THROWING A DIE WITH A COMPUTER ALGEBRA SYSTEM – IN THEORY AND PRACTICE

Karsten Schmidt
University of Applied Sciences Schmalkalden, Germany
kschmidt@fh-sm.de

Most management science and economics students attending an advanced statistics course appreciate a quick repetition of the concept of random variables. We describe the application of the computer algebra system Derive to the familiar "throwing a die" example which turns out to be a substantial time saver, e.g. since tedious manual calculations can be avoided. Our approach also allows straightforward application to the investigation of other gambling examples. Last but not least, using a computer algebra system also makes looking at the empirical side of gambling more comfortable by providing built-in functions for pseudo random number generation as well as for the computation of statistical parameters.

MOTIVATION

The idea for this paper emerged while teaching an undergraduate "Advanced Statistics" course at the Faculty of Management Science and Economics at the University of Applied Sciences Schmalkalden over several years. The main focus of this required second-year course (4 hours) is a thorough introduction to Linear Regression. The key prerequisites are "Introductory Statistics" (4 hours), and "Matrix Algebra" (2 hours). The latter course includes an introduction to, and intense use of, the computer algebra system Derive. The Derive license of our faculty covers all faculty computers (including two PC labs) as well as private PCs of all students (as long as they are enrolled in a faculty program).

At the beginning of the course students have to acquire some basic understanding of stochastic vectors and matrices, i.e., vectors and matrices containing random variables instead of real numbers. This includes rules on how the expected value of a stochastic matrix, or the dispersion matrix of a stochastic vector, are affected by certain transformations.

Since this portion of the course is pretty theoretical by nature students find it more difficult than other topics. Moreover, while students understand and can easily calculate the mean of a sample, and the majority of students have knowledge of the sample variance and its computation, many students do have problems when it comes to the application of random variables. Therefore, we were looking for a fairly simple example from the real world which

(1) can be used for the creation of a sample of values for which mean and sample variance can be calculated;
(2) can be used for the definition of an associated random variable for which expected value and variance can be calculated;
(3) allows us to demonstrate that the values of mean and sample variance are in general different from the values of expected value and theoretical variance, but that the former tend to get closer to the latter if the sample size is increased;
(4) and, last but not least, allows more or less meaningful transformations of the random variable in order to see how expected value and variance are effected in these cases.

One main advantage of using Derive for throwing a die empirically and theoretically is that everything can be done in the same (for our students: well-known) environment. Moreover, rolling a die, say, 10 times, writing down the scores, and calculating statistical parameters, sounds like an interesting activity, but doing it 1000 times certainly does not. For a comparable application of a computer algebra system in teaching statistics cf. Edwards (2004).

This paper is structured as follows: in the next section the throw of a single die is applied as an example for the definition of a random variable and hence covers the theory part of the paper by discussing points (2) and (4) of the above list. Section three contains the practice part by simulating series of throws of a single die with the help of a pseudo random number generator; it thereby discusses points (1) and (3) of the above list. After briefly extending the theory part by looking at other gambling examples in section four, some concluding remarks are given in the final section.

THROWING A DIE – THEORY

In teaching statistics, throwing a die is one of the most often adopted examples used to illustrate the concept of a random variable, as well as its expected value and variance (see, for example, Keller and Warrack, 1999, pp. 171-173, 263-267). This is so because from childhood on everybody is familiar with games involving dice. Applying the respective formulae for discrete random variables students can calculate the expected value of a random variable describing one throw of a single die (3.5), and its variance ($\frac{35}{12} \approx 2.92$).

We apply the computer algebra system Derive to demonstrate the calculation of these values and start by defining (cf. Figure 1)

- 2 functions for the calculation of the expected value, E(x,p), and the variance, VAR(x,p), of a random variable defined in 2 vectors x and p; note that x and p are the formal parameters;
- a function LT(x,β1,β2) which allows easy linear transformations of a vector x (containing the outcomes of a random variable) where β1 denotes a constant to be added to x and β2 denotes a factor by which x is to be multiplied; for example, y:=LT(x,2,3) results in a vector $y = 2 + 3x$;
- 2 vectors of dimension 6 which contain the 6 possible outcomes, xd, and their respective probabilities, pd; note that xd and pd are one pair of actual parameters (“d” stands for a single die; cf. section four for additional pairs of actual parameters).

#1: $\mathrm{E}(x,\ p) := \sum_{j=1}^{\mathrm{DIM}(x)} x_j \cdot p_j$

#2: $\mathrm{VAR}(x,\ p) := \sum_{j=1}^{\mathrm{DIM}(x)} (x_j - \mathrm{E}(x,\ p))^2 \cdot p_j$

#3: $\mathrm{LT}(x,\ \beta1,\ \beta2) := \beta1 \cdot \mathrm{VECTOR}(1,\ i,\ \mathrm{DIM}(x)) + \beta2 \cdot x$

#4: $xd := [1,\ 2,\ 3,\ 4,\ 5,\ 6]$

#5: $pd := \left[\frac{1}{6},\ \frac{1}{6},\ \frac{1}{6},\ \frac{1}{6},\ \frac{1}{6},\ \frac{1}{6}\right]$

#6: $[\mathrm{E}(xd,\ pd),\ \mathrm{VAR}(xd,\ pd)] = \left[\frac{7}{2},\ \frac{35}{12}\right]$

#7: $y := \mathrm{LT}(xd,\ 3,\ 1)$

#8: $y = [4,\ 5,\ 6,\ 7,\ 8,\ 9]$

#9: $[\mathrm{E}(y,\ pd),\ \mathrm{VAR}(y,\ pd)] = \left[\frac{13}{2},\ \frac{35}{12}\right]$

#10: $y := \mathrm{LT}(xd,\ 0,\ 2)$

#11: $y = [2,\ 4,\ 6,\ 8,\ 10,\ 12]$

#12: $[\mathrm{E}(y,\ pd),\ \mathrm{VAR}(y,\ pd)] = \left[7,\ \frac{35}{3}\right]$

Figure 1: First Screenshot of Derive Algebra Window

Students can then compute the expected value and variance of the single die random variable by using the respective functions. They can also investigate these values for different linear transformations of the single die outcomes, e.g. find out that

- adding a constant shifts the expected value by the same amount ($\mathrm{E}[a + X] = a + \mathrm{E}[X]$), but leaves the variance unchanged ($\mathrm{Var}[a + X] = \mathrm{Var}[X]$), or
- multiplying by a factor changes the expected value by this factor ($\mathrm{E}[aX] = a\mathrm{E}[X]$), and changes the variance by the squared factor ($\mathrm{Var}[aX] = a^2\mathrm{Var}[X]$).

THROWING A DIE – PRACTICE

We proceed by simulating several series of actual throws of a single die in Derive, based on a pseudo random number generator, and computing the mean and variance for each sample. The function D(n) generates an n-dimensional vector of random integers from the set $\{1, 2, 3, 4, 5, 6\}$. We use the built-in functions AVERAGE(x) and VARIANCE(x) to compute the arithmetic mean and (sample) variance of the elements of a vector x.

Students can now start to experiment with different sample sizes. Generating a sample of size 10, for example, could be done manually by actually throwing a die 10 times, writing down the results, and calculating their mean and variance. Students could then proceed to Derive, generate a random vector of size 10, print it on the screen, and then compute the statistics in Derive, which could in turn be checked manually. Since there is no point in watching increasingly large vectors on the screen when the sample size is increased, students quickly skip this step and compute only the statistics in Derive.

```
#1:   D(n) := VECTOR(RANDOM(6) + 1, i, n)
#2:   APPROX([AVERAGE(D(10)), AVERAGE(D(10)), AVERAGE(D(10))]) = [3.6, 4.7, 2.9]
#3:   APPROX([AVERAGE(D(1000)), AVERAGE(D(1000)), AVERAGE(D(1000))]) = [3.45, 3.578, 3.519]
#4:   APPROX([AVERAGE(D(100000)), AVERAGE(D(100000)), AVERAGE(D(100000))]) = [3.49537, 3.49879, 3.49821]
#5:   APPROX([VARIANCE(D(10)), VARIANCE(D(10)), VARIANCE(D(10))]) = [2.322222222, 3.833333333, 1.877777777]
#6:   APPROX([VARIANCE(D(1000)), VARIANCE(D(1000)), VARIANCE(D(1000))]) = [2.916747744, 2.911147147, 2.968727727]
#7:   APPROX([VARIANCE(D(100000)), VARIANCE(D(100000)), VARIANCE(D(100000))]) = [2.911643714, 2.909046338, 2.911424184]
```

Figure 2: Second Screenshot of Derive Algebra Window

Each time AVERAGE(D(n)) and VARIANCE(D(n)) are executed a new vector of n scores is generated for which the respective statistic is computed. Altogether, 18 different random vectors (of dimension 10, 1000, and 100000) are generated in Figure 2.

By comparing the sample means and variances with the expected value and variance of the random variable, students will see that they differ in general, but also that increasing the size of the sample generally results in diminishing deviations from the theoretical values, thereby illustrating the law of large numbers. See Mills (2002) for a critical overview of the literature concerning the use of computer simulation methods in statistics education.

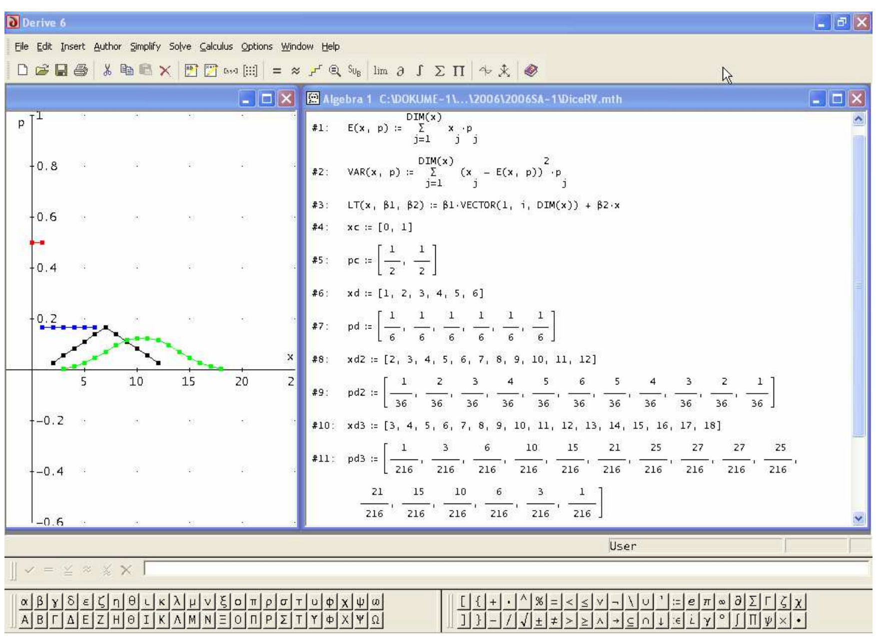

Figure 3: Screenshot of Derive Window

DICE AND COINS – MORE GAMBLING

The framework of section two, namely the two functions for expected value and variance of a discrete random variable, as well as the function for linear transformations of outcomes of a random variable, allow straightforward application to other random variables.

Figure 3 contains the definition of three more random variables in addition to the one from section two, as well as graphical representations of the distributions (actually bar charts) of all these random variables:

- the first pair of vectors (xc,pc) describes the toss of a coin, head coded as 1 and tail as 0;
- the second pair (xd,pd) is the above example of a throw of a single die;
- the third pair (xd2,pd2) describes the sum of scores of a simultaneous throw of 2 dice;
- the forth pair (xd3,pd3) describes the sum of scores of a simultaneous throw of 3 dice.

Students can then compute the expected value and variance of all these random variables by using the respective functions (cf. Figure 4). They could also investigate these values for different linear transformations of the outcomes.

#12: $$[E(xc,\ pc),\ VAR(xc,\ pc)] = \left[\frac{1}{2},\ \frac{1}{4}\right]$$

#13: $$[E(xd2,\ pd2),\ VAR(xd2,\ pd2)] = \left[7,\ \frac{35}{6}\right]$$

#14: $$[E(xd3,\ pd3),\ VAR(xd3,\ pd3)] = \left[\frac{21}{2},\ \frac{35}{4}\right]$$

Figure 4: Third Screenshot of Derive Algebra Window

CONCLUSION

It is a fact that a high percentage of management science and economics students attending an advanced statistics course are in need of a quick repetition of the concept of random variables, as well as their expected value and variance, to get them ready for things to come. For this we use the computer algebra system Derive since the students are familiar with it from a previous course and are allowed to use it on their private PCs under the license we have. The advantages of using a computer algebra system in this context include that

- students can perform tedious calculations of expected values and variances quickly and faultlessly based on functions provided;
- students can easily perform linear transformations of outcomes, and so check certain theorems on expected values and variances of random variables;
- it adds the illustrative power of graphs, thereby offering better understanding of the investigated probability distributions;
- it facilitates the obvious step to the empirical side of gambling, for example by easy simulation of throws of dice with increasing sample sizes using the built-in functions for pseudo random numbers, mean, and sample variance.

REFERENCES

Edwards, M.T. (2004). Emptying the bowl: An investigation of probability and mathematics using CAS as an inquiry-based counting tool. *International Journal for Technology in Mathematics Education*, 11, 91-100.

Keller, G. and Warrack, B. (1999). *Statistics for Management and Economics* (5th edition). Pacific Grove: Duxbury.

Mills, J. D. (2002). Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education*, 10(1).