Illustrating dependence between random variables using slot machines

Ian N. Durbach and Graham D. I. Barr
Department of Statistical Sciences, University of Cape Town

Summary
This paper illustrates the concept of statistical independence using the example of slot machines that may be played on multiple lines.

Keywords
Independence, binomial distribution, gambling, simulation

Introduction

One of the more subtle concepts to convey to students is the notion of statistical independence. Although students may superficially learn a mathematical definition of statistical independence, they often fail to properly apply this notion to a practical situation. In particular, they struggle to identify the circumstances under which independence is appropriate or inappropriate and to understand what costs might be incurred by assuming independence when it does not exist. Since analytical procedures are often greatly simplified by assuming statistical independence, it is of some importance to be able to identify situations in which independence holds. It is dangerous, however, simply to think that independence will hold in all or most cases and can be assumed as a matter of course. The importance of this kind of thinking is difficult to convey without examples of just how costly false assumptions around independence can be. In this paper, we use an example from slot machine gambling to explain the circumstances under which payouts are independent of one another and when they are not, and highlight quantitatively the considerable errors that can be made by incorrectly assuming independence.

The way in which a typical slot machine works is generally well known, but a brief outline is given here (more details can be found in Dowling (2005)). A slot machine generally contains a set of 5 reels, each with a set of symbols (e.g. apples, oranges, bars) appearing in random order along the reel. Symbols may appear more than once on each reel or may not appear at all on a particular reel. When the machine is played, the reels are "spun", either mechanically or digitally, and a position on each reel is chosen at random. The symbol in the selected position on each reel, as well as the symbols in the positions adjacent to the selected position on the reel, are displayed graphically to the player in a 3×5 grid i.e. the remainder of the reel is invisible to the player. Payouts are made when the same symbol appears along certain 'winning lines' across the grid. A winning line is a sequence of cells appearing from left to right in the 3×5 grid, with some typical winning lines given in Figure 1 for the online slot machine *Casino Roboto* (http://www.battlelinegames.com/). The first three winning lines are indicated by the horizontal line in each row of the display, and are shared by nearly all slot machine varieties, while winning lines 4 and 5 are indicated by the other two v-shaped lines. Lines beyond the third winning line are more likely to be specific to the particular game or slot machine manufacturer. Players elect how many lines they wish to make use of when playing; by default this option is set to the middle horizontal line in Figure 1 (a single-line play).

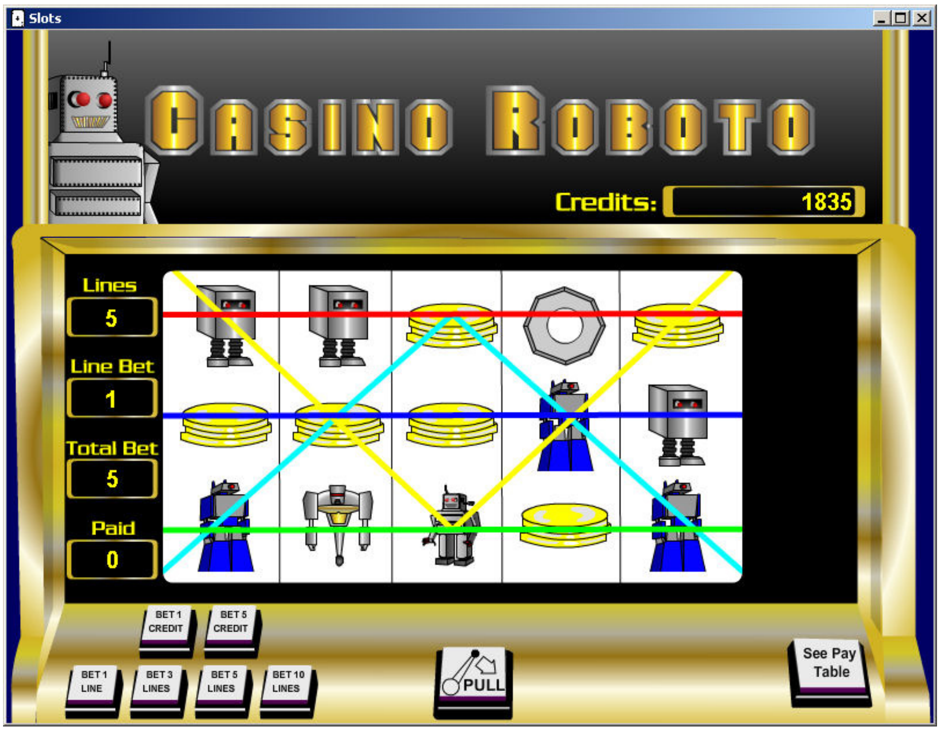

Figure 1: A sample of winning line configurations for a hypothetical slot machine

Two important features of the winning lines are: firstly, they must make use of one and only one position on each reel i.e. they may only move horizontally or diagonally left to right on the grid and not straight up or down; and secondly, for a payout to be won, the symbol must appear in the first *P* reels from left to right along this line, with the payout usually only being non-zero for $P \geq 3$ and increasing with *P*. For example, in Figure 1 three "coin" symbols appear in the first three positions of the middle horizontal line, which would result in some payout being won. There are several subtleties to the operation of real-world slot machines such as the appearance of 'wild' symbols, bonus features, and pooled cumulative jackpots, but these are not relevant to the aims of the current paper.

The main choices facing a slot machine player are (1) how many lines to play on, denoted *L* and (2) what amount to bet per played line, denoted *B*. Modern-day slot machines can typically be played on up to 25 lines, with choices offered in discrete sets and bets of up to 5 units can usually be placed on each played line (e.g. for *Casino Roboto*, $L \in \{1,3,5,10\}$ and $B \in \{1,5\}$). The total amount bet for a single play is thus *LB*. The typical outcomes which are affected by the choice of how many lines to play and how much to bet per line are the expected value of any winnings (always negative), the standard deviation of winnings, and the probability of any success on a single spin (regardless of amount won). In the remainder of the paper, we look at how choices of *L* and *B* affect these three outcomes with or without assumptions of independence.

When is an assumption of independence appropriate?

The expected value and standard deviation of payouts and the probability of success statistics can usually be easily calculated for a single-line, minimum-bet option. If we let $W_{L,B}$ be a random variable representing the payout from a single spin of *L* lines with a bet of *B* per line, then the expectation and standard deviation of $W_{1,1}$ are given by $\mathrm{E}\lfloor W_{1,1} \rfloor$ and $\sigma_{W_{1,1}}$ respectively, and the probability of any success is denoted ϕ_{hit}. In this paper we generally use the expected value and standard deviation of the absolute payouts themselves, net of the original bet LB i.e. if nothing is won, then $w_{L,B} = -LB$. In some instances it is also useful to refer to the expected payout and standard deviation per unit bet i.e. $\mathrm{E}\lfloor W_{L,B} / LB \rfloor$ and $\sigma_{W_{L,B}} / LB$. In the terminology of the casino industry, the expected payout per unit bet (which is invariably negative) is usually multiplied by -1 and termed the 'hold percentage', and indicates the fraction (or percentage) of all bets which are expected to be 'held' by the casino.

Since the past performance of a slot machine does not affect its future performance, successive plays are independent of one another, so that after n plays of a single-line, minimum-bet machine, one obtains

$$\mathrm{E}\left[\sum_{k=1}^{n} W_{1,1}^{(k)}\right] = n\mathrm{E}\lfloor W_{1,1} \rfloor$$

$$\mathrm{StdDev}\left[\sum_{k=1}^{n} W_{1,1}^{(k)}\right] = \sqrt{n}\sigma_{1,1}$$

Probability of any winning play = $1-(1-\phi_{hit})^n$

where we index the realisations with the superscript k. In this instance, the actual mechanical construction of the slot machine provides a clear justification for assuming that individual plays are independent of one another.

The cost of incorrectly assuming independence

Let us now consider the case where $L > 1$, keeping the bet per line constant at $B = 1$. At some stage in the evolution of slot machines, manufacturers decided that it would be more entertaining and time-saving for players to play on multiple lines rather than multiple times on a single line. Under the assumption that wins on different lines played in the same spin are independent of one another, the two formulations are identical and the above results hold. If we also allow the bet per line B to vary, then $W_{L,B} = BW_{L,1}$ and since under the assumption of line independence $W_{L,1}$ is equivalent to the sum of L independent realisations of $W_{1,1}$,

$$W_{L,B} = B\sum_{k=1}^{L} W_{1,1}^{(k)}$$

The expectation and standard deviation of $W_{L,B}$ are then obtained by substituting our earlier results for the sum of a series of independent realisation of $W_{1,1}$ into the equation above, giving

$$\mathrm{E}\lfloor W_{L,B} \rfloor = LB\mathrm{E}\lfloor W_{1,1} \rfloor$$

$$\sigma_{W_{L,B}} = B\sqrt{L}\sigma_{W_{1,1}}$$

The probability of achieving any winning play is independent of the amount bet and is therefore given by

$$\phi_{hit}(L,B)=1-(1-\phi_{hit})^{L}$$

so that the standard deviation of a particular game's payout is therefore linear in the bet per line *B* and follows the square root of the number of lines played *L*. This shows that if one wishes to increase the volatility of payouts (a desirable strategy when expected return is negative), one should increase the bet per line rather than the number of lines played, a similar message to the one identified in Croucher (2005). The probability of success is obviously not changed by the amount bet per line.

Let us more closely examine the assumption of line independence that was necessary to obtain the analytical results above. For most slot machines, the first three winning lines are independent of one another so that the above results hold; but for cases in which many lines are played, the winning lines are often neither completely dependent nor completely independent. In cases such as these we have

$$B\sqrt{L}\sigma_{W_{1,1}} \le \sigma_{W_{L,B}} \le LB\sigma_{W_{1,1}}$$
$$\phi_{hit} \le \phi_{hit}(L,B) \le 1-(1-\phi_{hit})^{L}$$

Modern slot machines make use of as many as 25 possible lines, and there is often substantial dependence between some of these lines. For games such as these, if line independence is assumed then variability is likely to be substantially understated, and probability of success substantially overstated, though expected monetary returns remain undisturbed.

Unfortunately, the dropping of the dependence assumption means that analytical results are no longer tractable. However, it is fairly easy to build a simulation model of a typical slot machine in order to compare the true moments and probability of success with those obtained under the assumption that lines are independent, and to see how the difference between the two increases as more lines are played. Since every slot machine is described by a few basic mechanisms – the placement of symbols on its reels, a schedule of payouts (the "paytable"), and a set of line configurations for the winning lines – a simulated slot machine can make use of a random number generator to select a position on each of the simulated reels and generate a simulated grid that would be observed by the player. The combination of symbols appearing along each line configuration is then checked against the schedule of payouts, and any winnings on different lines are added to the running total.

Table 1 shows the expected payout and the standard deviation of payouts both under and without the assumption of the independence of the random variables, where the 'without independence' results are obtained from 5 000 000 simulated plays of a simulated slot machine whose inputs – reels, symbols, lines and payouts – have been chosen so that it reflects as far as is possible a real-world slot machine. Note that the standard deviation of payouts is given first, with standard deviation per unit bet given in parentheses. Since the simulated expected payout statistics are guaranteed to converge to the theoretical values, only one set of values are shown; again, the expected payout per unit bet is given in parentheses, and is constant at -0.059 i.e. a hold percentage of approximately 6% for any number of played lines.

Although expected payout is left unaffected by the validity of the assumption of line independence, the volatility (or standard deviation) of payout is substantially underestimated and probability of success substantially overestimated if line independence is falsely assumed. These errors increase as the number of played lines increases, and in the case of this particular simulated slot machine become severe when $L > 9$. In reality, most slot machine players play on the maximum number of lines using the minimum possible bet per line, the so-called "maximin" bet. The increasing divergence in volatility and probability of success is clearly illustrated in Figure 3, with a clear 'kink' observable at $L = 9$. This occurs because for this particular slot machine, winning line 10 overlaps with winning line 2 in the first two positions, and thus dependence is high. Although for different machines the 'kink' may be located at other values of L, the important point is that in a real-world slot machine such dependencies becomes inevitable as L becomes large, but this inevitability is not picked up by the model assuming independence.

	$E[W_{L,1}]$	$\sigma_{W_{L,1}}$		$\phi_{hit}(L,B)$	
L	Indep/no Indep	Indep	No Indep	Indep	No Indep
1	-0.059 (-0.059)	16.422 (16.422)	16.422 (16.422)	0.146	0.146
3	-0.177 (-0.059)	28.444 (9.481)	28.142 (9.381)	0.377	0.269
5	-0.296 (-0.059)	36.720 (7.344)	38.967 (7.793)	0.546	0.377
9	-0.532 (-0.059)	49.266 (5.474)	58.238 (6.471)	0.758	0.544
13	-0.769 (-0.059)	59.210 (4.555)	78.161 (6.012)	0.871	0.575
19	-1.124 (-0.059)	71.581 (3.767)	105.67 (5.562)	0.950	0.616
25	-1.478 (-0.059)	82.109 (3.284)	132.752 (5.31)	0.981	0.654

Table 1: Theoretical results obtained by erroneously assuming lines are independent and simulation results modelling the true dependency between lines for a hypothetical slot machine

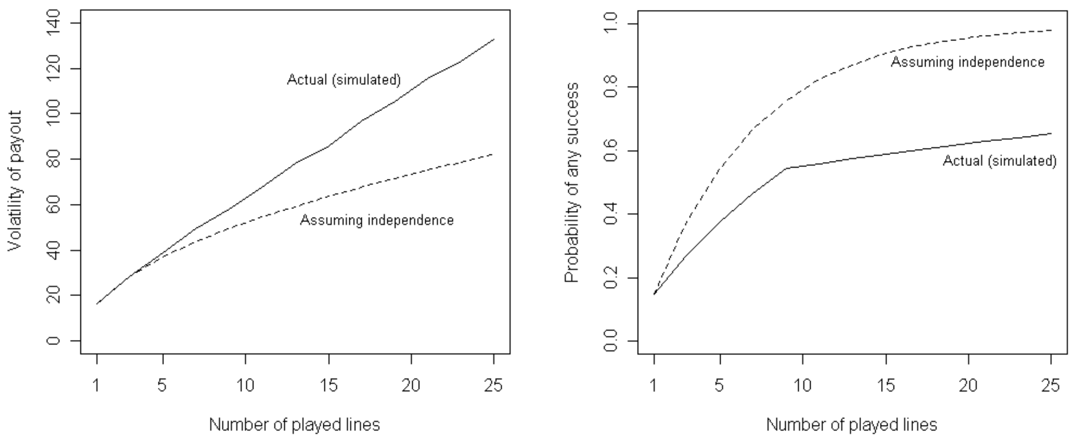

Figure 3: Effect of erroneously assuming lines are independent on (a) volatility of payout $\sigma_{W_{L,1}}$, (b) probability of success ϕ_{hit}

Incorrectly assuming independence in this case has several important implications. Firstly, underestimating the standard deviation of payouts results in confidence intervals around the expected win that are too narrow. The 95% confidence interval around the expected win per unit bet for 100 000 spins of a machine playing on 25 lines, is given by

$$\left[-0.059 \pm 1.96 \frac{3.284}{\sqrt{100\,000}}\right] = [-0.079; -0.039]$$

i.e. a 95% confidence interval around the hold percentage of [3.9%; 7.9%]. If line independence is not assumed, the 95% confidence interval obtained from the simulation results is

$$\left[-0.059 \pm 1.96 \frac{5.310}{\sqrt{100\,000}}\right] = [-0.092; -0.026]$$

i.e. a 95% confidence interval around the hold percentage of [2.6%; 9.2%], which is more than 2% wider (in absolute terms) than when independence is assumed. In practice, use of the independence assumption results in casino operators believing that machine payouts are too variable and that the machines are therefore faulty – when in fact, the payout percentage is entirely within what should reasonably be expected.

Conclusions

It is difficult to find real-world examples of dependence where the exact nature of that dependence does not need to be specified probabilistically, but exists because of the structure of the problem. In the case of playing a slot machine on more than one line, the dependence is built into the mechanical structure of the game, and requires only that the student understand the concept of winning line configurations. From this, a quite rich and detailed discussion of dependence can ensue, and some real subtleties can be explored. Furthermore, one can precisely identify the costs of ignoring dependence and how these vary as the assumption of independence becomes less and less tenable.

References

Croucher, J. (2005). A comparison of strategies for playing even money bets in roulette. *Teaching Statistics*, 27(1), 20 – 23.

Dowling, N; Smith, D and Thomas, T. (2005). Electronic gaming machines: are they the ‘crack-cocaine’ of gambling? *Addiction*, 100, 33 – 45.