PROBABILITY DENSITY FUNCTION FOR KLONDIKE SOLITAIRE

B. Hughes, 2017 August 27

1. Introduction

It has been determined that the theoretical probability of winning an open game of Klondike solitaire, i.e. all 52 cards in the foundation stacks at the end of the game, and all cards known to the player at all times in the game, is approximately[1] 0.8. Anyone who has played a traditional non-open game – cards face down in the tableau except for the end card in each column, and the remaining cards hidden from view except on turnover – knows that the probability of winning is more like a tenth of this[2]. The number of different card arrangements is 52! (approx. 10^{68}) and for the traditional Klondike games there are many more possibilities than this because of choices that arise during each game. The problem of determining the probability of winning in a traditional Klondike game is very difficult and has not yet been done. Nevertheless, by playing a large number of games estimates of the statistics *can* be determined.

The present report provides an estimate of the complete probability density function (pdf) for the number of cards in the foundation stacks at the end of a game, i.e. the probability of ending with zero cards, or one card, or two cards, etc. or a "win" of 52 cards. It is recognized that some players are more adept at the game than others but no special skill is assumed by the author – the player of the games. The results reported here simply represent one attempt at estimating the full pdf, and were determined purely out of curiosity and interest. It is likely that there is not a single pdf curve but a set of them ordered by the skill of the player[3]. For very skilled players the proportion of "wins" will be higher than for less skilled players and that will have a reducing effect on the probabilities of obtaining low numbers of cards in the foundation stacks at the end of the game. How wide a range the pdf curves will exhibit in this regard is unknown but it leads to the possible concept of a limiting pdf curve for highly skilled players. If there is one, perhaps the results obtained here will be indicative of the values and trends in that curve.

The games in this report comprise several sets: 220 games were played by cards a number of years ago; 197 games were played – also a number of years ago – using an early computer game (not now available); 823 games were played recently using a current game written for an Apple computer by Growly Software; and 40 more games were played recently by cards. That brings the total to 1280. Before combining these, tests were done to determine if there were any significant statistical differences in the results of the different modes of play.

As well as the basic pdf, the data set has been used to determine as much as possible the probability of runs of different lengths from one game with a particular number of cards resulting, to the next game with the same result, and this is compared to theory assuming the games are statistically independent and using the measured pdf's as parameters.

Finally, an estimate is provided for the precision of the estimated pdf values and from that an estimate of the number of games needed to obtain a given level of precision.

Many computer solitaire versions of Klondike have options to alter the degree of difficulty, and that indicates a weakening of the degree of randomization of the "deck" before layout, or a manipulation of the hidden "cards" during play, or something similar. For the majority of games reported here and played on computer, i.e. those played using Growly Solitaire, the software was specifically written with maximized randomness in mind[4] and with no options for reducing the degree of difficulty. For the early computer games these circumstances were not apparent but cannot be commented on. For the card games reported here, the deck was riffle-shuffled up to 6 or 7 times before the layout in an attempt to attain a high degree of randomness.

2. Method of Play

Once the tableau was laid out, 3 cards at a time from the remaining deck were turned over and play continued until no plays could be made. The number of cards in the foundation stacks were counted at that time and were recorded as the result of the game. The only special allowance made was in the computer games: if a new turnover was initiated before a missed move was seen but the missed move was seen while the turnover was being executed, that turnover could be "undone". The rationale was that if a deck of cards was being used, and the top three cards in the remaining stack were taken off but not yet turned over before a missed move was seen, that play would be stopped and the cards put back before they were seen. In the computer games once the turnover is initiated it can't be stopped before it completes, hence the allowance for "undoing" in this case.

Other rules that were followed besides the usual rules of play: cards could not be moved from the foundation stacks to the tableau; all the cards in a build of turned-up cards must be moved from one column in the tableau to another – moving part of a build could uncover a needed card for a foundation stack but this was not allowed; the cards being turned over from the remaining deck could be exposed while put on the talon but only the top card on the talon could be played; cards need not be played on the foundation stacks, and cards need not be played on the tableau.

The number of cards in the foundation stacks at the end of the game will be referred to as the number of cards "up".

3. Basic Statistics

With the original 220 games by cards the probabilities are shown in Figure 1. The horizontal axis is the number of cards "up" at the end of a game, i.e. zero to 52, and the vertical axis is the proportion, out of the 220 games, that pertained to each of the numbers of cards "up". It can be seen that the most probable number of cards in the foundation stacks at the end of a game was 3, and 52 cards resulted in about 1 in 10 games.

With the early computer game the probabilities are shown in Figure 2, and here it can be seen that the results are quite different from those in Figure 1 with a serious dip at 3 cards. Nevertheless, most resulting cards "up" are in the low count and about 1 in 12 games resulted in a "win" of 52.

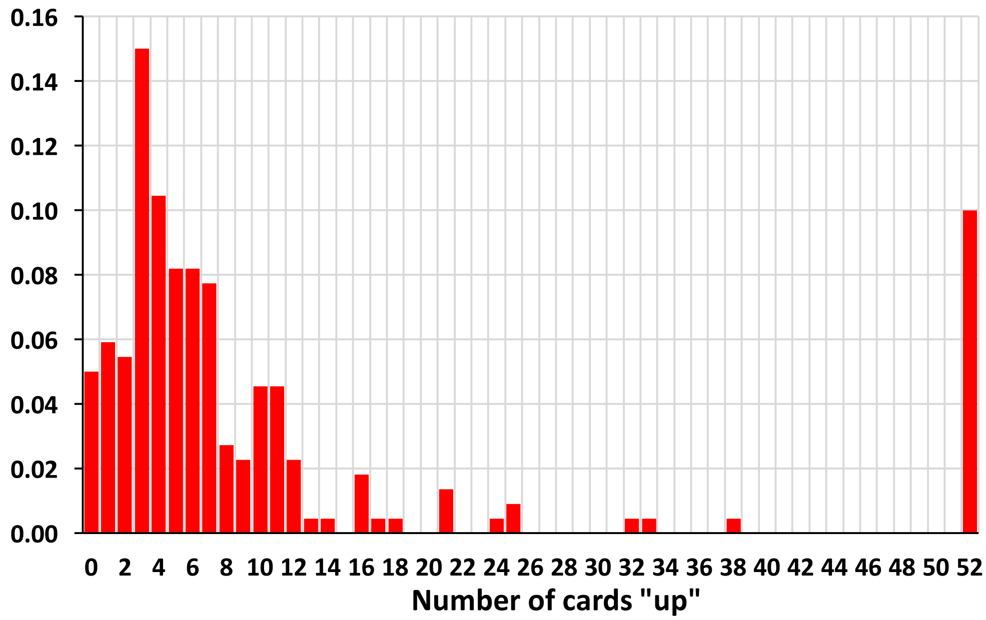

Figure 1. Probability of the number of cards “up” at the end of a game, from the results of 220 games played with a deck of cards. The most probable number “up” is 3, and a 52-card “win” shows a probability of 0.10.

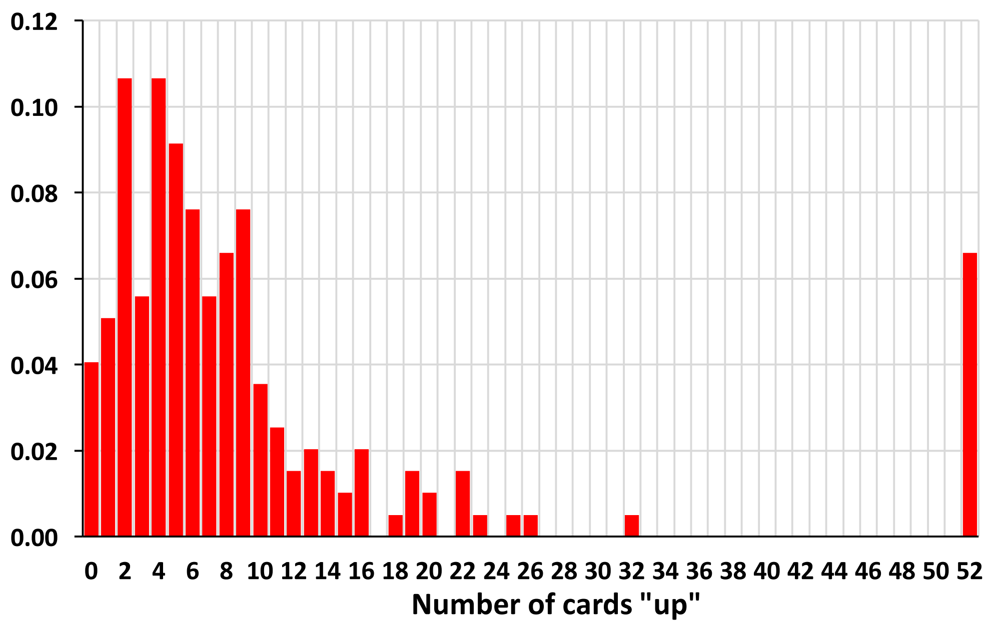

Figure 2. Probability of the number of cards “up” occurring at the end of a game, from the results of 197 games played on computer using an early solitaire game. The most probable number “up” is 2 and 4 with a dip at 3, and a 52-card “win” shows a probability of 0.066.

For the 823 games by Growly Solitaire, the results are shown in Figure 3. The results here are more similar to Figure 1 than Figure 2 with a clear peak at 3 cards “up” and about 1 in 8 games giving a “win” of 52.

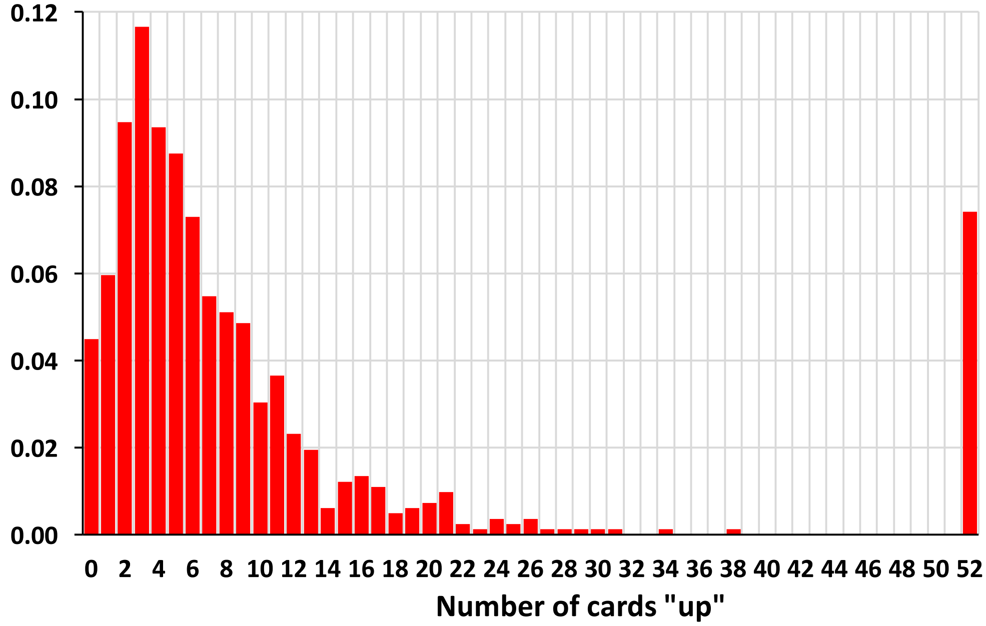

Figure 3. Probability of the number of cards “up” occurring at the end of a game, from the results of 823 games played on computer using Growly Solitaire. The most probable number “up” is 3, and a 52-card “win” shows a probability of 0.0741.

4. Statistical Tests On Different Modes And Strategies

During the course of playing the games, several questions arose regarding the similarities of the different computer games, and the similarity of those to the card games. There has been some interesting discussion on the internet regarding random number generators and their efficacy for this purpose. It has been suggested[5] by a simple statistical test that the early process used in Windows (1998) for solitaire was deficient unless the game was rebooted each time. The writer of the Growly software had doubts[4] about Apple’s internal random number generator so he decided to use ISAAC instead[6].

Before all 823 games using Growly software were played some tests were done concerning the need to reboot, namely tests of the similarity of the results using Growly software with a reboot each time compared to it being left “on” between each game, i.e. with only one reboot – one seed for the random number generator – for a complete set of games. The programmer[4] had used a cycle for the random number generator in Growly solitaire of order 10^{12} which is many orders of magnitude beyond being long enough for the present purposes.

Tests were also made regarding the similarity between the results of the card games and the computer games.

The tests were carried out with the null hypothesis being “all modes of play are statistically the same”, and the tests were done to determine if this was likely false. The probabilities (p-values) pertaining to the observed differences in the “up”-card counts using the χ^2 two-sample test were the final determinant of falseness, with values less than 5% chosen as the indicator of likely falsification.

The set of p-values for the various tests is given in Table 1. The overall result is that the p-values are *larger* than 5% for all cases and *much larger* for all but one. On the basis of this it was considered that there is little or no evidence for considering the null hypothesis false and so all modes of play were combined into one unified set. It can also be noted that the overall comparison between cards and computer, the last entry in Table 1, has a p-value of 34% indicating a high degree of statistical similarity.

Sample 1	Sample 2	p-value %
2017 cards	early cards	68
GS rebooted	non-rebooted	58
early computer	all cards	10
early computer	all GS	59
all cards	all GS	60
all cards	all computer	34

Table 1. Results of χ^2 two-sample tests to determine if the samples come from the same probability distribution. “All GS” is all Growly Solitaire games combined (rebooted each time plus not rebooted each time). “All cards” is early cards and 2017 cards combined. “All computer” is all Growly Solitaire and early computer games combined. (The tests were also carried out using a Kolmogorov-Smirnov process which provided larger p-values for the two smaller ones and similar values for the others.)

5. The Total Set

For all 1280 games, the resulting probabilities are shown in Figure 4 and Table 2. The peak at 3 cards is quite clear and the tapering off for larger numbers of cards is also well-delineated (with some minor dips and peaks). The statistics are generally not well-established for the infrequently occurring larger card numbers: 27 cards or more “up” occurred in only 1 or 2 games out of the 1280, with the exception of a complete “win” of 52 which occurred 100 times (about 1 in 13 games).

6. A Curve Fitted to the Data

The data of Figure 4 are reshown in Figure 5 in a log-linear format. It can be seen that the taper-off for more than 3 cards is quite closely exponential. An enlargement of Figure 4 for card numbers 0 to 3 shows a result that is closely linear. Based on these overall appearances, a linear least-squares regression line was fitted to data for 0 to 3 cards, and an exponential curve was fitted to the data for 4 to 40 cards. The latter was fitted by minimizing the average squared-difference between the data set and an exponential curve, and the minimization was performed by varying the decrement in the exponential while keeping the overall scale factor for the exponential set so that the probability curve over the full 0 to 52 cards normalized properly. The measured value for 52 cards was used without change. The results are shown in Figure 6 where the data of Figure 4 are reshown along with the fitted curve appearing as the dots.

The two equations for the fitted curve are

$$Probability(n) = 0.0236n + 0.0404 \quad \text{for } 0 \le n \le 3 \text{ cards}$$
$$Probability(n) = 0.1025e^{-0.1808(n-4)} \quad \text{for } 4 \le n \le 40 \text{ cards}$$

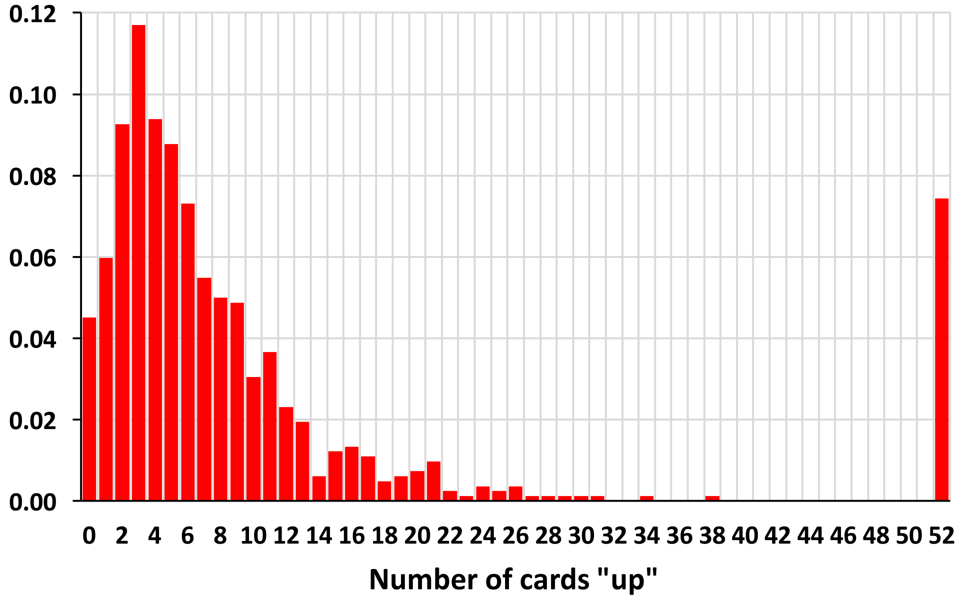

Figure 4. Probability of the number of cards “up” occurring at the end of a game, from the results of 1280 games that were played – computer games and card games. The most probable number “up” is 3, and a 52-card “win” shows a probability of 0.0781.

# of Cards	Freq.	Prob.	# of Cards	Freq.	Prob.	# of Cards	Freq.	Prob.	# of Cards	Freq.	Prob.
0	56	0.044	14	10	0.008	28	1	0.001	42	0	0.000
1	75	0.059	15	13	0.010	29	1	0.001	43	0	0.000
2	113	0.088	16	19	0.015	30	1	0.001	44	0	0.000
3	144	0.113	17	11	0.009	31	1	0.001	45	0	0.000
4	124	0.097	18	6	0.005	32	2	0.002	46	0	0.000
5	113	0.088	19	8	0.006	33	1	0.001	47	0	0.000
6	96	0.075	20	8	0.006	34	1	0.001	48	0	0.000
7	78	0.061	21	12	0.009	35	0	0.000	49	0	0.000
8	64	0.050	22	5	0.004	36	0	0.000	50	0	0.000
9	61	0.048	23	2	0.002	37	0	0.000	51	0	0.000
10	42	0.033	24	4	0.003	38	2	0.002	52	100	0.078
11	46	0.036	25	5	0.004	39	0	0.000	Total	1280	1
12	28	0.022	26	4	0.003	40	0	0.000			
13	22	0.017	27	1	0.001	41	0	0.000			

Table 2. For the full set of 1280 games, the frequency of occurrence and probability for the number of cards in the “up” stacks at the end of the game are shown.

and other fixed values are

$$Probability(n) = 0 \qquad \text{for } 41 \le n \le 51 \text{ cards}$$
$$Probability(n) = 0.0781 \qquad \text{for } n = 52 \text{ cards.}$$

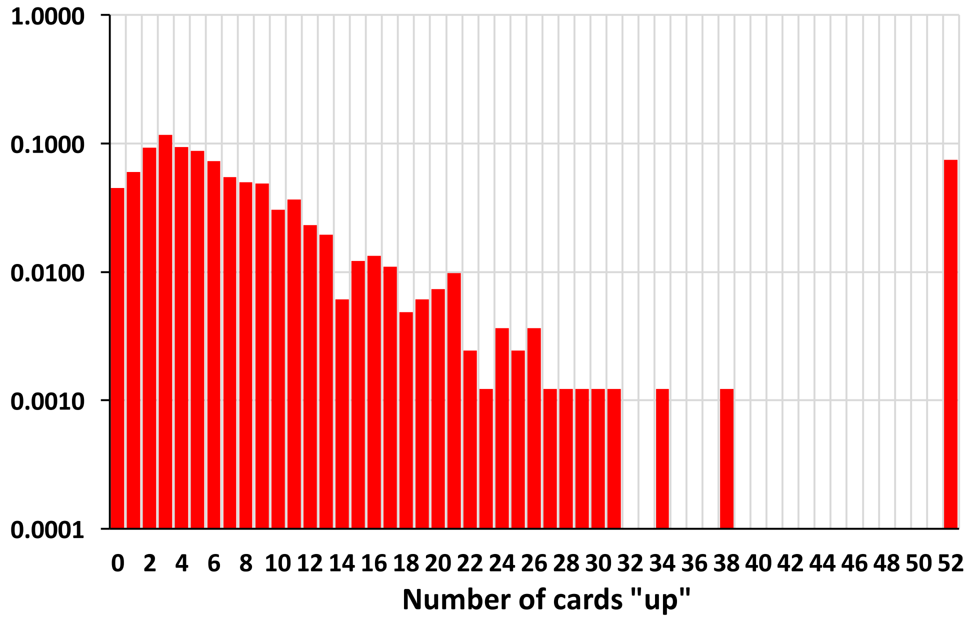

Figure 5. The data of Figure 4 reshown on a logarithmic vertical axis. The long decrease from 3 to 39 shows an approximately linear behaviour in this format.

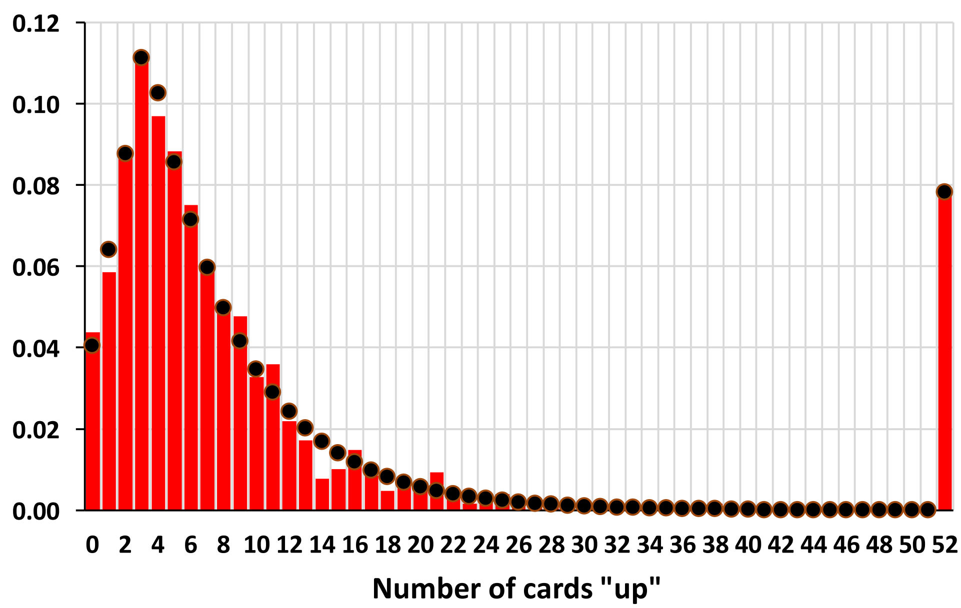

Figure 6. The results of the full 1280 games (as in Figure 4) with the fitted curve (dots) shown as well.

From the measured data several approximate average frequencies of occurrence stand out:

Average Frequency of 0 cards "up": Once in 23 Games

Average Frequency of 3 cards “up”:	Once in 9 Games
Average Frequency of 4 cards “up”:	Once in 10 Games
Average Frequency of 52 cards “up”:	Once in 13 Games

7. Number of Games between Wins, and Between Other Numbers “Up”

From the 1280 data set the number of games between a given number of cards “up” has also been determined with the intention of obtaining some estimate of the distribution for each number of “up” cards. The average for each number should be the same as the probability of occurrence given in Table 2 multiplied by the total number of games, 1280. The unexpected result is that the distribution of games “between” shows the largest values at the smallest number of games “between”, implying, for example, that for the number of games from one “win” to the next is more weighted to low numbers than high ones, even though a player may have the opposite impression.

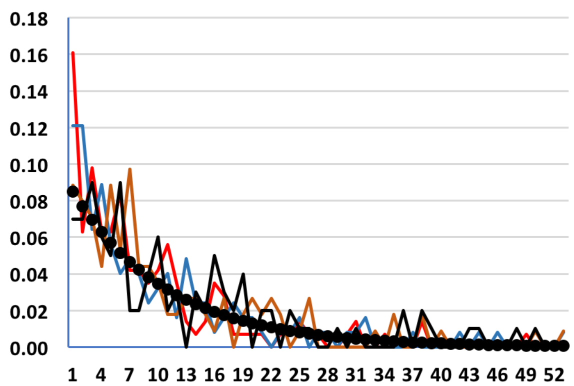

Figure 7. Probability of the number of games before the re-occurrence of given number of “up” cards.

The probabilities of re-occurrence are shown as solid lines in Figure 7 for several numbers of “up” cards: 3 (red), 4 (green), 5 (rust) and 52 (black). The horizontal axis numbers give the number of games to the *next* game with that same number of “up” cards so they are one more than the number of games *between*. It can be seen that all the curves show a drop-off toward large values with the maxima occurring in the 1 to 3 range. The black dots show a theoretical line which is an average for the 4 solid lines (3,4,5 and 52 cards up).

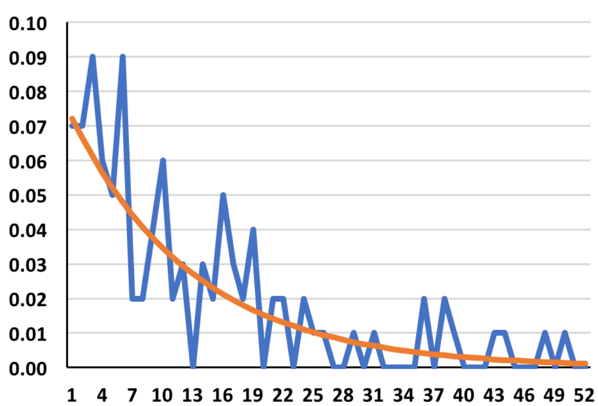

Figure 8. Probabilities of lengths of runs from one “win” of 52 cards “up” to the next: data (blue), $P_{52}(m)$ (red) using the overall probability of occurrence of 52 cards “up” from Table 2 (0.0781).

The theoretical values are obtained as follows. The full expression for the probability $P_n(m)$ of a length m of runs of non-re-occurrence of a given number of cards “up” n is given by the probability of *not* having n cards “up” m times multiplied by the probability of having n cards up on the $(m+1)^{th}$ time. Designating the probability of having n cards up as $Q(n)$, the probability of not having n cards up is $1 - Q(n)$, so $P_n(m)$ is given by

$$P_n(m) = Q(n)\{1 - Q(n)\}^m \qquad (1)$$

Here m is the length *between* re-occurrence, so $m = 0$ means the re-occurrence happens on the next game. The data for $n = 52$ and the curve from Eq. (1) are shown in Figure 8 with the measured value as given in Table 2 and shown in Figure 6 used for $Q(52)$.

8. Stability of the Measured Histogram Values

Let the histogram values from which the probabilities are obtained be given by $H(n)$ where n is the number of cards "up". Then

$$H(n) = \sum_{j=1}^{N} \delta_{nR_j} \tag{2}$$

where N is the total number of games that are included in producing the results, j indicates each game, R_j is the number of cards "up" for the j^{th} game and δ_{nR_j} is the Kronecker delta-function whose value is 1 if the subscripts are the same and zero if not. This is a sum of N independent random binomial variables and the variance $\sigma^2(n)$ of such a sum is given by[7]

$$\sigma^2(n) \;=\; NQ(n)\big(1 - Q(n)\big) \tag{3}$$

To obtain the variance of the *probability* estimates, $\sigma^2(n)$ as defined in Eq. (3) must be divided by N^2, so that the standard deviation of the probability estimates becomes

$$\frac{\sigma(n)}{N} = \sqrt{\frac{Q(n)(1 - Q(n))}{N}} \tag{4}$$

The confidence interval for $Q(n)$ based on N samples is

$$\text{Interval Bounds} = \pm Z(\alpha/2)\sqrt{\frac{\hat{Q}(n)\left(1 - \hat{Q}(n)\right)}{N}} \tag{5}$$

where the unknown $Q(n)$ is replaced by its estimate $\hat{Q}(n)$, and $Z(\alpha/2)$ is the standard normal value such that the area to the right is $\alpha/2$, and $\alpha = 1 -$ confidence level percentage/100.

For $N = 1{,}280$, and $n = 3$, and using the measured value 0.113 in Table 2 for$\hat{Q}(3)$, the estimate for the standard deviation of the measured probability from Eq.(4) is

$$\frac{\hat{\sigma}(3)}{N} = \; 0.0088.. \tag{6}$$

The implication from this is that the 90% error bar on the probability for $n = 3$, i.e. $Z(0.05) = 1.645$, is 0.028 in total length, or, the statistically expected value $Q(3)$ for $n = 3$ is such that

$$0.099 < Q(3) < 0.127 \tag{7}$$

with a confidence level of 90%. The 90% confidence bounds for all n-values for the present number of games are shown in Figure 9. Only at $n = 14$ is the fitted curve outside of the bounds.

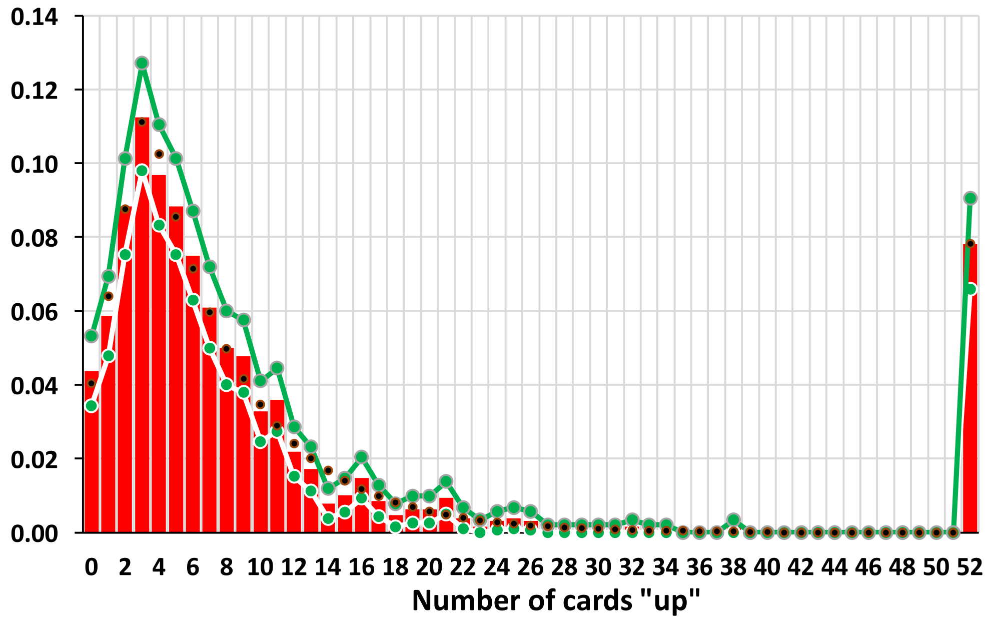

Figure 9. The measured *pdf* as shown in Figure 6 but with the 90% confidence level bounds shown in green and white.

The precision of each estimate of the probability density function can be defined as Eq. (4) divided by $\hat{Q}(n)$ and from this an estimate of the number of games needed to obtain a given level of precision can be determined. Using the 90% confidence level and Eq. (3) the precision ε is given by

$$\varepsilon(n) = \pm\frac{1.645\hat{\sigma}(n)}{N\hat{Q}(n)} = \pm 1.645\sqrt{\frac{1-\hat{Q}(n)}{N\hat{Q}(n)}} \tag{8}$$

or, on solving for N,

$$N = 2.71\frac{1-\hat{Q}(n)}{\varepsilon^2(n)\hat{Q}(n)} \tag{9}$$

To reach a precision of ± 5% for $n = 52$, and using the Table 2 value of 0.0781 for $\hat{Q}(52)$

$$N \geq \ 12{,}800 \tag{10}$$

For the number of games reported here, the precision for $n = 52$ is approximately ± 16%.

9. Summary and Conclusions

For the number of games played being about 200 there is no significant statistical difference in the histograms obtained from using cards compared to using computer games, and of two different computer games no significant statistical differences are apparent either. Rebooting each time to acquire a new starting seed makes no statistical difference compared to using one seed for a random number generator with a 10^{12} cycle or longer for a complete set of games.

The most probable number of cards put into the foundation stacks when the game is completed is 3 with a linear taper-off towards smaller numbers and an exponential taper-off towards larger ones. The exception for the larger numbers is a compete “win” of 52 for which the probability of occurrence is measured at 0.078 ± 0.012 (at the 90% confidence level).

The number of games played until a re-occurrence of a given number of “up” cards at the end of a game is distributed approximately exponentially with a taper-off towards large numbers and maximum values at zero, i.e. the next game is the most probable one for a re-occurrence, as non-intuitive as this seems. The average number of games to a re-occurrence is (by definition) the inverse of the probability of the number of cards “up” so the exponential distributions for runs of non-re-occurrence are completely determined by the probabilities of the number of cards “up” themselves. The trends of the measured data are fitted quite well by the theoretical curves.

To achieve a precision of ± 5% with a 90% confidence level in the estimated probability for a “win”, the number of games played must be at least 12,800, based on the measured probability estimates obtained here.

10. References

1. Wikipedia, *https://en.wikipedia.org/wiki/Klondike_(solitaire)*
2. It has been claimed that the “number of games a skilled player can probabilistically expect to win is at least 43%”, *ibid*., but perhaps it should be noted that it appears to have been based on wins by one person, *http://www.jupiterscientific.org/sciinfo/KlondikeSolitaireReport.html*
3. There is a plethora of anecdotal information on the internet stating winning percentages. *http://historymike.blogspot.ca/2010/04/on-winning-percentages-at-klondiek.html* provides a lengthy list of blog feedback entries encompassing a variety of Klondike apps and a variety of modes of play (the misspelling in the URL given here is correct). There have also been computer programs written to assist in winning at Klondike, and one such sophisticated program has been able to achieve a 35% win rate (see Bjarnason, Ronald, A. Fern and P. Tadepalli, *Lower Bounding Klondike Solitaire with Monte-Carlo Planning*, Oregon State University, 2009 *http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.151.7088&rep=rep1&type=pdf*)
4. Mason, Chris, *email communication*, Growlybird Software, 2017 April 27
5. Fricker, Ronald D. Jr., *Teaching Statistics*, Vol.20, 2, Summer 1998
6. Jenkins, Robert J. Jr. ISAAC and RC4, *http://burtleburtle.net/bob/rand/isaac.html*, 1993-1996, see also *http://burtleburtle.net/bob/rand/isaacafa.html*
7. Wikipedia, *https://en.wikipedia.org/wiki/Binomial_distribution - Variance*