ANNALS OF THE FACULTY OF ENGINEERING HUNEDOARA
2006, Tome IV, Fascicole 1, (ISSN 1584 – 2665)
FACULTY OF ENGINEERING HUNEDOARA,
5, REVOLUTIEI, 331128, HUNEDOARA

CONSIDERATIONS ABOUT POLYNOMIAL PROBABILITY DENSITY OF 3rd DEGREE

[1.] MAKSAY Ştefan, [1.] STOICA Diana, [2.] MAKSAY Elisabeta

[1.] UNIVERSITY "POLITEHNICA" OF TIMISOARA,
FACULTY OF ENGINEERING OF HUNEDOARA
[2.] HIGHSCHOOL "SIGISMUND TODUȚĂ" OF DEVA

ABSTRACT:
The present note tries to use the results of mathematic shaping to determine probability density of polynomial regression of 3rd degree. After the determination of a regression curve of 3rd degree, aided by the method of the smallest squares, it will be selected from this curve a domain of definition in which this function will satisfy the conditions of being a probability density.

KEYWORDS:
Polynomial regression, probability density, correlation coefficient.

1. DETERMINE THE REGRESSION CURVE

For typifying we will consider the following dates, in which x represent the independent variable and y the dependent variable.

$Tab^T =$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.663	1.791	1.81	1.864	1.905	1.912	1.976	2.009	2.111	2.135	2.162	2.172	2.183	2.197	2.338
2	0.334	0.38	0.314	0.306	0.361	0.36	0.29	0.314	0.244	0.268	0.211	0.262	0.261	0.259	0.226

To determine the regression curve of 3rd degree (g=3) will use the following MathCAD program [1]

$ORIGIN \equiv 1$

$g := 3 \qquad n := \mathrm{length}(x) \qquad i := 1..n$

$xm := \min(x) \qquad xM := \max(x)$

$ym := \mathrm{mean}(y) \qquad ym = 0.293$

$j := 1..2\cdot g + 1$

$$S_j := \sum_i (x_i)^{2\cdot g - j + 1}$$

$$S^T = \left(1.128\times 10^3 \;\; 539.048 \;\; 259.441 \;\; 125.756 \;\; 61.414 \;\; 30.228 \;\; 15\right)$$

$k := 1..g + 1 \qquad v := 1..g + 1 \qquad m_{k,v} := S_{v+k-1}$

$$m = \begin{pmatrix} 1.128\times 10^3 & 539.048 & 259.441 & 125.756 \\ 539.048 & 259.441 & 125.756 & 61.414 \\ 259.441 & 125.756 & 61.414 & 30.228 \\ 125.756 & 61.414 & 30.228 & 15 \end{pmatrix}$$

$$TL_k := \sum_i y_i \cdot (x_i)^{g-k+1} \qquad TL^T = (35.429\ \ 17.521\ \ 8.734\ \ 4.39)$$

resulting the coefficients of polynomial correlation

$$co := m^{-1} \cdot TL \qquad co^T = (1.554\ \ -9.52\ \ 19.103\ \ -12.253)$$

and also the regression curve from equation

$$y = 1.554 \cdot x^3 - 9.52 \cdot x^2 + 19.103 \cdot x - 12.253$$

which in the nodes has this values

$$kk := 1..\, g+1 \qquad Su_i := \sum_{kk} co_{kk} \cdot (x_i)^{g-kk+1}$$

$Su^T =$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.334	0.35	0.349	0.341	0.332	0.33	0.311	0.301	0.266	0.259	0.251	0.249	0.246	0.242	0.229

For the polynomial regression of 3rd degree will be obtained the following values for the correlation coefficient and respectively for the departure from the regression curve [2]

$$r := \sqrt{1 - \frac{\sum_i (y_i - Su_i)^2}{\sum_i (y_i - ym)^2}} \qquad r = 0.8752321444$$

$$St := \sqrt{\frac{1}{n} \cdot \left[\sum_i (y_i - Su_i)^2\right]} \qquad St = 0.024 \quad .$$

For this modeling we have the next graphic

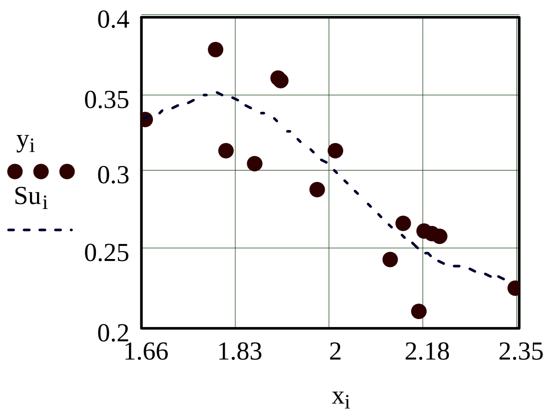

FIGURE 1. Polynomial regression of 3rd degree and the distribution of experimental points

Next we will give attention to a domain on which to choose the expression of probability density.

Distribution density must fulfill the conditions:

$$f(x) \geq 0, \qquad \int_{-\infty}^{\infty} f(x)\,dx = 1.$$

Because

$$Int := \int_{xm}^{xM} \sum_{kk} co_{kk} \cdot (x)^{g-kk+1}\,dx \qquad Int = 0.2$$

using the next program we will determine the limits of a definition domain for a positive function, restriction of the regression polynomial, so that the integral on this range will be equal with the unit

$$var1 := 0.3 \qquad var2 := 0.8$$

Given

$$\int_{xm-var1}^{xM+var2} \sum_{kk} co_{kk} \cdot (x)^{g-kk+1}\,dx = 1$$

$$tvar := Find(var1, var2) \qquad tvar = \begin{pmatrix} 0.3 \\ 0.864 \end{pmatrix}$$

We will obtain in this way the values

$$tvar_1 = 0.3 \qquad tvar_2 = 0.864$$

for which

$$\int_{xm-tvar_1}^{xM+tvar_2} \sum_{kk} co_{kk} \cdot (x)^{g-kk+1}\,dx = 1$$

where

$$xm - tvar_1 = 1.363 \qquad xM + tvar_2 = 3.202$$

In conclusion, the expression of probability density of 3rd degree is

$$f1(x) := if\left[xm - tvar_1 < x < xM + tvar_2, \sum_{kk} co_{kk} \cdot (x)^{g-kk+1}, 0\right]$$

and it's graphic is show in FIGURE 2 with the help of the adjoining program

$$nrnod := 500 \qquad is := 1..nrnod$$

$$xv_{is} := xm - (tvar)_1 - .7 + \frac{is - 1}{nrnod - 1} \cdot \left[xM + tvar_2 + .7 - \left(xm - tvar_1 - .7\right)\right]$$

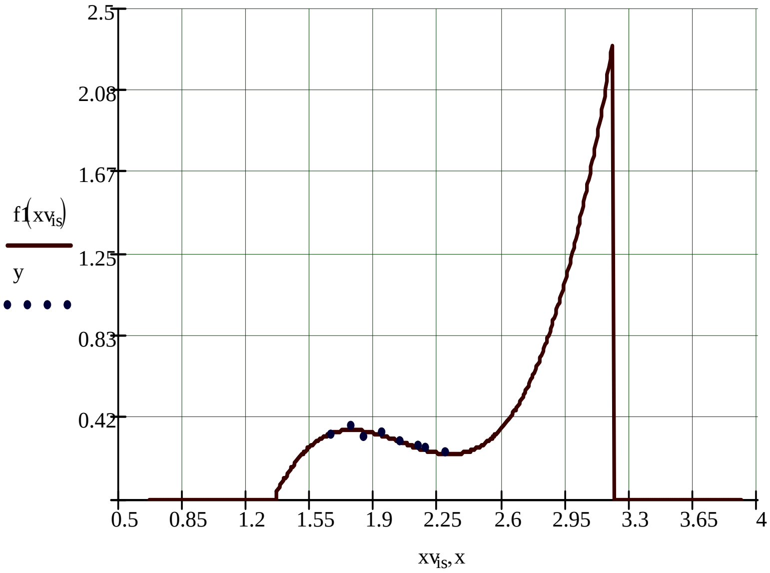

FIGURE 2. Polynomial probability density of 3rd degree and the distribution of experimental points

The expression of the repartition function is

$$\mathrm{Ftrunc3}(x) := \int_{xm-tvar_1}^{x} \mathrm{if}\left[xm - tvar_1 < u < xM + tvar_2, \sum_{kk} co_{kk}\cdot(u)^{g-kk+1}, 0\right] du$$

and it's graphic is shown in FIGURE 3.

$$\mathrm{wFtrunc3}_{is} := \mathrm{Ftrunc3}\left(xv_{is}\right)$$

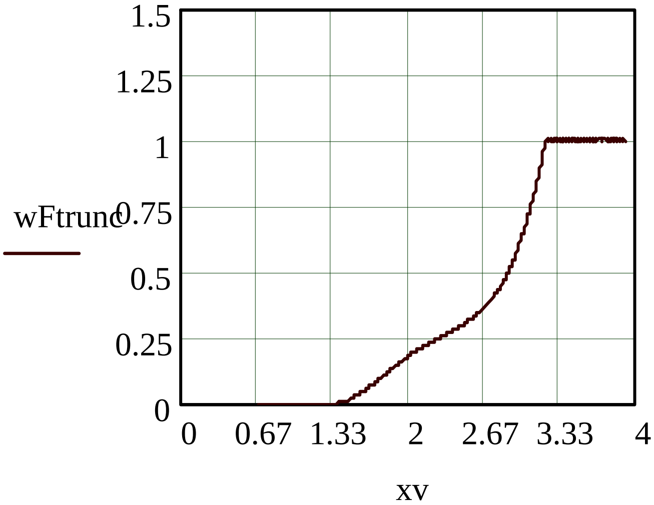

FIGURE 3. Repartition function for regression polynomial of 3rd degree

The expression of the characteristic function is

$$ex(t) := \int_{-1}^{6} e^{i\cdot t\cdot x}\cdot\left[\mathrm{if}\left[xm - tvar_1 < x < xM + tvar_2, \sum_{kk} co_{kk}\cdot(x)^{g-kk+1}, 0\right]\right] dx$$

and it's graphic is shown in Figure 4 respectively in Figure 5.

$t := -1, -0.9 .. 6$

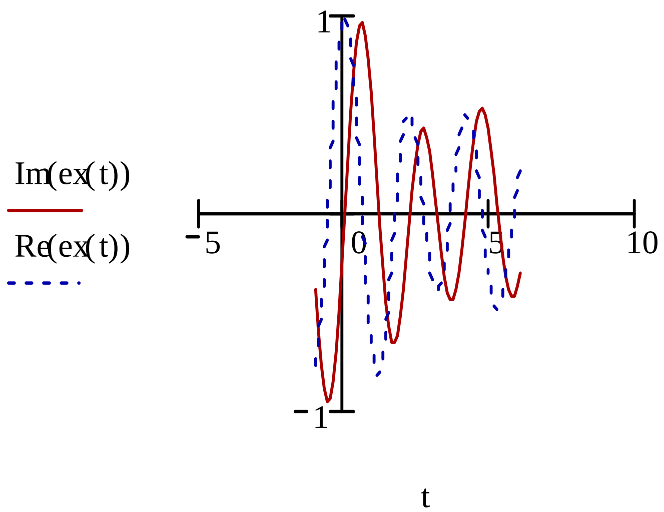

FIGURE 4. The real and imaginary part of the characteristic function for the probability density of 3rd degree

$t := -10, -9.9 .. 10$

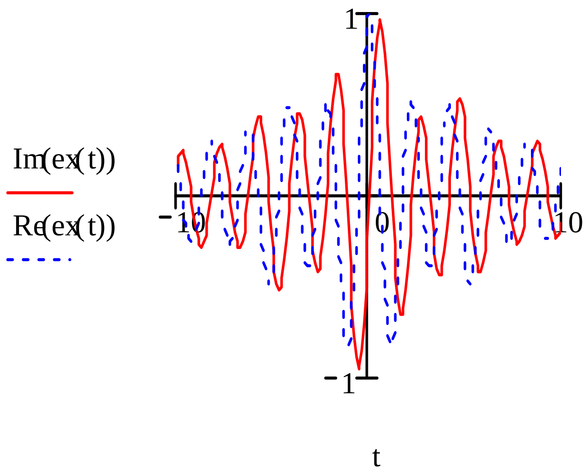

FIGURE 5. The real and imaginary part of the characteristic function for the probability density of 3rd degree on extended domain.

2. CONCLUSIONS

Because any of the classic probability densities admits a polynomial development, results that this method represents a unique way to treat a modeling of some experimental dates.

We specify the fact that, in generally we are not interested in modeling on relative big ranges, so the modeling is useful, for example, on a range centered in mean value of a independent variable and the length until the 3rd time standard deviation.

We specify also the fact that the condition that the integral from probability density must be equals with the unit, admission much more solutions regarding of the pair of the integrating limits (from which one can be predefine). All the obtained densities modeling the experimental domain dates identically.

BIBLIOGRAPHY

[1.] JALOBEANU C., RASA I., Mathcad. Probleme de calcul numeric şi statistic, Editura Albastră, 1995
[2.] MAKSAY, Şt., STOICA D., Calculul probabilităţilor, Editura Politehnica, Timişoara, 2005.
[3.] MAKSAY, Şt., STOICA D., Considerations on Modeling Some Distribution Laws, Applied Mathematics and Computation, AMC 10008, 2005 (in press).