

Treatment of Experimental Data with Discordant Observations: Issues in Empirical Identification of Distribution

G. Barbato¹, G. Genta¹, A. Germak², R. Levi¹, G. Vicario³

¹DIGEP, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy, gianfranco.genta@polito.it

²Istituto Nazionale di Ricerca Metrologica (INRIM), Strada delle Cacce 91, 10135 Torino, Italy, a.germak@inrim.it

³DISMA, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy, grazia.vicario@polito.it

Performances of several methods currently used for detection of discordant observations are reviewed, considering a set of absolute measurements of gravity acceleration exhibiting some peculiar features. Along with currently used methods, a criterion based upon distribution of extremes is also relied upon to provide references; a modification of a simple, broadly used method is mentioned, improving performances while retaining inherent ease of use. Identification of distributions underlying experimental data may entail a substantial uncertainty component, particularly when sample size is small, and no mechanistic models are available. A pragmatic approach is described, providing estimation to a first approximation of overall uncertainty, covering both estimation of parameters, and identification of distribution shape.

Keywords: Distribution shape, outlier, uncertainty, confidence region, gravimetry

1. INTRODUCTION

RECORDS OF MEASUREMENT results and experimental data often include discordant observations, that is outliers, bad data, mavericks. Their treatment implies choice among such options as *incorporation* along with bona fide values, *outright rejection*, and *accommodation* by adopting more comprehensive models; *robust methods* shielding estimated terms from adverse influences may also be adopted [1]. Identification of discordant observations, initially left to subjective evaluation, became the goal of a number of methods based upon some sort of statistical approaches [2], whose plurality caters for a variety of premises, and personal orientations. A model is usually assumed, covering among other aspects data distribution; should discordance from expected value either exceed some limit assessed in terms of sample statistics, or estimated probability of occurrence of some extravagant result fall below a given threshold, the observation concerned is conventionally tagged as discordant. A special cause may actually have produced the discordant observation, or it may correspond to a distribution tail, unlikely to show up but by no means less legitimate than the central mound of more frequently observed results.

Outlier treatment implies, as a support for decision making, assessing how unlikely observing such a deviation would be, owing to chance variation only. Features of underlying distribution are estimated in terms of those of a sample, whose size may entail getting a fairly good evaluation of central tendency, a reasonable one of spread, at least around middle values, and but a hazy picture of shape, particularly towards tails. Predicting probability of discordant values on such shaky foundations may be an awkward undertaking. Two approaches are sometimes practicable, one whenever a substantial data base affords modeling distribution adequately over a comprehensive range, and another when background information on the phenomena under examination provides theoretical support

for selection of a distribution, regardless of size of the data base. Often, however, neither is available.

Admittedly, there is a yawning gap between knowledge provided by a scanty amount of experimental evidence, and capability of making meaningful predictions concerning discordant values. Expert knowledge may be called upon to provide what is needed to complement statistical evaluation of experimental data, in the process of drawing a realistic conclusion from less than compulsory evidence, following the logical lines successfully developed for evaluation of measurement uncertainty [3], see also [4], [5]. All that is required is a mechanism allowing smooth blending of information provided by the sample at hand, and the body of specific knowledge accumulated and agreed upon by experts. Families of empirical distributions may provide an algebraic coat if required. Furthermore, combining together information gathered from both statistical evaluation and case specific knowledge may lead to inferences provisionally acceptable to qualified observers, until substantial additional information is added to the existing body, in the incremental, never ending sequential learning process typical of experimental science. Such an approach may well be disputed as arbitrary, a legitimate criticism. No less arbitrary is, however, a fairly popular axiomatic *modus operandi*, consisting in assuming on shaky evidence of a convenient distribution shape, and then dealing with it as revealed truth.

When dealing with distributions fitted to experimental data, the latter providing estimates of parameters, uncertainty is often assumed as affecting such estimates only, disregarding what pertains to identification of underlying distribution, whose shape may be tentative only for want of cogent evidence. Taking explicitly this into account, in the overall uncertainty evaluation process, also a component pertaining to identification of shape appears to make sense, as it may turn out to account for a substantial contribution. An empirical approach might first try to locate approximately the candidate distribution on the skewness-

kurtosis plane in terms of sample statistics ([6], [7]), and then assess bounds of a reasonable joint confidence region in the mean-standard deviation-skewness-kurtosis space, based upon estimates of the first four moments and relevant uncertainties. Should, e.g., symmetry be taken for granted, the confidence region on the skewness-kurtosis plane would shrink down to a segment on the ordinate, including Student's *t* distribution when the normal one is the first candidate. In the general case, an assumed array of candidate distributions may be modeled resorting to subgroups of empirical distributions such as Johnson's or Pearson's type families ([8], [9]), with a range of parameters consistent with bounds of confidence region.

An example is presented concerning treatment of a substantial set of measured values of free fall acceleration, recently obtained with a rise-and-fall absolute gravimeter [10] at Istituto Nazionale di Ricerca Metrologica (INRIM); several discordant observations were also recorded, due to a peculiar pattern of parasitic effects. The physical model considered takes into consideration, as main influence factors, gravity acceleration gradient and residual air drag, experiments being carried out in a vacuum chamber. However, other disturbances appear to affect flight pattern and motion-time measuring system, particularly rotation of the flying reflector and residual movement or vibration of interferometric system's stationary reference, provided by a reflector carried by a seismometer's pseudo-inertial mass, exploited to filter out the vertical component of parasitic motion imparted to the gravimeter's supports. Identification of the statistical distribution of effects induced by such disturbances, and estimation of relevant parameters, is affected by a number of hardly predictable factors, such as floor vibration induced by human activity, vehicles, machinery and small/remote earthquakes. Self-induced vibrations occurring during launch may also produce parasitic movement of the pseudo-inertial mass, and residual friction on the moving stage of the launch system may impart rotation on the flying interferometric reflector.

Application of some criteria currently used for outlier detection - and exclusion - is shown to lead to a fairly broad range of results, some of them implying sizable distortion of experimental distribution. A simple, two distribution model was found adequate for describing the pattern of results observed, enabling estimation of parameters pertaining either to substantially disturbance free operation of the gravimeter, or affected by parasitic influences as summarized above.

2. METROLOGICAL PROBLEMS IN ABSOLUTE MEASUREMENT OF GRAVITY ACCELERATION

In a modern absolute ballistic gravimeter, major contribution to data scatter is due to floor vibration, amounting typically to some tens of microgal ($1 \mu\text{Gal} = 1 \times 10^{-8} \text{ms}^{-2}$), according to gravimeter, and measurement site. Therefore, the estimate of the measurand (the free-fall acceleration) is routinely obtained by averaging some hundreds of individual measurement results. The present accuracy limit, due to reproducibility of results, is believed to lie at the microgal level, corresponding to about one part in 10^9 of the Earth's gravitational acceleration. An

absolute measurement of a physical quantity requires a detailed understanding of the influence factors affecting its realization. Special attention should be given to those effects that remain almost constant during the experimental activity and therefore are difficult, or almost impossible to identify and correct. A proper uncertainty evaluation should include also these effects, which do not contribute to data scattering. In ballistic gravimetry, the two dominating effects besides floor vibrations are the centripetal, and the Coriolis acceleration [11].

These effects are evaluated in the case concerning the ballistic gravimeter IMG-02, developed and tested at INRIM [10]. Absolute ballistic gravimeters are based on the reconstruction of the vertical trajectory followed by a test body in vacuum; IMG-02 instrument adopts the symmetrical rise and fall method, where the test body is thrown vertically upwards (Fig.1). Laser interferometry is used to determine the trajectory of the flying body, which acts as the moving reflector *M* in a vertically oriented arm of a modified Mach-Zehnder interferometer, whose reference reflector *R* is supported by a seismometer's inertial mass. Optical fringes at the interferometer are converted to electrical signals by a photodetector, while time values corresponding to selected positions of the test body during its flight are obtained by timing a recurring phase of the signal oscillation.

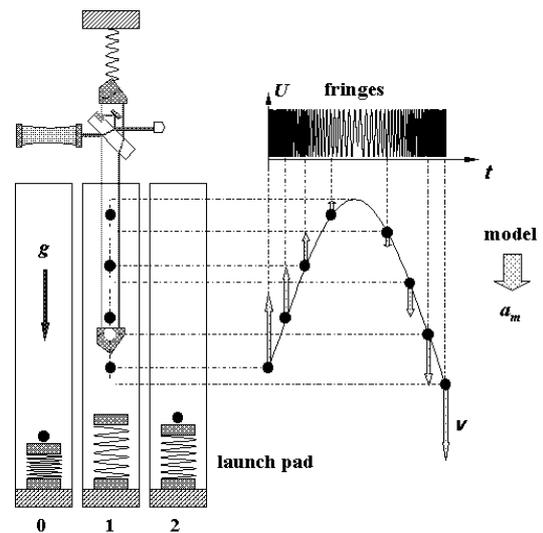


Fig.1. The symmetrical rise and fall method. Time 0, take off produced by the spring (as reaction floor vibration). Time 1, flight phase monitored in time with a rubidium clock and in position with a modified Mach-Zehnder interferometer with the reference reflector kept nearly steady on the arm of a seismometer.

Relating the space-time coordinates determined in this way to a number of parameters, a measurement model is obtained, from which the law of motion of the test body, and hence gravity acceleration *g*, is determined [12]. First and foremost among model parameters is the measurand, i.e. acceleration *g* experienced by the moving reflector during free rise and fall under the influence of gravity. The law of motion considers main known influence factors, described and thoroughly analyzed in the literature ([11], [13]), such as gravity acceleration gradient and residual air drag,

experiments being carried out in a vacuum chamber; however, to isolate hidden systematic factors, evidenced in international comparisons, residual variations - mainly due to floor vibration and centripetal and Coriolis accelerations of the test body - must be accounted for, and deducted from result variability. An analysis of some effects on gravity measurements [14] was performed, applying the Monte Carlo method under reasonable assumptions, starting from possible experimental signs of effects of Coriolis and centripetal accelerations (Fig.2).

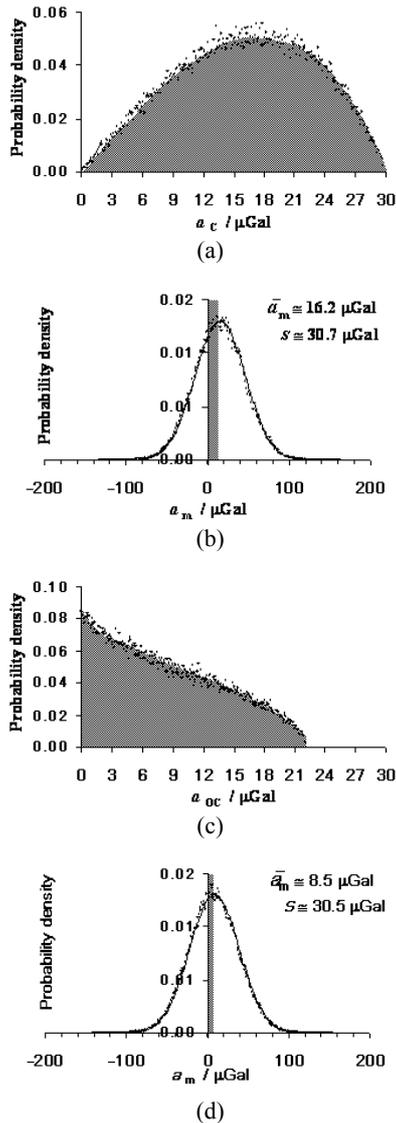


Fig.2. Expected distribution of the Coriolis acceleration (a) and the relevant experimental data (b). Expected distribution of the centripetal acceleration (c) and the relevant experimental data (d).

These effects are expected to produce a larger variability when compared to groups of results obtained in particularly favorable conditions. Experimental data exhibit the typical pattern shown in Fig.3. For the sake of expediency, data are coded by multiplication by 10^4 , after deducting the median (see section 2.A). An interpretation may be in terms of a sum of factors as previously described, justifying variation

of the central part in the range of about $\pm 100 \mu\text{Gal}$. However, tails appear to be mainly affected by other unidentified accidental effects, such as floor motion, anomalies in projectile's trajectory, data acquisition and processing, and other irregular events, producing results showing up as discordant observations.

Estimation of moments of experimental distribution is awkward owing to the striking similarity with Cauchy distribution, which turns out to provide the best fit among a number of theoretical and empirical distributions according to, e.g., Kolmogorov-Smirnov and Anderson-Darling tests. Inclusion or exclusion of a few data lead to inordinate variation in estimates of moments, furthermore, no physical justification for Cauchy distribution of data was found hitherto. An alternative explanation, consistent with experimental results, assumes contamination of a normal distribution, pertaining to the vast majority of data obtained under favorable conditions, by a fairly smaller set of data apparently affected by substantially larger disturbances, assuming again normal distribution, Fig.3 (a) and (b).

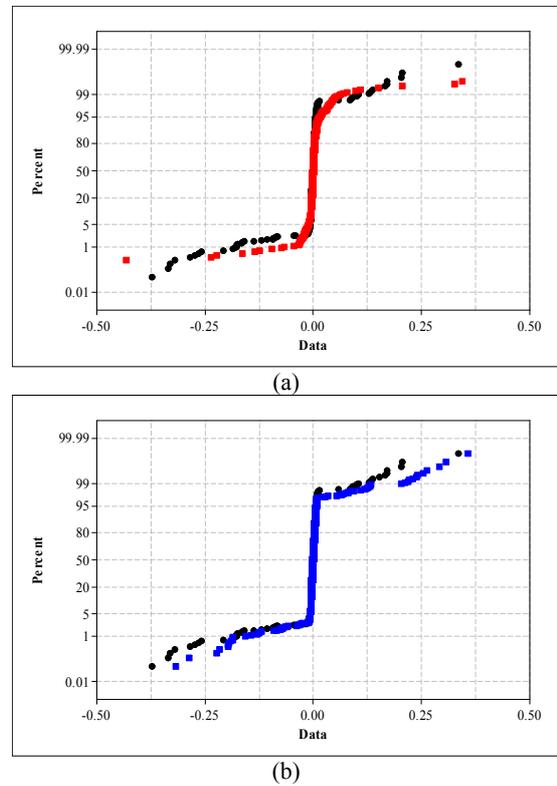


Fig.3. Normal probability plots of over a thousand measured values (coded) of gravity acceleration with some discordant observations (black), approximated in terms of a Cauchy distribution (red) with location and scale parameters equal to 0 and 2.3×10^{-3} respectively (a), and a mixture (in a proportion of about 1:16) of two normal distributions (blue) both with mean equal to 0, and standard deviations equal to 1.5×10^{-1} and 3.4×10^{-3} (b).

Identification and treatment of outliers

The data set referred to above deals with over a thousand measured values of the gravity acceleration recently obtained with INRIM's ballistic gravimeter IMG02. Values ranging from about 9.80531 to 9.80538 were coded

by multiplication by 10^4 after deducting the median of approximately 9.80534. Some descriptive statistics are shown in Table 1.

Tab.1. Descriptive statistics for a set of measured values of gravity acceleration (coded).

N	<i>mean</i>	<i>std. dev.</i>	<i>IQR</i>
1192	-0.00223	0.037	0.00501
$x_{(1)}$	Q_1	Q_3	$x_{(n)}$
-0.37329	-0.00271	0.00230	0.33432

Fig.3 shows normal probability plots of values of coded gravity acceleration; at either tail a number of discordant results may be observed, mainly traceable to such anomalies as referred to above. Outliers are identified according to Chauvenet's, Grubbs', and IQR (interquartile range) methods; another based upon extreme value distribution [15], and a modification of IQR method were also considered [16], see also [2], [17].

For both whisker limits, namely -0.0102 and 0.0098, being readily computed, IQR method yields 68 outliers. Modified IQR method yields -0.0150 and 0.0146 as limits, respectively, detecting 50 outliers. Applying Chauvenet's criterion, the probability of having an accident is approximated to 0.04%, so the acceptance interval, i.e. $[L_l, L_u] = [-0.132, 0.128]$, leads to identification of 30 outliers in the first iteration. Grubbs' test and the method based upon distribution of extremes are then compared adopting for both the risk of 0.04%, in order to make comparison meaningful, detecting in the first iteration 12 outliers in either case. Acceptance intervals are:

- $[L_l, L_u] = [-0.198, 0.194]$ for Grubbs' test;
- $[L_l, L_u] = [-0.190, 0.186]$ for the method of extremes.

Both methods appear to behave similarly, detecting a substantially lower number of outliers than Chauvenet's criterion.

When the outlier detection process is applied iteratively, with a fixed risk of error of 0.04%, the following results are obtained:

- both Chauvenet's criterion and method of extremes stop detecting additional outliers after the fifth iteration, while Grubbs' test stops after the sixth;
- all in all, Chauvenet's criterion detects 54 outliers, Grubbs' test 49 and the method of extremes 48.

IQR method (albeit at another level of risk), and to a lower extent Chauvenet's, appear to identify as outliers a number of values substantially larger than modified IQR method, Grubbs' test and the method of extremes.

Evidence of rather abusive exclusion of outliers appears for both IQR and Chauvenet's methods [16], suggesting an empirical distribution shape radically deviating from the original data set. On the other hand, by retaining a few borderline experimental values, modified IQR, Grubbs' and the extreme values method manage to preserve some typical features of the original data set, while at the same time effectively shielding estimates of main population parameters – central tendency, and scatter - from outlier induced bias.

3. EMPIRICAL IDENTIFICATION OF DISTRIBUTION SHAPE

When only a data set, and precious little else, may be exploited for identification of the shape of underlying distribution, substantial uncertainty applies particularly when a limited number of observations is available, as is not uncommon in experimental work. Such a component may account for a sizable part of type A uncertainty pertaining to inferences drawn in terms of that distribution, besides contributions pertaining to estimates, e.g., of moments. While the latter component is routinely estimated in terms of sample statistics using established procedures, the former is often conveniently dispensed with, assuming data distributed according to a family conveniently shaped and easy to handle, without further ado. Analytical evaluation may be awkward if at all possible, owing to a lack of specific information, typical of situations where no reliable mechanistic models are available; a solution is presented addressing the statistical issue on a pragmatic basis, following a simple way capable of leading to quantitative results. A realistic estimate of that uncertainty component, obtained with a single significant digit, may be considered an improvement over gut feelings only.

Sample size may be reasonably required to be large enough to allow at least crude estimation of the first four moments of the distribution. The first two sample moments lead to defining boundaries of a joint confidence region for estimates of parent population's location and spread parameters, e.g., on the $\mu - \sigma$ plane; another joint confidence region pertaining to some features of underlying distribution may be mapped on the $\beta_1 - \beta_2$ plane in terms of non-dimensional sample estimates of skewness and kurtosis ([6], [7]). Several theoretical distributions may then be found to fit the bill, as well as a range of empirical ones such as Johnson's or Pearson's type families ([8], [9]), within a parameter range consistent with the relevant confidence region at the level considered. Irrespective of whether a theoretical distribution is eventually fitted to data, or an empirical one, the goal is to find out expeditiously how to obtain realistic confidence bands, taking into account uncertainty in estimates of parameters as well as in identification of distribution shape.

According to the information available, and properties of the problem considered, one may end up in the simplest case with a three dimensional confidence region – e.g., for a symmetrical distribution defined in terms of location and scale parameters, and kurtosis – or with one in $k > 3$ dimensions in the general case. Points on confidence region's boundary surface identify distributions compatible with data; the envelope of their cumulative distribution functions defines a confidence band pertaining to type A overall uncertainty, comprehensive of components pertaining to estimation of parameters, and identification of model form.

Determination of a joint confidence region for substantially independent parameters concerning location, and spread, is simple enough as long as a number of convenient assumptions are accepted; to a first approximation a trapezoidal region in the $\mu - \sigma$ plane may be adopted [18], where every apex corresponds to a limit condition. Similar considerations do not, however, apply to

measures of asymmetry, and kurtosis, whose sample values may offer rather poor estimates of those pertaining to population, particularly for small sample size. For a symmetrical distribution the statistical index β_1 (and its square root, preferred in some instances) vanishes; the reverse is not necessarily true, as non-symmetrical distributions whose odd order moments vanish are known [19]. Sometimes $\gamma_2 = \beta_2 - 3$ is adopted as a measure of peakedness; while for a class of regular, symmetrical distributions $\gamma_2 > 0$ entails a more flat topped shape than the normal one, and a more peaked one for $\gamma_2 < 0$, such considerations may not necessarily apply to other symmetrical or skew distributions. Sample estimates b_1 and b_2 of β_1 and β_2 turn out to be biased and far from independent, particularly for moderate sample size; furthermore, their distribution – affected by that underlying experimental data – does not lend itself to be covered under the convenient cloak of the central limit theorem. Properties of estimators in terms of consistency, efficiency and unbiasedness being far from attractive, a pragmatic approach was preferred to lengthy analytical developments, resorting to numerical simulation performed on random samples drawn from normal distribution, frequently encountered dealing with measurement errors.

Monte Carlo simulation supplies readily approximations to single and joint distributions of estimators of interest. Not unexpectedly, considering a normal parent distribution, numerical estimates of β_1 turn out to be distributed with a reverse *J* shape, while those of β_2 tend towards approximately normal shape for sample size n , at least in a three digit range (see Fig.4, 5, 6). Empirical corrections were obtained numerically in order to offset bias in estimates, observed particularly for small sample size (see Fig.7), where kurtosis is systematically underestimated. Peculiar patterns may be observed in the empirical joint distribution of numerical estimates b_1 and b_2 , somehow slanted along a diagonal region over the $\beta_1 - \beta_2$ plane for small to medium sample size, with a slope slowly decreasing for larger values of sample size n (see Fig.8(a), 8(b), 9(a)); joint confidence regions may be derived at specified levels.

Coefficients of correlation between sample estimates b_1 and b_2 , obtained numerically for a normal parent distribution, are plotted versus n in Fig.9(b).

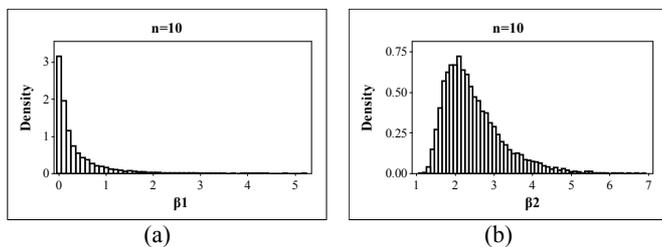


Fig.4. Histograms of numerical estimates of β_1 (a) and β_2 (b) for sample size n equal to 10.

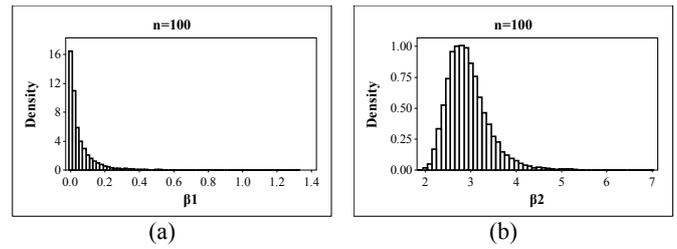


Fig.5. Histograms of numerical estimates of β_1 (a) and β_2 (b) for sample size n equal to 100.

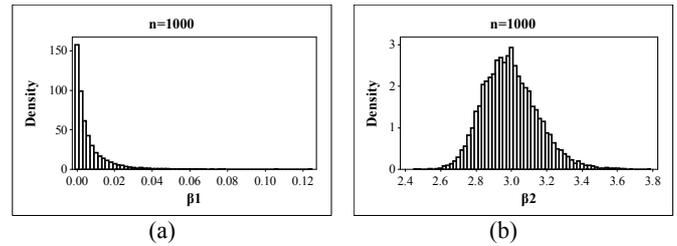


Fig.6. Histograms of numerical estimates of β_1 (a) and β_2 (b) for sample size n equal to 1000.

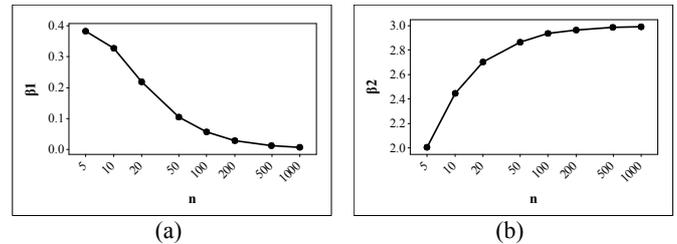


Fig.7. Empirical evaluation of bias in estimates of β_1 (a) and β_2 (b) plotted versus n .

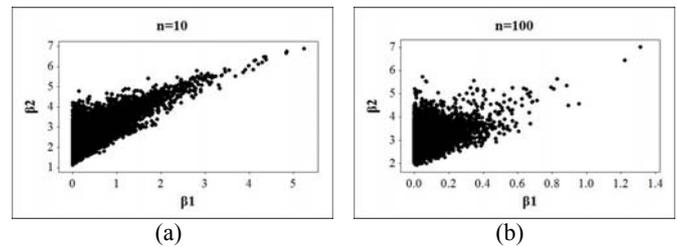


Fig.8. Empirical joint distribution of numerical estimates of β_1 and β_2 for sample size n equal to 10 (a) and 100 (b).

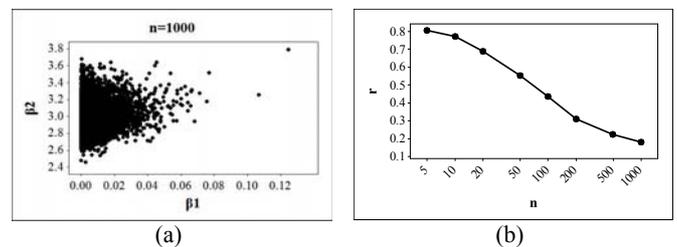


Fig.9. Empirical joint distribution of numerical estimates of β_1 and β_2 for sample size n equal to 1000 (a); coefficient of correlation r between estimates of β_1 and β_2 plotted versus sample size n (b).

Evaluation of confidence region

Let us consider first an example concerning measurements of battery capacity, with $n = 10$, $x_{ave} = 40$, $s = 3.5$ [20]. Confidence intervals at 90% level for μ and σ are readily obtained, leading to a joint confidence region drawn on the $\mu - \sigma$ plane with an approximately trapezoidal shape bounded by ordinates 2.5 and 5.7, and abscissas 38, 42 and 36, 44, respectively. A normal distribution may be fitted to the set of data at hand, and a confidence belt at 90% level for cumulative distribution function (cdf) obtained, given that the distribution shape assumed is the correct one – an educated guess only. Now let us map a rough numerical approximation of an overall 90% confidence region corresponding to our case, further simplified by restricting uncertainty to kurtosis only, namely by assuming symmetry, $\beta_1 = 0$, as consistent with the set of data at hand. Sample estimate of kurtosis, corrected for bias due to sample size ($b_2 = 2.6$), and a confidence interval 90% level for β_2 at the level considered ($1.6 \leq \beta_2 \leq 4$) are readily obtained by the Monte Carlo simulation. For $b_2 = 4$ a Johnson S_U distribution would fit the bill, and a Johnson S_B for $b_2 = 1.6$; other families of empirical distributions might of course be selected, e.g., Pearson’s type I and IV. Parameters of these distributions are first estimated according to the procedure described in [7], and further refined by heuristic methods. A joint confidence region for the three parameters considered in the μ - σ - β_2 space, roughly centered on point (40, 3.5, 3) corresponding to normal distribution, may be assumed to be bounded by a prism such as shown in Fig.10, whose apex coordinates A to H are listed in Table2.

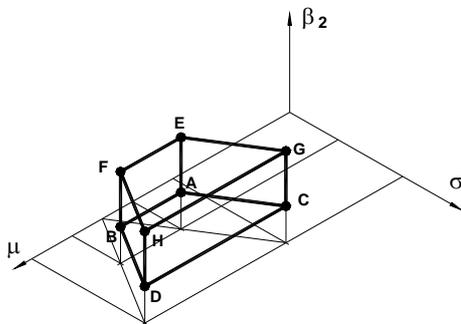
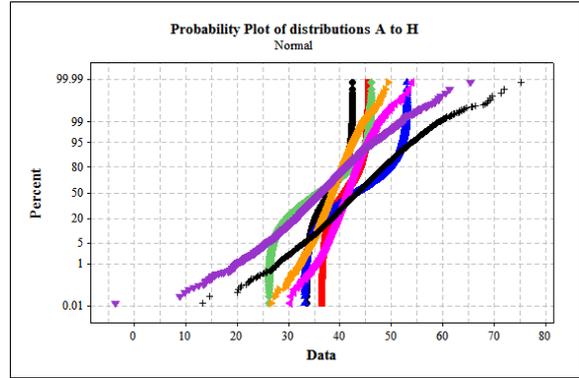


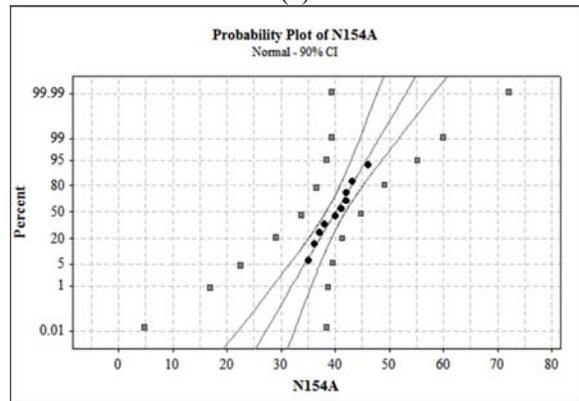
Fig.10. Prism enclosing a joint confidence region at 90% level in the μ - σ - β_2 space for location, spread and peakedness of distributions.

Tab.2. Apex coordinates of prism shown in Fig.10 for measurements of battery capacity based upon sample parameters.

apex	mean	st. dev.	b_2	distr. type
A	38	2.5	1.6	S_B
B	42			
C	36	5.7		
D	44			
E	38	2.5	4.0	S_U
F	42			
G	36	5.7		
H	44			



(a)



(b)

Fig.11. Cdf of the eight distributions corresponding to apexes A through H of prism shown in Fig. 10; normal probability plot of battery capacities with 90% confidence region pertaining to assumed normal distribution (inner curves), and envelope of distributions (dotted) corresponding to points A through H (b).

Cdf’s of the eight Johnson distributions corresponding to coordinates A to H listed in Table2 are plotted in Fig.11(a), their envelope defining to a first approximation a joint confidence region for the cdf of parent population, taking into account uncertainties associated both with estimation of parameters and identification of distribution form. The latter contribution appears by no means negligible, as for the case at hand it does substantially exceed the former, as shown in Fig.11(b), where the comprehensive envelope is plotted along with the confidence belt corresponding to unquestioning acceptance of normal distribution hypothesis.

In comparable instances, when little information is available besides sample values, the wider confidence region would be a sensible choice. In the light of expert knowledge, and specific, detailed information concerning expected shape of parent distribution, realistic confidence limits may get closer to those derived in terms of joint confidence region for estimates of parameters.

Coming back to INRIM’s gravimetric data, the empirical procedure described above was applied to the mixture model. Data were centered since comparison among average values – e.g., corresponding to data sets obtained on different sites, or on the same site under different conditions – would, as a rule, be considered apart from issues concerning scatter. While normal distribution is

substantially confirmed for the well-behaved data, sizable uncertainty concerns distribution underlying the subgroup more affected by scatter. A joint confidence region for the latter subgroup for scatter and kurtosis may be approximated by a rectangle on the σ - β_2 space, derived from the projection of the prism shown in Fig.10. The apexes, which identify four Johnson empirical distributions - two S_B and two S_U - as mandated by symmetry considerations ([7], [21], [22]), listed in Table3 along with their coordinates, lead to the 90% confidence region shown along with experimental values in the half-normal plot of Fig.12.

Tab.3. Apex coordinates of rectangle (derived from prism of Fig.10) for INRIM's gravimetric data, based upon sample parameters.

apex	st. dev.	b_2	distr. type
A	0.132	2.3	S_B
C	0.175		
E	0.132	3.9	S_U
G	0.175		

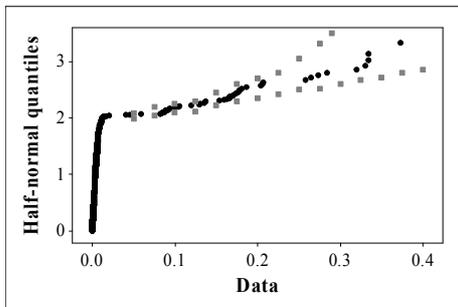


Fig.12. Half-normal probability plot of (coded) gravity acceleration with envelope of limiting empirical distributions identifying a 90% confidence region. No outlier was discarded.

Inclusion of uncertainty components covering estimation of parameters, and identification of distribution, may cater for more realistic estimation of type A contributions to uncertainty budget, thereby affecting evaluation of compatibility among data sets.

4. DISCUSSION

Handling discordant observations is affected by statistical and technical evaluation of the case at hand, as – but for gross blunders – the outcome of evaluation process is affected by personal orientation, and specific knowledge. No automatic procedure may completely substitute expert's judgment [23], just as in uncertainty evaluation [24]. Some measurement tasks exhibit a particular combination of sources of scatter, which may be modeled as a contamination of the bulk of "good" data by a small proportion of "bad" ones, or alternatively as a single, heavy tailed underlying distribution, the latter sometimes liable to entail peculiar problems.

Lacking dependable theoretical basis and proper mechanistic models, empirical identification of distribution underlying observed data implies a substantial uncertainty component, which may well exceed that pertaining to

estimation of parameters, routinely accounted for. A comprehensive evaluation may, however, be performed in a straightforward way, exploiting numerical simulation to supply information as required to obtain results with up to two significant digits, consistent with GUM guidelines [3].

Realistic confidence regions thus obtained may turn up broad enough to suggest adopting less conservative levels than usual to make appearance palatable, as not unusual in some applications of extreme value statistics [15]. In the case concerning measurements of battery capacity [20], for instance, the width of a 90% confidence band for cumulative distribution accounting also for uncertainty in identification of form is found to range between over twice that of confidence band accounting only for uncertainty in mean and standard deviation, normal form being taken for granted, to as much as three times as large. For the set of gravimetric measurements considered, whether identified outliers are discarded or not hardly affects a minor, yet substantial subset of data affected by a much larger scatter than the bulk. That subset, admittedly influencing but marginally the average value and its related confidence intervals, must, however, be taken into account if uncertainty in empirical identification of underlying distribution is to be accounted for.

Overly optimistic estimates of precision being of little if any use, it makes sense to face squarely the fact that data sets of finite size, routinely handled in the course of experimental work, particularly concerning underlying distribution carry but a limited amount of information, on whose basis only rather hazy predictions may be formulated. Realistic evaluation may afford some protection against such misfortunes as underestimating uncertainty by some two orders of magnitude [25].

ACKNOWLEDGMENT

Friendly cooperation and fruitful discussion with Dr. Giancarlo D'Agostino are gratefully acknowledged.

REFERENCES

- [1] Barnett, V., Lewis, T. (1978). The study of outliers: Purpose and model. *Applied Statistics*, 27, 242-250.
- [2] Barnett, V., Lewis, T. (1994). *Outliers in Statistical Data* (3rd ed.). New York, USA: Wiley.
- [3] Joint Committee for Guides in Metrology. (2008). *Evaluation of measurement data - Guide to the expression of uncertainty in measurement (GUM)*. JCGM 100:2008. Sèvres, France.
- [4] Zakrzewski, J. (2003). Error and uncertainty reduction – challenge for a measuring systems designer. *Measurement Science Review*, 3 (1), 31-34.
- [5] Song, J., Vorburger, T., Thompson, R., Renegar, T., Zheng, A., Ma, L., Yen, J., Ols, M. (2010). Three Steps towards metrological traceability for ballistics signature measurements. *Measurement Science Review*, 10 (1), 19-21.
- [6] Rhind, A. (1909). Tables to facilitate the computation of the probable errors of the chief constants of skew frequency distributions. *Biometrika*, 7 (1/2), 127-147.
- [7] Hahn, G.J., Shapiro, S.S. (1967). *Statistical Models in Engineering*. New York, USA: Wiley.

- [8] Johnson, N.L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, 36 (1/2), 149-176.
- [9] Pearson, K. (1895). Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London A*, 186, 343-414.
- [10] D'Agostino, G. (2005). *Development and metrological characterization of a new transportable absolute gravimeter*. Unpublished Doctoral Dissertation, Politecnico di Torino, Torino, Italy.
- [11] Niebauer, T.M., Sasagawa, G.S., Faller, J.E., Hilt, R., Klotting, F. (1995). A new generation of absolute gravimeters. *Metrologia*, 32 (3), 159-180.
- [12] Bich, W., D'Agostino, G., Germak, A., Pennechi, F. (2008). Reconstruction of the free-falling body trajectory in a rise-and-fall absolute ballistic gravimeter. *Metrologia*, 45 (3), 308-312.
- [13] Hanada, H., Tsubokawa, T., Tsuruta, S. (1996). Possible large systematic error source in absolute gravimetry. *Metrologia*, 33 (2), 155-160.
- [14] Bich, W., D'Agostino, G., Pennechi, F., Germak, A. (2011). Uncertainty due to parasitic accelerations in absolute gravimetry. *Metrologia*, 48 (3), 212-218.
- [15] Gumbel, E.J. (1958). *Statistics of Extremes*. New York, USA: Columbia University Press.
- [16] Barbato, G., Barini, E.M., Genta, G., Levi, R. (2011). Features and performance of some outlier detection methods. *Journal of Applied Statistics*, 38 (10), 2133-2149.
- [17] David, H.A., Nagaraja, H.N. (2003). *Order Statistics* (3rd ed.). New York, USA: Wiley.
- [18] Kendall, M.G., Stuart, A. (1977). *The Advanced Theory of Statistics* (4th ed.). London, UK: Griffin.
- [19] Ord, J.K. (1968). The discrete Student's t distribution. *The Annals of Mathematical Statistics*, 39 (5), 1513-1516.
- [20] Natrella, M.G. (1963). *Experimental Statistics*. NBS Handbook 91. Washington, USA: National Bureau of Standards.
- [21] Hill, I.D., Hill R., Holder R.L. (1976). Algorithm AS99: Fitting Johnson curves by moments. *Applied Statistics*, 25, 180-189.
- [22] DeBroda, D.J., Dittus R.S., Roberts, S.D., Wilson, J.R. (1989). Visual interactive fitting of bounded Johnson distributions. *Simulation*, 52, 199-205.
- [23] Gumbel, E.J. (1960). Discussion of the papers of Messrs. Anscombe and Daniel. *Technometrics*, 2 (2), 165-166.
- [24] Genta, G. (2010). *Methods for Uncertainty Evaluation in Measurement*. Saarbrücken, Germany: VDM Verlag.
- [25] Youden, W.J. (1972). Enduring values. *Technometrics*, 14 (1), 1-11.

Received February 2, 2012.

Accepted August 3, 2012.