BACK

Investigating Probability Distributions with DERIVE

by

T.A.Etchells

It is important that Derive is not just seen as a powerful tool for solving complex algebraic problems but also an environment in which students (and teachers!) can discover mathematics in an exciting and intellectually stimulating way. Certain mathematical topics and concepts are much easier to grasp and appreciate if it is possible to have a pictorial or graphical representation of them. There are many other mathematical graphics packages on the market, but Derive's superiority over these, is its ability to perform massive algebraic computations. It also has the facility for the student/teacher to program (declare) functions that will produce graphics that can be used to enhance the understanding of mathematics.

This work deals with the development of the Binomial , Normal and Poisson probability distributions (and their connections) with an upper sixth 'A' level Mathematics class (ages 18-19). Derive's easy handling of mathematics well above the standard of 'A'level was an important factor in this work, the students could develop ideas in a way that, previously ,would have been hindered by their (and their teachers) inability to perform certain calculations or operations.

The Binomial Distribution

The proof for the expressions of the *mean* and *variance* of a binomial distribution **B**(n,p) is somewhat confusing and laborious, leaving the average student perplexed and unconvinced. e.g. next, a typical text book proof.

$$\mu = E(x) = \sum_{r=0}^{n} r\binom{n}{r} p^r q^{n-r}$$

$$= \binom{n}{1} pq^{n-1} + 2\binom{n}{2} p^2 q^{n-2} + \cdots + r\binom{n}{r} p^r q^{n-r} + n\binom{n}{n} p^n$$

But

$$r\binom{n}{r} = r\frac{n!}{r!(n-r)!} = n\frac{(n-1)!}{(r-1)!(n-r)!} = n\binom{n-1}{r-1}$$

and $\therefore$,

$$\mu = npq^{n-1} + n\binom{n-1}{1} p^2 q^{n-2} + \cdots n\binom{n-2}{r-1} p^r q^{n-r} + \cdots np^n$$

$$= np\left\{q^{n-1} + \binom{n-1}{1} pq^{n-2} + \cdots + \binom{n-1}{r-1} p^{r-1} q^{n-r} + \cdots p^{n-1}\right\}$$

$$= np\,(p+q)^{n-1}$$

$$= np \qquad \text{since} \qquad p + q = 1$$

With the help of Derive's ***Calculus***, ***Sum*** facility, students can investigate the mean and variance of a binomial distribution for specific values of ***n***. The student should have developed some idea of a generalised binomial distribution from the consideration of simple Bernoulli trials.

The expression for general binomial probability can be authored and then used to build up further expressions. Students should be aware of how to find the expected value of a discrete distribution and author the generalised expected value using ***Calculus***, ***Sum***. Different values of ***n*** can be substituted into this expression via ***Manage***, ***Substitute***, the new expression can be simplified. Students can substitute many values of ***n*** into the general expression for the expectation, and deduce the mean (expected value) to be ***np*** (*fig. 1*), it is unfortunate that Derive (at the moment?) can not generalise for ***n***.

$$14:\quad r\left[\frac{n!}{(n-r)!\,r!}\,p^{r}\,(1-p)^{n-r}\right]$$

$$15:\quad \sum_{r=0}^{n} r\left[\frac{n!}{(n-r)!\,r!}\,p^{r}\,(1-p)^{n-r}\right]$$

$$16:\quad \sum_{r=0}^{10} r\left[\frac{10!}{(10-r)!\,r!}\,p^{r}\,(1-p)^{10-r}\right]$$

$$17:\quad 10\,p$$

$$18:\quad \sum_{r=0}^{67} r\left[\frac{67!}{(67-r)!\,r!}\,p^{r}\,(1-p)^{67-r}\right]$$

$$19:\quad 67\,p$$

```
COMMAND: Author Build Calculus Declare Expand Factor Help Jump soLve Manage
         Options Plot Quit Remove Simplify Transfer moVe Window approX
Enter option
Sub(15)                C:PROBTAE.MTH        Free:83%                Derive Algebra
```

fig. 1

Similarly $E(x^2)$ can be generalised by authoring the general term for this sum, substituting into the expression different values of ***n*** and generalising a result for $E(x^2)$ by inspection of the specific results. Hence the variance of the binomial distribution can be generalised for any ***n*** and ***p*** (see *fig.2*).

$$20:\quad r^{2}\left[\frac{n!}{(n-r)!\,r!}\,p^{r}\,(1-p)^{n-r}\right]$$

$$21:\quad \sum_{r=0}^{n} r^{2}\left[\frac{n!}{(n-r)!\,r!}\,p^{r}\,(1-p)^{n-r}\right]$$

$$22:\quad \sum_{r=0}^{10} r^{2}\left[\frac{10!}{(10-r)!\,r!}\,p^{r}\,(1-p)^{10-r}\right]$$

$$23:\quad 10\,p\,(9\,p+1)$$

$$24:\quad \sum_{r=0}^{67} r^{2}\left[\frac{67!}{(67-r)!\,r!}\,p^{r}\,(1-p)^{67-r}\right]$$

$$25:\quad 67\,p\,(66\,p+1)$$

```
COMMAND: Author Build Calculus Declare Expand Factor Help Jump soLve Manage
         Options Plot Quit Remove Simplify Transfer moVe Window approX
Compute time: 120.3 seconds
Simp(24)               C:PROBTAE.MTH        Free:86%                Derive Algebra
```

fig. 2

An alternative strategy to this method, is to exploit the ***Expand*** and ***Factor*** options by expanding the sum after substituting in a value for ***n*** and then factoring this expression. This shows a generalisation for the mean and variance that can be described against a formal proof to illuminate the meaning of each step of the proof. So that Derive does not do all the *hard* work automatically, re-author the binomial probability function with ***n***,***p*** and ***q***, where ***q=1-p***.

It is possible to ***Declare*** functions that will plot rectangles of specific widths and heights, and this can be used to display probability distributions graphically. The functions described below are written in full form at the end. If you wish to use these functions type them on the author line (or use a text editor), and save as a Derive (or .MTH) file for future use.

BINRECTS(n,p) will plot the binomial distribution in the form of rectangles of width 1. Strictly, a discrete distribution should be represented as straight vertical lines (this can be done by plotting **BINLINE(n,p)).** However, rectangles help to introduce the links between area and probability, and the importance of continuity correction when approximating a discrete to a continuous distribution . So that the graphics are plotted correctly, the user must change the ***State*** of the 2-D Plot window to ***Rectangular***, ***Connected*** and ***Small*** (this is done via the ***Options*** menu).

The function **BINRECTSBIG(n,p)** is used for large values of ***n***. If the value of ***n*** is bigger than 40, the vector produced can take a long time to compute and display, and many of the rectangles will not appear on the plot as they are far too small. **BINRECTSBIG(n,p)** plots the rectangles within 2.5 standard deviations of the mean.

To speed calculation of the vectors (in all cases), change the ***Notation*** to ***Decimal*** and ***Precision*** to ***Approximate*** ,and in each case the ***Significant Digits*** to 3 (via the ***Options*** menu).

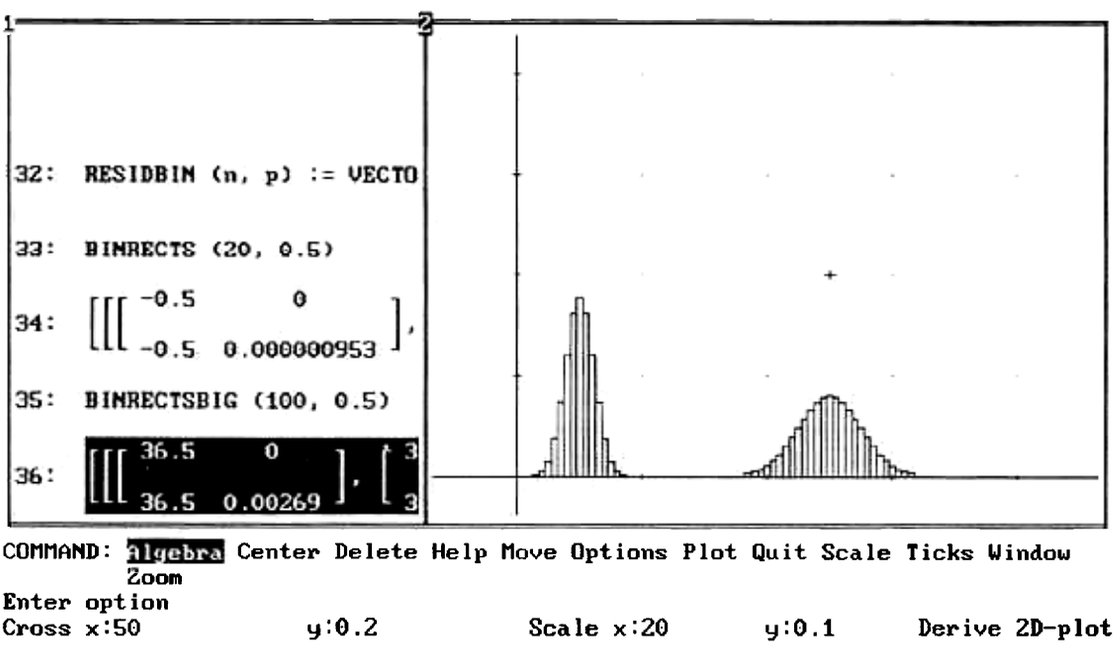


fig. 3

As the screen dumps (*fig.3* and *fig.4*) show, different distributions can be compared with each other with respect to a change in either p or n. Students can visualise for themselves how, with an increase in n, the distribution tends to a continuous bell like curve about its mean. Changing p to a lower value (e.g. 0.1) shows a skewed shape to the left with the peak at the mean (np) and by increasing n the distributions still tends to a bell shape about the mean (np).

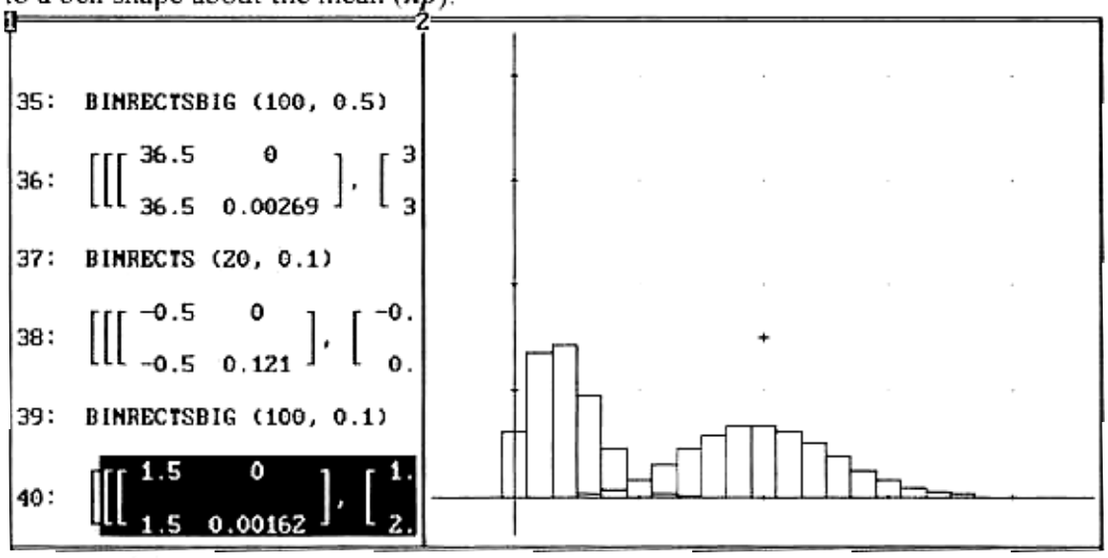


fig. 4

For illustrative purposes the normal approximation of the binomial distribution can be plotted over the rectangles to see how close the approximation is for different combinations of n and p, and decide for themselves the criteria for which normal distribution is a good approximation of the binomial distribution.

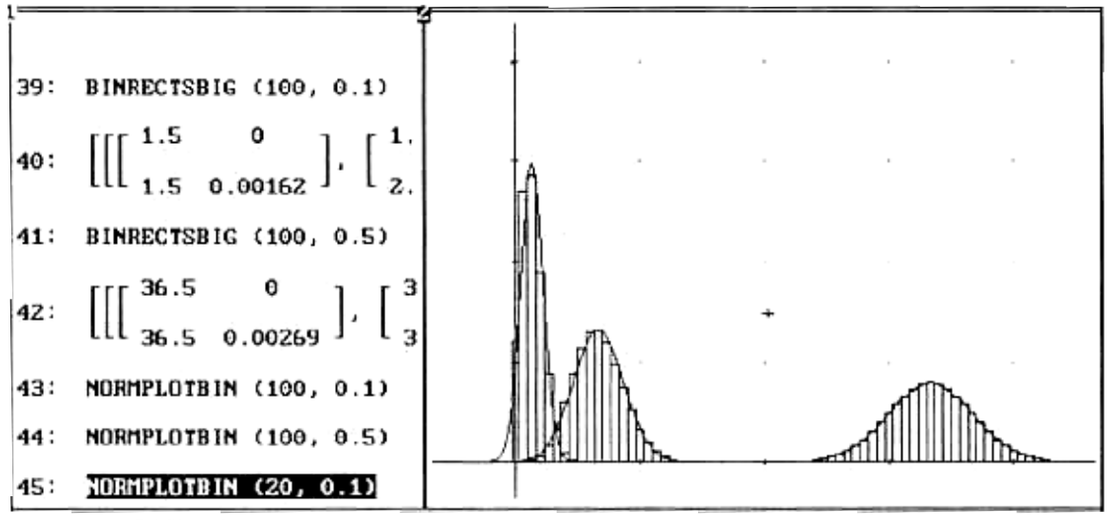
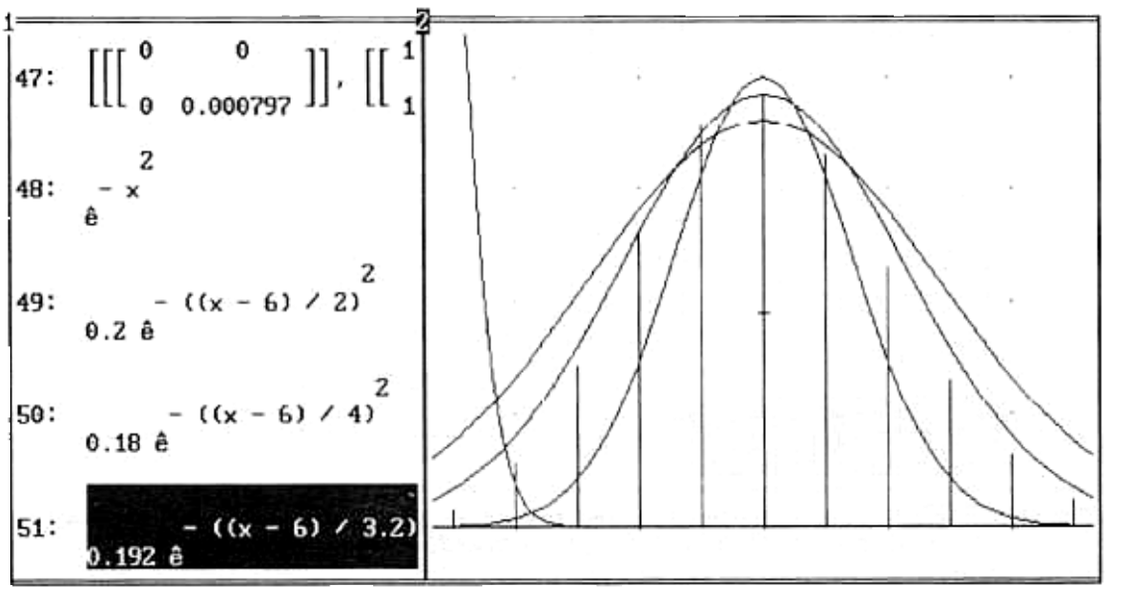


fig. 5

Using **BINLINE(n,p)** students can try to fit a curve over the endpoints of the binomial distribution. Being given a suitable starting point, by the teacher, and using appropriate stretches and transformations a function can be found that approximates to the binomial distribution (*fig. 6*). This is a good starting point for the development of the normal distribution



COMMAND: Algebra Center Delete Help Move Options Plot Quit Scale Ticks Window Zoom
Enter option
Cross x:6 y:0.0946428 Scale x:2 y:0.05 Derive 2D-plot

fig. 6

The Poisson Distribution

The Poisson distribution can be used as an approximation to the binomial distribution in certain circumstances, students can discover for themselves the criteria for a good approximation. **POISRECT**(μ) plots the rectangles for the Poisson distribution and by selecting different binomial distributions with the same mean as the Poisson distribution. An example of a mean (μ) of 10 is demonstrated in *fig. 7*. Although, in black and white, it is hard to distinguish between the different distributions , each distribution can be plotted in a different colour. When plotted in colour, the students can easily identify when the binomial distribution is a good approximation to the Poisson distribution. (To plot each distribution in a different colour use ***Options***, ***Colour*** , ***Next plot***, ensuring that ***Auto*** is set to ***No***.)

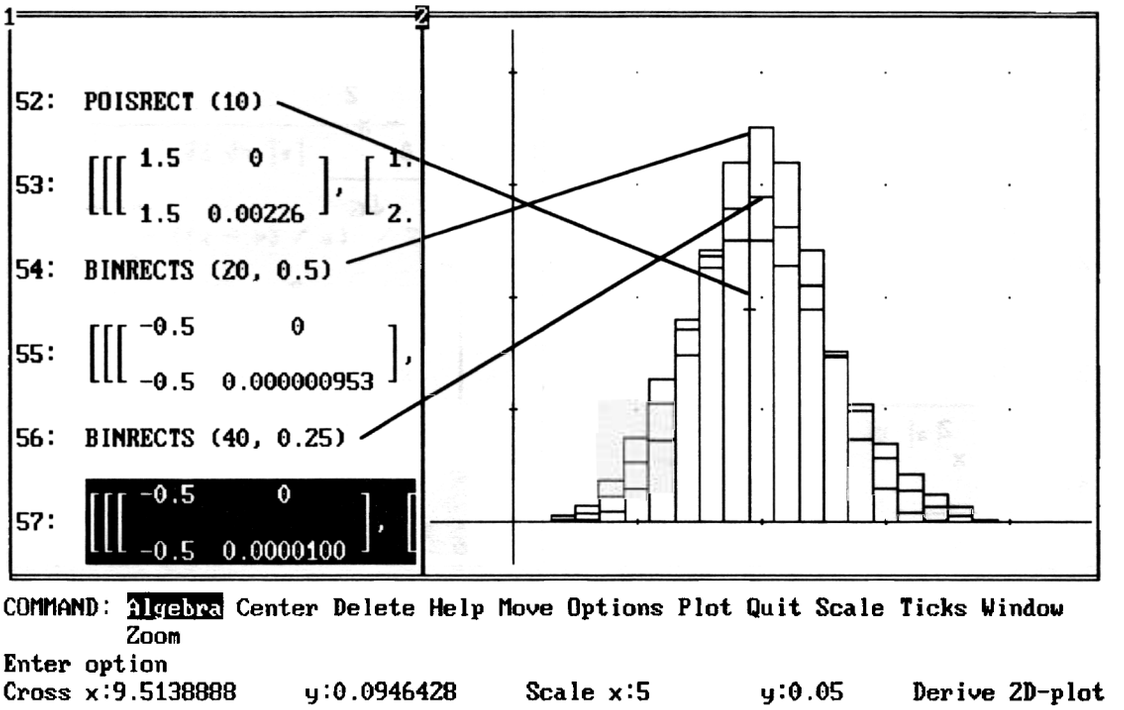


fig. 7

The Normal Distribution

An analytical investigation into the normal distribution is very difficult at this level as the integration is well beyond the scope of the course. However, to just present the probability density function is unhelpful, uninteresting and unmemorable. It also re-enforces passive, unquestioning mathematics that does not involve active participation by the learner. Derive enables the student (with guidance) to develop a function that is bell shaped and has a mean 0 and standard deviation 1. The aim of the lesson is to develop the p.d.f. not to learn complicated integration techniques, and as such Derive is the tool that performs the integration. The student uses Derive as a 14 year old uses a scientific calculator to solve a simple trigonometry problem.

Certain questions, that were previously difficult to answer, can be investigated with relative ease, e.g. why $e^{-x^2/2}$ is used instead of e^{-x^2} (*fig. 8* and *fig.9*).

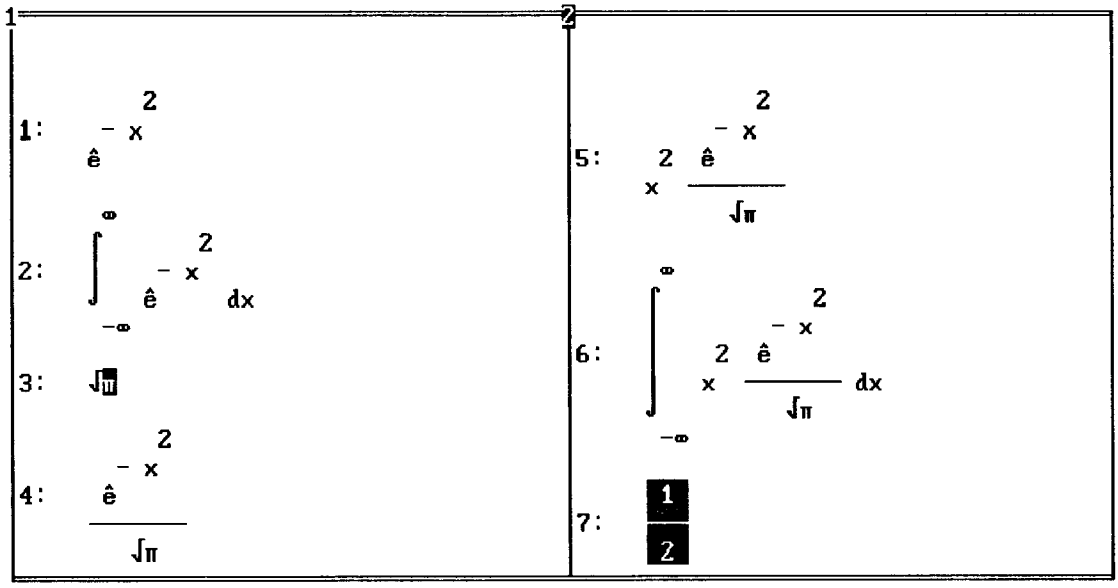


fig. 8

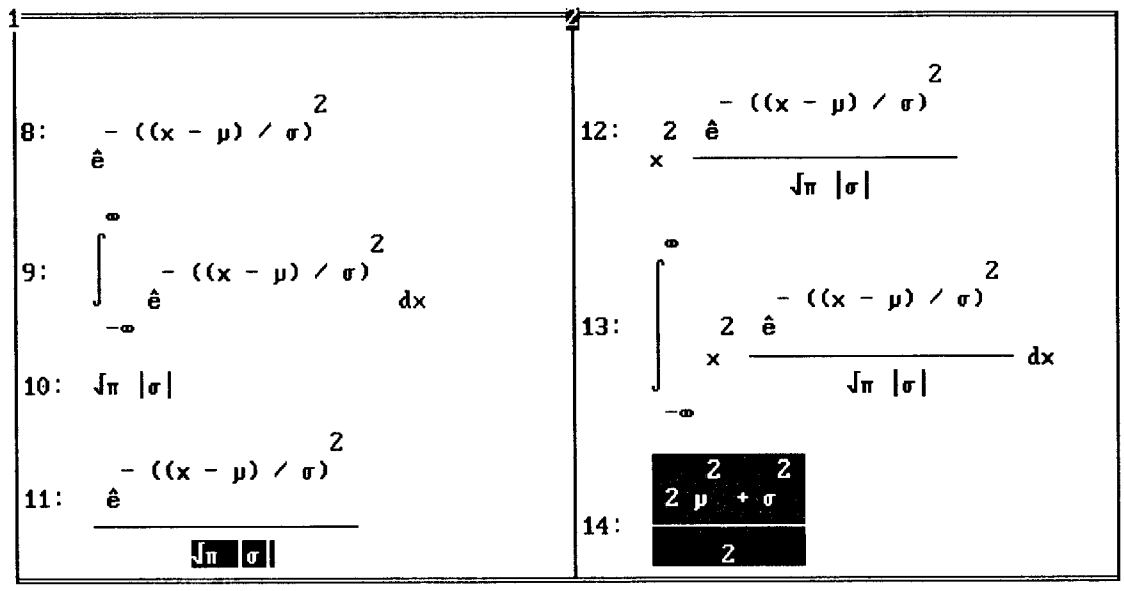


fig. 9

Students can develop for themselves the general p.d.f. for the Normal Distribution and prove that it has a mean μ and standard deviation σ (*fig. 10*).

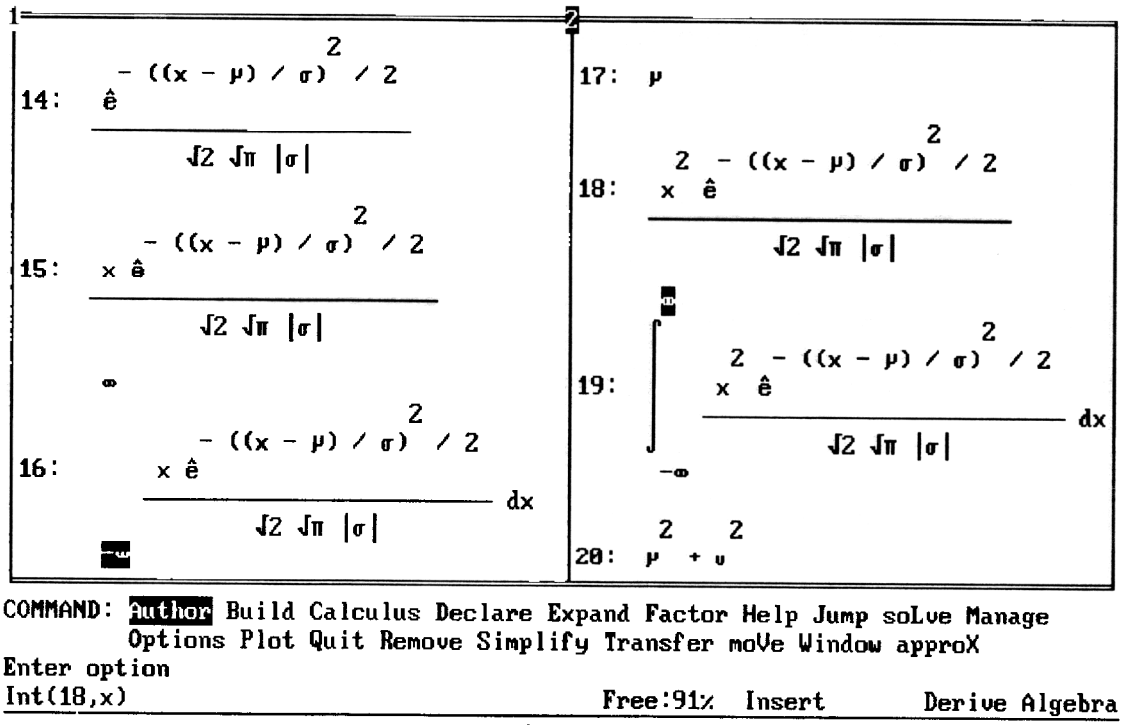


fig. 10

Derive is a '*box of delights*' that students can use to explore, discover, investigate and visualise mathematics. One of Derive's real strengths is that it was developed to do mathematics, not to *teach* mathematics. As such, Derive has mathematical power far beyond what is required for 16-19 year old students, this power can be harnessed to help students understand concepts and skills in ways that were not possible before. This new tool must make educationalists re-assess *what* mathematics is taught and *when* it is taught.

Derive is probably the most important tool for teaching mathematics, at the 16-19 age level, since the invention of pencil and paper. It is, indeed, an exciting time for teachers and students of mathematics.

Description of Derive Functions described in this chapter

$$\mathrm{RECT}(a,b,m,n) := \left[\begin{bmatrix} a & b \\ a & b+m \end{bmatrix}, \begin{bmatrix} a & b+m \\ a+n & b+m \end{bmatrix}, \begin{bmatrix} a+n & b+m \\ a+n & b \end{bmatrix}\right]$$

$$\mathrm{VLINE}(a,b,h) := \left[\begin{bmatrix} a & b \\ b & b+h \end{bmatrix}\right]$$

$$\mathrm{BINRECTS}(n,p) := \mathrm{VECTOR}\left[\mathrm{RECT}\left[\mathrm{j}-0.5, 0, \frac{\mathrm{n}!}{(\mathrm{n}-\mathrm{j})!\mathrm{j}!}\mathrm{p}^{\mathrm{j}}(1-\mathrm{p})^{\mathrm{n}-\mathrm{j}}, 1\right], \mathrm{j}, 0, \mathrm{n}\right]$$

$$\mathrm{BINRECTSBIG}(n,p) := \mathrm{VECTOR}\left[\mathrm{RECT}\left[j-0.5, 0, \frac{n!}{(n-j)!\,j!}p^{j}(1-p)^{n-j}, 1\right], j, \mathrm{FLOOR}(np-2.5\sqrt{np(1-p)}), \mathrm{FLOOR}(np+2.5\sqrt{np(1-p)})+1\right]$$

$$\mathrm{BINLINE}(n,p) := \mathrm{VECTOR}\left[\mathrm{VLINE}\left[j-0.5, 0, \frac{n!}{(n-j)!\,j!}p^{j}(1-p)^{n-j}, 1\right], j, 0, n\right]$$

$$\mathrm{POISRECT}(\mu) := \mathrm{VECTOR}\left[\mathrm{RECT}\left[j-0.5, 0, \frac{e^{-\mu}\mu^{j}}{j!}, 1\right], j, \mathrm{FLOOR}(np-2.5\sqrt{\mu}), \mathrm{FLOOR}(np+2.5\sqrt{\mu})+1\right]$$