SIMULATING EXPERIMENTS WITH THE GRAPHIC CALCULATOR TI83PLUS

Guido Herweyers
Katholieke Universiteit Leuven
Belgium

Simulating quickly a large number of repetitions of an experiment is a powerful tool in statistics. Simulations can be performed and explored individually by students in the classroom with graphics calculators like the TI83(Plus). Applications are estimating the probability of an event, predicting the profit in a game of chance, approximating the probability distribution of a random variable, drawing samples from a population, testing the quality of the random number generator. These applications will be illustrated with examples.

INTRODUCTION

When students have statistical software at their disposal in a graphic calculator, statistical tables become superfluous and applied statistics can be focussed on. It becomes easy to draw histograms and to study the influence of the class width in descriptive statistics. Simulations are important to explore the variability of samples and the connection with inferential statistics, and to estimate probabilities with relative frequencies. More information on the use of the graphic calculator TI83(Plus) in teaching statistics, can be found in the references. In the underlying text no attention is paid to the syntax of the calculator; the screens should be self-explaining.

ESTIMATING A PROBABILITY

Choose at random 5 different numbers from the set $\{1, 2, \ldots 40\}$. What is the probability P that there will be at least two neighbours (such as 21 and 22) in the chosen set?

Simulating the experiment several times results in:

Ignoring the 5 lists with equal numbers, we find 12 lists with neighbours on a total of 22 lists. This yields a first rough estimate $12/22 = 54.5\%$ for the probability P. Every student in the class-room can do 10 repetitions of the experiment, combining these results gives a new relative frequency as a (more reliable) approximation of P. A program to simulate the experiment N times can be written (see next page):

The exact probability is 42.7%. Remark that 100 simulations do not guarantee a better approximation than 50 simulations. To calculate the exact probability, imagine a sequence of marbles representing the sequence of the numbers 1 to 40. The 5 selected numbers correspond to black marbles and the other 35 numbers to white marbles. For the complement "no neighbours" of the event "at least two neighbours", we have to count the sequences with 5 separated black marbles between the 35 white marbles. To form such a sequence, we choose 5 places of the 36 possible places (to the left of each white marble and to the right of the last white marble) for the black marbles. Thus we find:

$$P(\text{at least two neighbours}) = 1 - P(\text{no neighbours}) = 1 - \frac{\binom{36}{5}}{\binom{40}{5}} = 0.427 .$$

PROGRAM: SIMUL
ClrHome
$0 \to T$
Prompt N
For (I , 1 , N)
Repeat sum (ΔList(L1) = 0) = 0   (ignore lists with equal numbers)
randInt (1 , 40 , 5) $\to$ L1
SortA (L1)
End
If sum (ΔList (L1) = 1) $\neq 0$ (are there neighbours in the sorted list?)
T+1 $\to$ T
End
Disp " THE ESTIMATED"
Disp " PROBABILITY IS "
Disp T / N

It is easy to transmit this program from one calculator to another with a cable. We can observe the students' results (taking $N = 50$) to study the variability of the relative frequencies and combine them to investigate that probability is the relative frequency in the long run.

A GAME OF CHANCE

Toss two dice and calculate the square of the difference of the numbers that appear. The result is paid out to you in Rand. But first you have to pay 7 Rand. Are you willing to play this game during the evening?

We simulate the game 500 times with the calculator to predict your profit. First observe the list of differences of the first and the second dice, then square this list and substract 7 to study your profit. Then calculate the total profit and the average profit. Four simulations of 500 games

gives an ominous average loss of 1.01 , 1.37, 1.16, 1.22 Rand.

Let X be the result of the first and Y the result of the second dice, we can calculate the expected value $E\left((X-Y)^2\right) = E\left(X^2 - 2XY + Y^2\right) = 2\left(E\left(X^2\right) - E(X)^2\right) = 2 \cdot Var(X) = \frac{35}{6} .$

The theoretical expected value of your profit is $35/6 - 7 = -1.17$ Rand. This means that your average long run loss will be 1.17 Rand. Indeed the results of our simulations give numbers in the neighbourhood of 1.17, a nice illustration of how predictable theoretical calculations can be!

EXPLORING DISTRIBUTIONS

Let X be a random variable, uniform distributed on the interval $[0,1]$. What is the distribution of X^2? As in the preceding examples, there are two possibilities to solve this problem. Simulation gives a quick and good idea of the distribution , theoretical calculation however yields the exact distribution of the random variable. Students should keep in mind the simulation technique to gain a better insight into the meaning of theoretical results.

The calculator generates a sample of 500 numbers taken from a uniform distribution and squares this list. An appropriate histogram of these data gives an idea of the distribution of X^2:
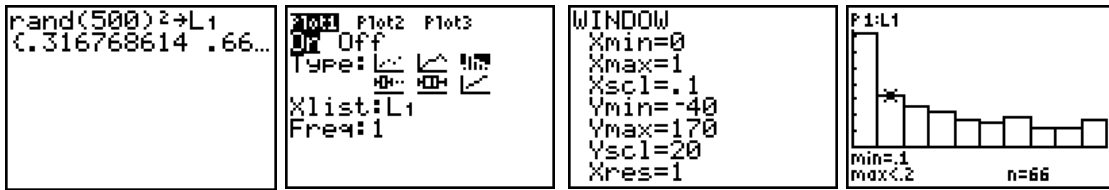

The cumulative distribution function $F(y)$ and the probability density function $f(y)$ of $Y = X^2$ are

$$F(y) = P(Y \le y) = P\left(X \le \sqrt{y}\right) = \int_0^{\sqrt{y}} 1 \cdot dx = \sqrt{y} \quad (0 \le y \le 1)$$

$$f(y) = F'(y) = \frac{1}{2\sqrt{y}} \quad (0 < y < 1)$$

To see the correspondence between the histogram and the density function $f(y)$ the vertical frequencies can be divided by $500 \cdot 0.1 = 50$ (the number of observations multiplied by the class width) to obtain the relative frequency density plot or, more easily, one can multiply $f(y)$ by 50:
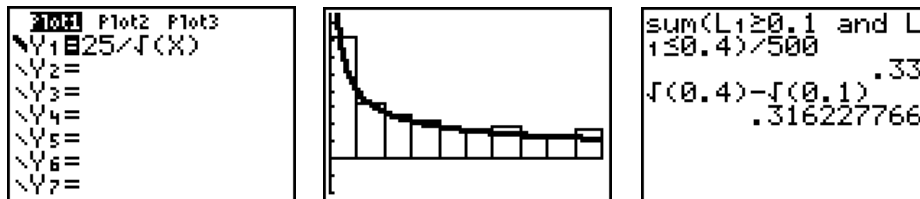

Observe that the relative frequency of the data between 0.1 and 0.4 is 0.33, a good approximation for the theoretical probability $P\left(0.1 \le X^2 \le 0.4\right) = 0.316$.

DRAWING SAMPLES FROM A DISTRIBUTION

Measure the length of the students in a college. The length of the girls is normally distributed with mean 172 cm and standard deviation 4.1 cm. The length of the boys is normally distributed with mean 178 cm and standard deviation 5.2 cm. Choose at random a girl and a boy from the students' population. Find the probability that the girl is smaller than the boy.

We simulate a sample of size 200 from the population of girls and the population of boys. We put the results into the lists L1 and L2. Then we verify how many times the girl is smaller than the boy:
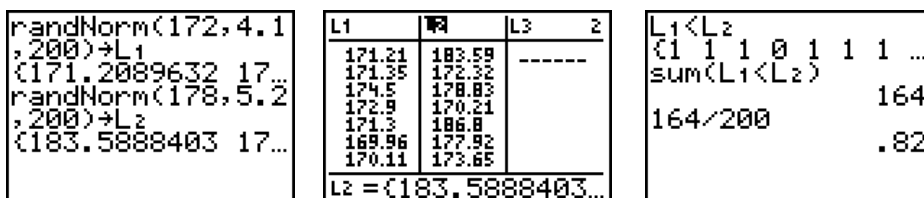
The estimated probability is 82%. We denote a normal distributed random variable X with mean μ and standard deviation σ as $X \sim N(\mu, \sigma)$. Let G be the length of the girl chosen at random and B the length of the boy, then we can investigate the distribution of the random variable $D = G - B$. The histogram of the 200 differences in list L3 = L1 - L2 indicate a normal

distribution for *D*, the mean -6.32 and the standard deviation 6.76 of these 200 data are estimations of the mean and the standard deviation of *D*.

Indeed, theory confirms that $D \sim N(172 - 178, \sqrt{4.1^2 + 5.2^2})$ or $D \sim N(-6, 6.62)$. The exact probability is $P(G < B) = P(G - B < 0) = P(D < 0) = 0.818$.
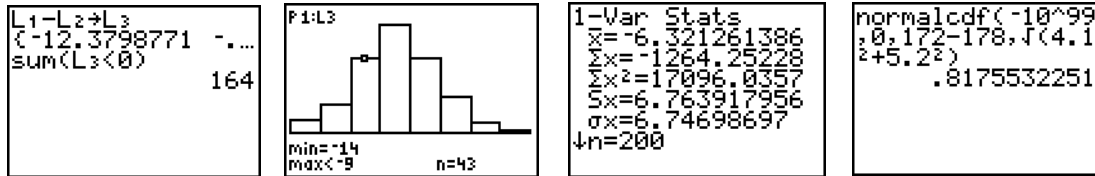
TESTING THE NUMBER GENERATOR

The function rand of the calculator generates a number from a uniform distribution on the interval $[0,1]$, at least it should be. Generating a sample of 100 data from this distribution yields 0.482 as the mean value. Is this a reason for looking for another number generator?

We test the null hypothesis $H_0 : \mu = 0.5$ against the alternative hypothesis $H_1 : \mu \neq 0.5$. Although the data are not from a normal population, the sample size 100 is large enough to conclude that the test statistic $T = \dfrac{\overline{X} - 0.5}{\dfrac{S}{\sqrt{100}}}$ nearly has the *t* distribution with 99 degrees of freedom. The p-value is 0.53, no reason to doubt the quality of the number generator!
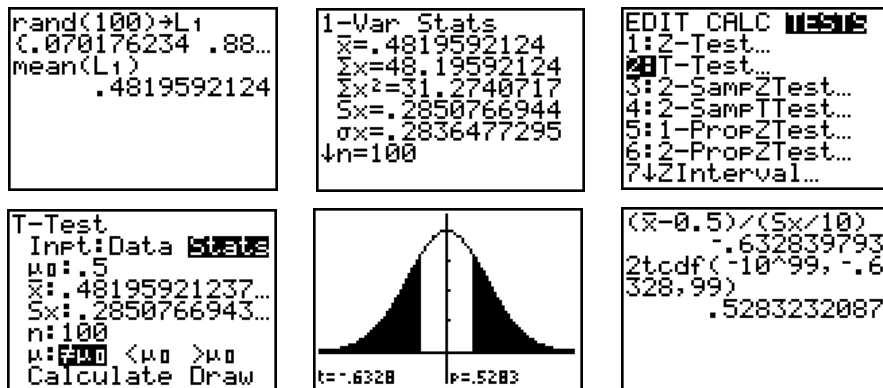
The last screen shows how students can verify the testing results generated by the calculator.

CONCLUSIONS

The graphic calculator TI83(Plus) illustrates the power of statistical software in the palm of the student's hand. He can investigate theoretical concepts by simulating experiments and drawing samples, making clear the connection between real data and theoretical calculations of parameters, density functions and probabilities. A tool to understand and enjoy statistics.

REFERENCES

Barret, G. (1997). *Statistics with the TI-83*. USA: Meridian Creative Group.

Herweyers, G., & Stulens, K. (2000). *Statistiek met een grafisch rekentoestel*. Leuven: Acco.

Yates, D., Moore, S., & McCabe, G. (1999). *The practice of statistics, TI-83 graphing calculator enhanced*. New York: W.H. Freeman and Company.