JOURNAL OF ENGINEERING annals OF FACULTY OF ENGINEERING HUNEDOARA

DENSITY OF POLYNOMIAL REPARTITIONS

MAKSAY Stefan, STOICA Diana

"Politehnica" University of Timisoara, Faculty of Engineering of Hunedoara, ROMANIA

Abstract:
The present note tries to use the results of mathematic shaping to determine some probability density of polynomial forms. After the determination of a regression curve by a degree previously given, 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:
Probability density, Density of polynomial repartitions, coefficient of correlation

1. PRELIMINARIES

The present note tries to use the results of mathematic shaping to determine some probability density of polynomial forms.

For example we will consider the following dates, in which the first line represent the independent variable x and second line represent the dependent variable y.

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	2.238
0.334	0.266	0.314	0.306	0.361	0.36	0.29	0.44	0.171	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]

$\mathrm{ORIGIN} \equiv 1 \qquad \mathrm{TOL} := 10^{-5} \quad g := 3$

$x^T =$

	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	2.238

$y^T =$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.334	0.266	0.314	0.306	0.361	0.36	0.29	0.44	0.171	0.268	0.211	0.262	0.261	0.259	0.226

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

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

$ym := \mathrm{mean}(y) \qquad ym = 0.289$

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

$$S_j := \sum_i (x_i)^{2\cdot g - j + 1} \qquad S^T = \begin{pmatrix} 1.09\times 10^3 & 525.332 & 254.647 & 124.186 & 60.957 & 30.128 & 15 \end{pmatrix}$$

$k := 1..g + 1 \qquad v := 1..g + 1$

$$m_{k,v} := S_{v+k-1} \qquad m = \begin{pmatrix} 1.09\times 10^3 & 525.332 & 254.647 & 124.186 \\ 525.332 & 254.647 & 124.186 & 60.957 \\ 254.647 & 124.186 & 60.957 & 30.128 \\ 124.186 & 60.957 & 30.128 & 15 \end{pmatrix}$$

$$TL_k := \sum_i y_i \cdot (x_i)^{g-k+1} \qquad TL^T = (34.754 \;\; 17.235 \;\; 8.607 \;\; 4.329)$$

$co := m^{-1} \cdot TL$

resulting the coefficients of polynomial correlation

$$co^T = (-0.843 \;\; 4.14 \;\; -6.637 \;\; 3.778) \tag{1}$$

and also the regression curve equation is

$$y = -0.843 \cdot x^3 + 4.14 \cdot x^2 - 6.637 \cdot x + 3.778 \tag{2}$$

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	0.311	0.325	0.327	0.329	0.328	0.327	0.321	0.315	0.282	0.271	0.257	0.251	0.244	0.236	0.206

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

$$r3 := \sqrt{1 - \frac{\sum_i (y_i - Su_i)^2}{\sum_i (y_i - ym)^2}} \qquad r3 = 0.615 \;, \tag{3}$$

$$St3 := \sqrt{\frac{1}{n} \cdot \left[\sum_i (y_i - Su_i)^2\right]} \qquad St3 = 0.051 \;. \tag{4}$$

Next we will give attention to a domain on which to choose the expression of probability density. Distribution density must fulfill the conditions [2]:

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

Because

$$\int_{xm}^{xM} \sum_{kk} co_{kk} \cdot (x)^{g-kk+1}\, dx = 0.174 \;, \tag{5}$$

using the next program we will determine the limits of a definition domain for a positive function, marked f3(x), restriction of the regression polynomial, so that the integral on this range will be equal with the unit

$var1 := 0.15$ $\qquad$ $var2 := 0.8$

Given

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

$tvar := Find(var1, var2)$ $\qquad$ $tvar = \begin{pmatrix} 1.28 \\ 0.141 \end{pmatrix}$.

We will obtain in this way the values

$tvar_1 = 1.28$ $\qquad$ $tvar_2 = 0.141$

for which

$$\int_{xm-tvar_1}^{xM+tvar_2} \sum_{kk} co_{kk} \cdot (x)^{g-kk+1}\, dx = 1 \;, \tag{6}$$

where

$xm - tvar_1 = 0.383$ $\qquad$ $xM + tvar_2 = 2.379$.

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

$$f3(x) := if\left[xm - tvar_1 < x < xM + tvar_2, \sum_{kk} co_{kk} \cdot (x)^{g-kk+1}, 0\right], \tag{7}$$

and it's graphic is show in figure 1 with the help of the adjoining program

$$nrnod := 1000 \qquad is := 1..nrnod$$

$$xv_{is} := xm - (tvar)_1 - .35 + \frac{is-1}{nrnod-1} \cdot \left[xM + tvar_2 + .1 - \left(xm - tvar_1 - .35 \right) \right]$$

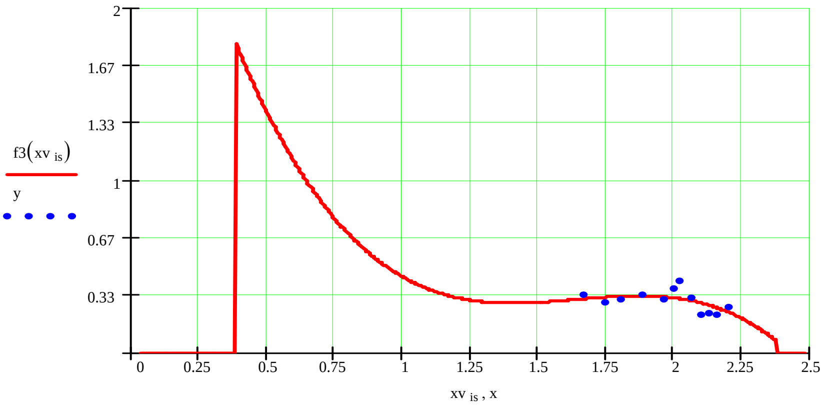

Figure. 1 Probability density of 3rd degree and the distribution of experimental points

The expression of the repartition function is

$$Ftrunc3(x) := \int_{xm-tvar_1}^{x} if\left[xm - tvar_1 < u < xM + tvar_2, \sum_{kk} co_{kk} \cdot (u)^{g-kk+1}, 0 \right] du \quad , \qquad (8)$$

and it's graphic is shown in Figure 2.

$$wFtrunc3_{is} := Ftrunc3\left(xv_{is}\right)$$

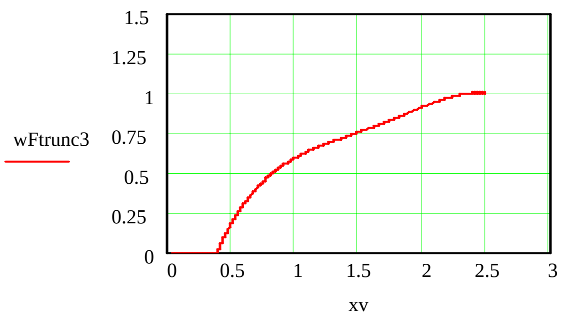

Figure. 2 Repartition function for regression polynomial of 3rd degree

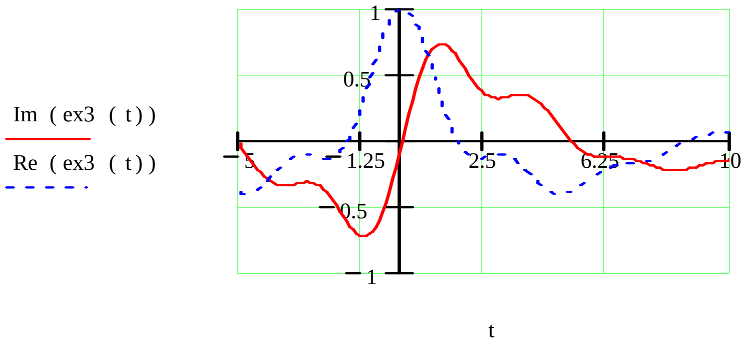

Figure. 3 The real and imaginary part of the characteristic function for the probability density of 3rd degree

The expression of the characteristic function is

$$ex3(t) := \int_{-1}^{6} e^{i \cdot t \cdot x} \cdot \left[if\left[xm - tvar_1 < x < xM + tvar_2, \sum_{kk} co_{kk} \cdot (x)^{g-kk+1}, 0 \right] \right] dx \qquad (9)$$

and it's graphic is shown in figure 3.

If we made a regression of 4th degree (g=4), we will use the program

$g := 4 \qquad j := 1..2\cdot g + 1 \qquad S_j := \sum_i (x_i)^{2\cdot g - j + 1}$

$k := 1..g + 1 \qquad v := 1..g + 1 \qquad m_{k,v} := S_{v+k-1} \qquad TL_k := \sum_i y_i \cdot (x_i)^{g-k+1}$

$co := m^{-1}\cdot TL \qquad co^T = \left(42.345 \ \ -332.126 \ \ 972.561 \ \ -1.26\times 10^3 \ \ 609.971\right)$

from which results the regression polynomial of the following form

$$y = 42.345\cdot x^4 - 332.126\cdot x^3 + 972.261\cdot x^2 - 1.26\cdot 10^3\cdot x + 609.971 \tag{10}$$

which in the nodes has the values

$Su^T =$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.333	0.282	0.293	0.328	0.348	0.35	0.354	0.343	0.269	0.251	0.235	0.232	0.229	0.229	0.254

For polynomial regression of 4th degree we will obtain the following values for the correlation coefficient and respectively for the deviation from the regression curve

$$r4 := \sqrt{1 - \frac{\sum_i (y_i - Su_i)^2}{\sum_i (y_i - ym)^2}} \qquad r4 = 0.736 \ , \tag{11}$$

$$St4 := \sqrt{\frac{1}{n}\cdot\left[\sum_i (y_i - Su_i)^2\right]} \qquad St4 = 0.044 \ . \tag{12}$$

Because

$$\int_{xm}^{xM} \sum_{kk} co_{kk}\cdot (x)^{g-kk+1}\, dx = 0.171 \ , \tag{13}$$

it will determine the limits of the definition domain for regression density of 4th degree. The program used is likewise with the one from previous case, and it will obtain

$tvar = \begin{pmatrix} 0.15 \\ 0.374 \end{pmatrix} \qquad tvar_1 = 0.15 \qquad tvar_2 = 0.374$

resulting

$$\int_{xm - tvar_1}^{xM + tvar_2} \sum_{kk} co_{kk}\cdot (x)^{g-kk+1}\, dx = 1 \tag{14}$$

where $xm - tvar_1 = 1.513 \qquad xM + tvar_2 = 2.612$.

In conclusion, the expression of probability density of 4th degree is

$$f4(x) := if\left[xm - tvar_1 < x < xM + tvar_2, \sum_{kk} co_{kk}\cdot (x)^{g-kk+1}, 0\right], \tag{15}$$

and it's graphic is shown in figure 4

$nrnod := 1000 \qquad is := 1..nrnod$

$$xv_{is} := xm - (tvar)_1 - .3 + \frac{is - 1}{nrnod - 1}\cdot\left[xM + tvar_2 + .3 - (xm - tvar_1 - .3)\right]$$

The expression of the repartition function is

$$Ftrunc4(x) := \int_0^x if\left[xm - tvar_1 < u < xM + tvar_2, \sum_{kk} co_{kk}\cdot (u)^{g-kk+1}, 0\right] du \ , \tag{16}$$

and it's graphic is shown in figure 5

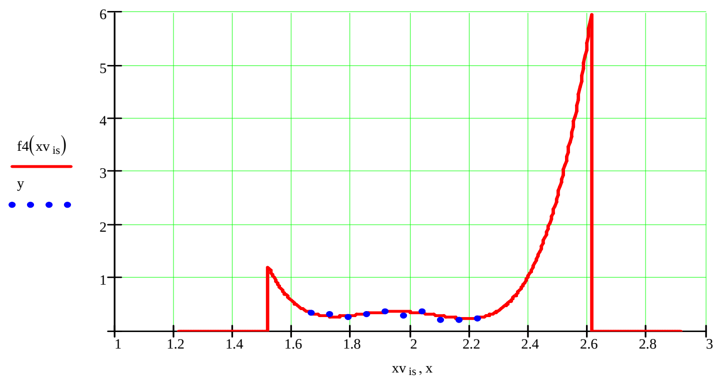

Figure 4. Probability density of 4th degree and the distribution of the experimental points

$$wFtrunc4_{is} := Ftrunc4(xv_{is})$$

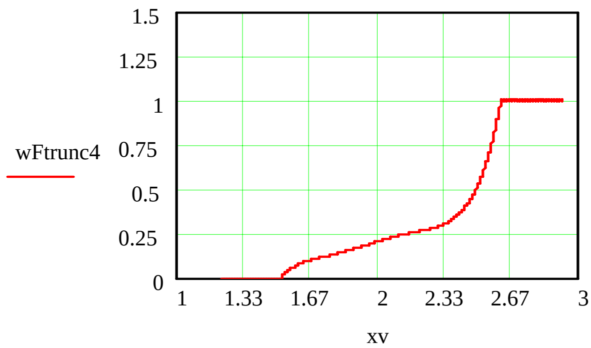

Figure 5. The repartition function for repression polynomial of 4th degree

The characteristic function has the expression

$$ex(t) := \int_{-1}^{6} e^{i \cdot t \cdot x} \cdot \left[if \left[xm - tvar_1 < x < xM + tvar_2, \sum_{kk} co_{kk} \cdot (x)^{g-kk+1}, 0 \right] \right] dx \tag{17}$$

and it's graphic is shown in figure 6

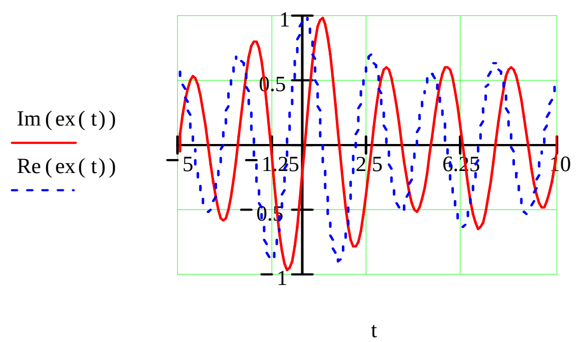

Figure 6. The real and imaginary part of the characteristic function for probability density of 4th degree

In case that we do a regression of 5th degree (g=5), it results de regression polynomial of the following form

$$y = -119.156 \cdot x^5 + 1212 \cdot x^4 - 4911 \cdot x^3 + 9912 \cdot x^2 - 9962 \cdot x + 3988 \tag{18}$$

which in the nodes has the values

$Su^T =$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.335	0.27	0.289	0.337	0.358	0.359	0.35	0.331	0.259	0.247	0.239	0.238	0.237	0.237	0.244

For polynomial regression of 5th degree we will obtain the following values for the correlation coefficient and respectively for the deviation from the regression curve

$$r5 := \sqrt{1 - \frac{\sum_i (y_i - Su_i)^2}{\sum_i (y_i - ym)^2}} \qquad r5 = 0.7462 \quad , \tag{19}$$

$$St5 := \sqrt{\frac{1}{n}\cdot\left[\sum_i (y_i - Su_i)^2\right]} \quad St5 = 0.043 \quad . \tag{20}$$

Because $$Int := \int_{xm}^{xM} \sum_{kk} co_{kk}\cdot(x)^{g-kk+1}\,dx \qquad Int = 0.167 \quad , \tag{21}$$

we will determine the limits of the definition domain for the regression density of 5th degree. The program used is likewise with the previous one, and we will obtain

$$tvar = \begin{pmatrix} 0.277 \\ 0.126 \end{pmatrix} \qquad tvar_1 = 0.277 \qquad tvar_2 = 0.126$$

Resulting
$$\int_{xm-tvar_1}^{xM+tvar_2} \sum_{kk} co_{kk}\cdot(x)^{g-kk+1}\,dx = 1 \tag{22}$$

where $\quad xm - tvar_1 = 1.386 \quad , \quad xM + tvar_2 = 2.364 \quad .$

The expression of probability density of 5th degree is

$$f5(x) := if\left[xm - tvar_1 < x < xM + tvar_2, \sum_{kk} co_{kk}\cdot(x)^{g-kk+1}, 0\right], \tag{23}$$

and the graphic is shown in figure 7

$nrnod := 1000 \qquad is := 1..nrnod$

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

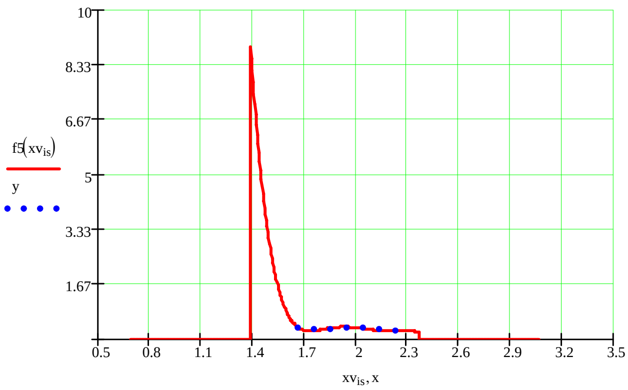

Figure 7. Probability density of 5th degree and repartition of the experimental points

The expression of the repartition function is

$$Ftrunc5(x) := \int_0^x if\left[xm - tvar_1 < u < xM + tvar_2, \sum_{kk} co_{kk}\cdot(u)^{g-kk+1}, 0\right] du \tag{24}$$

and it's grafic is shown in figiure 8.

The characteristic function has the expression

$$ex(t) := \int_{-1}^{6} e^{i\cdot t\cdot x}\cdot\left[if\left[xm - tvar_1 < x < xM + tvar_2, \sum_{kk} co_{kk}\cdot(x)^{g-kk+1}, 0\right]\right] dx \tag{25}$$

and it's graphics is shown in figure 9.

$$wFtrunc5_{is} := Ftrunc5(xv_{is})$$

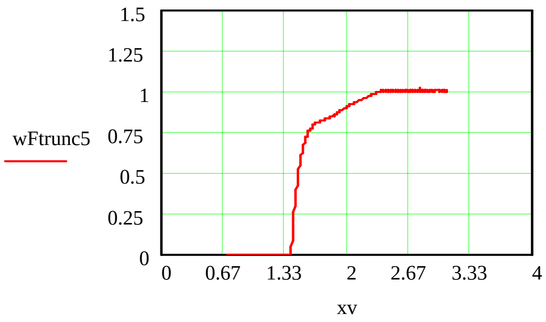

Figure 8. Repartition function for the regression polynomial of 5th degree

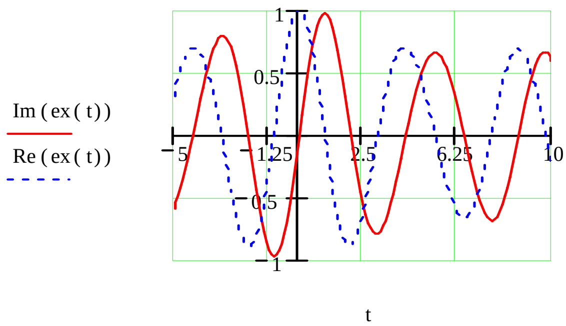

Figure 9. The real and imaginary part of the characteristic function for probability density of 5th degree

2. OBSERVATIONS

For the regression polynomial of 2nd degree we will obtain the regression parable of the following form

$$y = -0.778 \cdot x^2 + 2.874 \cdot x - 2.323 . \tag{26}$$

In this case we will obtain the following values for the correlation coefficient and respectively for the deviation from the regression curve

$$r2 := \sqrt{1 - \frac{\sum_i (y_i - Su_i)^2}{\sum_i (y_i - ym)^2}} \qquad r2 = 0.6131 \tag{27}$$

$$St2 := \sqrt{\frac{1}{n} \cdot \left[\sum_i (y_i - Su_i)^2 \right]} \qquad St2 = 0.051 \; . \tag{28}$$

Analyzing the form of the parable, shown in figure 10, it’s obvious that, on the domain in which the function is positive, the condition cannot be satisfied that integral from it to be equal with the unit.

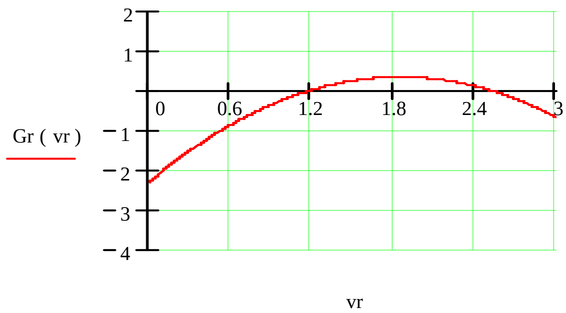

Figure 10. Regression parable

Results that this set of dates, which is obvious a modeling of 2nd degree, it does not accept an appropriate probability density. From the same reason does not accept also probability density of 6th degree.

3. CONCLUSIONS

The advantage of polynomial modeling is the fact that from passing from an inferior grade to a superior grade appears every time new coefficients, and by determines this coefficient the regression curve moulds better and better the experimental figures. By increasing the polynomial grade it will be more accurate the modeling, so it's increasing the value of correlation coefficient.

We specify also the fact that the condition that the integral from probability density must be equals with the unit admission, in possible cases, for a fast grade, much more solutions regarding of the pair of the integrating limits (from which one can be predefine). All the obtained densities mould the experimental domain dates identically. Although, in any form of the function of chosen density, the probability that the random variable to take a value into a range, is the same. 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 deviation from the square mean.

REFERENCES

[1] Jalobeanu, C., Rasa, I., *MathCAD. Probleme de calcul numeric si statistic*, Editura Albastra, Cluj-Napoca, 1995;

[2] Maksay, St., Stoica, D., *Calculul probabilitatilor*, Editura Politehnica, Timisoara, 2005.

[3] Maksay, St., Stoica, D., *Considerations on Modeling Some Distribution Laws*, Applied Mathematics and Computation, 175, 2006, pag. 238-246.