The Unit Histogram Concept for Scarce Statistical Information

Radu D. RUGESCU
University Politehnica of Bucharest
spl.Independentei nr.313, RO-060042 Bucharest
rugescu@yahoo.com

***Abstract*—The new unit histogram concept is described as increasing with one order of magnitude the amount of statistical information extracted from experimental data. The new statistical technology is particularly suited for very scarce population samples, when it dramatically increases the information extracted from the available data. Otherwise, existing features of the population would remain inaccessible. The method could prove useful also for signal processing, when a high level of accuracy in detail rendition is required. The efficiency of the method is demonstrated in examples of experimental measurement of the combustion heat delivered by rocket propellants. Populations as small as of six readings, where the regular statistics is useless, are successfully processed and characterized.**

***Index Terms*—Unit histogram, scarce population statistics, squeezed information, unit filter**

I. Introduction

Histograms were introduced for long to provide a means of estimating the character of a statistical distribution of data (population), especially in connection with the Gaussian, normal distribution [26]. They are the regular histograms with fixed bins, almost exclusively used in different types of data and signal processing [11], [2], [28], [4], [27]. We address here a common set of such data from measurements or items, composing a univariate population. For bivariate data, the same regular technique is used [19]. For multivariate problems, the main concern [29] is the connectivity of individual columns, especially when their variance is different. Ingenious techniques were imagined [32], all of those being based on the fixed bin sampling method [14]. The compressed histograms for selectivity estimation in [14] are also of regular type. These applications are only oriented towards large sample populations, where data grouping does not disturb the analysis [25], [5], [15]. They seem unavoidable for a correct statistical image of the population. The authors in [23] only apply their mobile windows in updating the population.

Filtering techniques [4] extract the main frequencies of variable signals [24], [7]. The known methods are based on digital filters of Hamming type [18], [21], Hann or Blackman [24]. Good presentations of these techniques are found in [7], [10], [1], [20], [17], [9].

These methods are based entirely on the superposition of modifying windows, with a synthesized transfer function of given weighing laws that exaggerate the frequencies around an arbitrary basic, or *desired* frequency, ω, in order to underline the main modes and diminish the others. The start point in developing the filtering techniques is the requirement to reduce noise.

Even when dealing with large populations and with data vectors [22], [16], from images or sound [16] processing, like in person identification algorithms, for example [9], the construction of the histograms is performed on the same classical rules of data grouping. Filters, including digital, are imagined for real time application. Each moment the signal is then intentionally distorted within a band of frequencies.

The Hamming and Hann windows [6], [18], [10] multiply the harmonics around the desired frequency ω by the set of coefficients, depending on the parameter α,

$$w_H(n) = \begin{cases} \alpha - (1-\alpha)\cos\dfrac{2\pi}{N}(n+\dfrac{1}{2}), & n \in [\overline{1, N-1}] \\ 0, & \textit{in the rest}. \end{cases}$$

The number of covered harmonics N is the length of the filtering window, only related to frequencies from the signal. The slight difference between the two filtering windows is that the Hamming corresponds to α=0.54 while the Hann is for α=0.5 and nothing more [10]. Despite this minor difference, they were attributed different names.

The new method we present here induces, in fact, true reversed effects, each value of the statistical set being reproduced independently and in its real size, without any alteration, amplification or adjustment, opposite to the case of usual digital filters. Perhaps it represents mainly a *true-filtering* technique, by which the influence of other neighbouring values is dismissed and the true personality of each encountered measurement (reading) is immediately considered in the histogram, without waiting to be grouped with other items. Once the very detailed histogram is accomplished, it remains to be further analyzed and the effect of noise to be segregated by extra means, depending on the resulted, real statistical distribution.

All goes well with large amounts of inputs, when data bulking into fixed collector bins to determine the frequency of appearance is unimportant. The reduction of individual information does not matter too much.

II. Regular histograms

The statistical information is regularly processed through the crisp membership allocation method, by grouping the data of the set into a number of m equally sized collecting bins of width $2b$. The number of readings that fall within each bin j counts for the *frequency of appearance* h_i of that bulk of readings v_i, crisp appearance which could only equal either the value 0 or 1 for each reading.

Namely, the *integer* value h_i for the bulk within the given bin is,

Digital Object Identifier 10.4316/AECE.2009.03013

$$h_i(v_i;\, c_j, b) = \mathrm{int}\frac{v_i}{(c_j - b)} - \mathrm{int}\frac{v_i}{(c_j + b)}\ . \tag{1}$$

The symbol *int* is used for integer truncation. Besides, its given width $2b$, each j bin is defined by its centreline position c_j, starting with the first, given centreline c_1,

$$c_j = c_1 + 2b\,(j-1) \tag{2}$$

The hierarchy $v_n + a < 2\,(v_1 + a - b)$, based on a constant a, conveniently chosen, is assumed.

Referring to scalar sets, which are the basic stuff of usual experimental measurements, after summing the individual histograms (1) over the entire set

$$h(c_0, b) = \sum_{i=1}^{n}\sum_{j=1}^{m} h_i(v_i\,; c_j, b) \tag{3}$$

a bar diagram results (Figure 1), called the *histogram* of the set. Its aspect hangs on the width of the bins $2b$ and the location of the first bin, namely its centreline, c_1 [30], which are in fact two empirical constants.

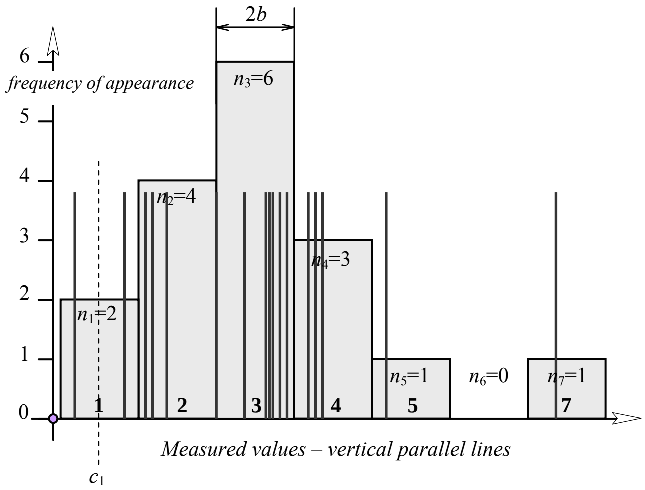

Figure 1. Grouped histogram with 7 bins for n=17 readings.

While histograms play a central role in experimental data evaluation for errors and confidence, in medical care statistics or in image and sound transmission, extensively used today in deep space communications and in person, signature, fingerprint, or voice recognition [11], [2], [12], [22], [9], [16], their proper definition is decisive. To have clear information by grouped histograms, the size of the population n must exceed 40 items [11], [8], due to the loss induced by grouping. The mean and standard deviation, computed over all values of the set, are given by:

$$\mu_0 = \sum_{i=1}^{n} v_i \Big/ n\,, \quad \sigma_0 = \sqrt{\sum_{i=1}^{n} (v_i - \mu_0)^2 \Big/ n} \tag{4}$$

either with initial data or with weighted centrelines of the bins, show the magnitude of errors and the confidence in the experiment [3]. The grouping of data into fixed bins wastes however a good part of the existing information and denatures the remainder. Identical histograms could prove belonging to basically different distributions (Figure 2), engaging, this way, inherent inaccuracy. For large populations, when corrections are introduced [11], the error is acceptable.

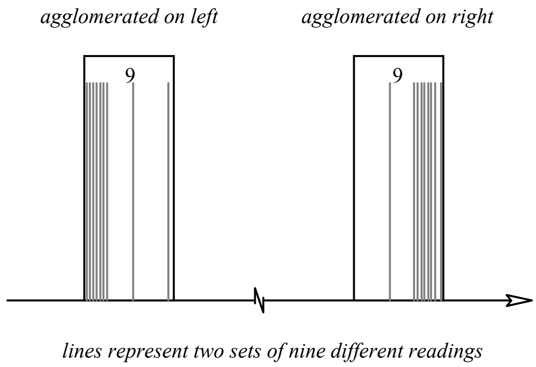

Figure 2. Grouping induced distortion into fixed bins.

Different spacing is hidden by grouping.

When the data go scarce however, grouping is no more possible and the regular histograms become unachievable. Consequently the representation of a scarce data set by a histogram comes out of practical reach. Data sequencing remains hidden.

III. Unit Histograms

A considerable improvement is introduced when, instead of assuming given positions to m fixed collecting bins, a unique sliding bin of equal width $2b$ is continuously moved along the data set. Each time a new reading enters the travelling bin window, the frequency is increased with one unit and vice versa, when the window exits. The individual histogram of every value v_i, centred on its very own reading v_i, namely the dual function of unit-step type with the edges on v_i-b and v_i+b,

$$h_i(v; v_i, b) = \mathrm{int}\frac{v}{v_i - b} - \mathrm{int}\frac{v}{v_i + b}\ \Big|\, v \in \Re,\ i = 1, n \tag{5}$$

will be added up to the current position v of the sliding bin,

$$h(v; b) = \sum_{j=1}^{i} \mathrm{int}[v/(v_j - b)] - \mathrm{int}[v/(v_j + b)] \tag{6}$$

to give finally the total *centreline wake* (CLW) histogram $h(b)$. Again the symbol *int* denotes truncation and the hierarchy $v_n + b < 2\,(v_1 - b)$ is considered. When this condition is not met from the beginning, the set $\{v_i \,|\, i = 1, n\}$ will be translated to the right with a constant value $a > v_n - 2v_1 + 3b$ [30] to comply with the given hierarchy, fact always possible.

The alternative of raising h_i right at the left margin of the travelling bin (not at centreline) for a new v_i reading encountered at its right side (*sliding bin pattern SBP*) proves less useful [30]. The new histogram is sensible to any change in an individual position of a reading, eliminating the intrinsic inaccuracy of the grouped histogram completely. Furthermore, when the mean and standard deviation are short-computed by replacing the original set with the centreline of the fixed bins, the intrinsic inaccuracies transmit into them. This doesn't happen with the new method and the standard deviation is computed more accurately. The following numerical examples show that the amount of statistical information revealed through the CLW histogram doubles the original scarce data information, besides eliminating the intrinsic inaccuracy.

Thus the travelling bin histogram produces no alteration of the initial data and secures a clearly enriched and personalized statistical picture.

IV. EXPERIMENTAL EXAMPLE WITH 17 PROBES

The new statistics was first used to analyze scarce laboratory measurements of the isochoric heat of combustion Q_v of a solid homogeneous rocket propellant, performed in a common 282 cm^3 calorimetric bomb under vacuum [30]. Samples from a nominal blend of triple base propellants of genuine origin were subjected to the investigation. Due to technical difficulties and other inevitable causes, the population v_i was relatively scarce. To induce enough confidence in the measurements, a strong statistical analysis was required, confronted however with the problem of the very limited amount of data: 6 readings in the first run and 17 readings in the second run. Under these circumstances, regular histograms with grouping are of no use and the new approach was required. The method of the unit histogram or of the travelling bin had thus emerged, proving the only tool for that difficult application.

The set of n=17 delivered values (vertical, short grey lines in Figure 2) are used here first to compare the new with the classical method. The ordered values (increasing) of the experiment are given in Table I.

TABLE I. MEASUREMENT WITH 17 READINGS
(*Calorimetric heat of combustion*)

Measurement No. i	Reading v_i (*cal/g*)
1	853.62
2	855.10
3	855.74
4	855.95
5	856.37
6	857.85
7	858.70
8	859.33
9	859.44
10	859.55
11	859.76
12	859.97
13	860.60
14	860.81
15	861.03
16	862.93
17	868.01

This scarce set is considered as a test for the efficiency of the new statistics, which are compared, first by grouping the data in seven collecting bins, with the equal width of 2b=2.33 units (cal/g in this case). The first bin starts at its centreline location c_1=854.36 namely at position 853.20 cal/g. As usual, the resulting histogram has a visibly coarse aspect (Figure 2), where the intrinsic details are hidden, due to this process of grouping. Additionally, variation of the fixed bins in size and position produce visible variations in the aspect of the regular histogram, due to variations of the number of recordings that fall within each bin.

The image of the distribution with unit histograms is completely different. With a travelling bin of 2.33 units in width, a new CLW histogram type is now built in Figure 3. Hidden details in the first, grouped histogram are, thus, impressively revealed.

Among these details, for example, the layout of a bimodal statistical distribution of measurements is well revealed in Figure 3, as a mixture of two separate normal distributions that may be analyzed independently, although the data set was that small.

The minute distribution (Figure 3), where 27 bins are formed, is typical for ordinary histograms with more than 50 readings; still they belong to a population of only 17.

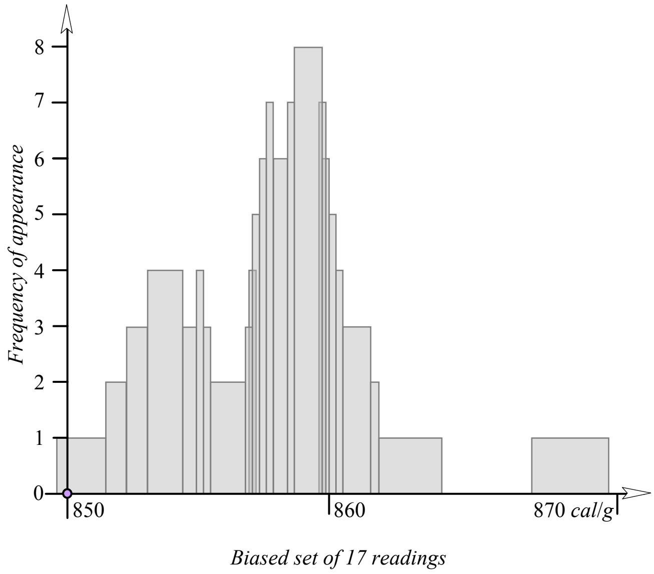

Figure 3. Multiple unit-CLW histogram for the data set in Table I.

To perform this separate weighing for the bimodal analysis, the set is split into the left six readings and the right twelve ones. The two Gauss-distributions are revealed and both are represented into the drawing in Figure 4.

Considering the right part of the set only of twelve items, the isolated reading from its right margin seems to be clearly affected by noise and two distributions were consequently computed and drawn. The first, comprising 12 elements, includes the extreme right element, while the other, with 11 elements only, excludes the right hand, heretic reading. This isolated value proves to produce a minor shift of 0.668 units of the mean of the partial set, while the modification of the standard deviation looks very large, almost double in the presence of the irregular reading. However, the dismissing of the isolated value remains rather arbitrary.

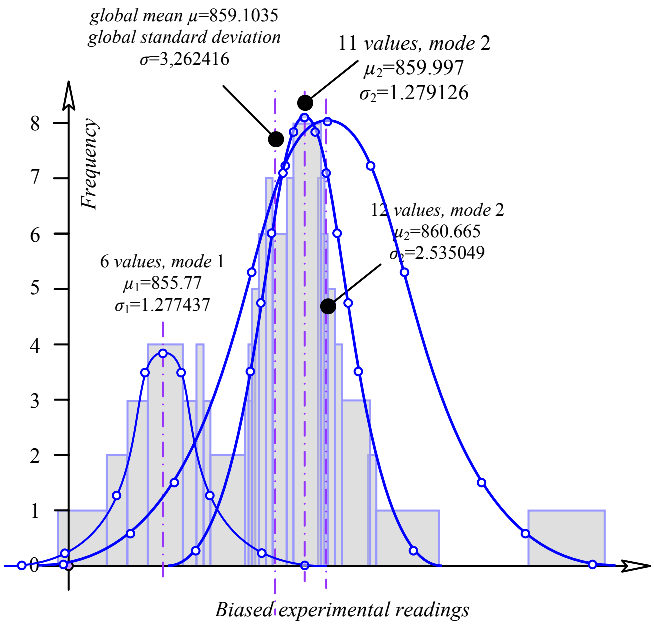

Figure 4. CLW histograms reveal the bimodal behaviour.

Returning to the general aspect of the new unit histogram in Figure 3, the comparison with the old, grouped histogram in Figure 2, gives the real scale of the improvement induced. The additional information of one order of magnitude is transformed into impressive pictures.

As a matter of consequence, the accuracy of computer thermo chemical simulations was also dramatically improved when these values for the heat of combustion were used to determine the standard enthalpy of formation of the cellulose ester with various degrees of nitration, largely used in the colloidal rocket propellants. The accuracy in the prediction of the specific impulse has reached the unusually low margin of 0.5%.

This comparison is a limit case, when for 17 readings the classical grouped histogram may still be built. For scarcer populations of data, the classical histograms can not be built at all, and we reach a new field regarding the full reign of the new unit histograms, as shown in the next example.

V. EXTREME EXAMPLE WITH ONLY SIX READINGS

A second example shows further how a very good CLW histogram can be built for a set of an incredible small number of only 6 readings, where the standard histogram of Gauss can not be built any more.

As no usual histogram can be conveniently attached to this very small amount of data, due to the very limited stuff available, it brings into attention the readings in Table II, containing only 6 elements of information. No other measurements could have been added to the recovered the data or could multiply the measurements.

Neither the known condition to have at least six collecting bins in the histogram [11], nor the condition to have at least 40 readings could have been met with this small amount of only six values and no information could thus be driven from the regular histograms. The new statistics enters in action as the only support for the successful analysis of the experiment.

TABLE II. HEAT OF COMBUSTION WITH 6 MEASURED VALUES

Sample no.	N_2, %	Q_v, cal/g
1	12,04	987,5
2	12,03	987,8
3	11,95	983,6
4	11,96	970,0
5	11,86	961,0
6	11,89	971,2

The mean value of the set equals 976.85 cal/g and the standard deviation divided by N for the set is equal to 10,0743 cal/g. The new histogram is built now for this scarce set (Figure 5), with different values for the width $2b$.

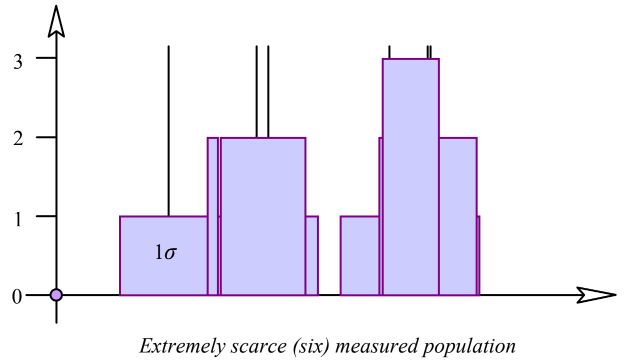

Figure 5. Detailed CLW histogram of the n=6 data set (*unit histograms with* 1σ *width*).

This is the new CLW histogram type that unexpectedly offers detailed information about this very small distribution. It reveals for example that a dual mode distribution is again present, which suggests that a repeated error in the measurements had been embedded. The left and right parts of the histogram resemble two adjacent normal distributions superposed. Considering each of them in the overall distribution up to an acceptable degree as a Gaussian normal distribution, two distinct sets are considered within the diagram. This example could say a lot about the extracting power of the CLW histogram. It proves especially suited as a tool for very scarce statistical populations. Each individual reading produces two modifications in the CLW histogram, one at its entrance and the second on its exit from the travelling bin window. As far as a group within a bin of the standard histogram involves at least five readings, the CLW histogram is a more detailed magnitude order.

Data in Figure 5 seem to belong to two distinct sets, where the conclusion regarding the dual mode population comes from the situation when the unit histogram, or the sliding bin, have the width of 2σ for the whole representation.

Further reduction or amplification of the width of the unit histogram has a strong effect upon the transformed image of the population. Some different values were added and the subsequent analysis had shown improvements or degradations of the result. The estimation of the result is still a matter of subjective evaluation and a fully automated quality evaluator is still under investigation. The same situation is valid for a lot of current methods of statistical analysis in image restoration [34], where the input normalized mean square error is used.

The effect of smaller widths upon the quality of the rendered histogram is first shown below.

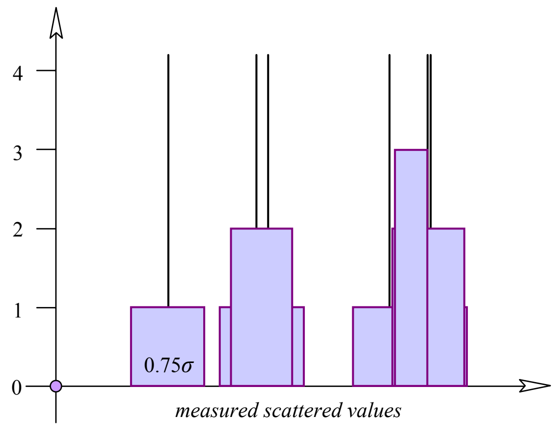

Figure 6. CLW histogram with 0.75σ width of the bin.

Figure 6 shows that for a width of 0.75σ the data are splitting into three groups, that means a lower correlation appears. The same observation preserves and even sharpens for a width of 0.5σ of the travelling bin (Figure 7).

In this limit case of data scarcity, the impact of the travelling bin width is very important, as the image of the CLW histogram is visibly changing with its width. Consequently, the question on the optimal size of the unit histogram to be recommended remains to be solved. Other aspects remain to be developed. One of those is the correlation properties of the new CLW histograms.

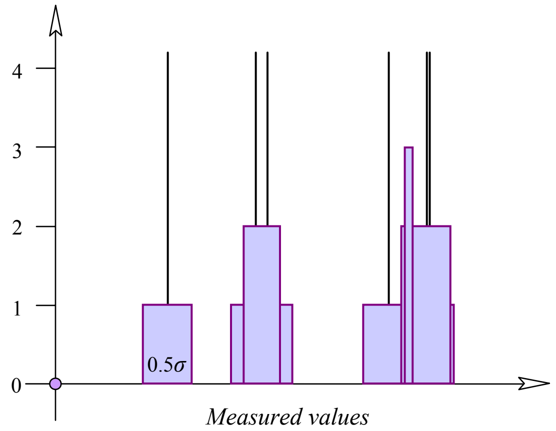

Figure 7. CLW histogram with the 0.5σ width of the bin.
Data appear as non-correlated.

VI. Wider bins and the distance to normal

The effect of a wider than 1σ collector bin upon the aspect of the travelling bin distribution, reveals that an optimum width may occur. Two sample distributions involving travelling bins of 1.22σ and 1.5σ were operated and they are drawn in Figures 8 and 9. They show that the variation of the density in the horizontal direction is approaching that encountered into a normal distribution with bimodal character. The 10 new points of reference, derived from the variable bins, are drawn in magenta in Figure 8.

These new points that enhance the representation are typically aligned over a bimodal appearance, as shown by the magenta curve that winds along the ten points of reference (Figure 8). This aspect does not require first attention for the moment and the entire set of six readings is treated as an entity.

Provided the set was purely Gaussian, its distribution would have been represented by the bell curve depicted in blue, in Figure 8. But the set is far from being normal; consequently, an appreciable distance rises up to the normal distribution.

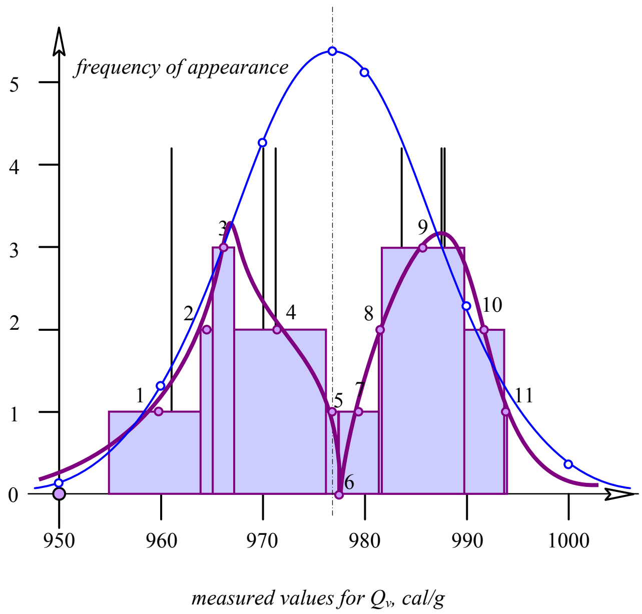

Figure 8. The 11 reference points and the distance to normal.
(*travelling bin with the width* 2*b*=12,3 *cal/g*)

As an objective measure of that distance we use the concept of the *distance between histograms*, first suggested by Shannon [41] and specifically introduced in the classical works of Kullback & Leibler [42] and Matusita [43]. However, different other approaches were elaborated [39], [40], all over the fixed bins histograms. An additional and strong inconvenience of these definitions is the fact that the distance is computed by multiplying the discrepant reading with the value of the measurement, which induces irrelevant inequity among the items of the histogram. When a different number of readings is encountered for some arbitrary value of the measurement, say 970 cal/g in the present example, this very value is in fact unimportant and if the same difference is recorded for 950 cal/g, the difference between the measurements is neither greater nor smaller, contrary to the evoked definitions [38]. We shall correct the statement.

On the other hand, all present definitions suppose that the measurements are quantified, in other words they are grouped in some wider or tighter fixed bins. In the new treatment introduced here, no quantification is envisaged however, and this sets a major challenge in finding the convenient correspondence between two different travelling bin histograms. To solve this conflict, we chose to first compare each CLW histogram with its normal distribution, obviously continuous. Consequently, the difference in occurrences between the actual, discrete measurement and the normal distribution curve is computable in any point.

This way, we introduce a new distance definition, valid for arbitrarily dissipated values of the samples,

$$D = \sqrt{\sum_{r=1}^{R}[H(v_r) - N(v_r)]^2 \Big/ R} \tag{7}$$

and actually based on the density of probability rather than on the values of the readings. Here, $H(v_r)$ denotes the occurrences of the node value v_r, while $N(v_r)$ is the Gaussian distribution in the same node v_r

$$N(v_r) = \frac{A}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x_i - \overline{x})^2}{2\sigma^2}\right) \tag{8}$$

In order to level inequities, the distance is normalized with respect to the total number of reference nodes R. For the case study in Figure 8, the operating values are given in Table III.

TABLE III. REFERENCE NODES FOR THE HISTOGRAM 8.
(HALF WIDTH OF THE TRAVELLING BIN B=6.15)

r node	Measured Q_v, *cal/g*	v_{margin}, *cal/g*	v_r, *cal/g*	$H(v_r)$	$N(v_r)$
1	961,0	954,85	959,35	1	1,54833
2	970,0	963,85	964,45	2	3,28188
3	971,2	965,05	966,10	3	3,96139
4	983,6	967,15	971,65	2	6,12696
5	987,5	976,15	976,75	1	6,99966
6	987,8	977,35	977,40	0	6,98958
7		977,45	979,40	1	6,77931
8		981,35	981,50	2	6,29268
9		981,65	985,70	3	4,75909
10		989,75	991,70	2	2,36201
11		993,65	993,80	1	1,69982
		993,95			

The position of the margins $v_{margin} \equiv v_m$ and of the middle v_r (the reference node) of the variable width bins j result from

$$v_{m,j} = Q_{v,i} \pm b \qquad v_r = (v_{m,j} + v_{m,i+1})/2 \tag{9}$$

The value 3.817338 results for the distance between the travelling bin histogram and the normal distribution of set.

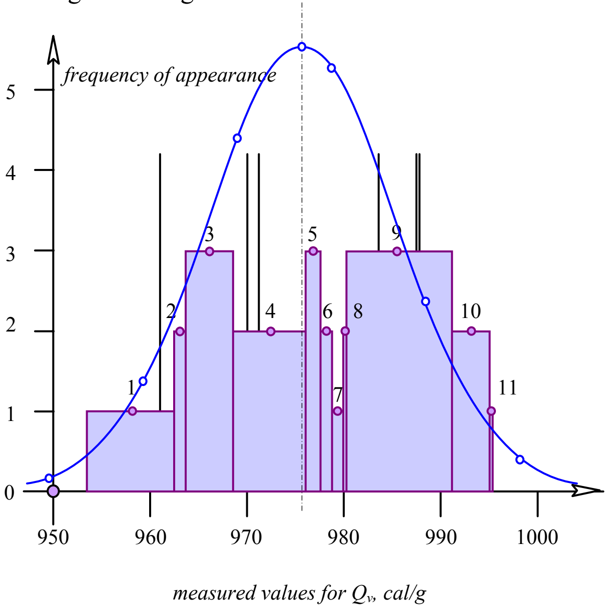

Figure 9. The travelling bin histogram with b=1.5σ.

A distance of 4.201127 results in the latter case in Figure 9, after yielding the computations. The representation in Figure 8, with the moving bin of 1.22s is more appropriate for the set.

VII. CONCLUSION

We obtained 859 *cal/g* for the combustion heat of A-100 triple base propellant by using laboratory measurements and the enhanced statistical tool of *travelling bin histogram*. A good standard deviation of σ_0=0,791 *cal/g* appeared. A grouped histogram should not have been structured so far. Not only the new histogram was clearly structured, but it also revealed a slightly bi-modal distribution, which suggests that a small systematic error was introduced into the measurements.

The source could not have been established. Including this uncertainty, the accuracy of the measurements entered well within the requested technical accuracy. The new type of travelling bin histograms behaves as a special relevant statistical tool, opening new general applicability doors in the technology of measurements. The resulting histogram doubles the amount of information of the primary data set and rises the information on classical, regular histograms with one order of magnitude, transforming this tool into a recommendable means for the statistical analysis in general.

This method was successfully used in processing statistical sets as scarce as of six readings only, producing a clear CLW histogram, where the classical statistics is impossible. As the CLW histogram is a variable problem, the only thing to do is to properly choose the width of the collecting bin as a problem of optimum value. This aspect is under current development by the author, based on the so-called *distance between histograms* [35], [8], [22], [37]. Similar extensions are expected with applications to vector or multivariate data sets, typical for complex signal processing, where similar advantages are envisaged when convenient assumptions are set. The demonstration of the utility of this natural histogram is thus considered for unvaried problems, as simpler and more relevant. Results from application of the new technology are obviously welcome.

As far as the randomness in collecting bins positioning is removed with the new travelling bin, this method reduces the statistical analysis to a variable problem. The correlation of data is strongly emphasized, on the basis of the very value of the individual readings. Consequently, the new travelling bin histograms are sensible to the slightest change in reading values, rendering a really *personalized problem*.

Extensions are expected in other fields like the spectral entropy application, mainly in important medical care data processing [36], [13] and to bi-varied S-distributions utilizing copulas [37] or to multi-varied S-distributions [28], [32], [8] and vectors of data for example, where the new method could prove largely profitable.

ACKNOWLEDGMENT

The publication of the new unit histogram concept by the author was encouraged by the fruitful discussions, in 2007, with the fellow researchers from Texas A&M University, the Aerospace Engineering Department, whom he addresses thanks.

REFERENCES

[1] Allen, J.B. and L.R. Rabiner, "A unified approach to short-time Fourier analysis and synthesis", Proceedings of the IEEE, ISSN 0018-9219, 65, Issue 11(1977), pp. 1558-1564.
[2] Amabile, T., "Describing distributions, Consortium for Mathematics and Its Applications", Chedd- Angier Production Co., 59 min. Videorecording, KAMU-TV, College Station, Texas USA, 1989.
[3] Apostolescu, N. and D. Taraza, "Basics of experimental research in thermal machines", pp. 443-472. Ed. D. P., Bucharest 1979.
[4] Austin, J.D., D.W. Scott, "Beyond histograms: average shifted histograms", Math. Comput. Educ., 25, 1(1991), pp. 42-52.
[5] Bajramovic, F., Ch.Grassl and J. Denzler, "Efficient Combination of Histograms for Real-Time Tracking Using Mean-Shift and Trust-Region Optimization", in Kropatsch, W. Sablatnig, R., Hanbury, A. (Eds.): DAGM 2005, LNCS 3663, Springer-Verlag Berlin Heidelberg 2005, pp. 254-261.
[6] Basseville, M., "Signal Processing", 18(1989), pp. 349-369.
[7] Berghaus, D., "Numerical Methods for Experimental Mechanics", Kluwer Acad. Publ., ISBN 0-7923-7403-7Dordrecht, 312 p. 2001.
[8] Brdiczka, O., J. Maisonnasse, P. Reignier and J.L. Crowley, "10th Int. Conf. KES 2006", Bournemouth, UK, Oct. 9-11, 2006, Part II, Springer ISBN 978-3-540-46537-9, p. 162.
[9] Brunelli, R. and D. Falavigna, "Person Identification Using Multiple Cues", 17, 10, pp. 955-966, 1995.
[10] Ciochina, S., "Numerical Processing of Signals, Chapter 10: Fast Algorithms to Perform Convolution and the Discrete Fourier Transform", University Politehnica of Bucharest (in Romanian), 2006, site address www.comm.pub.ro/curs/pns/cursuri/capitolul2p1.pdf.
[11] Cook, N.H. and E. Rabinovicz, "Physical Measurement and Analysis", p.29-68. Addison-Wesley Publishing Co., Reading, Massachusetts, 1963.
[12] Coolidge, F.L., "Statistics: a gentle introduction", 2nd ed., Sage Publications ISBN 1412924944, Thousand Oaks, CA, 2006.
[13] El-Samahy, E., M. Mahfouf and D.A. Linkens, "Artificial Intelligence in Medicine 38", pp. 257-274, 2006.
[14] Fuchs, D., Z. He and B.S. Lee, "Info. Sci. 177", pp. 680-702, 2007.
[15] Gibbons, Ph. B., Y. Matias and V. Poosala, "ACM Transactions on Database Systems", 27, 3(Sept. 2002), pp. 261–298.
[16] Gish, H., J. Makhoul and S. Roucos, "Vector quantization in speech coding", Proc. IEEE, 73, 11, pp. 1551-1588, 1985.
[17] Griffin, D. and L. Jae., "Signal estimation from modified short-time Fourier transform", IEEE Transactions on Signal Processing, ISSN 0096-3518, 32, Issue 2(1984), pp. 236-243.

[18] Hamming, R.W., "Numerical methods for scientists and engineers. International Series in Pure and Applied Mathematics", McGraw-Hill Book Co., Inc., New York-San Francisco, Calif.-Toronto-London, xvii+411 p., 1962.
[19] Hamalainen, J.S., "IEE Proc.-Sci. Meas.", Technol., 151, 6 (Nov. 2004), pp. 492-495.
[20] Harris, F.J., "On the use of windows for harmonic analysis with the discrete Fourier transform", Proc. IEEE, 66, 1, pp. 51-83, 1978.
[21] Jackson, L.B., "Digital Filters and Signal Processing", Boston, MA: Kluwer A.P, pp. 85-92, 1986.
[22] Juhee, S., "Bootstrapping in a high dimensional but very low sample size problem", Ph. D. Thesis, Texas A&M University, USA, 2006.
[23] Lam, E. and K. Salem, "Proceedings of the 9th International Database Engineering & Application Symposium", (IDEAS-2005).
[24] Lobos, T. and J. Rezmer, "Real-Time Determination of Power Systems", Proceedings of the lEEE Instrumentation and Measurement Technology Conference, Brussels, Belgium, June 4-6, 1996, pp. 756-761.
[25] Lu, Z.-M., H. Burkhardt, "Electronics Letters", 18th August 2005 41, 17.
[26] Matusita, K., "The Annals of Mathematical Statistics", 26, 4 (Dec. 1955), pp. 631-640.
[27] Obraztsov, S.M., Yu.V. Konobeev, G.A. Birzhevoy and V.I. Rachkov. J. "Nucl." Mat., 359(2006), pp.263-267.
[28] Plackett, R.L. J. Amer, "Statistic. Assoc.", 60, No. 310 (June 1965), p. 516.
[29] Rubner, Y., C. Tomasi, and L.J. Guibas, "A Metric for Distributions with Applications to Image Databases", pp. 59-66, 1998.
[30] Rugescu, R. D., "Revista de Chimie", ISSN 0034-7752, Ed. Chiminform Data S. A., 59, 6(2007), pp. 579-581.
[31] Rugescu, R. D., V. Silivestru, and S. Aldea, "Natural Histogram for High Definition Data Statistics", Paper KD-005681-4610, Proceedings of the 6th IEEE International Conference on Industrial Informatics, INDIN-2008, 13-16 July, Daejeon, Korea (CD, pp. NA).
[32] Serratosa, F. and A. Sanfeliu, "N. Computer Sci.", 3773; 10th Iberoamerican Congress on Pattern Recognition, CIARP-2005, p. 1027.
[33] Su, Y.T. and Ju-Ya Chen, " ", IEEE Transactions on communications, 48, 8(Aug. 2000), pp. 1338-1346.
[34] Wang, Jung-Hua, Liu, Wen-Jeng, and Lin, Lian-Da, "Histogram-Based Fuzzy Filter for Image Restoration", IEEE Transactions on Systems, Man, and Cybernetics—Part B, Cybernetics, Vol. 32, No. 2, April 2002.
[35] Wilcox, R.R., "Applying contemporary statistical techniques", Academic Press ISBN 0127515410, San Diego, CA.
[36] Xiaoli, Li., "Lecture Notes in Computer Science/Neural Information Processing", Springer-Verlag, Volume 4234/2006, (The 13th International Conference on Neural Information Processing-ICONIP2006, HK. 2006 Oct.3-6.), pp. 66-73.
[37] Yu, L. and E.O. Voit, "Computational Statistics & Data Analysis", 51, 2006, pp. 1822-1839.
[38] Cha, Sung-Hyuk, Srihari, Sangur N., "On measuring the distance between histograms", Pattern recognition, 35(2002), pp. 1355-1370.
[39] Serratosa, F., and Sanfeliu, A., "A Fast and Exact Modulo-Distance Between Histograms", D.-Y. Yeung et al. (Eds.): SSPR&SPR 2006, LNCS 4109, pp. 394-402, 2006.
[40] Serratosa, F., and Sanfeliu, A., "Signatures versus histograms: Definitions, distances and algorithms", Pattern recognition, 39(2006), pp. 921-934.