On the Statistical Distribution of the Number of Word Groups whose Numerical Values are Divisible by Prime Numbers

Ola Hössjer

January 10, 2018

1 Problem Statement

Let $i = 1, 2, 3, \ldots, n$ be labels of n Hebrew words of a text, and let $Y_1, Y_2, \ldots$ be their numerical values. Each finite subset I of $\{1, \ldots, n\}$ defines a word group that consists of the words in I. The total numerical value

$$Y_I = \sum_{i \in I} Y_i$$

of word group I is obtained by adding the numerical values of all its words. Let p be a prime number. Our objective is to find the statistical distribution of the number of word groups among the n given words, whose numerical values are divisible by p. To this end, let

$$X_I = Y_I \bmod p$$

be the remainder of the division of Y_I by p. We will make the simplifying assumption that $X_1, X_2, \ldots$ are independent random variables with a uniform distribution on $\{0, 1, \ldots, p-1\}$. A uniform distribution means that all values of X_i are equally likely, so that the probability of obtaining a specific remainder j modulo p, is

$$H_0 : \ P(X_i = j) = \frac{1}{p}, \quad j = 0, 1, \ldots, p-1. \tag{1}$$

Equation (1) is our null hypothesis (H_0); that all words are formed simply by chance. Let

$$S_n = \sum_{I; I \subset \{1, \ldots, n\}} K_I \tag{2}$$

be the number of word groups formed by the n available words, whose total numerical values are divisible by p. Here $K_I = 1(X_I = 0)$ is a word group count that equals 1 if Y_I is divisible by p, and 0 otherwise. We need to find (an approximation of) the probability distribution

$$p_n(l) = P(S_n = l | H_0), \quad l = 0, \ldots, 2^n - 1$$

of S_n under the null hypothesis. Clearly $0 \leq S_n \leq 2^n - 1$, since S_n cannot exceed the total number $2^n - 1$ of possible word groups formed by n words. In particular, we want to find the

$$p\text{-value } = P(S_n \geq s) = \sum_{l \geq s} p_n(l), \tag{3}$$

where s is the observed value of S_n, when testing the null hypothesis against the alternative hypothesis

H_1 : Numerical values of word groups divisible by p more often than expected under H_0.

Consequently, the p-value is the probability to observe by chance at least s word groups with a numerical value divisible by p, among all word groups that can be formed by n words.

2 Different Ways to Find the Distribution of S_n

In this section we consider one exact and two approximate ways of finding the statistical distribution of S_n under the null hypothesis (1) that the numerical values of words are formed simply by chance.

2.1 Simplified Binomial Approach

In order to describe the first approximate approach, it will be convenient to introduce the probability $\pi = 1/p$. It can be shown that

$$\begin{aligned} P(X_I = 0|H_0) = E(K_I|H_0) &= \pi, \\ \mathrm{Var}(K_I|H_0) &= \pi(1-\pi), \\ \mathrm{Cov}(K_I, K_J|H_0) &= 0, \quad I \neq J. \end{aligned} \tag{4}$$

This means that the word group counts K_I are uncorrelated random variables with the same expected value π and variance $\pi(1-\pi)$. In particular, it follows from (4) that

$$\begin{aligned} E(S_n|H_0) &= \textstyle\sum_I E(K_I|H_0) = (2^n - 1)\pi, \\ \mathrm{Var}(S_n|H_0) &= \textstyle\sum_I \mathrm{Var}(K_I|H_0) = (2^n - 1)\pi(1-\pi). \end{aligned} \tag{5}$$

If, in addition, all the word group counts K_I were independent, then

$$S_n|H_0 \sim \mathrm{Bin}(2^n - 1, \pi) \tag{6}$$

would have a binomial distribution under the null hypothesis. That is, S_n would then have the same distribution as the number of heads when we flip a coin $2^n - 1$ times, and the probability of heads is π. According to (3), this would give us a

$$p - \text{value} = \sum_{l=s}^{\infty} \binom{2^n - 1}{l} (1-\pi)^l \pi^l. \tag{7}$$

Symbol	Meaning
n	Number of words
i	Word label ($\in \{1, \ldots, n\}$)
Y_i	Numerical value of word i
p	Prime number
X_i	Numerical value of word i modulo p
I	Label for a word group (a collection of words), i.e. a subset of $\{1, \ldots, n\}$
J	Another symbol for a word group label
Y_I	Total numerical value of all the words in word group I
X_I	Total numerical value of all the words in word group I modulo p
K_I	Word group count that equals 1 if the total numerical value of the words in word group I is divisible by p, and 0 otherwise
S_n	Number of word groups formed by n words, whose total numerical values are divisible by p
s	Observed value of S_n
M_n	Number of words, among n words, whose numerical values are divisible by p
q	Observed value of $S_n - M_n$, the number of words among n, whose numerical values are *not* divisible by p
T_n	Number of word groups whose total numerical values are divisible by p, among those that are formed by the q words (out of n) that have a numerical value not divisible by p
H_0	Null hypothesis that the numerical values of all words have a remainder modulo p that is uniformly distributed on $0, 1, \ldots, p-1$
H_1	Alternative hypothesis that word groups are divisible by p more often than predicted under H_0
H_0'	Modified null hypothesis that no words are divisible by p, with numerical values modulo p that are uniformly distributed on $1, \ldots, p-1$
p-value	Probability of obtaining by chance (H_0) at least as many word groups with total numerical values divisible by p, as the observed number s of such word groups
$\mathcal{I}$	Word group configuration (collection of word groups)
$K_{\mathcal{I}}$	Word group configuration count that equals 1 if the total numerical values of all word groups in $\mathcal{I}$ are divisible by p, and 0 otherwise

Table 1: List of some of the symbols used.

At a first glance (6) looks as a reasonable approximation of the distribution of S_n under the null hypothesis, since the expected value and variance of $\mathrm{Bin}(2^n-1,\pi)$ agree with (5). However, (6) is *false*, since overlapping word groups I and J ($I \cap J \neq 0$) have common words, and therefore their count variables K_I and K_J are *not* independent random variables. This dependency between the counts word groups with common words is present in spite of the fact that these variables are uncorrelated, according to (4). There is even more dependency between some larger sets of word groups. For instance, if a number of words have numerical values divisible by p, then all the word groups formed by these words will also have numerical values divisible by p. Therefore, such word groups tend to be divisible by p at the same time more often than if their count variables would have been independent. This will imply that the distribution of S_n under the null hypothesis has a much heavier tail to the right than that of the binomial distribution in (6).

2.2 Exact Recursive Approach

Introduce the vector $\boldsymbol{S}_n = (S_{n0}, \ldots, S_{n,p-1})$, where

$$S_{nj} = |\{I;\, I \subset \{1,\ldots,n\}, X_I = j\}|$$

is the number of word groups formed by the n given words, whose numerical value has a remainder of j modulo p. Notice that $S_n = S_{n0}$, and $0 \leq S_{nj} \leq 2^n - 1$, since S_{nj} cannot exceed the total number $2^n - 1$ of possible word groups formed by n words. For each vector $\boldsymbol{l} = (l_0, \ldots, l_{p-1})$ with components $0 \leq l_j \leq 2^n - 1$, we define the joint probability distribution

$$p_n(\boldsymbol{l}) = P(\boldsymbol{S}_n = \boldsymbol{l}|H_0) = P\left(S_{n0} = l_0, \ldots, S_{n,p-1} = l_{p-1}|H_0\right),$$

of $S_{n0}, \ldots, S_{n,p-1}$ under the null hypothesis. It is possible to compute this distribution recursively with respect to n. To this end, we will use (1) and the key relation

$$S_{nj} = S_{n-1,j} + S_{n-1,j-X_n} + 1_{\{X_n=j\}}, \tag{8}$$

where $j - X_n \in \{0, 1, \ldots, p-1\}$ is taken modulo p. The word groups with a remainder j for their numerical values, have been split into three parts on the right hand side of (8); those word groups that are formed by the first $n-1$ words, those word groups that contain the n:th word and at least one word among the $n-1$ first, and finally the word group formed by one single word, the n:th one.

It follows from (8) that

$$p_n(\boldsymbol{l}) = \frac{1}{p}\sum_{j=0}^{p-1}\sum_{\boldsymbol{h}} p_{n-1}(\boldsymbol{h}) p_{n-1}(\boldsymbol{l} - \boldsymbol{h} - \boldsymbol{e}_j), \tag{9}$$

where the sum is over all vectors $\boldsymbol{h} = (h_0, h_1, \ldots, h_{p-1})$ whose components satisfy $0 \leq h_j \leq 2^{n-1} - 1$, whereas $\boldsymbol{e}_j$ is a vector of length p with 1 in position j and zeros elsewhere.

In principle, it is possible to use (9) in order to compute the distribution of $\boldsymbol{S}_n$ under the null hypothesis, recursively with respect to n. This also gives us the null distribution

$$p_n(l) = \sum_{\substack{\boldsymbol{l}=(l_0,\ldots,l_{p-1}) \\ l_0=l}} p_n(\boldsymbol{l}) \tag{10}$$

of $S_n = S_{n0}$. The exact p-value (3) is then obtained by summing (10) for $l = s, \ldots, 2^n - 1$. However, this approach is infeasible for all but very small n and p, since we have to evaluate the probability function in (9) for 2^{pn} arguments $\boldsymbol{l}$.

2.3 Conditioning on Number of Divisible Words

Let M_n be the number of words among the n given ones, that have a numerical value Y_i divisible by p. Let also T_n be the number of word groups divisible by p which are formed by the remaining $n - M_n$ words. In particular, $T_n = 0$ when $M_n = n$. Then

$$S_n = 2^{M_n} - 1 + (2^{M_n} - 1)T_n + T_n = 2^{M_n}(T_n + 1) - 1, \tag{11}$$

where the first term $2^{M_n} - 1$ is the number of word groups formed by the words whose numerical values are divisible by p, the second term $(2^{M_n} - 1)T_n$ is the number of word groups with numerical values divisible by p that contain at least one word divisible by p and at least one word not divisible by p, and the third term T_n is defined as above. In order to find the null hypothesis distribution of S_n we will condition on the value of M_n. This enables us to express the p-value in (3) as

$$\begin{array}{rcl} p-\text{value} & = & P(S_n \geq s|H_0) \\ & = & \sum_{m=0}^{n} P(M_n = m|H_0)P(S_n \geq s|M_n = m, H_0) \\ & = & \sum_{m=0}^{n} P(M_n = m|H_0)P(2^m(T_n + 1) - 1 \geq s|M_n = m, H_0) \\ & = & \sum_{m=0}^{n} P(M_n = m|H_0)P(T_n \geq \lceil 2^{-m}(s+1) - 1)\rceil|M_n = m, H_0), \end{array} \tag{12}$$

where $\lceil x \rceil$ is the smallest integer greater or equal to x. We thus need to know the joint distribution of M_n and T_n under the null hypothesis in order to simplify (12). It follows from (1) that M_n has a binomial distribution

$$M_n|H_0 \sim \text{Bin}(n, \pi) \Longrightarrow P(M_n = m|H_0) = \binom{n}{m}\pi^m(1-\pi)^{n-m}$$

under the null hypothesis. Since none of the remaining $n - M_n$ words have a numerical value divisible by p, it follows that the distribution of T_n under the hull hypothesis H_0 when $M_n = m$, is the same as the distribution of S_{n-m} under the adjusted null hypothesis

$$H_0' : P(X_i = j) = \frac{1}{p-1}, \quad j = 1, 2, \ldots, p-1 \tag{13}$$

that a word's numerical value is not divisible by p, but otherwise its remainder modulo p is uniformly distributed. If $q = n - m$ refers to the number of words with a numerical value not divisible by p when $M_n = m$, we rewrite (12) as

$$\begin{array}{rcl} p-\text{value} & = & P(S_n \geq s|H_0) \\ & = & \sum_{q=0}^{n} \binom{n}{n-q}\pi^{n-q}(1-\pi)^q P\left(S_q \geq \lceil 2^{-m}(s+1) - 1)\rceil|H_0'\right). \end{array} \tag{14}$$

In order to evaluate (14), it remains to compute or find good approximations of the distribution of S_q under H_0' for $q = n - m = 0, \ldots, n$. Since S_q is the sum of a number of word group counts K_I formed by q words (cf. (2)), we will first look at the expected value and variance of K_I as well as the covariance between K_I and K_J under H_0'. Since no words have a numerical value divisible by p under H_0' we loose some symmetry, and therefore formulas get more complicated. The expected value and variance

$$\begin{array}{rcl} E(K_I|H_0') & = & \pi'_{|I|}, \\ \mathrm{Var}(K_I|H_0') & = & \pi'_{|I|}(1 - \pi'_{|I|}) \end{array} \tag{15}$$

of a word group count, will, in contrast to (4), depend on the number of words $|I|$ of the word group I. Since $\pi'_{|I|}$ is the probability that word group I is divisible by p under H_0', by conditioning on whether a sub-word group of length $|I| - 1$ is also divisible by p or not, it can be seen that $\pi'_{|I|}$ satisfies the recursive relation

$$\pi'_{|I|} = 0 \cdot \pi'_{|I|-1} + \frac{1}{p-1}(1 - \pi'_{|I|-1}) = \frac{1}{p-1}(1 - \pi'_{|I|-1})$$

for $|I| \geq 2$. This is a linear recursion, with a boundary condition $\pi'_1 = 0$ that follows from (13). The explicit solution of this recursion is

$$\pi'_{|I|} = \pi \left[1 - (\frac{-\pi}{1-\pi})^{|I|-1} \right]. \tag{16}$$

It can be seen that (16) quickly approaches the limit π when the number of words $|I|$ of the word group grows. Moreover, it follows from (16)

$$\begin{array}{rcl} E(S_q|H_0') & = & \sum_{I;I\subset\{1,\ldots,q\}} E(K_I|H_0') \\ & = & \sum_{m=1}^{q} \binom{q}{m} \pi'_m \\ & = & \pi \sum_{m=2}^{q} \binom{q}{m} \left[1 - (\frac{-\pi}{1-\pi})^{m-1} \right]. \end{array} \tag{17}$$

The expected value in is smaller than $E(S_q|H_0) = (2^q - 1)\pi$. This is not surprising, since in (17) we condition on that none of the q words are divisible by p, and this will to some extent decrease the expected number of word groups divisible by p.

We recall from (4) that the covariance between two different word group counts K_I and K_J is always 0 under H_0. Their covariance under H_0' is more complicated though when I and J overlap; it can be either negative, zero or positive. When none of I and J is a subset of the other, then

$$\begin{array}{rcl} \mathrm{Cov}(K_I, K_J|H_0') & = & E(K_I K_J|H_0') - E(K_I|H_0')E(K_J|H_0') \\ & = & \pi'_{|I\setminus J|}\pi'_{|J\setminus I|}\pi'_{|I\cap J|} + \frac{\pi^2}{(1-\pi)^2}(1 - \pi'_{|I\setminus J|})(1 - \pi'_{|J\setminus I|})(1 - \pi'_{|I\cap J|}) \\ & - & \pi'_{|I|}\pi'_{|J|}. \end{array} \tag{18}$$

In order to evaluate $E(K_I K_J|H_0')$ in (18), we had to split $I \cup J$ into three disjoint word groups $I \setminus J$, $I \cap J$, and $J \setminus I$. Then we used that $E(K_I K_J|H_0')$ is the sum of two terms, the first of which is the probability that these three word groups are all divisible by p,

whereas the second term is the probability that the remainder of word group $I \cap J$ modulo p is nonzero and minus the remainders of the other two word groups $I \setminus J$ and $J \setminus I$.

When $I \subset J$ things are simpler, since $E(K_I K_J | H_0')$ is the probability that word groups I and $J \setminus I$ are both divisible by p. Due do independence between the numerical values of disjoint word groups, this gives

$$\begin{aligned} \mathrm{Cov}(K_I, K_J | H_0') &= E(K_I K_J | H_0') - E(K_I | H_0') E(K_J | H_0') \\ &= E(K_I | H_0') E(K_{J \setminus I} | H_0') - E(K_I | H_0') E(K_J | H_0') \\ &= \pi'_{|I|} \pi'_{|J \setminus I|} - \pi'_{|I|} \pi'_{|J|}. \end{aligned} \tag{19}$$

Finally, when $I \cap J = \emptyset$ we similarly obtain

$$\begin{aligned} \mathrm{Cov}(K_I, K_J | H_0') &= E(K_I K_J | H_0') - E(K_I | H_0') E(K_J | H_0') \\ &= \pi'_{|I|} \pi'_{|J|} - \pi'_{|I|} \pi'_{|J|} \\ &= 0. \end{aligned} \tag{20}$$

Using (18)-(20), it is possible to compute

$$\mathrm{Var}(S_q | H_0') = \sum_{I,J} \mathrm{Cov}(K_I, K_J | H_0'),$$

by summing over all ordered pairs I, J of word groups that are drawn with replacement from $\{1, \ldots, q\}$.

3 Numerical Example: $n = 7$, $p = 37$, $s = 23$

The expected value and variance of S_7 under the null hypothesis are easily obtained from (4), as

$$\begin{aligned} E(S_7 | H_0) &= 3.4324, \\ \mathrm{Var}(S_7 | H_0) &= 3.3397 \end{aligned}$$

respectively. An observed value $s = 23$ of S_7 is almost seven times as high as $E(S_7 | H_0)$. It is therefore reasonable to assume that the p-value will be small. If we use the binomial distribution (6)-(7), we actually get a

$$\begin{aligned} p\text{-value} &= P(S_7 \geq 23 | H_0) \\ &\approx \textstyle\sum_{l=23}^{127} \binom{127}{l} \pi^l (1-\pi)^{127-l} \\ &= \textstyle\sum_{l=23}^{127} \binom{127}{l} (1/37)^l (1-1/37)^{127-l} \\ &= 6.3505 \cdot 10^{-13} \end{aligned} \tag{21}$$

that is far *too* small. In order to get a better approximation of the p-value, we use formula (14), and find that

$$\begin{aligned} p\text{-value} &= P(S_7 \geq 23|H_0) \\ &= (1/37)^7 \\ &+ 7(1-1/37)(1/37)^6 \\ &+ 21(1-1/37)^2(1/37)^5 \\ &+ 35(1-1/37)^3(1/37)^4 \cdot P(S_3 \geq 1|H_0') \\ &+ 35(1-1/37)^4(1/37)^3 \cdot P(S_4 \geq 2|H_0') \\ &+ 21(1-1/37)^5(1/37)^2 \cdot P(S_5 \geq 5|H_0') \\ &+ 7(1-1/37)^6(1/37) \cdot P(S_6 \geq 11|H_0') \\ &+ (1-1/37)^7 \cdot P(S_7 \geq 23|H_0') \\ &> 2.8935 \cdot 10^{-7}. \end{aligned} \tag{22}$$

The lower bound of the p-value in the last step was obtained from the fact that all probabilities $P(S_q \geq k|H_0')$ are non-negative for $q = 3, 4, 5, 6, 7$. It turns out that this lower bound is also far too small, although it is several orders of magnitude larger than (21). In order to obtain a better upper bound of the p-value, we need to find upper bounds of all the terms $P(S_q \geq k|H_0')$ that appear in (22). This is not easy, but we get an idea of the size of all S_q under H_0' by computing the expected values $E(S_q|H_0')$. Inserting $p = 37$ into (17), we find that

$$E(S_q|H_0') = \frac{1}{37}\sum_{m=2}^{q}\binom{q}{m}\left[1-(\frac{-1}{36})^{m-1}\right] = \begin{cases} 0.0278, & q=2, \\ 0.1103, & q=3, \\ 0.3017, & q=4, \\ 0.7100, & q=5, \\ 1.5514, & q=6, \\ 3.2583, & q=7. \end{cases} \tag{23}$$

This gives an upper bound

$$P(S_q \geq k|H_0') \leq \frac{E(S_q|H_0')}{k} \tag{24}$$

through Markov's Inequality, which for $q = 3$ and $k = 1$ equals

$$P(S_3 \geq 1|H_0') \leq 0.1103. \tag{25}$$

However, (24) is too crude to use for the other four terms of (22) with $q \geq 4$. Instead we will use the definition of S_q in (2) to find another upper bound

$$\begin{aligned} P(S_q \geq k|H_0') &= P\left(\cup_{I_1,\ldots,I_k}\{K_{I_1} = \ldots = K_{I_k} = 1\}|H_0'\right) \\ &\leq \textstyle\sum_{I_1,\ldots,I_k} P(K_{I_1} = \ldots = K_{I_k} = 1|H_0') \\ &= \textstyle\sum_{I_1,\ldots,I_k} E(K_{I_1} \cdot \ldots \cdot K_{I_k}|H_0') \\ &= \textstyle\sum_{\mathcal{I}} E(K_{\mathcal{I}}|H_0'), \end{aligned} \tag{26}$$

where the sum ranges over all word group configurations $\mathcal{I} = \{I_1, \ldots, I_k\}$ that consist of k distinct word groups formed by q words, and $K_{\mathcal{I}} = K_{I_1} \cdot \ldots \cdot K_{I_d}$ is a word group

configuration count, that equals 1 if the word groups within $\mathcal{I}$ are divisible by $p = 37$, and 0 otherwise.

We only need to include those word group configurations $\mathcal{I}$ in (26) for which $E(K_{\mathcal{I}}|H_0')$ is positive. We know from (13) that $K_{I_j} = 0$ under H_0' for single words ($|I_j| = 1$). It therefore suffices to include those configurations $\mathcal{I}$ in (26) whose k word groups I_j have at least two words. Likewise, we don't include any terms in (26) where two of the word groups of $\mathcal{I}$ satisfy $I_h \subset I_j$ and $|I_j \setminus I_h| = 1$ for some $h, j \in \{1, \ldots, k\}$. The reason is that at least one of K_{I_h} and K_{I_j} must be zero under H_0' for such pairs of word groups.

When $q = 3$ and $k = 1$, each configuration $\mathcal{I} = \{I\}$ consists of one single word group. Formula (26) then gives an upper bound

$$\begin{array}{rcl} P(S_3 \geq 1|H_0') & \leq & \sum_I P(K_I = 1|H_0') \\ & = & P(K_{\{1,2\}} = 1|H_0') + P(K_{\{1,3\}} = 1|H_0') \\ & + & P(K_{\{2,3\}} = 1|H_0') + P(K_{\{1,2,3\}} = 1|H_0') \\ & = & 3\pi_2' + \pi_3' \\ & = & 0.1103, \end{array} \tag{27}$$

by summing over all word groups I that contain at least two words from $1, 2, 3$. In the last step we used (16) to deduce that

$$\begin{array}{rcl} \pi_2' & = & \pi/(1-\pi) = 1/(p-1) = 1/36 = 0.02778, \\ \pi_3' & = & \pi(1-2\pi)/(1-\pi)^2 = 0.02701. \end{array}$$

Notice that (27) gives the same upper bound for $P(S_3 \geq 1|H_0')$ as in (25) .

When $q = 4$ and $k = 2$, formula (26) gives the upper bound

$$\begin{array}{rcl} P(S_4 \geq 2|H_0') & \leq & \sum_{I_1,I_2} E(K_{I_1}K_{I_2}|H_0') \\ & \leq & 12 \cdot (\pi/(1-\pi))^2 \\ & + & 6 \cdot (\pi/(1-\pi))^2(1-\pi_2') \\ & + & 12 \cdot (\pi/(1-\pi))^2(1-\pi_2') \\ & + & 6 \cdot \pi_2' \cdot \pi_2' \\ & + & 3 \cdot \pi_2' \cdot \pi_2' \\ & = & (12+6+3) \cdot (1/36)^2 + (6+12) \cdot (1/36)^2(1-1/36) \\ & = & 0.02971, \end{array} \tag{28}$$

where the sum ranges over all word group configurations $\mathcal{I} = \{I_1, I_2\}$ that consist of two word groups $I_1, I_2 \subset \{1, 2, 3, 4\}$. Notice that the upper bound of $P(S_4 \geq 2|H_0')$ in (28) is a lot smaller than the one ($=0.3017/2 = 0.1508$) obtained from (23)-(24).

Formula (26) was derived by summing over those word group configurations I_1, I_2 in (28) whose word groups contain at least two words each, and if $I_1 \subset I_2$, we must have $|I_2 \setminus I_1| \geq 2$, i.e. $|I_1| = 2$ and $|I_2| = 4$. There are 39 such word group configurations, and in the the second step of (28) we divided them into five sets, as shown in Table 2. For the first three sets (with 12, 6, and 12 word group configurations), neither I_1 nor I_2 is a subset of the other. Therefore we applied formula (18) in order to compute $E(K_{I_1}K_{I_2}|H_0')$.

q	k	r	$\mathcal{I}$	$r(\mathcal{I})$	$N(\mathcal{I})$	N_r
3	1	1	$\{1,2\}$	1	3	4
			$\{1,2,3\}$	1	1	
4	2	2	$\{\{1,2\},\{1,3\}$	2	12	39
			$\{\{1,2,3\},\{2,3,4\}\}$	2	6	
			$\{\{1,2,3\},\{1,4\}\}$	2	12	
			$\{\{1,2\},\{1,2,3,4\}\}$	2	6	
			$\{\{1,2\},\{3,4\}\}\}$	2	3	
5	5	3	$\{\{1,2\},\{2,3\},\{3,4\},\{1,4\},\{1,2,3,4\}\}$	3	15	75
			$\{\{3,5\},\{4,5\},\{1,2,3\},\{1,2,4\},\{1,2,3,4,5\}\}$	3	60	

Table 2: List of different compatible word group configurations $\mathcal{I}$. For each $\mathcal{I}$, we also give the number of linearly independent and compatible restrictions $r(\mathcal{I}) = r$ that it imposes on the numerical values of the words, and the total number $N(\mathcal{I})$ of equivalent configurations (i.e. those that are obtained from $\mathcal{I}$ by a permutation of word labels). When $q = k = 5$, there are 2070 additional compatible word group configurations with $r(\mathcal{I}) = 4$. In the rightmost column, N_r is obtained by adding $N(\mathcal{I})$ for all rows with $r(\mathcal{I}) = r$.

Notice that the first term of (18) vanishes of for all these word group configurations, since at least one of $\pi'_{|I_1 \setminus I_2|}$, $\pi'_{|I_2 \setminus I_1|}$, and $\pi'_{|I_1 \cap I_2|}$ equals zero. For the fourth set of 6 word group configurations in (28), we have that $I_1 \subset I_2$, $|I_1| = 2$, and $|I_2| = 4$. For them we used (19) in order to find $E(K_{I_1} K_{I_2} | H'_0)$. Finally, for the last set of 3 configurations I_1, I_2 in (28), we have that $I_1 \cap I_2 = \emptyset$ and $|I_1| = |I_2| = 2$. For them we used (20) in order to compute $E(K_{I_1} K_{I_2} | H'_0)$.

For the remaining three terms of (22), it is too complicated to list all possible word group configurations $\mathcal{I} = \{I_1, \ldots, I_k\}$ in (26). As mentioned above, we only need to consider those $\binom{2^q - 1 - q}{k}$ word group configurations whose words have length at least 2. We refer to such a word group configuration as *incompatible*, if the requirement $K_{\mathcal{I}} = 1$, that all its word groups are divisible by $p = 37$, requires that at least one of its words is divisible by 37 as well ($X_i = 0$ for at least one $i \in I_1 \cup \ldots \cup I_k$). We know from (13) that this is not possible under H'_0, and therefore each such word group configuration has $E(K_{\mathcal{I}} | H'_0) = 0$. Examples of incompatible word group configurations are those that contain two word groups $I_h \subset I_j$ with $|I_j \setminus I_h| = 1$. All word group configurations that contain a triangle of three word groups $I_h = \{i_1, i_2\}$, $I_j = \{i_2, i_3\}$, and $I_l = \{i_1, i_3\}$ of length 2, are incompatible as well.

The remaining compatible word group configurations $\mathcal{I}$ with $E(K_{\mathcal{I}} | H'_0) > 0$ will be divided into a number of classes, depending on how many linearly independent restrictions $r(\mathcal{I})$ on the numerical values $\{X_i;\, i \in I_1 \cup \ldots \cup I_k\}$ of the words in the configuration that they impose. Each such restriction is a sum of a subset of all X_i, $i = 1, \ldots, q$.

Let N_m the number of word group configurations $\mathcal{I}$ with $E(K_{\mathcal{I}} | H'_0) > 0$ and $r(\mathcal{I}) = m$. An incompatible word group configuration $\mathcal{I}$ can be viewed as having an infinite number

of linearly independent restrictions ($r(\mathcal{I}) = \infty$). We therefore refer to N_∞ as the number of incompatible word group configurations. From this it follows that

$$\binom{2^q - 1 - q}{k} = N_\infty + \sum_{m=1}^{\min(q-1,k)} N_m. \tag{29}$$

For the upper bound of m in (29), we used that q linearly independent restrictions on $\{X_i;\, i \in I_1 \cup \ldots \cup I_k\}$, for a certain word group configuration, requires $X_1 = \ldots = X_q = 0$, since there are at most q different X_i in the word group configuration. But this is not possible in view of (13), and therefore $N_q = 0$. Below we will motivate that

$$E(K_{\mathcal{I}}|H_0') \leq \left(\frac{1}{36}\right)^{r(\mathcal{I})}. \tag{30}$$

By inserting (30) into (26), we find that

$$P(S_q \geq k|H_0') \leq \sum_{m=1}^{\min(q-1,k)} N_m \left(\frac{1}{36}\right)^m. \tag{31}$$

Let us first apply (31) to the two cases (27) and (28) for which we already have an upper bound of $P(S_q \geq k|H_0')$. Starting with $q = 3, k = 1$, it follows from Table 2 that $N_1 = 4$, and hence (31) implies

$$P(S_3 \geq 1|H_0') \leq 4 \cdot (1/36) = 1/9 = 0.1111,$$

which is only slightly larger than upper bound in (27). When $q = 4, k = 2$, we have that all 39 compatible configurations $\mathcal{I} = \{I_1, I_2\}$ have $r(I_1, I_2) = 2$ linearly independent restrictions (cf. Table 2). Therefore $N_1 = 0$ and $N_2 = 39$, and formula (31) gives an upper bound

$$P(S_3 \geq 1|H_0') \leq 39 \cdot (1/36)^2 = 0.0301$$

that is only slightly larger than the one in (28).

In order to motivate that all compatible word group configurations $\mathcal{I}$ have $r(\mathcal{I}) = 2$ linear restrictions when $q = 4, k = 2$, consider for instance one that consists of $I_1 = \{1, 2, 3\}$ and $I_2 = \{3, 4\}$. The requirement $K_{I_1} K_{I_2} = 1$ that both its word groups are divisible by $p = 37$ then imposes two restrictions

$$\begin{aligned} X_1 + X_2 + X_3 &= 0, \\ X_3 + X_4 &= 0 \end{aligned} \tag{32}$$

on $\{X_1, X_2, X_3, X_4\}$, with the additions in (32) being modulo p. We can rewrite these two restriction in matrix form as

$$\boldsymbol{AX} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \mathbf{0}.$$

We notice that the two rows of $\boldsymbol{A}$ are linearly independent, and therefore we have two independent restrictions on X_1, X_2, X_3, X_4. Since (32) has at least one solution for which all $X_m \neq 0$, it follows that $\mathcal{I}$ is compatible, and therefore $r(\mathcal{I}) = \text{rank}(\boldsymbol{A}) = 2$.

This can be developed into a general procedure of finding $r(\mathcal{I})$ for any word group configuration $\mathcal{I}$. The idea is that each $\mathcal{I}$ generates a $k \times q$ matrix $\boldsymbol{A} = \boldsymbol{A}(\mathcal{I})$. If the numerical values of all words are put into the column vector $\boldsymbol{X} = (X_1, \ldots, X_q)^T$, we can formulate the requirement that all the word groups of $\mathcal{I}$ have numerical values divisible by p, as a system of linear equations $\boldsymbol{AX} = \boldsymbol{0}$. This system of equations has a null space

$$\mathcal{N}(\mathcal{I}) = \{\boldsymbol{X};\ \boldsymbol{X} = \sum_{j=1}^{d} c_j \boldsymbol{v}_j\}$$

that is spanned by d linearly independent column vectors $\boldsymbol{v}_j = (v_{1j}, \ldots, v_{qj})^T$, where d is the dimension of the null space of $\boldsymbol{A}$. Put $r(\mathcal{I}) = \infty$ if either $d = 0$ or if $d > 0$ and there is at least one index m for which $X_m = 0$ is enforced, i.e. $v_{m1} = v_{m2} = \ldots = v_{md} = 0$. Otherwise let $r(\mathcal{I}) = q - d = \text{rank}(\boldsymbol{A})$.

The matrix formulation can be used in order to motivate (30). Suppose there are $r = r(\mathcal{I})$ linearly independent restrictions of $\mathcal{I}$, so that $\boldsymbol{A}$ has rank r. Assume without loss of generality that the first r rows $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r$ of $\boldsymbol{A}$ are linearly independent. Then write the left hand side of (30) as an iterated conditional probability

$$\begin{array}{rcl} E(K_{\mathcal{I}}|H_0') & = & P(\boldsymbol{a}_1\boldsymbol{X} = 0|H_0') \cdot P(\boldsymbol{a}_2\boldsymbol{X} = 0|\boldsymbol{a}_1\boldsymbol{X} = 0, H_0') \cdot \ldots \\ & \cdot & P(\boldsymbol{a}_r\boldsymbol{X} = 0|\boldsymbol{a}_1\boldsymbol{X} = \ldots = \boldsymbol{a}_{r-1}\boldsymbol{X} = 0, H_0') \\ & \leq & (1/36) \cdot (1/36) \cdot \ldots \cdot (1/36) \\ & = & (1/36)^{r(\mathcal{I})}. \end{array} \tag{33}$$

In the second step of (33) we imposed an upper bound of $1/36$ for each conditional probability. This can be shown similarly as $\pi_m' \leq \pi/(1-\pi) = 1/36$ was proved in Section 2.3 (cf. (16)), for all $m = 1, 2, \ldots$.

The automated procedure is convenient to use in order to find $r(\mathcal{I})$ for all compatible word group configurations $\mathcal{I}$ when $q = k = 5$, since there are far too many of them to be listed manually. We have written a computer program, where $r(\mathcal{I})$ is evaluated for all $\binom{2^5-1-5}{5} = 65780$ configurations with word groups of size at least 2. It was found that 2145 of them were compatible configurations ($r(\mathcal{I}) < \infty$), with their number or linearly independent restrictions distributed as $N_1 = N_2 = 0$, $N_3 = 75$, and $N_4 = 2070$. Among these, the 75 configurations with $r(\mathcal{I}) = 3$ can be divided into two sets of size 15 and 60 (see Table 2). From (31) it follows that

$$P(S_5 \geq 5|H_0') \leq 75 \cdot \left(\frac{1}{36}\right)^3 + 2070 \cdot \left(\frac{1}{36}\right)^4 = 0.00284. \tag{34}$$

When $q = 6$ and $k = 11$, it can be shown that no word group configuration $\{I_1, \ldots, I_{11}\}$ has 5 or fewer linearly independent restrictions on $\{X_1, X_2, \ldots, X_6\}$. We therefore put $N_1 = N_2 = \ldots = N_5 = 0$ in (31), and consequently

$$P(S_6 \geq 11|H_0') = 0. \tag{35}$$

Similarly, when $q = 7$ and $k = 23$, there is no word group configuration $\{I_1, \ldots, I_{23}\}$ with 6 or fewer linearly independent restrictions on $\{X_1, X_2, \ldots, X_7\}$, so that $N_1 = N_2 = \ldots = N_6 = 0$. From this it follows that

$$P(S_7 \geq 23 | H_0') = 0. \tag{36}$$

Putting things together, by inserting (25), (28), and (34)-(36) into (22) we arrive at the upper bound

$$\begin{aligned} p\text{-value} &\leq (1/37)^7 \\ &+ 7(1-1/37)(1/37)^6 \\ &+ 21(1-1/37)^2(1/37)^5 \\ &+ 35(1-1/37)^3(1/37)^4 \cdot 0.1103 \\ &+ 35(1-1/37)^4(1/37)^3 \cdot 0.02971 \\ &+ 21(1-1/37)^5(1/37)^2 \cdot 0.00284 \\ &= 5.857 \cdot 10^{-5}. \end{aligned} \tag{37}$$