ANNALS OF FACULTY ENGINEERING HUNEDOARA
- INTERNATIONAL JOURNAL OF ENGINEERING
Tome IX (Year 2011). Fascicule 3. (ISSN 1584 - 2673)

1. Ştefan MAKSAY, 2. Diana STOICA

CONSIDERATIONS ON THE RECTANGLE TRUNCATED BIDIMENSIONAL NORMAL MODELING

1,2. UNIVERSITY "POLITEHNICA" TIMISOARA, FACULTY OF ENGINEERING HUNEDOARA, ROMANIA

ABSTRACT: In this article, which generalizes the paper [1], we started from the expression of the classic normal distribution density white two variables, we made a few considerations about the process of experimental data, by means of the distribution density that generalizes the bidimensional classical normal law. The bidimensional distribution densities are rectangle truncated [2], obviously keeping the properties of a density. The new expressions of the density, having one or more extra parameters, can better approximate the experimental data. A.M.S.(2000) Subject Classification: 33B15; 00A71
KEYWORDS: Modeling truncated, normal distribution density, coefficient of correlation

❖ INTRODUCTION

In this article, which generalizes the paper [1], we started from the expression of the classic normal distribution density with two variables [3], we made a few considerations about the process of experimental data, by means of the distribution density that generalizes the bidimensional classical normal law. Thus, we further on present a few distinct modeling variants.

The first two expressions for the distribution density will introduce one, respectively two parameters, fact that permits an optimal modeling of the experimental data.

This article proposes to model the experimental data presented in the next table, which shows the coordinates of points in the space with tree dimensions, where the first two lines represent the independent variables x and y, and the tree lines represent the dependent variable u.

var =

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.483	1.06	1.144	1.389	1.574	1.605	1.891	2.039	2.501	2.606	2.728	2.776	2.824	2.887	3.973
2	2.093	2.132	2.153	2.227	2.38	2.706	2.782	3.002	3.083	3.107	3.67	3.836	4.071	4.329	4.514
3	0.03	0.064	0.073	0.105	0.18	0.196	0.231	0.229	0.229	0.215	0.149	0.1	0.084	0.05	0.02

Further on we note

$$mx := \mathrm{mean}(x) \qquad my := \mathrm{mean}(y) \qquad mu := \mathrm{mean}(u),$$

$$sx := \mathrm{stdev}(x) \qquad sy := \mathrm{stdev}(y) \qquad su := \mathrm{stdev}(u),$$

where *mean* and *stdev* represent the mean value and respectively the standard deviation of the corresponding variable. For the given measurements, we have

$$mx = 2.0986 \qquad my = 3.0724 \qquad mu = 0.1304$$

$$sx = 0.8795 \qquad sy = 0.8042 \qquad su = 0.0748 .$$

Note by *fclas(x,y)* the classic normal distribution density with two variables [3]

$$fclas(x,y) := \frac{1}{2\cdot\pi\cdot sx\cdot sy}\cdot\exp\left[\frac{-1}{2}\cdot\left[\frac{(x-mx)^2}{sx^2}+\frac{(y-my)^2}{sy^2}\right]\right]. \tag{1}$$

❖ PRACTICAL CASE

Instead of the classic distribution density (1) we consider the function:

$$ftrunc1(x,y,cx,cy)=\begin{cases} K\cdot\dfrac{1}{2\pi\cdot sx\cdot xy}\exp\left[-\dfrac{1}{2}\cdot\left[\dfrac{(x-mx)^2}{sx^2}+\dfrac{(y-my)^2}{sy^2}\right]\right], & \left(|x-mx|<cx\cdot sx\right)\wedge\left(|y-my|<cy\cdot sy\right) \\ 0, \quad \text{in rest} & \end{cases} \tag{2}$$

where K, *cx* and *cy* are positives constants.

For the function *ftrunc1* to be a distribution density we impose the conditions

$$ftrunc1(x,y,cx,cy) \geq 0 \quad \text{and} \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ftrunc1(x,y,cx,cy)\,dx\,dy = 1,$$

hence results the value for K

$$K := \frac{1}{\displaystyle\int_{mx-cx\cdot sx}^{mx+cx\cdot sx}\int_{my-cy\cdot sy}^{my+cy\cdot sy} if\left[\left[(|x-mx| < cx\cdot sx) \wedge (|y-my| < cy\cdot sy)\right], \frac{1}{2\cdot\pi\cdot sx\cdot sy}\cdot \exp\left[\frac{-1}{2}\cdot\left[\frac{(x-mx)^2}{sx^2}+\frac{(y-my)^2}{sy^2}\right]\right], 0\right] dy\,dx}$$

We remark the fact that for $cx \to \infty$ and $cy \to \infty$, the function *ftrunc1(x,y,cx,cy)* coincides with the function *fclas(x,y)*.

The constants *cx* and *cy* will be determined by imposing the condition of minimizing the sum of the difference squares between of the value of theoretic function *ftrunc1(x,y,cx,cy)* and experimental value of independent variable u, that is we minimize the function *F(cx,cy)*

$$F(cx, cy) := \sum_{i=1}^{n} \left(ftrunc1\left(x_i, y_i, cx, cy\right) - u_i\right)^2 .$$

In order to achieve this, we will use the next program, written in language MathCAD.

$ORIGIN \equiv 1$

$x := (0.4829\ \ 1.0601\ \ 1.1437\ \ 1.3885\ \ 1.574\ \ 1.6049\ \ 1.8914\ \ 2.0392\ \ 2.5008\ \ 2.6057\ \ 2.7279\ \ 2.776\ \ 2.824\ \ 2.8866\ \ 3.9726)$

$y := (2.0933\ \ 2.1323\ \ 2.1534\ \ 2.2267\ \ 2.3801\ \ 2.7062\ \ 2.7818\ \ 3.0022\ \ 3.0829\ \ 3.1074\ \ 3.6696\ \ 3.836\ \ 4.0711\ \ 4.3291\ \ 4.5142)$

$u := (0.03\ \ 0.0639\ \ 0.0733\ \ 0.1055\ \ 0.1802\ \ 0.1956\ \ 0.2314\ \ 0.2294\ \ 0.2288\ \ 0.2149\ \ 0.1492\ \ 0.1003\ \ 0.0836\ \ 0.0501\ \ 0.02)$

$x := x^T \qquad y := y^T \qquad u := u^T$

$n := length(x) \qquad i := 1..n \qquad nrdivx := 5 \qquad nrdivy := 7$

$dxi := 1 \qquad dxf := 2.$

$dyi := 1 \qquad dyf := 3$

$$
prog1 := \left|
\begin{array}{l}
px \leftarrow nrdivx \\
py \leftarrow nrdivy \\
\text{for } j \in 1..px+1 \\
\quad \left| \begin{array}{l}
cx_j \leftarrow dxi + (j-1)\cdot\dfrac{(dxf-dxi)}{px} \\
\text{for } k \in 1..py+1 \\
\quad \left| \begin{array}{l}
cy_k \leftarrow dyi + (k-1)\cdot\dfrac{(dyf-dyi)}{py} \\
K \leftarrow \dfrac{1}{\displaystyle\int_{mx-cx_j\cdot sx}^{mx+cx_j\cdot sx}\int_{my-cy_k\cdot sy}^{my+cy_k\cdot sy} if\left[\left[\left[(x-mx)^2 < (cx_j\cdot sx)^2\right] \wedge \left[(y-my)^2 < (cy_k\cdot sy)^2\right]\right], fclas(x,y), 0\right] dy\,dx} \\
fCT_{j,k} \leftarrow \displaystyle\sum_{i=1}^{n} \left[if\left[\left[\left[(x_i-mx)^2 < (cx_j\cdot sx)^2\right] \wedge \left[(y_i-my)^2 < (cy_k\cdot sy)^2\right]\right], K\cdot fclas(x_i,y_i), 0\right] - u_i\right]^2
\end{array} \right.
\end{array} \right. \\
fmin \leftarrow min(fCT) \\
\text{for } j \in 1..px+1 \\
\quad \left| \begin{array}{l}
\text{for } k \in 1..py+1 \\
\quad \left| \begin{array}{l}
indx \leftarrow j \\
indy \leftarrow k \\
\text{break if } (fCT_{j,k} - fmin) = 0
\end{array} \right. \\
\text{break if } (fCT_{j,k} - fmin) = 0
\end{array} \right. \\
\left[cx_j\ \ cy_k\ \ fCT\ \ (indx\ \ indy\ \ fmin)\right]
\end{array}
\right.
$$

$prog1 = (2\ \ 1.8571\ \ \{6,8\}\ \ \{1,3\})$

$cxf := prog1_{1,1} \qquad cxf = 2$

$cyf := prog1_{1,2} \qquad cyf = 1.8571$

$fCT := prog1_{1,3}$

$(indx\ \ indy\ \ fmin) := prog1_{1,4}$

$prog1_{1,4} = (6\ \ 4\ \ 2.477\times 10^{-3})$

$$\mathrm{indx} = 6 \qquad \mathrm{indy} = 4 \qquad \mathrm{fmin} = 2.477\times 10^{-3}$$

$$\mathrm{fCT} - \mathrm{fmin} = \begin{pmatrix} 0.2904 & 0.1436 & 0.0864 & 0.0623 & 0.0512 & 0.046 & 0.0436 & 0.0425 \\ 0.1823 & 0.0708 & 0.0341 & 0.0203 & 0.0143 & 0.0117 & 0.0105 & 9.9963\times 10^{-3} \\ 0.1283 & 0.0399 & 0.0152 & 7.3245\times 10^{-3} & 4.353\times 10^{-3} & 3.1694\times 10^{-3} & 2.6787\times 10^{-3} & 2.4737\times 10^{-3} \\ 0.0993 & 0.0248 & 7.1072\times 10^{-3} & 2.6071\times 10^{-3} & 1.3245\times 10^{-3} & 9.7519\times 10^{-4} & 8.8255\times 10^{-4} & 8.5732\times 10^{-4} \\ 0.0831 & 0.017 & 3.5837\times 10^{-3} & 1.1204\times 10^{-3} & 8.4481\times 10^{-4} & 9.9182\times 10^{-4} & 1.1357\times 10^{-3} & 1.2172\times 10^{-3} \\ 0.0737 & 0.012 & 1.2018\times 10^{-3} & 0 & 3.4867\times 10^{-4} & 8.0354\times 10^{-4} & 1.0942\times 10^{-3} & 1.2419\times 10^{-3} \end{pmatrix}$$

Substituting these values of *cx* and *cy* in expression of *ftrunc1(x,y,cx,cy)* leads to the expression of the truncated distribution density of first form that approximate better the experimental data.

Thus, the graphs of the classic and truncated distribution density, for the given experimental values, are shown in Figure 1 and respectively in Figure2.

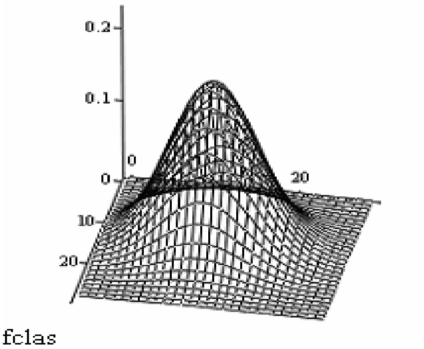

Figure 1. The classical distribution density

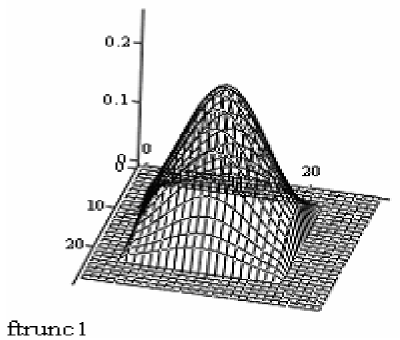

Figure 2. The truncated distribution density *ftrunc1(x,y,cx,cy)*

The correlation coefficients corresponding of the two modeling's are

$$\mathrm{rclas} := \sqrt{1 - \frac{\sum_i \left(u_i - \mathrm{fclas}(x_i, y_i)\right)^2}{\sum_i \left(u_i - \mathrm{mean}(u)\right)^2}}, \qquad \mathrm{rclas} = 0.9631 ,$$

$$\mathrm{rtrunc1} := \sqrt{1 - \frac{\sum_i \left(u_i - \mathrm{ftrunc1}(x_i, y_i, \mathrm{cxf}, \mathrm{cyf})\right)^2}{\sum_i \left(u_i - \mathrm{mean}(u)\right)^2}}, \qquad \mathrm{rtrunc1} = 0.9852 .$$

The values of dependent variable u, of the distribution density *fclas, ftrunc1* in the points (x_i, y_i), are given, comparatively, in the next table.

$$\mathrm{fclas}_i := \mathrm{fclas}(x_i, y_i), \qquad \mathrm{ftrunc1}_i := \mathrm{ftrunc1}(x_i, y_i, \mathrm{cxf}, \mathrm{cyf}),$$

$$\mathrm{vb} := \mathrm{augment}(x, y, u, \mathrm{fclas}, \mathrm{ftrunc1})^T .$$

vb=

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.4829	1.0601	1.1437	1.3885	1.574	1.6049	1.8914	2.0392	2.5008	2.6057	2.7279	2.776	2.824	2.8866	3.9726
2	2.0933	2.1323	2.1534	2.2267	2.3801	2.7062	2.7818	3.0022	3.0829	3.1074	3.6696	3.836	4.0711	4.3291	4.5142
3	0.03	0.0639	0.0733	0.1055	0.1802	0.1956	0.2314	0.2294	0.2288	0.2149	0.1492	0.1003	0.0836	0.0501	0.02
4	0.0198	0.0566	0.065	0.0934	0.13	0.1733	0.205	0.2236	0.2026	0.1904	0.1322	0.1066	0.0741	0.0444	$601\cdot 10^{-3}$
5	0.0222	0.0633	0.0727	0.1045	0.1454	0.1938	0.2293	0.2501	0.2267	0.2129	0.1479	0.1192	0.0828	0.0497	0

The distribution rate function Fclas(x,y) is given by

$$\mathrm{Fclas}(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} \mathrm{fclas}(u, v)\, dv\, du .$$

The rate of the truncated distribution functions is

$$\mathrm{Ftrunc1}(x, y, cx, cy) = \int_{-\infty}^{x}\int_{-\infty}^{y} \mathrm{ftrunc1}(u, v, cx, cy)\, dv\, du .$$

Using the following program, we showed in Figure 5 the function *Ftrunc1* and in Figure 6 the corresponding contours lines.

$$\mathrm{nrn} := 15$$

$$k := 1..nrn \qquad xv_k := (\min(x)-1) + \frac{(\max(x)+1)-(\min(x)-1)}{nrn-1}\cdot(k-1)$$

$$h := 1..nrn \qquad yv_h := (\min(y)-1) + \frac{(\max(y)+1)-(\min(y)-1)}{nrn-1}\cdot(h-1)$$

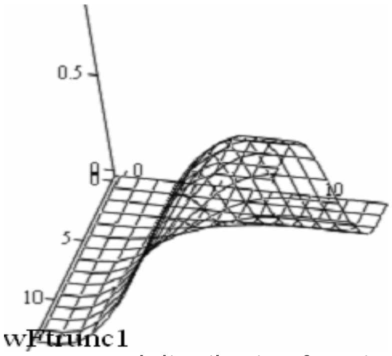

Figure 5.The truncated distribution function *Ftrunc1*

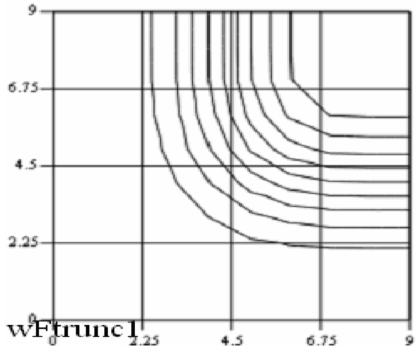

Figure 6. The contour lines of the truncate distribution function *Ftrunc1*

Inserting the notations

$$Fclas_i := Fclas(x_i, y_i), \qquad Ftrunc1_i := Ftrunc1(x_i, y_i, cxf, cyf),$$

$$vb := augment(Fclas, Ftrunc1)^T,$$

the constants that appear as having been previously determined, results the distribution of the values of distribution functions in points corresponding to the experimental system point.

vb=	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	$54\cdot10^{-3}$	0.0134	0.0165	0.0295	0.0519	0.0904	0.143	0.2161	0.3373	0.367	0.5817	0.6389	0.7024	0.7587	0.9393
2	$43\cdot10^{-4}$	$71\cdot10^{-3}$	0.0123	0.024	0.0461	0.0866	0.1406	0.2184	0.3461	0.3776	0.6121	0.6746	0.7441	0.8056	0.9948

CONCLUSIONS

We notice that, for the experimental data presented in the paper, the optimal modeling is given by the distribution density *ftrunc1(x,y,cx,cy)* ($rtrunc1 = 0.9852$).

This modeling are better than the non-truncated modeling fclas(x,y) ($rclas = 0.9631$).

In the practical case these results show us, that is more advantage to use the truncated modeling, because the classical modeling is the practical case of these.

REFERENCES

[1.] Şt. Maksay, D. Stoica – Considerations on Modeling Some Distribution Laws, Applied Mathematics and Computation, Volume 175, Issue 1, 1 April 2006, Pages 238-246;

[2.] Johnson, A.C. (2001), "On the Truncated Normal Distribution: Characteristics of Singly- and Doubly-truncated Populations of Application in Management Science," PhD. Dissertation, Stuart Graduate School of Business, Illinois Institute of Technology, Illinois.

[3.] Şt. Maksay, D. Stoica – Calculul probabilităţilor, Editura Politehnica, Timişoara, 2005;