatrix}$$

Histograms of observed ordered results with theoretical densities overlaid:



Histogram of simulated Q distribution with observed Q marked:



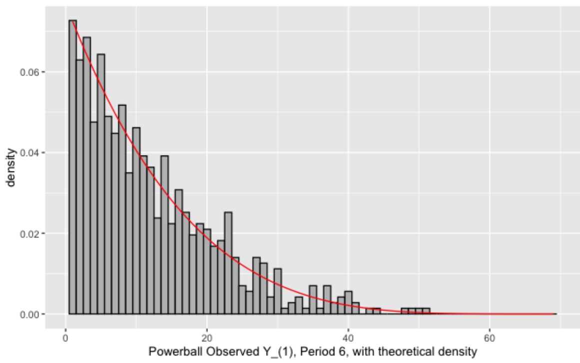
9.6 Powerball Period 6, 7-Oct-2015 - 13-Apr-2022, 5/69 ($m = 715$):

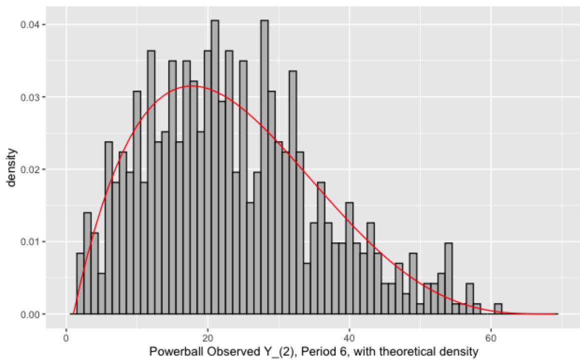
$$\boldsymbol{\mu} = [11.667,\ 23.333,\ 35,\ 46.667,\ 58.333]$$

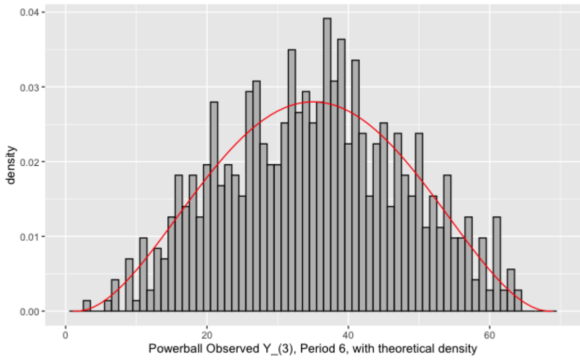
$$\mathbf{V} = \begin{bmatrix} 800/9 & 640/9 & 160/3 & 320/9 & 160/9 \\ 640/9 & 1280/9 & 320/3 & 640/9 & 320/9 \\ 160/3 & 320/3 & 160 & 320/3 & 160/3 \\ 320/9 & 640/9 & 320/3 & 1280/9 & 640/9 \\ 160/9 & 320/9 & 160/3 & 640/9 & 800/9 \end{bmatrix}$$

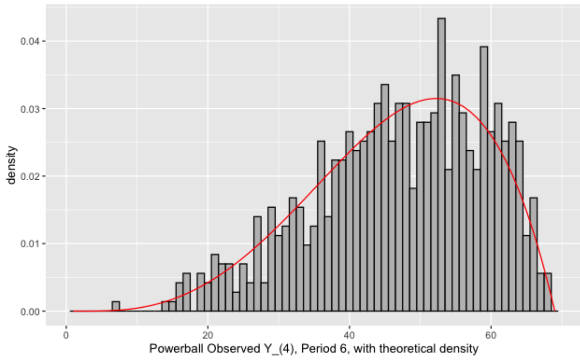
$$\mathbf{V}^{-1} = \begin{bmatrix} 3/160 & -3/320 & 0 & 0 & \\ -3/320 & 3/160 & -3/320 & 0 & 0 \\ 0 & -3/320 & 3/160 & -3/320 & 0 \\ 0 & 0 & -3/320 & 3/160 & -3/320 \\ 0 & 0 & 0 & -3/320 & 3/160 \end{bmatrix}$$

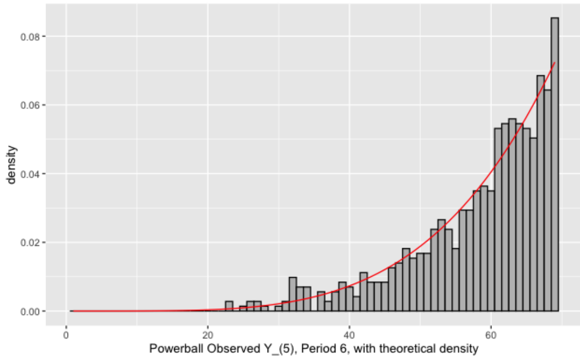
Histograms of observed ordered results with theoretical densities overlaid:



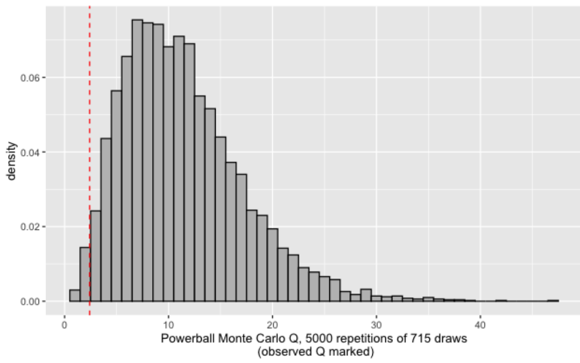








Histogram of simulated Q distribution with observed Q marked:



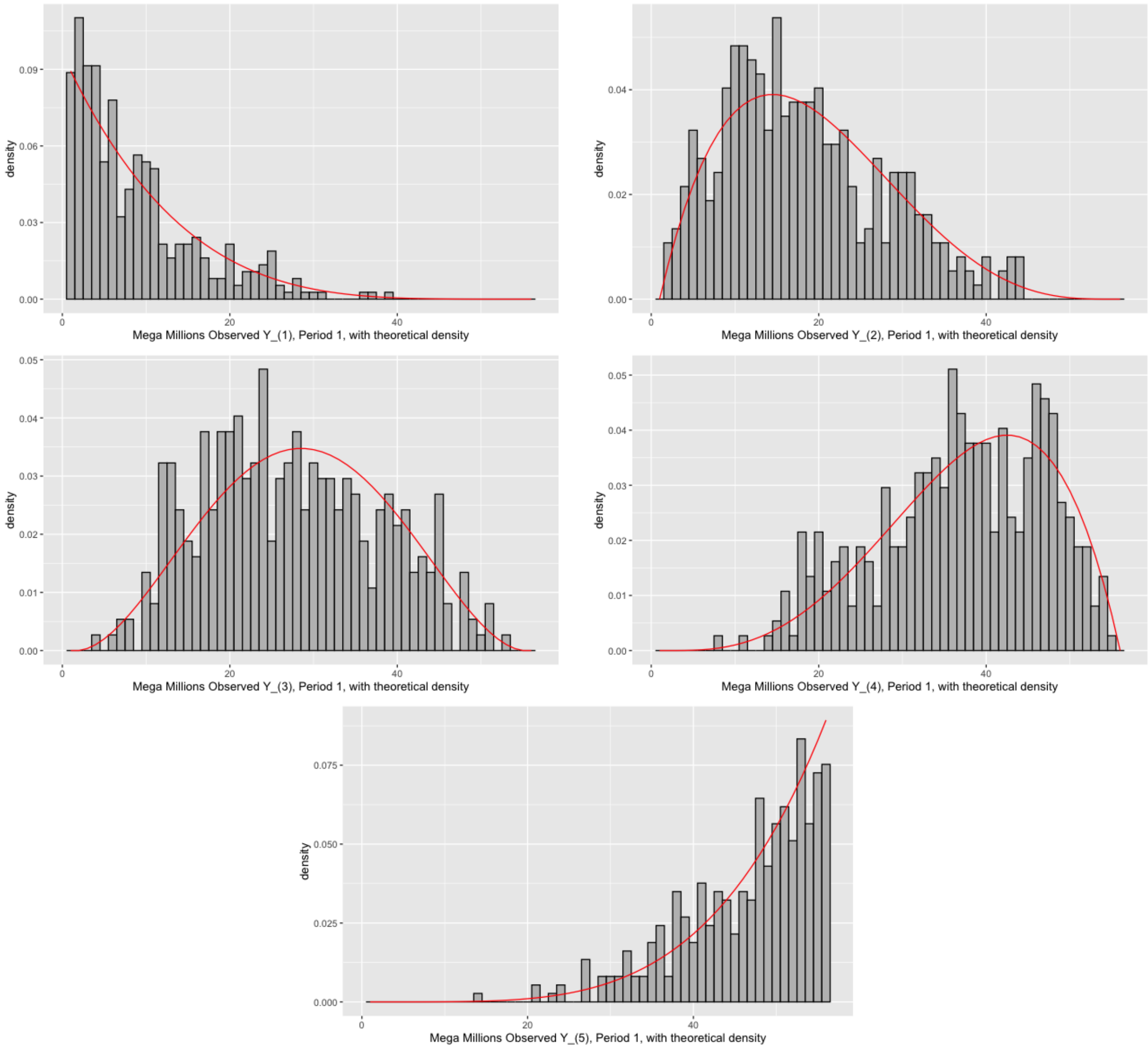
9.7 Mega Millions Period 1, 30-Mar-2010 - 18-Oct-2013, 5/56 ($m = 372$):

$$\boldsymbol{\mu} = [9.5,\ 19,\ 28.5,\ 38,\ 47.5]$$

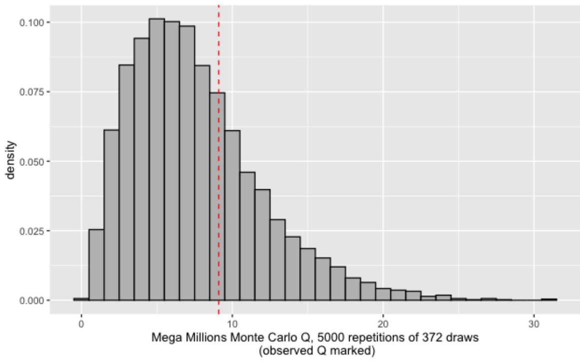
$$\mathbf{V} = \begin{bmatrix} 1615/28 & 323/7 & 969/28 & 323/14 & 323/28 \\ 323/7 & 646/7 & 969/14 & 323/7 & 323/14 \\ 969/28 & 969/14 & 2907/28 & 969/14 & 969/28 \\ 323/14 & 323/7 & 969/14 & 646/7 & 323/7 \\ 323/28 & 323/14 & 969/28 & 323/7 & 1615/28 \end{bmatrix}$$

$$\mathbf{V}^{-1} = \begin{bmatrix} 28/969 & -14/969 & 0 & 0 & \\ -14/969 & 28/969 & -14/969 & 0 & 0 \\ 0 & -14/969 & 28/969 & -14/969 & 0 \\ 0 & 0 & -14/969 & 28/969 & -14/969 \\ 0 & 0 & 0 & -14/969 & 28/969 \end{bmatrix}$$

Histograms of observed ordered results with theoretical densities overlaid:



Histogram of simulated Q distribution with observed Q marked:



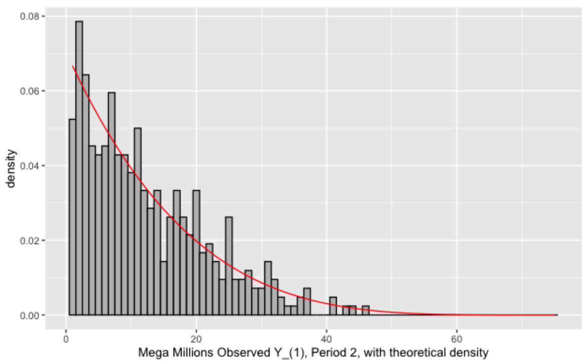
9.8 Mega Millions Period 2, 22-Oct-2013 - 27-Oct-2017, 5/75 ($m = 420$):

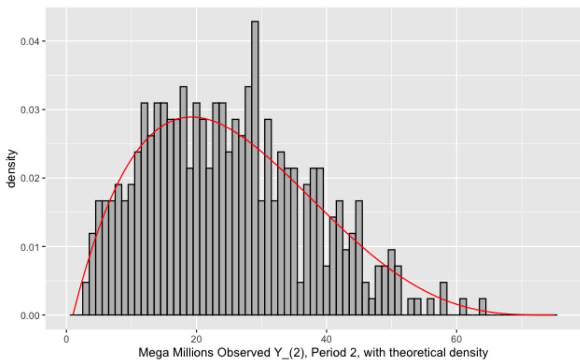
$$\boldsymbol{\mu} = [12.667,\ 25.333,\ 38,\ 50.667,\ 63.333]$$

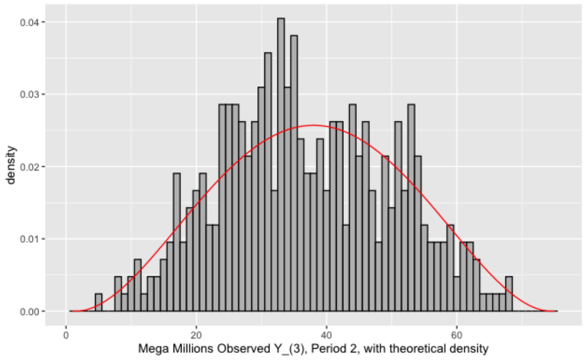
$$\mathbf{V} = \begin{bmatrix} 950/9 & 760/9 & 190/3 & 380/9 & 190/9 \\ 760/9 & 1520/9 & 380/3 & 760/9 & 380/9 \\ 190/3 & 380/3 & 190 & 380/3 & 190/3 \\ 380/9 & 760/9 & 380/3 & 1520/9 & 760/9 \\ 190/9 & 380/9 & 190/3 & 760/9 & 950/9 \end{bmatrix}$$

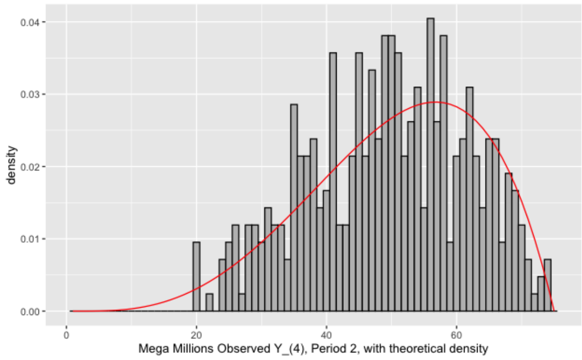
$$\mathbf{V}^{-1} = \begin{bmatrix} 3/190 & -3/380 & 0 & 0 & \\ -3/380 & 3/190 & -3/380 & 0 & 0 \\ 0 & -3/380 & 3/190 & -3/380 & 0 \\ 0 & 0 & -3/380 & 3/190 & -3/380 \\ 0 & 0 & 0 & -3/380 & 3/190 \end{bmatrix}$$

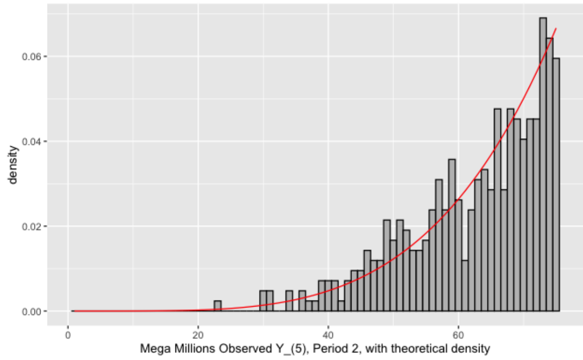
Histograms of observed ordered results with theoretical densities overlaid:



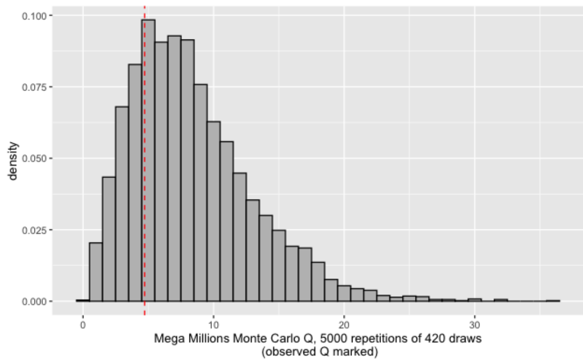








Histogram of simulated Q distribution with observed Q marked:



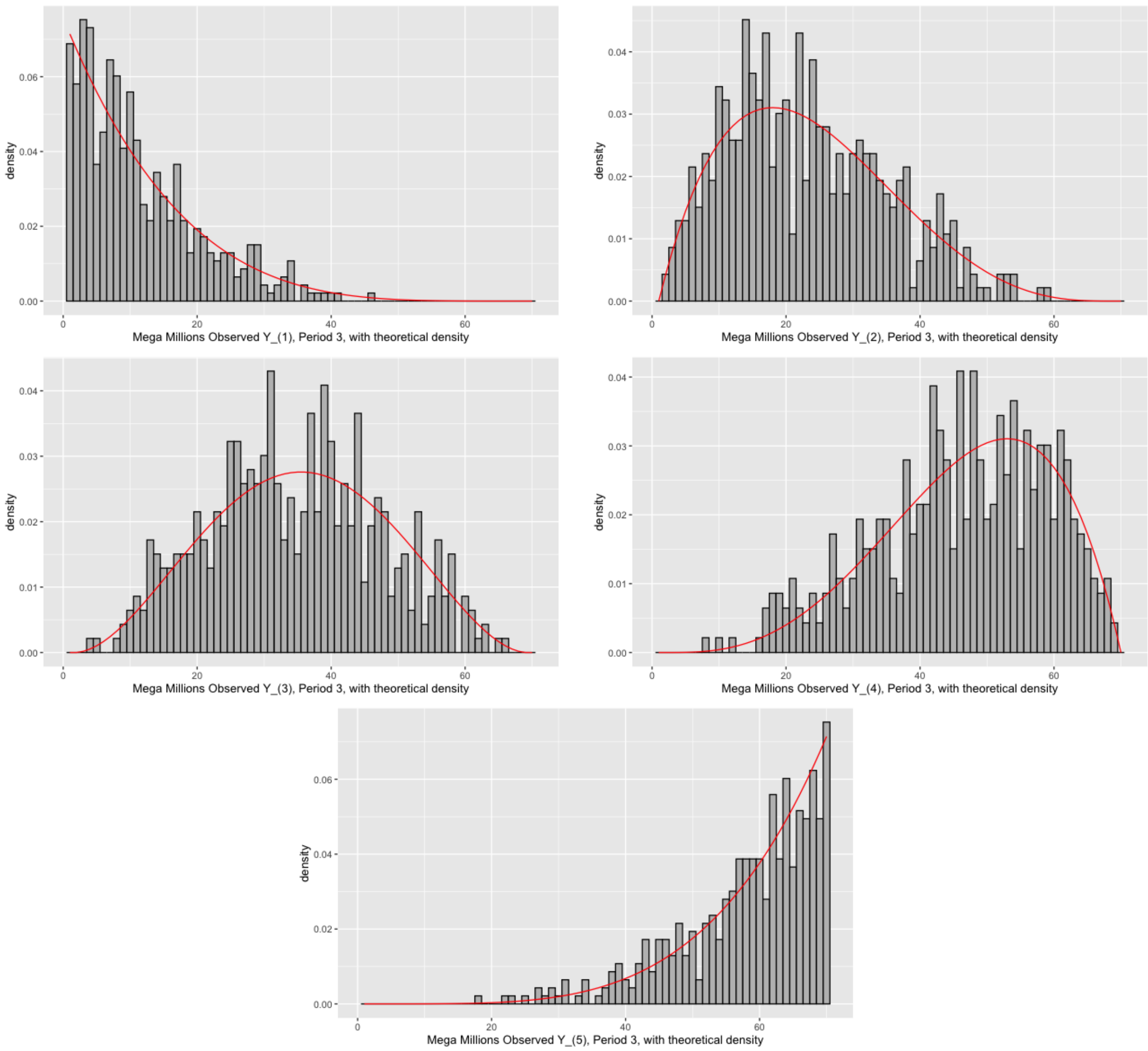
9.9 Mega Millions Period 3, 22-Oct-2013 - 27-Oct-2017, 5/75 ($m = 465$):

$$\boldsymbol{\mu} = [11.833, 23.667, 35.5, 47.333, 59.167]$$

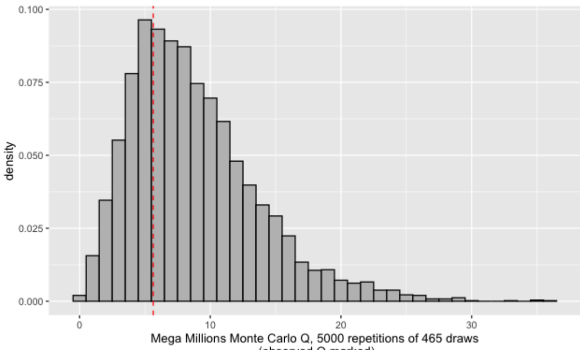
$$\mathbf{V} = \begin{bmatrix} 23075/252 & 4615/63 & 4615/84 & 4615/126 & 4615/252 \\ 4615/63 & 9230/63 & 4615/42 & 4615/63 & 4615/126 \\ 4615/84 & 4615/42 & 4615/28 & 4615/42 & 4615/84 \\ 4615/126 & 4615/63 & 4615/42 & 9230/63 & 4615/63 \\ 4615/252 & 4615/126 & 4615/84 & 4615/63 & 23075/252 \end{bmatrix}$$

$$
\mathbf{V}^{-1} = \begin{bmatrix} 84/4615 & -42/4615 & 0 & 0 & \\ -42/4615 & 84/4615 & -42/4615 & 0 & 0 \\ 0 & -42/4615 & 84/4615 & -42/4615 & 0 \\ 0 & 0 & -42/4615 & 84/4615 & -42/4615 \\ 0 & 0 & 0 & -42/4615 & 84/4615 \end{bmatrix}
$$

Histograms of observed ordered results with theoretical densities overlaid:



Histogram of simulated Q distribution with observed Q marked:



Mega Millions Monte Carlo Q, 5000 repetitions of 465 draws
(observed Q marked)