SCRATCHING THE SURFACE OF PROBABILITY

Robert J. Russell, University of Paisley, UK

Scratch cards are widely used to promote the interests of charities and commercial concerns. They provide a useful mechanism for introducing many key concepts typically met by students in a first or second level under-graduate course involving probability modelling. The work described here stems from a real marketing problem. Simple card layouts lead quickly to consideration of combinatorial counting techniques, the probability of a player winning, and the expected value of a win. From a company viewpoint, assessing the risk of exceeding its prize money budget requires modelling of the distribution of total payout. The Binomial and Normal models are relevant here, while establishing expected payout and variation in payout highlight properties of expectation and variance. Simulation may also be used to support theoretical results.

INTRODUCTION AND BACKGROUND

This paper considers the development and implementation of part of a module entitled "Applied Probability, Simulation and Experimental Design (APSED)", given to a mixed group of undergraduate students mainly from the areas of Mathematical and Management Sciences. Experience with other similar modules suggested that students found the material in probability and random variables rather dull. They tended to avoid this area of work in conventional examinations, in favour of more routine data analytic questions.

Quite fortuitously, during the first session in which the new module was offered, a consultancy arose concerning the use of a scratch card entitled "Go for Gold" intended to be used as a promotional tool for a finance company. The original situation involved the use of a 5 x 4 card layout. One of the covered entries in each row (which contains the fictitious time of runners in an Olympic 100m race) is revealed. The individual playing the card assumes the time revealed in the first row, while those revealed in each of the other rows represent the times of each of four well-known 100 metre runners. About 15,000 cards would be produced (and used), and there was a request that the *average* win per card should be about £10.

Several requirements were stipulated at the outset by the marketing company, the most important of which were that:

(i) each player should have a chance of winning the maximum prize of £100,

(ii) the total prize money to be paid out should not exceed £150,000.

A major aim was to promote customer goodwill, but, at the same time, a critical consideration for the company involved was to keep the total prize payout to within its budget. Specifically there was a request to determine the *chance* that the budget would be exceeded. In other words *assess the risk* of this event to the company.

Although rather obvious, when it was pointed out that if each of 15,000 customers were to have some chance of winning the maximum prize of £100 that it would then be possible for the total payout to be as much as £1,500,000 the company were somewhat perturbed. The chance of this occurring is, of course, rather remote - but this fact did little to reassure the company. Indeed the perception seemed to be that if there was *any* chance of a payout of £1,500,000 the promotion should be dropped! Also, since this sum is so much larger than the budgeted figure of £150,000 the company took the view that somehow there must be a considerable chance that the budget would be exceeded.

This type of situation provides a rich vein for introducing and reinforcing the relevance of much of the material covered in the early part of the APSED module. The main issue is that of establishing the *probability distribution* associated with the total payout.

A SIMPLIFIED VERSION OF THE GAME

Much benefit and insight is gained by considering a simplified version of the problem. For example, consider a 2 x 2 card layout. The four (covered) entries on the card can, for convenience, be labelled 1, 2, 3 and 4 in some random order, indicating finish position in a race. Restricting the card size in this way makes it easy to write out all possible 4! = 24 variations of the card. Suppose that 14,400 cards in total are produced, in 600 blocks, where each block is a complete set of each of the 24 possible distinct card variations.

A particularly simple version of the game might be,

Game A: *a player selects any one of the four entries on the card; if this shows a 1st place finish, the player wins £20, otherwise nothing is won.*

From a player's perspective, the amount which can be won, W say, is a random variable with a simple probability structure: P(W = 0) = ¾ and P(W = 20) = ¼ . The expected value of the amount to be won and the variance in this amount are easily established: E(W) = 5, and V(W) = 75. The last result is rather less intuitive than the others but is readily justified by carrying out simple *simulations* of the game, perhaps even all 14,400, and calculating the mean and standard deviation of the actual prize amounts won, using the usual data analytic formulas with which the students are typically more familiar.

From the company's perspective, the concept of variation in the total payout, denoted as T say, is of primary importance. Variation dictates the risk to the company of exceeding its budget for the total payout. Suppose there is a maximum prize money budget of £73,250. Assuming that all 14,400 cards are played, the *range space* of the random variable T is easily established. Any amount between 0 and 288,000 in multiples of £20 may be won. Furthermore, since $T = \sum W_i$ where W_i is the prize won by the i'th player, the expected value and variance of T are easily obtained using the standard *properties of expectation*. Also, since T may be viewed as the sum of a large number of *independent, identically distributed* (Bernoulli) random variables, the Central Limit Theorem may be invoked to establish that T, to a (very) good approximation will be Normally distributed with expected value and variance given by,

$$E(T) = E\left(\sum W_i\right) = \sum E\left(W_i\right) = 14400 \text{ x } 5 = 72{,}000 \text{ , and}$$

$$V(T) = V\left(\sum W_i\right) = \sum V\left(W_i\right) = 14{,}400 \text{ x } 75 = 1{,}080{,}000$$

Again all of these results can be justified by simple simulation of card plays.

An alternative approach is to view T as a function of the total number of winners in the 14,400 plays, N say. Since each player either wins with probability ¼ , or loses with probability ¾ , the total number of winners N ~ Bi (n = 14,400, p = ¼). Standard results give E(N) = np = 3600 and V(N) = np (1 – p) = 2700. For such a large n , the distribution of N will again be well approximated by a Normal distribution with mean $\mu = 3600$ and variance $\sigma^2 = 2700$. Then, since T = 20N, it follows that T will also be well approximated by a Normal distribution with E(T) = E (20N) = 20 x E(N) = 72000 and V(T) = V (20N) = 20^2 V(N) = 400 x 2700 = 1,080,000 , as before.

The chance that budget will be exceeded is then a routine Normal probability calculation,

$$P\ (T > 73{,}250) = P(U > \frac{73{,}250 - 72{,}000}{\sqrt{1{,}080{,}000}}) = P\ (U > 1.20) = 0.1151$$

i.e. there is, approximately, an 11.5% chance that the prize budget will be exceeded.

VARIATIONS ON A THEME

Careful choice of other simple alternatives can serve to illustrate different aspects of the original problem, increasing the richness, complexity and interest of the situation. Three variations worth considering are:

Game B: a player selects one entry from the four available; he/she wins £15 if a 1st place is revealed and £5 if a 2nd place is revealed; otherwise no prize.

Game C: a player selects one of the two entries from row 1 of the card (the player's finish position) and one entry from row 2; if the player finishes ahead of the rival the prize is £10; otherwise no prize.

The last variation, D, is played with a slightly enlarged version of the card with three rows and two columns, the rows representing the player and two competitors.

Game D: a player selects one entry from each of the three rows (the entry from the first row representing the player's finish position); if the player finishes ahead of both competitors s(he) wins £15; otherwise no prize.

In game B, the prize structure is a little more complicated. This invites discussion, from a player's viewpoint, about the relative attractiveness of prize structure weighed against the chance of a win. Results for game B are readily established along the lines already outlined for game A. The results are summarised in Table 1 below. Most noticeably, the variance of the total payout is halved for game B compared to game A, with a consequent reduction in the risk of exceeding budget from 11.5% to 4.5%.

Game C introduces a more obvious competitive element to the situation - the player has to beat a rival! The chance of winning is now *conditional* on the particular arrangement of finish positions (1,2,3,4) on the card played. A little thought reveals that there are now different *types of card*. Type here is taken to mean cards which have the same probability of producing a win for the player. Exactly how many types there are, and how many of each type there are, is now of critical importance. Listing all possible 24 arrangements and partitioning by type provides useful insight. The reduction in variance of total payout for game C is quite striking - see Table 1. Establishing this variance, however, is more difficult since the payouts on individual cards are *no longer independent.* This difficulty can be circumvented by focussing on the number of winners for each type of card - which is Binomially distributed with appropriate parameters n_j and p_j for type j. The mean and variance of total payout are then established by utilising the results for sums of independent (Binomial) variates and scaling the results by the prize money paid to a winner.

Variation D introduces an extra row of two entries, representing a third runner in the race. This increases substantially the difficulty of partitioning by type the possible (i.e. 6! = 720) arrangements of finish positions. Once established, however, the same approach as taken with game C can be repeated to obtain the results shown in Table 1.

Table 1. Summary of results for games A - D.

Game	A	B	C	D
Player:				
P(W)	0.25	0.5	0.5	0.33
E(W)	5	5	5	5
V(W)	75	37.5	25	50
Company:				
Min.payout	0	0	24,000	14,400
Max. payout	288,000	216,000	120,000	158,400
E(T)	72,000	72,000	72,000	72,000
V(T)	1,080,000	540,000	210,000	441,000
SD(T)	1039	735	458	664
P(T > budget)	0.115	0.045	0.003	0.030

Clearly other variations of the game can be contrived - a 2x3 rather than a 3x2 layout is a further natural candidate - but no new concepts are likely to be required. Of course, as the size of the card increases, computational feasibility becomes a more serious issue. Indeed for the original 5x4 card layout there would be $20! \approx 2.4 \times 10^{18}$ possible arrangement of 20 distinct finish positions, a number far in excess of the required number of cards to be used - and certainly rather costly and time consuming to produce! The whole complexion of the problem and the way in which it might be tackled takes on a new dimension of difficulty - but again this provides an opportunity for further discussion of alternative, perhaps more pragmatic approaches, to the company's original problem.

STUDENT EXPERIENCE AND REACTION

The above game scenarios were presented to students in the form of coherent case study, which can be supplied on request. Over a period of 3 weeks or so there were about 5 or 6 formal lectures covering the relevant topics, and during which students were split into groups of 3 or 4 to discuss the case study (in and out of class time). The reaction and participation of students was generally very positive - perhaps encouraged by the knowledge that the case study would be given a high weighting (40%) in the overall assessment of the module. Following the 3 week period, students were examined on an individual basis using a written class test taken on an open-book basis.

The mean mark achieved by the cohort of 36 students who were subjected to the case study approach was 58.6%. This may be contrasted with the results of another cohort of 40 students who undertook the same module in a more traditional fashion with a conventional end of module examination. This examination contained 5 questions of which 1 covered material of a similar nature to the case study. The students were asked to attempt 3 of the 5 questions. Of the 40 students involved, 14 (35%) did not attempt the relevant question; the 26 who did, achieved a mean mark of only 36.9%. Although any direct comparison should be made with considerable caution, the apparent improvement in performance and certainly the different reactions observed for the two groups of students are more than encouraging.