JOURNAL OF ENGINEERING
annals OF
FACULTY OF ENGINEERING HUNEDOARA
TOME VI (year 2008), FASCICULE 3, (ISSN 1584 – 2673)

AN ALGORITHM FOR TRUNCATED 4 DIMENSIONAL MODELLING USING MATLAB COMPUTER ALGEBRA SYSTEM

MAKSAY Ştefan, BISTRIAN Diana Alina

UNIVERSITY POLITEHNICA TIMISOARA
FACULTY OF ENGINEERING HUNEDOARA

ABSTRACT: In this article we present a computational method for a parallelepipedically truncated modeling of the four-dimensional normal distribution. Our goal was to prepare a functional algorithm that transforms the normal distribution in its truncated counterpart. We suggest a method of truncation in which the limits are determined applying the least squares method, under the conditions in which the corresponding probability density approximates ever better the numeric data. Calculations have been done in Matlab Computer Algebra System.

KEYWORDS:
probability distribution, normal distribution, least squares method, order statistics.

1. INTRODUCTION

According to the three-sigma rule, in certain problems in probability theory and mathematical statistics, an event is considered to be practically impossible if it lies in the region of values of the normal distribution of a random variable at a distance from its mean value of more than three times the standard deviation (see [7]).

Let X be a normally $N(\mu,\sigma^2)$ distributed random variable, where μ is the mean value and σ is the standard deviation. For any $k>0$,

$$P\{|X-\mu|<k\sigma\}=2\Phi(k)-1 \quad (1)$$

where $\Phi(.)$ is the distribution function of the standard normal law.

Whence, in particular, for $k=3$ it follows that

$$P\{\mu-3\sigma<X<\mu+3\sigma\}=0.99730 \quad (2)$$

and, obviously

$$P\{|X-\mu|>3\sigma\}=0.00270 \quad (3)$$

This circumstance is sometimes used in certain problems of probability theory and mathematical statistics, by assuming that the event $\{|X-\mu|>3\sigma\}$ is practically impossible and, consequently, the event $\{|X-\mu|<3\sigma\}$ is practically certain.

Starting from this rule, arose the idea of truncating the normal distribution (see [4]), the new function being null without a finite interval, which includes the mean value of the variable, and preserves the properties of probability density. The aim of this paper is to optimize the limits of truncation, so as the correlation coefficient of the function should be as close as possible to one.

We are going to give in Section 2 the rule of four-dimensional classic normal distribution, and the way in which it can be truncated. In order to obtain an optimal parallelepipedic truncation corresponding to a set of numerical data, we elaborated an algorithm under Matlab Computer Algebra System, which is detailed presented in Section 3. The rule of truncated normal distribution, as it is built in this paper, can be used in data modeling more efficiently than the classic rule. Section 4 presents an analysis of the results we obtained and includes the conclusions of the paper.

2. THE TRUNCATED 4D NORMAL DISTRIBUTION AGAINST THE CLASSIC ONE

The classic normal distribution rule represents one fundamental rule in the theory of probabilities, frequently used in the study of natural, social and economical phenomena [1, 2].

Let $fclas: R \times R \times R \times R \to R$ be the four-dimensional classic normal distribution rule, expressed by

$$fclas(x,y,z,t) = \frac{1}{\sigma_x \cdot \sigma_y \cdot \sigma_z \cdot \sigma_t \cdot (2\pi)^2} e^{-\frac{1}{2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}+\frac{(z-\mu_z)^2}{\sigma_z^2}+\frac{(t-\mu_t)^2}{\sigma_t^2}\right]} \quad (4)$$

where σ_x, σ_y, σ_z, σ_t are the standard deviations of variables x, y, z, respectively t, and μ_x, μ_y, μ_z, μ_t represent the mean values of the respective variables.

We are going to approximate this rule by a null function in the exterior of a parallelepiped, situated in the vicinity of the mean values of variables x, y, z, respectively t.

Let $ftrunc: R^6 \to R$, be of form

$$ftrunc(x,y,z,t,\alpha x,\alpha y,\alpha z,\alpha t) = \begin{cases} \frac{1}{K} \cdot \frac{e^{-\frac{1}{2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}+\frac{(z-\mu_z)^2}{\sigma_z^2}+\frac{(t-\mu_t)^2}{\sigma_t^2}\right]}}{\sigma_x \cdot \sigma_y \cdot \sigma_z \cdot \sigma_t \cdot (2\pi)^2}, & \begin{cases} |x-\mu_x| < \alpha x \cdot \sigma_x \\ |y-\mu_y| < \alpha y \cdot \sigma_y \\ |z-\mu_z| < \alpha z \cdot \sigma_z \\ |t-\mu_t| < \alpha t \cdot \sigma_t \end{cases} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where K will be determined by imposing the condition that this function be a probability density.

In order to meet our purpose, the function $ftrunc(x,y,z,t,\alpha x,\alpha y,\alpha z,\alpha t)$ has to be a probability density [3], and therefore has to meet the conditions

$$ftrunc(x,y,z,t,\alpha x,\alpha y,\alpha z,\alpha t) \geq 0 \quad (6)$$

and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ftrunc(x,y,z,t,\alpha x,\alpha y,\alpha z,\alpha t)\,dx\,dy\,dz\,dt = 1 \quad (7)$$

hence, for K results the expression

$$K = \int_{\mu_x-\alpha x\cdot\sigma_x}^{\mu_x+\alpha x\cdot\sigma_x}\int_{\mu_y-\alpha y\cdot\sigma_y}^{\mu_y+\alpha y\cdot\sigma_y}\int_{\mu_z-\alpha z\cdot\sigma_z}^{\mu_z+\alpha z\cdot\sigma_z}\int_{\mu_t-\alpha t\cdot\sigma_t}^{\mu_t+\alpha t\cdot\sigma_t} \frac{e^{-\frac{1}{2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}+\frac{(z-\mu_z)^2}{\sigma_z^2}+\frac{(t-\mu_t)^2}{\sigma_t^2}\right]}}{\sigma_x \cdot \sigma_y \cdot \sigma_z \cdot \sigma_t \cdot (2\pi)^2} dt\,dz\,dy\,dx \quad (8)$$

In order to determine an optimal modeling of the data, we processed them both classically and in the truncated way. Constants αx, αy, αz and αt are going to be determined by minimizing the error resulted from the use of the truncated function. The necessary condition is that the sum of the squares of the differences between the theoretical values of the function $\mathrm{ftrunc}(x, y, z, t, \alpha x, \alpha y, \alpha z, \alpha t)$ and the experimental values u, be minimum [5, 6]. That is, function

$$F(\alpha x, \alpha y, \alpha z, \alpha t) = \sum_{i=1}^{n} \left(\mathrm{ftrunc}(x_i, y_i, z_i, t_i, \alpha x, \alpha y, \alpha z, \alpha t) - u_i\right)^2 \quad (9)$$

shall be minimized with respect to αx, αy, αz and αt.

3. AN ALGRITHM FOR TRUNCATED 4D MODELING USING MATLAB COMPUTER ALGEBRA SYSTEM

Our goal was to prepare a functional algorithm that transforms the normal distribution in its truncated counterpart. In this method of truncation, the limits are determined applying the least squares method, under the conditions in which the corresponding probability density approximates ever better the numeric data. The algorithm is presented hereinafter.

```
clc,clear;
```

First, the input data set that is going to be processed, where the first four lines represent the values of the independent variables x, y, z and t, and the last line represents the independent variable u.

```
x=[-0.3775 0.3148 0.9409 -0.7420 -0.6355 0.5690 0.9863 -1.0039...
    1.4725 -1.3493 -1.1678 0.9312 -0.9898 -0.6841 0.4978];
y=[-0.2959 1.4435 -0.9921 1.0823 -0.5596 -0.8217 -0.5186 -0.9471...
    0.0557 -0.2611 -0.4606 0.0112 1.3396 -1.2919 1.4885];
z=[-1.4751 -0.3510 0.2120 -0.1315 0.4437 -0.2656 0.3274 -0.3744...
    -1.2173 0.9535 -0.2624 -0.6451 0.2895 -0.0729 -0.5465];
t=[-0.2340 0.6232 0.2379 0.3899 -0.9499 -1.1878 0.2341 -1.1859...
    -0.0412 0.1286 -1.2132 0.8057 1.4789 -0.3306 -0.8468];
u=[0.6630 -0.8542 -1.2013 -0.1199 -0.0653 0.4853 -0.5955 -0.1497...
    -0.4348 -0.0793 1.5352 -0.6065 -1.3474 0.4694 -0.9036];
```

```
nv=length(x);
```

The program calculates the mean value and the standard deviation of the considered variables.

```
mx=mean(x);my=mean(y);mz=mean(z);mt=mean(t);mu=mean(u);
sx=std(x,1);sy=std(y,1);sz=std(z,1);st=std(t,1);su=std(u,1);
```

The probability density function of the four-dimensional normal distribution is given now.

```
fc=@(x,y,z,t)(normpdf(x,mx,sx).*normpdf(y,my,sy).*normpdf(z,mz,sz).*...
normpdf(t,mt,st));
```

We have started the procedure of calculating the truncation limits with eight initial guess values, far away from the solution and the program compute the limits in four-dimensional space.

```
fmin=10^10;F=0;
```

```
    alfa_xi=0.0; alfa_xs=0.6; alfa_yi=0.15; alfa_ys=0.35; alfa_zi=0.05;
alfa_zs=0.65; alfa_ti=0.01; alfa_ts=0.25;

    fx=@(x)normpdf(x,mx,sx);
    fy=@(y)normpdf(y,my,sy);
    fz=@(z)normpdf(z,mz,sz);
    ft=@(t)normpdf(t,mt,st);

    nt=4;
    for i=1:nt+1
        alfa_x(i)=alfa_xi+(i-1)*(alfa_xs-alfa_xi)/nt;
        for j=1:nt+1
            alfa_y(j)=alfa_yi+(j-1)*(alfa_ys-alfa_yi)/nt;
            for k=1:nt+1
                alfa_z(k)=alfa_zi+(k-1)*(alfa_zs-alfa_zi)/nt;
                for s=1:nt+1
                    alfa_t(s)=alfa_ti+(s-1)*(alfa_ts-alfa_ti)/nt;
                      xinf=mx-alfa_x(i)*sx;
                      xsup=mx+alfa_x(i)*sx;
                      yinf=my-alfa_y(j)*sy;
                      ysup=my+alfa_y(j)*sy;
                      zinf=mz-alfa_z(k)*sz;
                      zsup=mz+alfa_z(k)*sz;
                      tinf=mt-alfa_t(s)*st;
                      tsup=mt+alfa_t(s)*st;
```

Next, the four dimensional integral is calculated and the truncated probability density function for 4D normal distribution is introduced.

```
    inte=quad(fx,xinf,xsup).*quad(fy,yinf,ysup).*quad(fz,zinf,zsup).* ...
quad(ft,tinf,tsup);

     ftr=@(x,y,z,t)(normpdf(x,mx,sx).*normpdf(y,my,sy).* ...
       normpdf(z,mz,sz).*normpdf(t,mt,st)).*...
       logical((alfa_x(i)*sx).^2-(x-mx).^2>0).*...
       logical((alfa_y(j)*sy).^2-(y-my).^2>0).*...
       logical((alfa_z(k)*sz).^2-(z-mz).^2>0).*...
       logical((alfa_t(s)*st).^2-(t-mt).^2>0);
```

Applying the least squares method, the algorithm finds the optimal truncation limits.

```
                         for ks=1:nv
                         F=F+(ftr(x(ks),y(ks),z(ks),t(ks))/inte-u(ks))^2;
                    end;
                    if F<=fmin
                        alfa_xf=alfa_x(i);
                        alfa_yf=alfa_y(j);
                        alfa_zf=alfa_z(k);
                        alfa_tf=alfa_t(s);
                        fmin=F;
                        intef=inte;
                        xinff=mx-alfa_xf*sx;
                        xsupf=mx+alfa_xf*sx;
                        yinff=my-alfa_yf*sy;
                        ysupf=my+alfa_yf*sy;
                        zinff=mz-alfa_zf*sz;
                        zsupf=mz+alfa_zf*sz;
                        tinff=mt-alfa_tf*st;
                        tsupf=mt+alfa_tf*st;
```

```
                    end;
                    F=0;
                end;
            end;
        end;
    end;
```

The final results are:

```
                    alfa_xf, alfa_yf, alfa_zf, alfa_tf
        xinff, xsupf, yinff, ysupf, zinff, zsupf, tinff, tsupf.
```

This values substituted in expression (8) and respectively (5), lead to the calculation of the K parameter and to the four-dimensional normal probability density expression.

```
    K=quad(fx,xinff,xsupf).*quad(fy,yinff,ysupf).*quad(fz,zinff,zsupf).*...
quad(ft,tinff,tsupf)

    ft=@(x,y,z,t)(1/K).*(normpdf(x,mx,sx).*normpdf(y,my,sy).*...
      normpdf(z,mz,sz).*normpdf(t,mt,st)).*...
      logical((alfa_xf*sx).^2-(x-mx).^2>0).*...
      logical((alfa_yf*sy).^2-(y-my).^2>0).*...
      logical((alfa_zf*sz).^2-(z-mz).^2>0).*...
      logical((alfa_tf*st).^2-(t-mt).^2>0);
```

The correlation coefficient and the standard deviation of the four-dimensional normal probability density function are

```
     for i=1:nv
         Asup(i)=(u(i)-fc(x(i),y(i),z(i),t(i)))^2;
         Ainf(i)=(u(i)-mu)^2;
     end;
     r_fclas=sqrt(1-sum(Asup)/sum(Ainf))
     sigma_fclas=sqrt(sum(Asup)/nv)

     for i=1:nv
         Asuptr(i)=(u(i)-ft(x(i),y(i),z(i),t(i)))^2;
     end;
     r_ftrunc=sqrt(1-sum(Asuptr)/sum(Ainf))
     sigma_ftrunc=sqrt(sum(Asuptr)/nv)
```

The graphic reprezentation of the probability density function of the four-dimensional normal distribution and its truncated counterpart can be obtained using the next program code.

```
    [XC,YC]=meshgrid(xinff-1:0.1:xsupf+1,yinff-1:0.1:ysupf+1);
    Z=fc(XC,YC,mz,mt);
    figure,mesh(XC,YC,Z,'EdgeColor','black'),alpha(0.5),grid on

    [XT,YT]=meshgrid(xinff-1:0.1:xsupf+1,yinff-1:0.1:ysupf+1);
    T=ft(XT,YT,mz,mt);
    figure,mesh(XT,YT,T,'EdgeColor','black'),alpha(0.5),grid on

    return
```

4. DISCUSSIONS AND CONCLUSIONS

The domain in which the truncated function we thus determined has non-null values is

$$V = \left\{ (x, y, z, t) \;\middle|\; \begin{array}{l} |x - mx| < alfa\_xf \cdot sx,\ |y - my| < alfa\_yf \cdot sy, \\ |z - mz| < alfa\_zf \cdot sz,\ |t - mt| < alfa\_tf \cdot st \end{array} \right\} \tag{10}$$

As both the normal classic function and the truncated one we obtained are 4-variable functions, for a more suggestive graphical representation, we successively substituted two variables by their mean values, which resulted in four partial functions.

Figure 1 presents in part **a,** the representation of the classic normal distribution and in the part **b**, the partial truncated normal distribution, considering variables x and y, variables z and t being substituted by their mean values mz, respectively mt. The scale may vary depending on the data set which have been computed.

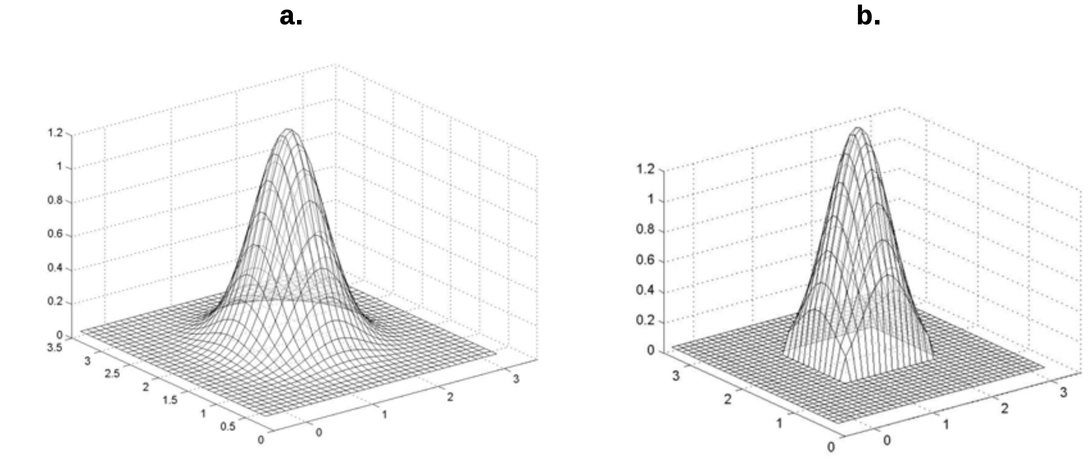

Figure 1. The Probability Density Function of the 4-dimensional normal distribution (a), and for the truncated normal distribution (b). The scale depends on the considered input data set.

Runing the program presented in the paper, one can conclude that the rule of parallelepipedically truncated normal distribution, obtained under the given conditions, preserving the properties of a probability density, leads to a higher correlation coefficient and to a smaller deviation than in the case of the classic distribution.

The 4-dimensional normal truncated distribution is a continuous probability distribution and it can be used in modeling certain problems.

REFERENCES/BIBLIOGRAPHY

[1] Abramowitz, M., Stegun, I.A., *„Handbook of Mathematical Functions"*, National Bureau of Standards, 1964.
[2] Cowan, G., *„Computing and Statistical Data Analysis"*, University of London Lectures.
[3] Feller, W., *„An Introduction to Probability Theory and Its Applications"*, 3rd ed., John Wiley & Sons, Inc., 1968.
[4] Maksay, Şt., *„A probabilistic distribution law with practical applications"*, Mathematica – Revue. D'Analyse numerique et de Theorie de L'Approximation, Tome 22 (45), Nr. 1, pp. 75-76, 1980.
[5] Maksay, Şt., *„Calcul numeric"*, Editura Politehnica, Timişoara, 2002.
[6] Maksay, Şt., Bistrian, D., *„Considerations upon the von Mises 2-dimensional distributions"*, Journal of Engineering, Faculty of Engineering Hunedoara, Tome V, Fascicole 3, 175-180, 2007.
[7] Smirnov, N.V., Dunin-Barkovskii, I.V., *„Mathematische Statistik in der Technik"*, Deutsch. Verlag Wissenschaft, 1969.