Dedicated to Academician Aureliu Săndulescu's 80th Anniversary

A DEMONSTRATION FOR THE STATISTICAL NORMAL DISTRIBUTION OF EXPERIMENTAL RESULTS

E. L. GRIGORESCU

"Horia Hulubei" National Institute of R&D for Physics and Nuclear Engineering - IFIN-HH,
P.O. Box MG-6, Magurele, Ilfov, Romania

Received July 2011

The demonstration is based on one side on the well known Bell Type shape of the experimental results, for a variable *X*, with continuous variation. In this case the distribution is represented by the "density of probability", *f(x)*.
The demonstration considers also the classical statistical characterization of each distribution by the normalization condition and by the calculus of dispersion D.
These two considerations impose for *f(x)* in a logic way, five conditions, as it follows:

1. *f(x)* must depend only on even powers of the deviations from the mean *m*, $t = x - m$; *f(t)* is resulting
2. *f(t)* and its derivative must tend to zero for t tending to $\pm\infty$
3. the derivative must be annulated for $t = m$
4. *f(t)* must have no more than two parameters
5. the two integrals which appear in the normalization condition and the calculus of dispersion must have analytical solutions from which the two parameters could be determined.

The 5th condition is the most restrictive:

For conditions 2 and 4 and (0, +∞) interval, the elementary algebra offers two simple formulae, the exponential and the Lorentz type. Starting from them and fulfilling the conditions 1 to 4, for the (-∞, +∞) interval, two types of more complex formulae for *f(t)* of general form are considered for analysis

$f_1(t) = ae^{-bP_1(t)}$ $\qquad f_2(t) = \dfrac{1}{c + dP_2(t)}$ with P_1, P_2 as polynomials.

Condition 5 restricts $f_1(t)$ at ae^{-bt^2}

And so, the famous formula is obtained: $f(t) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}}$

The $f_2(t)$ is not a solution.
The combinations as $[(f_1(t) \pm f_2(t)]$ or $[f_1(t)\ f_2(t)]$ do not fulfill the 5th condition. However the ratio $[f_1(t) / f_2(t)]$ in the form $a(1+t^2)e^{-bt^2}$ is a solution!

The demonstration is based on experiment!
This demonstration has no mathematical approximations unlike some previous demonstrations from literature.

Key words: statistics, normal distribution function.

1. INTRODUCTION

Demonstration is based on the well-known “Bell Type” shape of the experimental distribution results for *X* variable with continuously variation.
In this case the distribution is characterized by the “density of probability”, $f(x)$. Product $f(x)\ dx$ represents the probability of obtaining a value between x and (x+dx). There are, of course, non symmetrical distributions too. Demonstration takes also into consideration classical statistical characterization of any distribution by normalization condition and D dispersion calculus. These two considerations require, in a logic way, the following five conditions for $f(x)$:

(1) to depend only on even values of “t” deviations from “m” mean value, $t = x - m$; $f(t)$ is resulting

(2) $f(t)$ to tend to zero together with the derivative value for “t” tending to $\pm\infty$

(3) derivative value to be canceled for t=0

(4) function has only two parameters to be determined from normalization condition and dispersion calculus

(5) integrals for normalization conditions and dispersion calculus should have analytical solutions, otherwise the two parameters could not be determined

The last condition is the most restrictive and it would finally determine the concrete form of $f(t)$ function.

It should be mentioned that $\pm\infty$ limits represent exaggerations, acceptable in the literature, because “Bell Type” shape is tending to zero more rapidly than for $\pm\infty$.

It should also be mentioned that, for a symmetrical distribution, “non symmetry” and "excess” are zero by definition.

2. POSSIBLE VARIANTS FOR f(t)

For (2) and (4) conditions and (0, +∞) interval ordinary algebra provides two simple formula, exponential function with negative exponent and Lorenz type function.

$$a\mathrm{e}^{-bt} \text{ and } \frac{1}{c + dt^2} \tag{1}$$

Starting with these two simple formulae and according to (1) and (3) conditions for (–∞, +∞) interval, two more complex general functions could be considered for analyses, respectively:

$$\text{a) } f_1 = a\mathrm{e}^{-bP_s(t)} = a\mathrm{e}^{-b(t^2+t^4+...+t^{2k})} \tag{2}$$

with derivative

$$f'_1(t) = -abt\mathrm{e}^{-bP_1(t)}(2 + 4t^2 + ... + 2kt^{2k-2}) \tag{3}$$

and function

$$\text{b) } f'_2(t) = \frac{1}{c + dP_2(t)} = \frac{1}{c + d(t^2 + t^4 + \ldots + t^{2l})} \tag{4}$$

with derivative

$$f'_2(t) = -\frac{dt(2 + 4t^2 + \ldots + 2lt^{2l-2})}{[c + dP_2(t)]^2} \tag{5}$$

Where $P_1(t)$ and $P_2(t)$ are ascending polynomials.
"k" and "l" parameters are not limited but they do not appear in final solutions.
It could be noticed that derivatives are zero for $\pm\infty$ and $t=0$.
Polynomials form avoids derivatives cancellation for other values.
Even if polynomials assign numeric coefficients, final solutions would not be changed.

3. CONDITION 5 FOR $f_1(t)$

There are analytical solutions only for $a\mathrm{e}^{-bt^2}$ and the well known formula is obtained.

Integral from the normalization condition

$$1 = a\int_{-\infty}^{+\infty} \mathrm{e}^{-bt^2}\,\mathrm{d}t^6 \tag{6}$$

is calculated through an elegant calculation due to Laplace (annex 1) and results:

$$a = \sqrt{\frac{\pi}{b}} \tag{7}$$

Integral from dispersion calculation is integrated by parts (annex 1) and results:

$$b = \frac{1}{2\sigma^2} \text{ and } a = \frac{1}{\sigma\sqrt{2\pi}} \tag{8}$$

The well known formula results

$$f_1(t) = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\rho}}\mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}} \tag{9}$$

Formula could be noticed from the initial expression of $f_1(t)$ but $f_1(t)$ was not artificially introduced, resulting directly from the 5 conditions naturally occurred.

4. CONDITION 5 FOR $f_2(t)$

If only the first term of $P_2(t)$ polynomial is considered, an integrating expression is obtained.
Normalization condition is written:

$$1=\int_{-\infty}^{+\infty}\frac{\mathrm{d}t}{c+\mathrm{d}t^2}=\frac{1}{\sqrt{cd}}\left|\operatorname{arctg}\sqrt{\frac{d}{c}}t\right|_{-\infty}^{+\infty}=\frac{\pi}{\sqrt{cd}} \tag{10}$$

But dispersion calculus provides an infinite value.
Indeed, $\frac{t^2}{c+\mathrm{d}t^2}$ function tends to $\frac{1}{d}$ for "t" tending to $\pm\infty$
The following expression considers two terms of $P_2(t)$, respectively

$$\frac{1}{c+d(t^2+t^4)} \tag{11}$$

For integration the roots of denominator polynomial from a bipolar equation should be determined

$$\mathrm{d}t^4+\mathrm{d}t^2+c=0 \tag{12}$$

replacement $t^2 = y$ is made
Equation

$$\mathrm{d}y^2+\mathrm{d}y+c=0 \tag{13}$$

has 2 negative roots:

$$y_{1,2}=\frac{d\pm\sqrt{d^2-4cd}}{2d}=\frac{-1\pm\sqrt{1-4\frac{c}{d}}}{2} \tag{14}$$

There are four imaginary roots for "t"

$$t_1=\mathrm{i}\beta;\quad t_2=-\mathrm{i}\beta;\quad t_3=\mathrm{i}\gamma;\quad t_n=-\mathrm{i}\gamma \tag{15}$$

with:

$$\beta=\sqrt{\frac{1-\sqrt{1-4\frac{c}{a}}}{2}};\quad \gamma=\sqrt{\frac{1+\sqrt{1-4\frac{c}{d}}}{2}};\quad \beta^2+\gamma^2=1\quad \frac{c}{d}=\beta^2(1-\beta^2)=\beta^2\gamma^2 \tag{16}$$

β,γ parameters facilitate the calculation. Expression (11) could be written:

$$\frac{1}{d}\left[\frac{A}{t^2+\beta^2}+\frac{B}{t^2+\gamma^2}\right] \tag{17}$$

By obtaining a common denominator, the following relations result:

$$A = \frac{1}{\gamma^2 - \beta^2} = -B \tag{18}$$

Normalization condition is written:

$$\begin{aligned} 1 &= \frac{1}{d(\gamma^2 - \beta^2)} \int_{-\infty}^{+\infty} \left(\frac{1}{t^2 + \beta^2} - \frac{1}{t^2 + \gamma^2}\right) \mathrm{d}t = \\ &= \frac{1}{d(\gamma^2 - \beta^2)} \left[\frac{1}{\beta} \left| \operatorname{arctg} \frac{t}{\beta} \right|_{-\infty}^{+\infty} + \frac{1}{\gamma} \left| \operatorname{arctg} \frac{t}{\gamma} \right|_{-\infty}^{+\infty} \right] = \\ &= \frac{\pi}{d(\gamma^2 - \beta^2)} \left(\frac{1}{\beta} - \frac{1}{\gamma}\right) = \frac{\pi}{d(\gamma + \beta)\beta\gamma} \end{aligned} \tag{19}$$

Through an analogous calculation the following expression is resulting for dispersion:

$$\sigma^2 = \frac{\pi}{d(\gamma^2 - \beta^2)} (\beta\gamma + \gamma^2 - 2\beta^2) \tag{20}$$

By dividing (20) equation to (19) equation, results:

$$\sigma^2 (\gamma - \beta) = \beta^2\gamma + \gamma^2\beta - 2\beta^2 \tag{21}$$

And through the elimination of γ parameter, a 6 grade equation in β function of σ is resulting:

$$10\beta^6 - 6(\sigma^2 + 1)\beta^5 - (2\sigma^2 + 1)\beta^4 + (\gamma^2 + 1)^2 \beta^2 - \sigma^2 = 0 \tag{22}$$

So, even if expression (11) could be integrated, "*c*" and "*d*" parameters can not be determined. Going back to (c/d) does not simplify the problem.
So, using more terms from $P_2(t)$ has no sense.
In conclusion, $f_2(t)$ function does not represent a solution.

5. COMBINATIONS OF $f_1(t)$ AND $f_2(t)$ FUNCTIONS

[$f_1(t) \pm f_2(t)$] or $f_1(t)\, f_2(t)$ combinations do not verify No. 5 condition.
Instead, [$f_1(t)/f_2(t)$] relation using only the first terms of $P_1(t)$ and $P_2(t)$ polynomials and only *a* and *b* parameters represents a solution.

$$f_3 = a\left(1 + t^2\right) \mathrm{e}^{-bt^2} \tag{23}$$

Normalization condition is written applying integration by parts:

$$1=\int_{-\infty}^{+\infty} a(1+t^2)\mathrm{e}^{-bt^2}\,\mathrm{d}t = I_1+\frac{1}{2b}I_1 = I_1\frac{2b+1}{2b} = a\sqrt{\frac{\pi}{b}}(1+\frac{1}{2b});$$
$$a=\frac{1}{\sqrt{\frac{\pi}{b}}(1+\frac{1}{2b})} \tag{24}$$

with

$$I_1=\int_{+\infty}^{-\infty} a\mathrm{e}^{-bt^2}\,\mathrm{d}t \neq 1 \tag{25}$$

I_1 is different of 1 because it does not represent the normalization condition for $f_3(t)$!
Dispersion is calculated using also integration by parts:

$$\sigma^2=\int_{-\infty}^{+\infty} at^2(1+t^2)\mathrm{e}^{-bt^2}\,\mathrm{d}t = \frac{1}{2b}I_1 - \frac{a}{2b}\int_{-\infty}^{+\infty} t^3(-2bt\mathrm{e}^{-bt^2})\mathrm{d}t =$$
$$=\frac{1}{2b}I_1-\frac{a}{2b}\left|t^3\mathrm{e}^{-bt^2}\right|_{-\infty}^{+\infty}+\frac{3}{2b}\int_{-\infty}^{+\infty} at^2\mathrm{e}^{-bt^2}\,\mathrm{d}t = \tag{26}$$
$$=\frac{1}{2b}I_1+\frac{3}{4b^2}I_1 = I_1\frac{2b+3}{4b^2}$$

By dividing (27) by (26) it results:

$$\sigma^2=\frac{2b+3}{4b^2+2b} \tag{27}$$

A 2nd grade equation in b function of σ is obtained:

$$4\sigma^2b^2+(2\sigma^2-2)b-3=0 \tag{28}$$

with roots:

$$b_{1,2}=\frac{(2-2\sigma^2)\pm\sqrt{(2-2\sigma^2)^2+48\sigma^2}}{8\sigma^2}=\frac{(1-\sigma^2)+\sqrt{(1-\sigma^2)^2+12\sigma^2}}{4\sigma^2} \tag{29}$$

where "–" sign in front of radical has been eliminated because b could not have negative value.
σ value depends on measurement unit used in experiment so that the following limit situations could be considered:

$$\text{a) } \sigma \ll 1;\ b=\frac{1}{2\sigma^2} \tag{30}$$

and formula $\frac{(1+t^2)e^{-\frac{t^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}(1+\sigma^2)} \approx \frac{(1+t^2)e^{-\frac{t^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$ is obtained (31)

like the classical formula

$$\text{b) } \sigma >> 1;\ b = \frac{1}{2} \tag{32}$$

and formula $\frac{1+t^2}{2\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ is obtained (33)

like the reduced classical formula; appropriate graphic does not comply with No. 3 condition (see derivative for f_3).

A function using more then one terms of $P_2(t)$ could also be integrated but equation for b determination is more complex!

6. $f_3(t)$ FUNCTION DERIVATION

$f_3(t)$ function derivation should also be analyzed.

$$f\,'(t) = 2ate^{-bt^2} - 2ab(1+t^2)e^{-bt^2} = 2ate^{-bt^2}(1-b-bt^2) \tag{34}$$

Derivation is zero for t = 0, maximum of “Bell Type”. Derivation has two more roots obtained from:

$$bt^2 - b - 1 = 0\,; \quad t_{2,3} = \pm\sqrt{\frac{1-b}{b}} \tag{35}$$

In order to avoid a derivative with 3 real roots, $t_{2,\,3}$ should be imaginary roots, so $b > 1$.
From (29) formula is resulting:

$$\sqrt{(1-\sigma^2)^2 + 12\sigma^2} > 5\,\sigma^2 - 1 \tag{36}$$

By squaring and simplifying with σ^2 it results:

$$\sigma < \sqrt{\frac{5}{6}}\,;\ \sigma < 0.913 \tag{37}$$

Inserting $\sigma = \sqrt{\frac{5}{6}}$ in (29) formula it results:

$$b = 1 \tag{38}$$

and the 3 roots of derivative are equal to zero!

For $\sigma > \sqrt{\frac{5}{6}}$ a strange solution is, of course, obtained consisting in a graphic with two maxims being equal distanced from a minimum realized with $t = 0$. So, in this situation, $f_3(t)$ function does not comply with No. 3 condition.

In conclusion, even for $\sigma^2 < \frac{5}{6}$, it should be experimentally tested if a $f_3(t)$ type distribution density is ever obtained.

CONCLUSIONS

1. Demonstration starts from the experimental distribution in "Bell Type" form.
2. The five used conditions are naturally assigned.
3. Conditions provide two formulae:

   $a\mathrm{e}^{-bt^2}$ classical formula

   $a(1+t^2)\mathrm{e}^{-bt^2}$ to be experimentally tested
4. From physical point of view demonstration is simple; associated mathematic is a little complicated.
5. Demonstration is not based on the Binomial Distribution or Poisson Distribution.
6. Demonstration does not use mathematical approximations unlike some previous demonstrations from literature (DeMoivre, Laplace, Maxwell, Stevenson, Hagen, Roe).

REFERENCES

1. A. DeMoivre, *The Doctrine of Chance*, 1718.
2. P. S. Laplace, *Théorie analytique des probabilites*, 1812.
3. C. F. Gauss, *Méthode des moindres carrées*, 1820.
4. A. G. Worthing, F. Geffner, *Treatment of Experimental Data*, 1955, John Wiley & Sons, New York.
5. Sir J. Jeans, *An Introduction to the Kinetic Theory of Gases*, 1962, Cambridge at the University Press.
6. W. Feller, *An Introduction to Probability Theory and its Applications*, 1966, John Wiley & Sons, New York.
7. P. C. Stevenson, *Processing of Counting Data*, National Academy of Science – National Research Council, Report NAS-NS 3109, 1966, National Technical Information Service, Springfield, Virginia.
8. B. P. Roe, *Probability and Statistics in Experimental Physics*, 1992, Springer-Verlag New York.
9. N. Valcov, *Statistica in fizica experimentala*, 1998, ed. "Horia Hulubei", Bucuresti.
10. N. Ghiordanescu, *Metode de calcul şi simulare în fizica nucleară*, 1999, ed. Universitatii Bucuresti.

Annex 1

The following relation could be written:

$$1 = I^2 = \left[\int_{-\infty}^{+\infty} a\mathrm{e}^{-bt^2}\mathrm{d}t\right]\left[\int_{-\infty}^{+\infty} a\mathrm{e}^{-bn^2}\mathrm{d}n\right] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} a^2\mathrm{e}^{-b(t^2+n^2)}\mathrm{d}t\mathrm{d}n \qquad (1)$$

Change of variable in polar coordination is made:

$$t = r\cos\theta \qquad (2)$$

$$n = r\sin\theta$$

and it is obtained:

$$t^2+n^2 = r^2;\ \mathrm{d}t\mathrm{d}n = r\mathrm{d}r\mathrm{d}\theta \qquad (3)$$

It results:

$$I^2 = \int_0^{2\pi}\mathrm{d}\theta\int_{-\infty}^{+\infty} a^2\mathrm{e}^{-br^2} r\mathrm{d}r = 2\pi\left|-\frac{a^2}{2b}\mathrm{e}^{-br^2}\right|_0^{+\infty} = 2\pi\frac{a^2}{2b} \qquad (4)$$

so

$$a = \sqrt{\frac{b}{\pi}} \qquad (5)$$

Dispersion calculus is made by integrating by parts:

$$\sigma^2 = \int_{-\infty}^{+\infty} at^2\mathrm{e}^{-bt^2}\mathrm{d}t = -\frac{a}{2b}\int_{-\infty}^{+\infty} t\left(-2bt\mathrm{e}^{-bt^2}\right)\mathrm{d}t = \frac{a}{2b}\left|t\mathrm{e}^{-bt^2}\right|_{-\infty}^{+\infty} + \frac{a}{2b}\int_{-\infty}^{+\infty}\mathrm{e}^{-bt^2}\mathrm{d}t =$$
$$= \frac{1}{2b}I = \frac{1}{2b} \qquad (6)$$

It results

$$b = \frac{1}{2\sigma^2} \quad \text{and} \quad a = \frac{1}{\sigma\sqrt{2\pi}} \qquad (7)$$