

# Investigation of probability distributions using dice rolling simulation

Stanislav Lukac & Radovan Engel  
Pavol Jozef Safarik University, Slovakia  
<stanislav.lukac@upjs.sk>

This paper deals with the evaluation of dice outcomes in two problem-solving situations. Dice are considered one of the oldest gambling devices and thus many mathematicians have been interested in various dice gambling games in the past. We can take the famous mathematicians as an example: Blaise Pascal and Pierre de Fermat applied the theory of probability to solve problems connected with dice rolling. Dice have been used to teach probability ever since.

Dice rolls can be effectively simulated using technology. The National Council of Teachers of Mathematics (NCTM, 2000) recommends that teachers use simulations to give students experience with problem situations that are difficult to create without technology. Computer technologies furnish visual images of mathematical ideas; they facilitate organising and analysing data; and they compute efficiently and accurately. Similarly, in accordance with the *Standards for Excellence* of The Australian Association of Mathematics Teachers (AAMT, 2006), a variety of appropriate teaching strategies is incorporated in the intended learning experiences, enhanced by available technologies and other resources. Real dice rolling and recording of outcomes are crucial in an introductory phase of the problem-solving process. These activities develop students' abilities to organise data for effective analysis. Students can draw different conclusions by carrying out a small number of random experiments. Based on these differences, students should realise that a large number of random experiments is necessary to discover general relationships.

A spreadsheet program can be used for simulation of random events because it enables users to generate random numbers into the selected cells of a table (e.g., Microsoft Excel offers several tools for generating random numbers). In this activity we use the basic mathematical function RAND for the simulation of dice outcomes. Values of this function are random numbers from the interval  $\langle 0, 1 \rangle$  generated according to a uniform probability distribution. As we need random natural numbers from 1 to 6, we multiply the values of the function RAND with the number 6. We get random numbers from the interval  $\langle 0, 6 \rangle$ . To cut off the decimal part of real numbers, functions INT or TRUNC can be used. By applying one of these functions to random numbers from the interval  $\langle 0, 6 \rangle$ , we obtain integers

from the set {0, 1, 2, 3, 4, 5}. Therefore, it is necessary to add 1 to the partial result to simulate dice rolling. The final formula looks as follows:

$$=1 + \text{INT}(6*\text{RAND}()) \quad (1)$$

## Rolling the number 6

The first problem-solving situation can be presented to students as a game. One student from a group rolls a die; the other students guess the number of trials needed to roll the number 6 for the first time. Having repeated the game at least 10 times, students are asked to answer the following question:

Imagine that you have the following two options:

1. The number 6 will be rolled in the second trial.
2. The number 6 will be rolled in the fifth trial.

If you were to bet money, which one would you choose?

Students are required to consider which option is more probable. After the introductory phase of experimentation in a paper-and-pencil environment, they then proceed to a computer simulation.

Students create an MS Excel table to model outcomes of a series of dice rolls. The outcome 6 has to occur exactly once in each series, whereby rolling the number 6 terminates each series. Frequencies of rolling the number 6 will be evaluated in the table. In order that the obtained frequencies correspond to theoretical probabilities, it is essential to simulate a large number of random experiments. This connection between probability and frequency will be used to evaluate the outcomes of series of dice rolls. A large number of series of rolls will be modeled in the table, and the outcomes of each series will be placed into the rows of the table. Since it is impossible to determine in advance the number of trials necessary to roll the number 6 for the first time, the table should include generated rows with a variable number of cells completed by the number 6. Visual Basic for Applications (VBA) can be used to generate such a table. The basis of the VBA macro is the function RND where values are repeatedly entered into cells using cycles. The macro can be executed by means of a button placed on a sheet. The macro listed below is designed to generate the table with the first row beginning in cell A1. The table contains the given number of rows

VBA macro for generating the table with random outcomes of dice rolls

```
Private Sub CommandButton1_Click()  
Dim Count As Integer, I As Integer, J As Integer, R As Integer  
Count = InputBox("Specify number of series")  
Range("A1").Select  
For I = 1 To Count  
    J = 0  
    Do  
        J = J + 1  
        R = Int(6 * Rnd) + 1  
        ActiveSheet.Cells(I, J).Value = R  
    Loop Until R = 6  
Next I  
Range("A1").Select  
End Sub
```

(variable Count) with random outcomes of dice rolls completed by the number 6.

Another method—based on entering and copying the formulas—can be used to simulate several series of dice rolls. To make the copying of the formula easy, we suggest creating a table with a constant number of cells in rows. Each row will contain 30 random numbers from the set {1, 2, 3, 4, 5, 6} and will represent one series of dice rolls. We suggest that students generate at least 5000 series of dice rolls. It is necessary to ensure that no other outcomes of dice rolls will be written into the cells of each row after rolling the number 6. The number 0 will be entered in these cells and it will indicate the state after rolling the number 6 in an individual series of rolls. It may happen that the number 6 will not appear even after 30 rolls. However, these cases are rare and will not be included in the final evaluation due to counting the number 6 in the table columns.

Having created the table header (see Figure 2) we enter the following formula into cell B2:

$$=IF(OR(A2=6,A2=0),0,1 + INT(6*RAND())) \quad (2)$$

Since a text value is stored in cell A2, the condition in the function IF is not fulfilled and therefore a random number from the specified set is entered into cell B2. By copying this formula into a continuous area containing 5000 rows and 30 columns, students can generate outcomes of 5000 series of dice rolls very quickly. If the number 6 has not been rolled in an individual series of dice rolls, a new outcome will be generated, otherwise a value of 0 will be entered into the cell. Values 0 will repeat until the end of the series after the first occurrence of value 0. Zero values, gathered in the generated table, do not represent outcomes of dice rolls and therefore should not be displayed. Setting of a conditional format in the table can improve visual orientation in the outcomes. To hide zero values in the table, the text colour can be set to white by using the <Format> command for each cell containing value 0 in the area B2:AE5001. Figure 1 shows the dialogue



Figure 1 Conditional formatting for cells containing the zero value.

	A	B	C	D	E	F	G	H	I
1		1st r.	2nd r.	3rd r.	4th r.	5th r.	6th r.	7th r.	8th r.
2	1st s.	2	1	4	6				
3	2nd s.	3	4	3	4	1	6		
4	3rd s.	1	2	4	5	1	1	1	6
5	4th s.	6							
6	5th s.	1	6						
7	6th s.	1	2	3	6				
8	7th s.	3	4	1	1	4	2	4	6
9	8th s.	6							

Figure 2. The part of the table with outcomes of series of dice rolls.

window <Conditional Formatting> with the corresponding settings. The left upper cut-out of the table is shown in Figure 2.

Let  $A_i$  be the random event that the number 6 is rolled for the first time in the  $i$ -th roll. We can determine the number of occurrences of the number 6 in each column of the table using the function COUNTIF for comparing probabilities  $p(A_i)$ ,  $i = 1, \dots, 30$ . The bottom part of the table containing summary results for the first eight rolls in series is shown in Figure 3.

The frequency of rolling the number 6 in the first roll of a simulated series of dice rolls is calculated in cell B5002, where the following formula is entered: =COUNTIF(B2:B5001,"=6"). We copy this formula into the adjacent cells in the row until cell AE5002. The sum of frequencies in the area B5002:AE5002 does not need to be equal to 5000, because it may happen that the number 6 will not appear in some series of dice rolls. The number of "unsuccessful" series is calculated in cell I5003 according to the formula, which is shown in the formula bar in Figure 3.

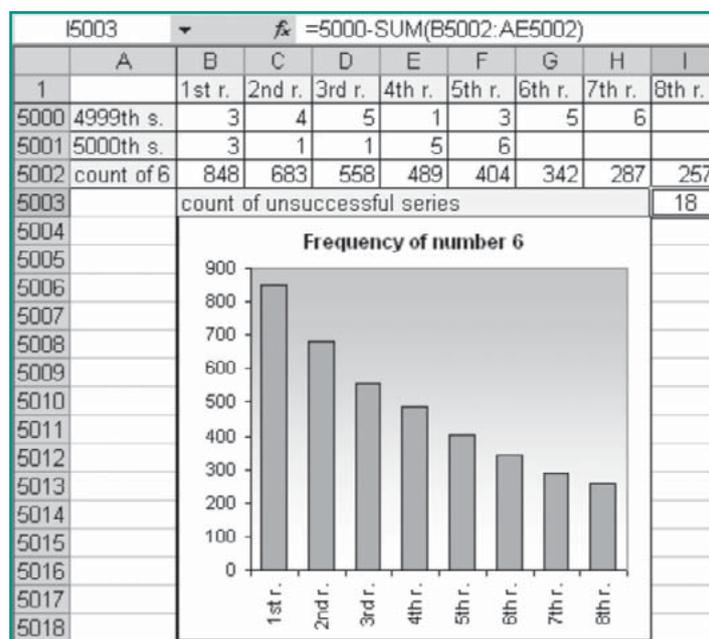


Figure 3. Summary results and the frequency column graph.

The frequencies of rolling the number 6 in the first eight rolls are displayed in the column graph in Figure 3. We can generate new random numbers using the F9 key. New dice outcomes enable students to investigate changes in summary results and also in the graph. The histogram shows that the frequency of rolling the number 6 is the largest in the first roll and then gradually decreases. This distribution of frequencies can be justified by the calculation of probabilities:

$$p(A_1) = \frac{1}{6}, p(A_2) = \frac{5}{6} \cdot \frac{1}{6}, \dots, p(A_i) = \frac{5^{(i-1)}}{6} \cdot \frac{1}{6}, i \in N$$

This probability distribution is an example of geometric probability distribution.

## Sum of the scores on the dice

The second problem is closely related to the dice game called Craps. Rolls of two dice and different values of the scores on the dice are the basis of this game. Considering the fact that the probability of the individual dice roll

outcomes is the same, it might seem that individual sums of the scores are also equally probable. Even the mathematician Gottfried Wilhelm Leibniz wrongly supposed that the sum of 11 is as probable as the sum of 12. However, practical experience differs. Students should decide which sum, 9 or 10, occurs more frequently when rolling the three dice.

Formula (1) can be used to generate dice roll outcomes. To begin, we create a table with three columns representing 5000 outcomes of rolls of the three dice. Then we add a new column, Sum, containing the sums of the scores on the dice. Considering the fact that operations with the data in the table change the values of the RAND function, it is necessary to transform generated data to constant values. Having selected the whole table by the key combination <Ctrl + \*>, we copy the table to the clipboard and then we paste the content of the clipboard as constant values to a new worksheet using the command <Paste Special>. In order to determine the frequencies of the individual sums, it is necessary to sort the table by values in the Sum column using the corresponding icon on the Standard Toolbar, and then use the command <Subtotal> from the <Data> menu. This command enables us to determine the numbers of successive equal values after each change in the Sum column. The dialogue window with the corresponding settings is displayed in Figure 4.

Figure 5 shows several rows of the ordered table after the application of

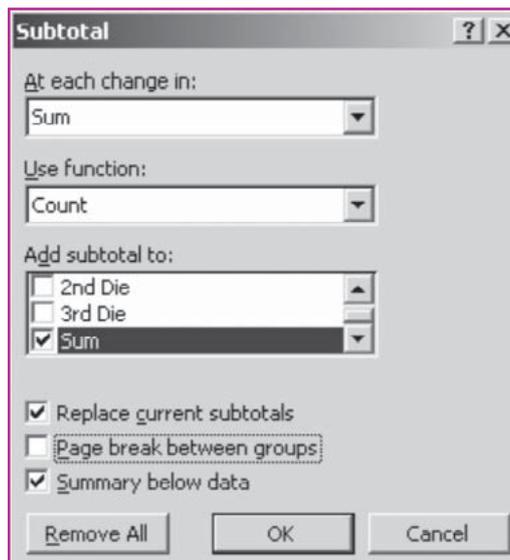


Figure 4. The determination of frequency of sums using the command Subtotal.

	1	2	3	A	B	C	D
	1	<b>1st Die</b>	<b>2nd Die</b>	<b>3rd Die</b>	<b>Sum</b>		
•	2	1	1	1	3		
•	3	1	1	1	3		
•	4	1	1	1	3		
•	5	1	1	1	3		
•	6	1	1	1	3		
•	7	1	1	1	3		
•	8	1	1	1	3		
•	9	1	1	1	3		
•	10	1	1	1	3		
•	11	1	1	1	3		
•	12	1	1	1	3		
-	13			<b>3 Count</b>	11		
•	14	1	1	2	4		

Figure 5. The part of the table after application of the command Subtotal.

the command <Subtotal>. The buttons 1, 2, and 3 in the left upper corner change the level of the overview of the calculated data in the table. Figure 5 shows the lowest level of the overview marked with the number 3. Level 2 displays the frequencies of individual sums and the total number of rolls. This summary table of the frequencies of individual sums (see Figure 6) can be used to create the corresponding column graph shown in Figure 7.

3 Count	11
4 Count	67
5 Count	145
6 Count	227
7 Count	345
8 Count	510
9 Count	561
10 Count	692
11 Count	601
12 Count	556
13 Count	501
14 Count	354
15 Count	214
16 Count	133
17 Count	60
18 Count	23

Figure 6. The frequency of individual sums.

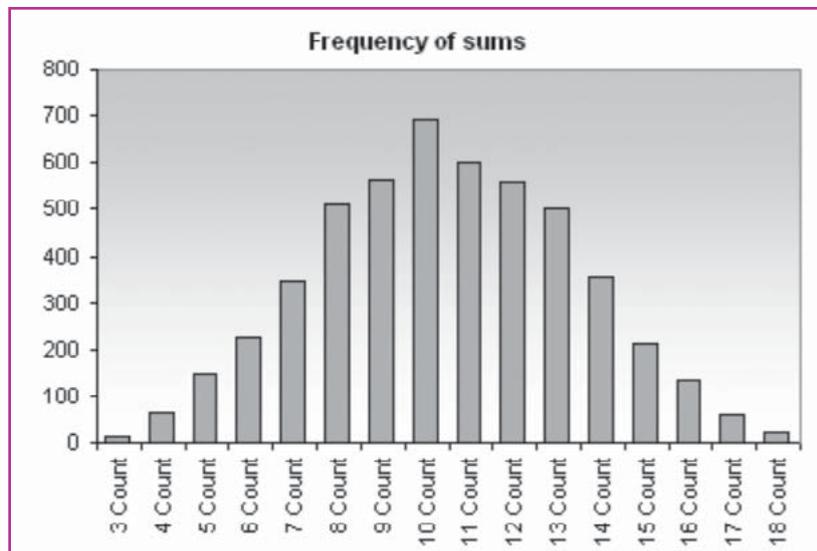


Figure 7. The frequency column graph of individual sums.

The sequence of steps described above (beginning with copying the random numbers using the command <Paste Special>) can be repeated to process another series of 5000 rolls of the three dice. The reason is that random numbers generated by means of the function RAND in the original table were changed when editing the copied table. Advanced students could record the above sequence of steps in the macro to automatise repeated dice rolling simulations.

Finally, the teacher should explain how the experiment results can be theoretically justified. Galileo Galilei was interested in the sums of the scores on the dice and explained why the sum 10 occurs more frequently than the sum 9 when rolling the three dice. The justification is based on the list of all possibilities of how the sums 9 and 10 can be obtained with the three dice. There are more possibilities for the sum 10 than for the sum 9 and this is the reason why the sum 10 occurs more frequently than the sum 9 in a large number of repeated random experiments. Analogously, median sums are the most probable when summing the scores on several dice, because there are more possibilities to obtain them. An increasing number of dice and an increasing number of repeated random experiments result in the fact that the distribution of frequencies of the score sums on the dice approaches to a normal probability distribution.

## References

- The Australian Association of Mathematics Teachers Inc. (2002, 2006). *Standards for excellence in teaching mathematics in Australian schools*. Adelaide: Author.
- National Council of Teachers of Mathematics [NCTM] (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- Walkenbach, J. (2003). *Excel 2003 Bible*. John Wiley & Sons.