A Graphic Probability Calculator

Ricardo Vélez
UNED   (SPAIN)

Introduction

Basic courses on Probability and Statistics are taught nowadays to many undergraduate students in mathematics –of course– but also in physics, engineering, economics, medicine, biology, psychology, etc. Even, sometimes the main topics in this area are introduced at secondary school level.

The mathematical backgrounds of these kind of students are very different, but the main concepts that they must understand and learn how to use are always, more or less, the same. A brief list of these topics is enumerated in section 2.

As a consequence of this situation, one can observe that many students have often a lot of difficulties in using the more basic notions and bump into some frequent errors, due to a bad comprehension of their knowledge.

I think that the key to these teaching problems is that the students should be able to visualize the meaning of the main concepts and work graphically with them. This must allow the pupils, instead of remember a plain mathematical definition, to get a sharp mental picture about it, and to be able to perform correct calculations based on it.

Nowadays it is possible to reach this purpose by means of a computer program with suitable abilities. Such a program, that I call a *graphic probability calculator*, will be described in section 3.

The basic concepts

The main content in a basic course on Probability and Statistics deals with the idea of **random variables** and the principal ways to study their probabilistic behaviour.

Usually the random variables X are classified as **discrete** or **continuous,** according to the set of values that it can assume. While a discrete random variable takes values in a countable set of real numbers, a continuous variable can take any value in a finite or infinite interval of real numbers.

In both cases, the **distribution function** associated to X is defined as

$$F(x)=P\{X \text{ £ } x\} \qquad \text{for all } x \hat{\textbf{I}} R$$

but in the discrete case it is a step function with a jump in any possible value of X, while in the continuous case it is a continuous function. More precisely, the continuous distribution functions can be generally characterized by its **density function,** $f(x)$ such that

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

whereas the discrete distribution functions are characterized by means of their **probability function:**

$$p(x) = P\{X = x\} = F(x) - F(x^-)$$

For a detailed description of these topics see, for instance, [1] Chapter 2.

Some special discrete and continuous distributions have a great interest due to the fact that they frequently occur in applications. They are enumerated in the following table:

Discrete distributions	Continuous distributions		Distributions with statistical interest
Discrete uniform	Normal	Exponential	Chi-square
Binomial	Beta	Gamma	Student's T
Geometric	Laplace	Logistic	Snedecor's F
Hipergeometric	Lognormal	Uniform	Noncentral Chi-square
Negative binomial	Pareto	Cauchy	Noncentral T
Poisson	Weibull	Rayleigh	Noncentral F
	Triangular		

The description of almost all these distributions can be found in [1] Chapters 5 and 7; the most specific ones are described in [2] and [3].

For practical purposes, one needs usually to compute:

1. the value of the distribution function $F(x)$ for some fixed value of the variable x.
2. reciprocally, the value of x for which $F(x)$ has a given value p.
3. more generally, the probability of a given interval $(a , b]$:

$$P\{a < X £b\}= F(b) - F(a).$$

4. in statistical applications, such as the construction of **confidence intervals** or in **hypotheses testing** (see [1] Chap. 9-11), the shortest interval (a , b) with a given probability p.

Furthermore there is a large number of parameters associated to any distribution that resume some of its properties, such as

- position parameters: **mean, mode, median.**
- variability parameters: **variance, standard deviation, and coefficient of variation**.
- **skewness and kurtosis coefficients.**

The **moments** associated to any distribution are essential in the definition of most of the preceding parameters. In turn the moments of the distribution can be calculated from its **characteristic function** or, in the case of discrete distributions, from the **generating function**. All this concepts are explained in [1].

The Graphic Probability Calculator

As a help for the students which must confront with the foregoing basic concepts, I have designed a computer program whose aim is to provide them with a graphical and numerical experience about the main distributions listed in the previous section.

It has a graphical interface, shown at figure 1, in which one can select the desired distribution (by means of A).
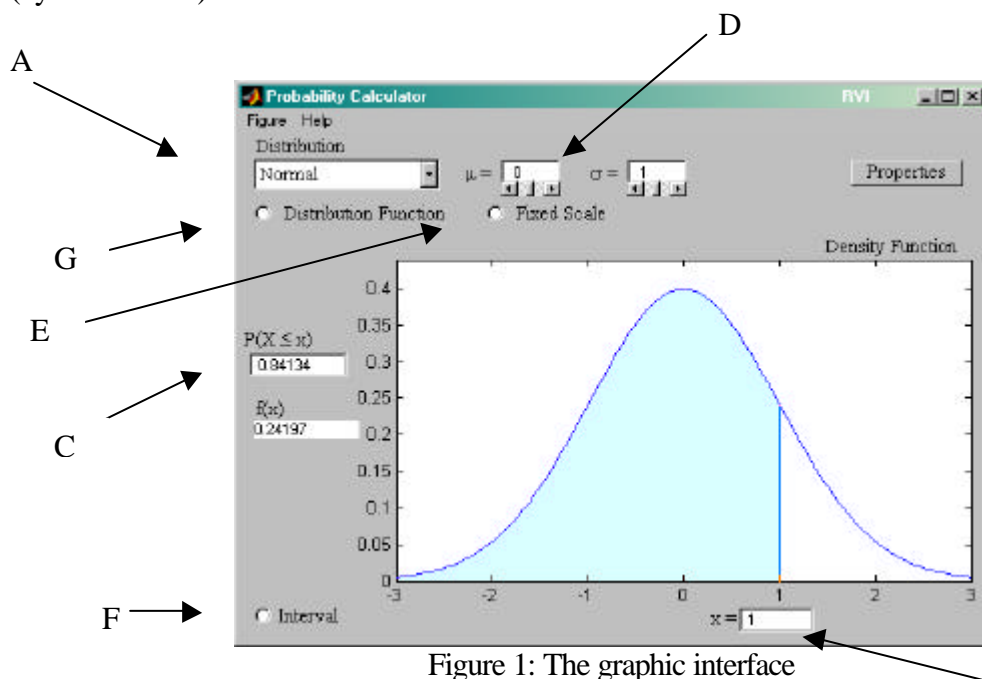

Figure 1: The graphic interface

This choice automatically displays the graph of the density function for continuous distributions or of the probability function in discrete cases (see figure 2). The variable value can be fixed in B, or with the mouse over the figure, and the corresponding value of the distribution function is shown in C. A value of p can also be given in C and the associated value of x is displayed in B.

The parameters of the observed distribution can be fixed in D, and the scale of the plot is automatically adapted in order to get the most representative figure.

Occasionally, if we want to observe the variation of the plot with the change of the parameter values, the scale may remain fixed by selecting option E. When this option is released, the scale automatically fits to the new parameter values.

Option F allows the alternating of the described situation with the one that gives the probability of a fixed interval *(a,x)*, as shown in figure 3. In this case, when a value of *p* is fixed in C, the program gives the shortest interval with that probability; this is extremely useful for statistical purposes.
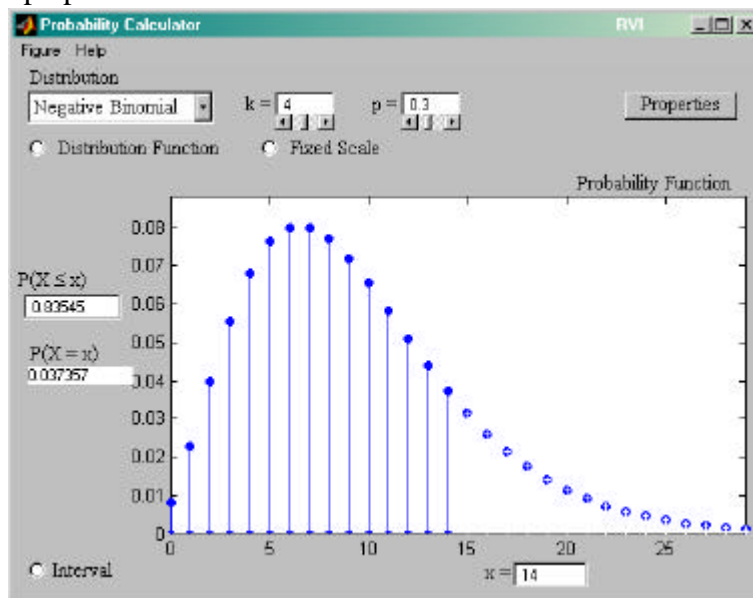

Figure 2: The probability function for a discrete distribution

Option G is designed to switch between the density/probability plot and the distribution plot, shown in figure 4, in the case of a continuous distribution. All the described functionalities remain working for this plot as well. A similar graph is obtained when the option "Interval" is chosen. The variable values can be fixed with the mouse under the graph, while the probability levels are fixed with the mouse above the graph.
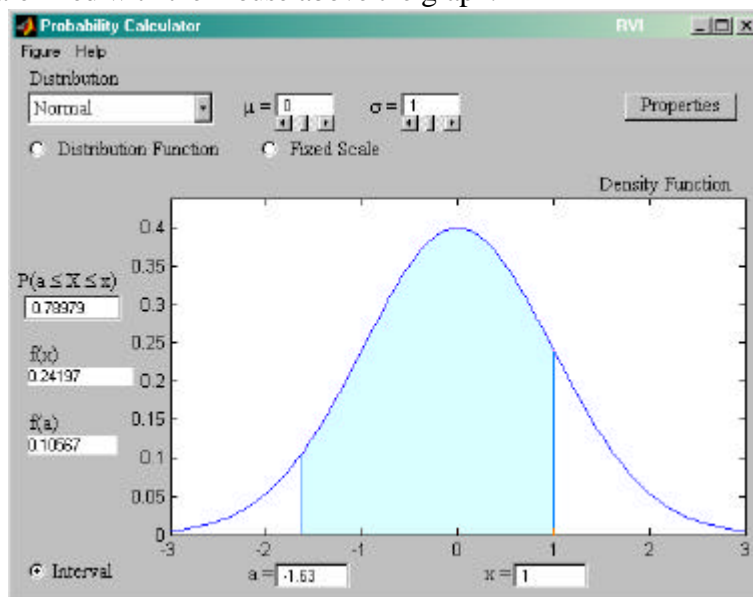

Figure 3: Probability of an interval

Finally, the button termed "Properties" displays a new window containing the main characteristics of the selected distribution, as figure 5 shows. The numerical values of these characteristics are automatically updated when the parameters of the distribution are changed.
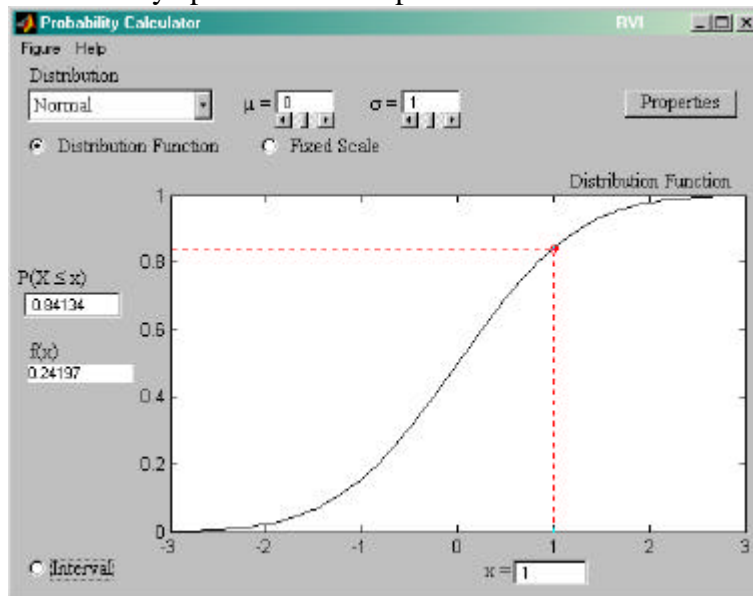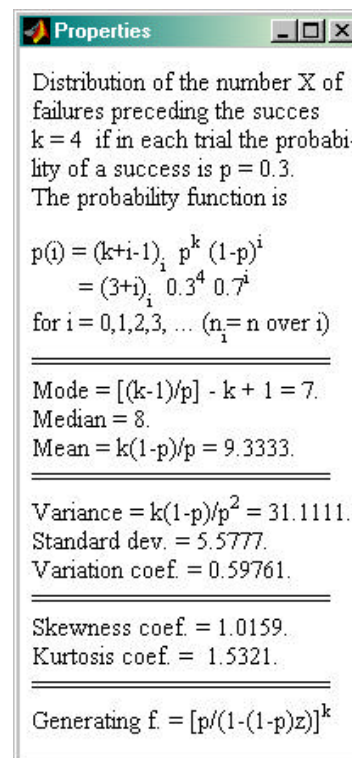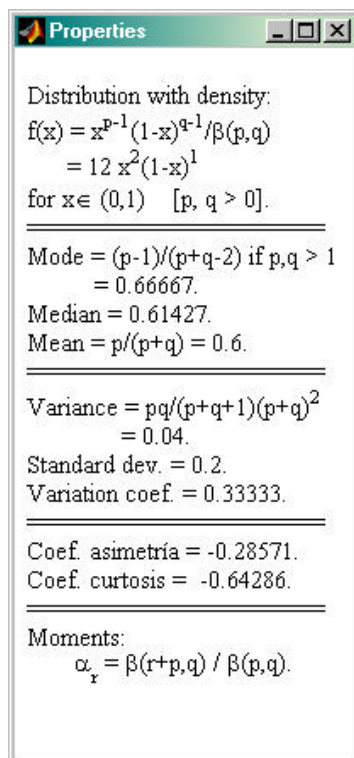

Figure 4: The distribution function plot.

Figure 5: Properties of a beta distribution and a negative binomial distribution

References:

[1] Rohatgi V. K. : *An introduction to probability theory and mathematical statistics,* John Wiley, 1976.
[2] Johnson, N.L. – Kotz, S. – Balakrisnan. N. : *Continuous univariate distributions*, John Wiley, 1994.
[3] Johnson, N.L. – Kotz, S. : *Distributions in statistics Vol. 2,* John Wiley, 1970.