February 2004

Randomness tests for large samples of Keno Numbers

Stephen J. Sugden
Bond University, ssugden@bond.edu.au

Chris Noble

Randomness tests for large samples of Keno numbers

Chris Noble Steve Sugden*

(Received 8 August 2003; revised 4 February 2004)

Abstract

We describe a study in which a comprehensive set of statistical tests verify the randomness of a large sample of pseudo-random numbers. These were derived from random electronic noise produced by a hardware random number generator used by Jupiter's Network Gaming in their state-wide Keno game. The procedure developed is suited to testing large samples of supposedly random numbers under various conditions. The procedure incorporates, for nine different aspects of randomness testing, probability and frequency calculations in Borland's Delphi, as well as test statistic structure and significance calculations in Excel or SPSS. A brief description and summary table of results illustrates the customisation of each test conducted to the particular set of supposedly random numbers under consideration. A

*School of Information Technology, Bond University, Queensland, Australia. mailto:ssugden@staff.bond.edu.au

See http://anziamj.austms.org.au/V45/CTAC2003/Nobl/home.html for this article, Published February 18, 2004. ISSN 1446-8735

user-friendly interface implemented the application of all nine tests in Borland's Delphi. We developed algorithms suitable for dealing with the special problems concerning potential overflow due to large samples and large numbers of outcome categories as well as for the calculation of large Stirling numbers.

Contents

1 Introduction: the need for a randomness testing package

Gambling on computerised lotteries and betting games using computerised draws of random numbers flourished in the last decade and will likely increase in the near future. It has been claimed that Australia has the largest percentage of personal income devoted to gambling of any country in the world. Even if this statement is false, there is little doubt that Australians are globally among the highest per capita spenders on gambling. According to Cooke [4], Australians spend more than $11 billion annually on poker machines alone! Note also the following statement published by the Australian Bureau of Statistics [3]:

> The total net takings from gambling during 2000–01 were $13,839 million, which represented an increase of 26% since 1997–98.

The general public gambles daily on a vast array of games such as Lotto, Qwik-pik, Scratch-it, all of which require this random selection of numbers from an eligible set, in order to determine the winner. Results and prizes are disseminated through the mass media and small retail outlets. Independent private newsagencies and grocery stores are the main retail outlets for the lottery industry. Hotels (pubs), casinos and Government totalisator betting agencies are the most popular off-line outlets for the section of this industry involving the gambling game marketed as Jackpot Keno [6, 7, 8, 9]. According to [3], total takings in 2000–01, from Keno in Australia were $181.3 million. Allowing for an average household of 25%, this gives a total revenue estimate of $725 million annually for Keno in Australia, and this figure is growing.

Clearly, state governments and corporations such as Jupiter's Network Gaming (JNG), rely heavily on revenue generated from this rapidly expanding industry. The potential income from on-line betting and any associated taxes is considerable. Accordingly, verification of randomness of any winning numbers drawn is crucial to the licensing and legal implementation of these games. A bank of statistical randomness tests that is fairly standard for the gaming industry is typically used to perform this verification. These were required by the state regulatory authority, Queensland Office of Gaming Review (QOGR).

Algorithms developed by Knuth [1] to implement these tests are well known. However, application of these algorithms to verifying randomness in gambling games requires analysis specific coding to deal with unique problems resulting from three factors. First, for each different game type, there is a different but usually large number of category options in the multinomial distribution of potential outcomes. Second, very large samples must be used to satisfy government requirements pertaining to these randomness checks. Both these factors dictate consideration of problems concerning potential overflow. These must be dealt with when encoding the algorithms in whatever computer language is chosen by the analyst. Finally, one of the specified randomness tests requires use of extended tables of Stirling Numbers [5]. The existing literature provides only a small subset of the Stirling Numbers needed. For Keno, and most games of its type, the number of potential outcomes requires far more Stirling Numbers (specifically those of a higher degree) than are available in this subset. Accordingly, a special algorithm for calculating larger (unpublished) Stirling Numbers was developed.

In summary, this paper describes the results of the real life application, using Borland's Delphi, of this package of randomness tests to two large random samples of more than 50,000 JNG Keno numbers. Customization of this package to any investigator's special problems, simply requires changing the numbers of categories to match the dimension of the set of eligible numbers to be tested.

It is thought that the coding for all the algorithms presented is easily converted to other currently popular programming languages, and as such form the basis of a useful testing package with wide applicability to the ever increasing number and variety of electronically generated gambling games. The special features for this application should be of interest to this rapidly growing industry.

2 Summary of analysis and random numbers tested

In KENO as played in Australia at JNG outlets, 20 two-digit (decimal) integers are drawn by the house, from an eligible set of the integers 1 to 80. These numbers are displayed on a large TV type screen in the outlet. A player has previously bought a ticket also displaying up to 11 such integers. To the extent that the player's ticket has numbers which match between 6 and 11 of the house drawn numbers, prizes of varying sizes are won. This process is repeated once every 3 minutes.

The source of these KENO numbers is a hardware pseudo-random number generator, in which random noise is converted to bit strings, representing numbers 0 through 127, and subsequently mapped onto the numbers 1 through 80. So as to verify randomness through all stages of the process, it was required that random samples of both sets of numbers should be tested. The pseudo-random number generator is discussed in [6]. The algorithm for the above mapping is proprietary information held in confidence by JNG. However, it is not necessary to know the algorithm in order to test its output for randomness.

First, 72,000 of these pseudo-randomly generated bit strings representing the numbers 0 through 127, used as the source for subsequent mapping onto the set of 80 integers were tested for randomness.

Second, 57,000 end result mapped decimal integers (1 through 80), as calculated and used by JNG, were tested for randomness from 9 different aspects.

For these very large data sets, special algorithms were developed and implemented in Borland's Delphi (Figure 1), in order to calculate theoretical probabilities and theoretical and observed frequencies in each case. These probability data were then imported into either Microsoft Excel or SPSS data files in each case. Finally, the value of the relevant test-statistic and its significance level were calculated in each case, in EXCEL or SPSS as appropriate. The algorithms used were based either on those recommended by Knuth [1], or were developed from scratch. The results are summarized in Table 1. Of note are the results of the serial correlation tests, which are inconsistent with all other results.

Note that for this particular study the QOGR independently performed a corresponding analysis written in C, that verified the results of the described analyses above, on the same sample.

3 Probability distribution of Keno random numbers and sample structure

In practice, a sequence of eight-bit random bit strings, representing numbers between 00000000 and 01111111 (inclusive) is first generated before use in Keno draws of 20 numbers per game. The decimal equivalents are from 0 to 127, so that, assuming replacement in the selection of these numbers, each number is the outcome of a random experiment with 128 equally likely outcomes. The probability distribution for this set of 128 outcomes is "multinomial" with $n = 128$ and $p = 1/128$ for each outcome [2].

$$f(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots p_k) = \frac{n!}{x_1! x_2! \ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$$

QOGR required that a large sample (very long string) of these source binary numbers be tested for randomness. Such a sample was provided by JNG and 9 randomness tests specified by QOGR conducted.

Each of these numbers is then mapped onto a corresponding integer between 1 and 80. Again, assuming replacement, the distribution of these numbers would be multinomial with 80 equally likely outcomes ($n = 80$, $p = 1/80$ for all outcomes). However, in practice, the numbers are actually selected for a KENO draw of 20 *different* numbers, (no replacement), so binary numbers that map onto any decimal (1 to 80) which had been selected previously in the current draw, are discarded from the draw. The next 7-bit number generated in sequence is mapped onto its 1-to-80 decimal equivalent using the same algorithm, but retained only if is different from all other numbers in subset chosen so far for the current draw. This process is repeated until 20 different decimals from 1 to 80 have been chosen, at which point the draw is complete. When a new game starts (15 seconds later), a new draw is begun, and at the start all 80 decimal integers are eligible, and so on. Within each game, this is sampling without replacement, and the distribution of these numbers is hypergeometric [2]. Across many games in a very large sample, it is approximately multinomial.

QOGR also required that this sample of decimal integers to be tested for randomness actually should be real data used in live games and thus reflect this grouping. The sample was a number set string that was lifted directly from the on-site pseudo-random number generator output selected for use in real games. Accordingly, two data files were provided by JNG: the first consisted of approximately 72,000 binary integers (0 through 127), and the second consisting of 57,600 integers (1 through 80, in sets of 20), each different within a set, but free to recur across sets.

4 Overall testing approach

4.1 Testing philosophy

For the purposes of testing for randomness, for both samples, the analysis philosophy adopted by the authors and approved by both QOGR and JNG management, was to test the significance of the net effect of any disturbances from the overall patterns described above, assuming multinomial distributions. Various aspects of randomness were tested using nine common tests that are well described in Knuth [1], and summarized below. A different selection of these tests was used for each of the two samples.

4.2 Selection of tests on string of binary numbers

The government authority (QOGR) and JNG both required that all the generated numbers (0 to 127) provided, constitute a truly random sample. Its overall distribution, as well as that of various separate sequences, was tested. For 9 different hypotheses, observed frequencies were compared with those expected under the randomness assumption, using a single overall statistic for the whole sample. The entire sample was thus assumed to be a random sequence of 50,000 binary numbers between 00000000 (0) and 01111111 (127), with all 128 numbers equally likely to be selected each time a number was generated. This method proved satisfactory for all 9 randomness tests performed. Relevant aspects of the applications of these tests are detailed along with their results in Section 5.

4.3 Tests on mapped decimal numbers

The Government and JNG both required only that the overall distribution of decimal integers, not individual separate sequences, should be tested. This is due to the nature of the sampling without replacement within each group of 20 selected and drawn in any game. In all tests of randomness within small sequences, true random sequencing would be distorted by the "without replacement" nature of the numbers chosen in each group of 20. Only those tests inspecting the overall randomness, but not those looking at specific sequences of 20 decimal integers from 1 to 80, were thus applicable. QOGR required that only those applicable tests be performed.

The entire sample was thus first tested to see if, when treated as an end product unit of 50,000 decimal integers between 1 and 80, it constituted a truly random sample across all draws, with not significantly unequal representations of all 80 eligible integers. For the other required applicable tests, hypotheses using various aggregations of these numbers were also tested. In each test, observed frequencies of the decimal integers were compared with those expected under a randomness hypothesis, using a single overall statistic for the whole sample.

5 Statistical tests and results

5.1 Details of tests on string of binary numbers from 00000000 (0) to 01111111 (127)

Except for serial correlation, for each of the following tests, the summary statistic measuring overall disturbance from the predicted pattern under a randomness assumption is chi-square [2]. This is the distribution for random variables that are the squares or sums of squares of normally distributed random variables. It is a special case of the Gamma distribution. Details of the these tests are in Knuth [1, Ch 3.3.2.]. Their application to this sample and results are in Figure 1 and Table 1.

Where noted, modifications have been made to the basic structure of the

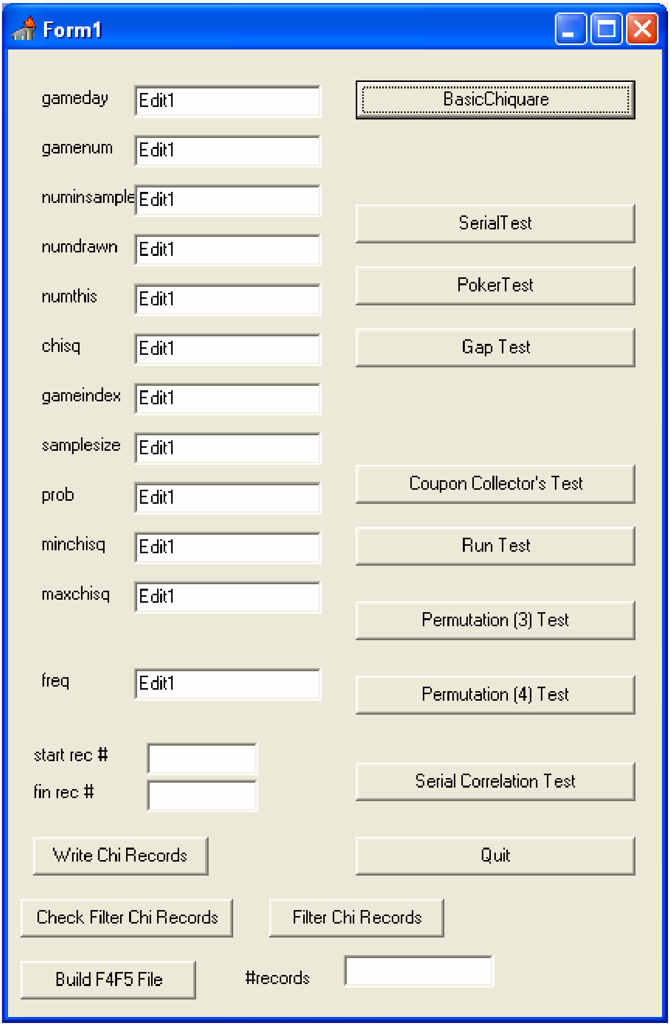

Figure 1: Delphi Analysis program user interface

test statistic to accommodate the special features of this sample, with regard to both the large number of different numbers being tested, and the large sample size. In each case the null hypothesis is that population proportions across all categories are equal.

Frequency test Is the overall sample random? The null hypothesis is that the proportions of each of 128 random numbers are equal. The distribution of numbers should be uniform.

Serial test Is the occurrence of pairs of numbers in sequence random? The null hypothesis is that the proportions of each of all pairs of random numbers are equal. Note: to ensure a minimum expected frequency of 5, 5461 categories combining 3 pairs (original categories) were created.

Gap test "Gaps" are strings of numbers greater than the median, which separate strings of numbers smaller than the median. Occurrences of such strings should be random with probability distributions according to Knuth [1]. The null hypothesis is that the proportions of each of 13 gap length categories are as specified in Tables 1 and 2. Note: to ensure a minimum expected frequency of 5, gaps of lengths 12 through 16 were combined into a single category. The largest gap was 16.

Poker test Occurrences in successive groups of 5 numbers of 1, 2 or 3 different, 4 different and 5 different integers should be random, according to Knuth [1]. The null hypothesis is that the proportions for each of all possible different, 5 number "poker" hand categories are as in Table 2. Note: poker hands of 1, 2 and 3 different were combined into one category.

Coupon collector's test Coupon collector segments are number strings which include all 128 integers. Occurrences of such strings should be random with probabilities based on Knuth [1]. The null hypothesis is that the proportions of 16 different coupon collector segment (string) lengths are as specified in Table 2. Note: First, the numbers 0 through 127 were mapped onto the set of integers 0 to 4. Next, 16 categories were created by considering the 15 segments of lengths 5 through 20, along with a single category for lengths 20 through 61 (largest). Note: tabulated values of Stirling Numbers from Abramowitz and Stegun [5], and also from an algorithm especially developed for this problem, were used to estimate the relevant probabilities. The special algorithm was for the larger Stirling numbers.

Permutation tests (2) Occurrences of permutations of groups of 3 consecutive integers (triplets) and groups of 4 consecutive integers (quadruplets) should be random. The null hypothesis is that the proportions for each of 6 different permutations of triplets are equal and each of 24 different permutations of quadruplets are equal. Note: 571 triplets containing duplicate values and 871 quadruplets were excluded from each test respectively. Final records were ignored when no triplet or quadruplet could occur.

Run test "Runs up" are monotonic increasing sub-sequences. Occurrences of "runs up" should be random with probabilities based on Knuth [1] and as specified in Table 2. The null hypothesis is that the proportions for each of 6 different run length categories ("runs up") are as specified in Table 2. Note: Run lengths of 6 and 7 (largest found in sample) were combined into a single category.

Serial correlation test Serial correlation should be very small. The null hypothesis is that the population serial correlation is "very close to" zero [1, pp70–71]. According to Knuth, a "good" value (close to zero) of the serial

Table 1: Summary of test results

Tests	$0 \rightarrow 127$		$1 \rightarrow 80$	
χ^2 goodness of fit	**Categories**	α	**Categories**	α
Frequency	128	0.073	80	0.9996
Serial	5461	0.151	6320[a]	0.83
Gap	13	0.155	13	0.274
Poker	3	0.3615	-	-
Coupon Collector's	16	0.961	-	-
Permutation (triplets)	6	0.239	6	0.864
Permutation (quadruplets)	24	0.437	24	0.884
Run	6	0.42	6	0.648
Serial Correlation	N/A	< 1%[b]	N/A	< 5%[c]

[a] Categories for Pairs of Duplicates have been removed, since they are not possible within games.

[b] The sample serial correlation coefficient was outside Knuth's preferred 99% confidence interval.

[c] The sample serial correlation coefficient was outside Knuth's preferred 95% confidence interval.

correlation coefficient is within 2 standard errors (95%) of the population mean coefficient.

5.2 Details of tests on sets of mapped decimal numbers from 1 to 80

Owing to the sampling without replacement present within each Keno draw (game) of 20 numbers, only 6 of the χ^2 tests were applicable to these numbers. Suitable modifications were made for the 80 eligible numbers, where any test was applicable. For the applicable tests, the sampling without replacement within games would imply that any non-significant results are conservative with regard to Type II errors, see Table 1.

Table 2: Theoretical probabilities for χ^2 tests with non-uniform distributions across categories.

Gap

gap length	expected prob
0	0.5000
1	0.2500
2	0.1250
3	0.0625
4	0.0313
5	0.0156
6	0.0078
7	0.0039
8	0.0020
9	0.0010
10	0.0005
11	0.0002
>11	0.0002

Poker

no. different	expected prob
1 to 3	0.0015
4	0.0745
5	0.9240

Run

run length	expected prob
1	0.5000
2	0.3333
3	0.1250
4	0.0333
5	0.0069
>5	0.0014

Coupon

segment length	expected prob
5	0.0384
6	0.0768
7	0.0998
8	0.1075
9	0.1045
10	0.0955
11	0.0838
12	0.0716
13	0.0601
14	0.0498
15	0.0409
16	0.0333
17	0.0270
18	0.0218
19	0.0176
20-61	0.0714

5.3 Summary of results

No significant non-randomness was detected at 10% by any of the goodness of fit tests. Accordingly, these tests showed no evidence of non-randomness in the population. In each case, the large sample size reflects high power, so the population data are probably random. Evidence of serial correlation was found, but QOGR did not consider this serious enough to delay its approval. Tables 1 and 2 show the results.

References

[1] Knuth, D. E. *Seminumerical Algorithms*, Vol 2 The Art of Computer Programming. Addison-Wesley, 2nd printing, 1998. C53, C55, C57, C58, C60, C61

[2] Freund, J. E., *Mathematical Statistics*, 6th edition, 1999, Prentice-Hall. C55, C56, C58

[3] ——, 8684.0 Gambling Industries, Australia. ABS document, available on-line at http://www.abs.gov.au/Ausstats/abs. C52

[4] Cooke, P. What pushes our buttons? *Sydney Daily Telegraph*, July 28, 2001, p. 16. C52

[5] Abramowitz , M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions*, Dover, 10th printing, with correction, ISBN 0-486-61272-4. C53, C61

[6] McMahon, G. B., Noble, C. and Sugden, S. J., *Keno Jackpots and Randomness.* A consulting report for Jupiters Network Gaming, June 1997. C52, C54

[7] Noble, C. and Sugden, S. J., *Statistical Tests on Random Numbers I (1 through 80)*. A consulting report for Jupiters Network Gaming, July 1997. C52

[8] Noble, C. and Sugden, S. J., *Statistical Tests on Random Numbers II (0 through 127)*. A consulting report for Jupiters Network Gaming, July 1997. C52

[9] Noble, C., Sugden, S. J., Stochastic Recurrences of Jackpot KENO. *Computational Statistics & Data Analysis*, 2002. 40: p. 189–205. C52