Accred Qual Assur (2009) 14:587–592
DOI 10.1007/s00769-009-0525-8

GENERAL PAPER

Comparison of calibration curves using the L_p norm

Luca Callegaro · Francesca Pennecchi ·
Pier Giorgio Spazzini

Received: 23 January 2009 / Accepted: 6 April 2009 / Published online: 25 April 2009

Abstract Interlaboratory comparisons are a fundamental task in order to provide measurements with traceability. The simplest possible scenario implies that a single traveling standard of a quantity is measured at various laboratories. A more complex scenario arises when the laboratories measure a large set of standard values pertaining to a given physical quantity or when the traveling standard is not a realization of the quantity of interest but a measuring instrument. In the last case, it might be convenient to globally compare the calibration curves provided by the laboratories. We will introduce a distance between two generic analytical curves based on the Least Power L_p norm of their difference. The properties of such distance will be presented, with particular attention to its dependence on the parameter p.

Keywords Interlaboratory comparison · Calibration curve · Least Power norm · Distance

Introduction

Interlaboratory comparisons are a fundamental task in metrology, as they are the main tool to check and maintain the compatibility between the various National Standards, and make them referable to an International Standard [1]. The original and principal aim of metrology, as stated since the introduction of the "Convention du mètre" in the nineteenth century, was and still is to obtain a situation in which all measurements performed anywhere in the world can provide comparable results, hence it could even be said that, nowadays, there can be no metrology without interlaboratory comparisons.

The simplest possible scenario for an interlaboratory comparison implies that a single traveling standard of a quantity is measured at various laboratories. If sufficient precautions are taken to avoid damages or modifications to the standard, the same stimulus will be provided to all measurement chains, which will therefore provide their various responses to it. Once these results are collected, point wise statistical methods can be employed for evaluating an appropriate reference value and for determining a degree of equivalence for each participating laboratory. A great number of publications and guidelines have been devoted to the subject (see, among many, [2]). This situation applies in "classical" cases, i.e., when a single, well-defined standard can be circulated and measured. The prototype application of this scenario is the comparison of the mass standard.

Though, with the evolution of metrology and the request for definition and maintenance of standards for various quantities, more complex scenarios have started to emerge. One more general scenario arises when the laboratories measure a large set of standard values pertaining to a given physical quantity. This situation, that still needs to be addressed by appropriate guidelines, typically occurs either when "several traveling standards (are) circulated [and (need) to be treated] together" or when a standard has "to be measured at each of a number of stipulated values of a parameter, such as wavelength or frequency" [2].

Another situation of practical importance occurs when the traveling standard is not a realization of the quantity of interest but a measuring instrument. Examples of such

Presented at IMEKO TC8 Workshop, November 2008, Torino, Italy.

L. Callegaro · F. Pennecchi (✉) · P. G. Spazzini
Istituto Nazionale di Ricerca Metrologica (INRIM),
Strada delle Cacce 91, 10135 Turin, Italy
e-mail: f.pennecchi@inrim.it

comparisons include those concerning mass flow rate, airspeed, ionizing radiation, photometry and so on. Since in such cases the quantity of interest must be generated by an appropriate test rig, it might be inconvenient or impossible for the laboratories to calibrate the instrument on the very same set of specified points. Therefore, the various data sets are usually made equivalent by comparing "approximately coincident" points, but this method appears quite questionable. One possible solution is to compute best fitting curves and re-compute the expected value at the reference points, but it would be clearly better to compare the calibration curves themselves as a whole.

Comparing calibration curves directly would have the additional advantage of allowing to globally assess the compatibility of the measurements; moreover, should a few measurement points be evident outliers, the comparison in those points would not be completely lost but at least some indicative results could be recovered through the global evaluation.

Another possible application of a method for globally comparing curves can be found in the chemical/spectroscopic field. Actually, it is often necessary to compare absorption spectra, as measured by a spectrograph on unknown compounds, to analogous reference spectra. Presently, this comparison is performed using distance-to-model methods [3], such as Euclidean norm and Pearson correlations, or more sophisticated tools, such as Principal Components Analysis (PCA) and Soft Independent Modeling by Class Analogy (SIMCA). The method proposed in the following might constitute an alternative, based on the direct computation of the distance between the measured spectrum and the reference one. Notice that the method, in the case $p = 2$, is a generalization of the Least Squares Algorithm described in [3].

In order to develop the ability to compare the results in such a way, several steps are required. Obviously, the first requirement is to be able to compare two different curves, irrespectively of their mathematical formulation (i.e., it is desirable to be able to compare a straight line to a parabola, or an hyperbola and so on). This comparison must also be quantitative and objective.

The following step will be to define a reference curve, which will be appropriately computed based on data from all the laboratories and to which all the single curves will be compared.

The present paper is concerned only with the first of these steps. We will introduce a distance between two generic analytical curves defined on a closed and bounded interval. Such distance is based on the Least Power L_p norm of the differences between the curves. The properties of such distance will be presented, with particular attention to its dependence on the parameter p.

The distance can also be generalized to discrete data sets, although there are some mathematical details to be worked over; this work is in progress, in parallel to the definition of reference curves/data sets, and will be the subject of future papers.

An exhaustive overview of the L_p norm estimation theory, as well as dedicated algorithms for its implementation, are available in the statistical literature [4–7].

L_p norms

The L_p norm (with $p \geq 1$) is a general concept that can be defined on a large class of abstract mathematical objects. In this context, however, we restrict its definition to the case of bounded functions $f(x)$ which are defined on a closed and bounded subset of the real numbers, for example on the interval $[a, b]$, having length $\Delta = b - a$.

If the Lebesgue integral of the p power of the absolute value of f

$$\int_a^b |f(x)|^p \mathrm{d}x \quad (1)$$

is defined and finite, then

$$||f||_p = \left(\frac{1}{\Delta} \int_a^b |f(x)|^p \mathrm{d}x \right)^{1/p} \quad (2)$$

is the (normalized) L_p norm of $f(x)$ on $[a, b]$ [8]. If $0 < p < 1$, Eq. 2 is no longer a norm, since it does not satisfy the triangular inequality. If $f(x)$ describes a dimensional physical quantity, its L_p norm is defined to have the same dimension as $f(x)$. If $f(x) \equiv k$, for k constant, then $||f||_p \equiv |k|$, irrespectively of the interval $[a, b]$ and the parameter p. For $p = 1, 2, \infty$, respectively, Eq. 2 leads to the following norms:

$$||f||_1 = \frac{1}{\Delta} \int_a^b |f(x)| \, \mathrm{d}x \quad (3)$$

$$||f||_2 = \sqrt{\frac{1}{\Delta} \int_a^b f(x)^2 \mathrm{d}x} \quad (4)$$

$$||f||_\infty = \sup_{[a,b]} \{|f(x)|\}, \quad (5)$$

where sup is the supremum of $|f(x)|$ (the supremum of a bounded function $g(x)$ is its least upper bound, that is, the smallest value s such that $g(x) \leq s$, for every x; if the function is continuous on a close interval $[a, b]$, its supremum coincides with its maximum). Consider, for

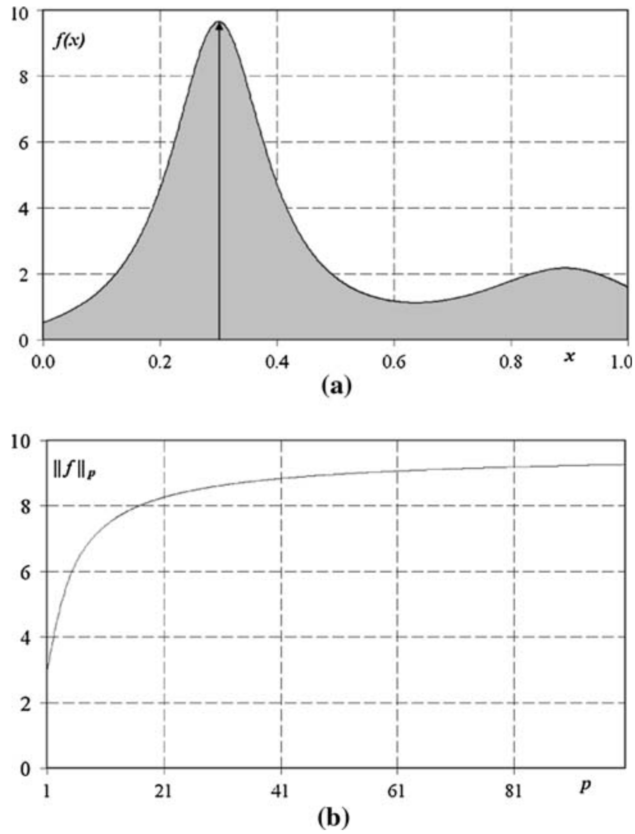

Fig. 1 **a** A generic function and **b** the evolution of its L_p norm as a function of p

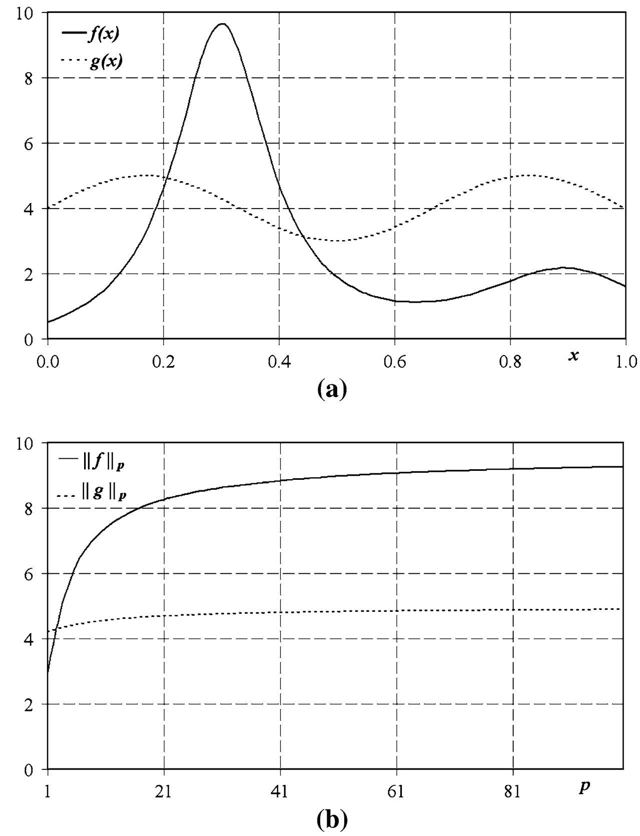

Fig. 2 **a** Two functions and **b** the evolution of their L_p norm as a function of p

example, function $f(x)$ in Fig. 1a) (for simplicity, the function has been chosen to be positive on the considered interval [0, 1]). The L_1 norm $\|f\|_1$ is proportional to the area of the region bounded by the function graph (the shaded area in the figure), and is exactly the mean value of $f(x)$ on the interval. $\|f\|_2$ is proportional to the inertia moment of the function graph with respect to the x axis. $\|f\|_\infty$ is the maximum value of $f\,(x)$ (indicated by the arrow in the figure).

By varying p, the norm behavior has a continuous transition through the three special cases given by Eqs. 3, 4 and 5. Figure 1b) depicts such dependence of $\|f\|_p$ on p. The monotonic increase of the $\|f\|_p$ norm for increasing p is a general property [8]. It can been seen, as expected, that $\|f\|_p$ tends, for increasing p, to the maximum value of $f\,(x)$.

Figure 2a, b show function $f(x)$ together with another function $g(x)$ and the corresponding L_p norms. It is clear that the ordering of $\|f\|_p$ and $\|g\|_p$ is dependent on the value of p. Since the mean value of $g(x)$ is larger than that of $f(x)$, for low p values one has that $\|f\|_p < \|g\|_p$ while when increasing p, the higher peak of $f(x)$ becomes more important and the order between norms is inverted.

A distance between curves

In order to evaluate the distance between two functions $f(x)$ and $g(x)$ defined over the same domain $[a, b]$, we propose to consider the L_p norm, defined by Eq. 2, of the difference between the two curves:

$$D^p_{f,g} = ||f - g||_p. \tag{6}$$

From now on we will drop the subscript "f, g" if it is enough clear from the context which are the relative functions.

Obviously, the same properties hold for this distance as for the L_p norm of a single function treated in the previous section. In particular, if the curves differ from each other only by an additive term (simple translation), i.e. $f(x) - g(x) \equiv k$, then $D^p \equiv |k|$ for any interval $[a, b]$ and any p. If the two curves coincide everywhere on $[a, b]$, then $D^p \equiv 0$. If $g(x) \equiv 0$ all over $[a, b]$, then $D^p \equiv \|f\|_p$.

Consider again the same functions plotted in Fig. 2a). These are plotted again in Fig. 3a), while Fig. 3b) reports the L_p norms of their difference, i.e., their distance D^p; two special cases of this distance are also graphically represented in Fig. 3a), namely $p = 1$ (the shaded area) and p

tending to infinity (the maximum distance between the curves, indicated by the arrow).

The distance provides a *global* evaluation of the difference between the curves. Different properties of this difference can be emphasized, depending on the choice of p. For small values of p the distance takes into account the overall similarity of the curves, while for large values of p it highlights the maximum local differences between the curves.

An appropriate choice of p can therefore be very useful to evidence different properties of the curves and their relationships. For instance, a small variation of the distance all over the range of p will indicate that the curves can be considered similar in shape, although shifted with respect to each other (see the above-mentioned property of the distance for curves that are translated one with respect to the other). In this case, the question is whether such (almost) constant distance represents a small or a significant amount with respect to the actual context.

On the other hand, a small value of the distance for small p values followed by a strong increase for large p values will indicate that the curves are relatively close to each other except in some reduced regions where large differences may exist. In this case, one should decide which of the two aspects is more important for the aim of the actual analysis: the good overall agreement between the curves or their few but high discrepancies.

Therefore, depending on the specific purposes of the evaluation, a specific value of p will be chosen that will underline some particular characteristics of the curves. Of course, this task involves a certain arbitrariness, but an appropriate knowledge of the mathematical tool should be a sufficient guide to avoid misleading results.

Comparing curves on the whole p range is the main reason for studying the behavior of their distance as a function of p. Moreover, if a maximum value (e.g., a prescribed tolerance) for the distance between two curves is set, it is possible to check which range of p satisfies such condition; this result would allow to check whether the curves are within tolerance only in a global sense or also locally. For instance, assume that it is required that the curves in Fig. 3a have a difference lower than some value, say 3.5; it can be seen that this condition is satisfied when $p < 6.2$. This means that the curves are within prescription up to that value of p, i.e., only in a global sense. If the prescribed tolerance was larger, say 5.5, the curves would be within prescription for all values of p, hence also locally at each point. Intermediate values would indicate intermediate situations. Conversely, it could be possible to decide which tolerance should be considered for the curves to be within tolerance only in a global sense or even locally.

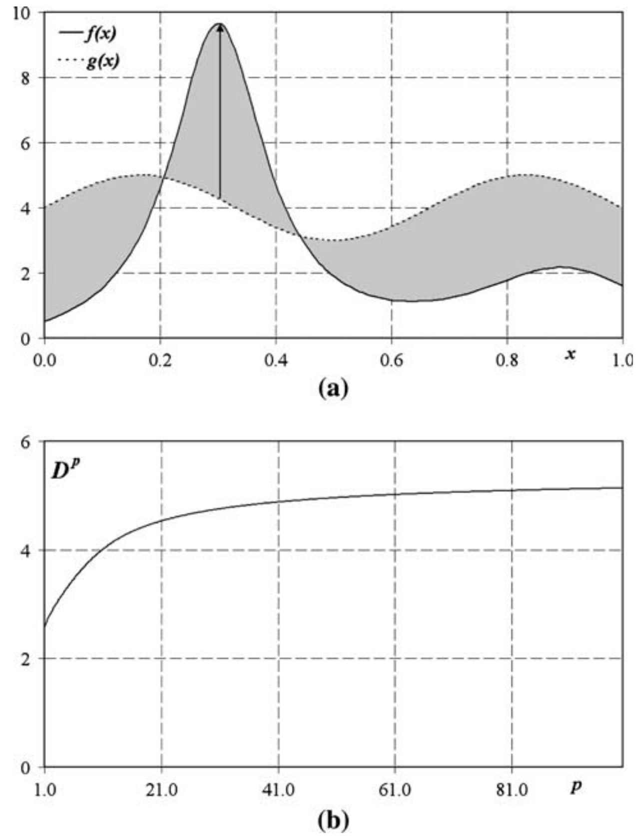

Fig. 3 **a** The same curves as in Fig. 2a and graphical representation of two special cases of their distance D^p; **b** relative distance D^pas a function of p

Finally, it was shown in [9] that, within the context of parameter estimation, various values of p can provide L_p estimators with different properties. In perspective, an extension to the case discussed here could be developed.

Calibration curves versus a reference curve: an example

The distance introduced in the previous sections can be used for the comparison of calibration curves. For the sake of discussion, we introduce a very specific application example (based on an anemometer, due to the specific experience of one of the authors on this type of instruments), but many others can be conceived.

Anemometers' calibration curves are usually nonlinear for the low speeds; in these cases, an hyperbolic function of first or second order can usually have a good fit to the data. Assume now that the calibration curve of a specific anemometer has the form of a second order hyperbola

$$f(x) = a_0 + a_1/x + a_2/x^2, \quad (7)$$

and that the anemometer is calibrated five times at five different laboratories. The result of each calibration is the estimate of the parameters a_0, a_1, a_2 of the curves; the values are reported in Table 1.

Let us further assume that a reference curve, having the parameter set marked as REF in Table 1, is available. For the purposes of the present work, this curve will be considered as given, e.g., as coming from a primary laboratory when performing an audit. The problem of constructing a reference curve from the available data will be discussed in future works.

The curves are plotted in Fig. 4a) in the range $1 \le x \le 25$. Units of measurement are immaterial for the aims of the present work, but for the sake of completeness, since the example is based on an anemometer, it can be imagined that x is in units of m s^{-1}, whereas $y = f(x)$ represents the calibration coefficient, which is often chosen to be dimensionless.

Table 1 Pseudo-calibration curves parameters

Curve number	a_0	a_1	a_2
1	0.95	−0.015	0.030
2	0.95	−0.020	0.033
3	0.95	−0.0185	0.0365
4	0.95	−0.010	0.022
5	0.95	−0.023	0.040
REF	0.95	−0.0162	0.0305

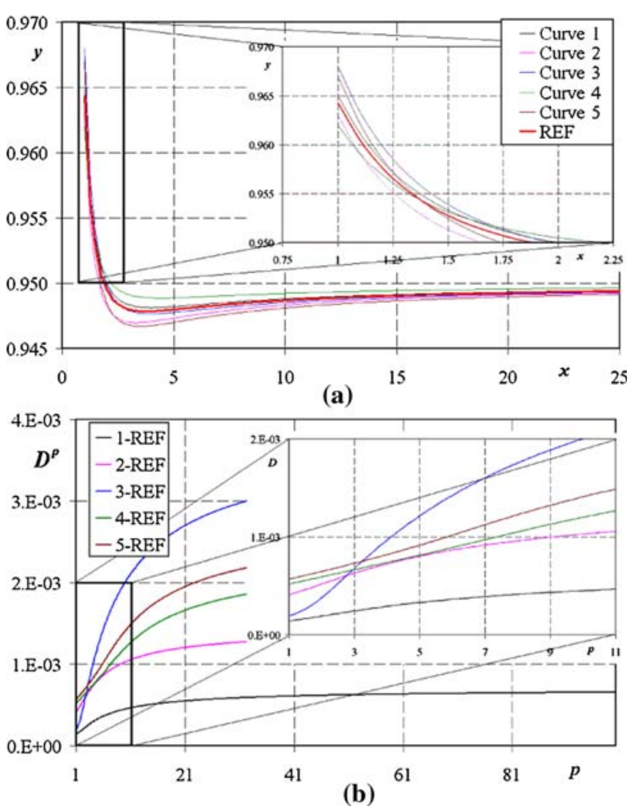

Fig. 4 **a** Diagrams of five pseudo-calibration curves (*thin lines*) and of a reference curve (*thick red line*); in the box, a zoom of the region with the highest slope. **b** Distances of the five curves from the reference curve; *in the box*, a zoom of the region close to origin (color figure online)

Consider first the calibration curves 1–5. It can be observed that the curves are relatively similar to each other, but some differences exist. In particular, for large x the curves tend to be essentially parallel to each other and to keep their distances essentially constant, while for small x there are several curve crossings. This behavior is typical of anemometer calibration curves.

Figure 4b) shows the distances, defined by Eq. 6, of curves 1–5 from REF, as functions of p. It can be noticed that the order of the distances is dependent on the value of p. In particular, for low values of p, curves 1 and 3 have the smallest distance from REF. Actually, from Fig. 4a), it is quite evident that the curves 1 and 3 seem on the average closer to REF than the others.

On the other hand, looking again at Fig. 4b), the distance of curve 3 from REF rises quickly and becomes the largest, for large p. In fact, curve 3 has the largest local distance from REF among all curves, as can be seen by carefully observing the zoom of the region with the highest slope in the box of Fig. 4a).

Conclusions

The distance defined by Eq. 6 can be employed to rank calibration curves with respect to a reference one. The ranking strategy can be varied by choosing different ps. If the interest is focused on the overall differences between the curves, a small p value is appropriate, for example, when dealing with a global assessment of a calibration laboratory in an audit. Otherwise, if large (although local) differences are important for the context, a large p value should be used, for example, to identify the presence of outliers in a single curve or a malfunctioning of the traveling instrument in a subrange.

The work described in the present paper must be considered as a first step towards a complete analysis of the comparison between two calibration curves; two major problems, which are the subject of work in progress, are the definition of a reference curve and the evaluation of the uncertainty to be associated to the distance defined in the present paper (which of course should take into account the uncertainties associated to the original data points and, consequently, to the various calibration curves).

References

1. CIPM Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes, BIPM, Sèvres, France, 1999

2. Cox MG (2002) The evaluation of key comparison data. Metrologia 39(6):589–595. doi:10.1088/0026-1394/39/6/10
3. http://www.thermo.com/com/cda/resources/resources_detail/1,13229,00.html
4. Sposito VA (1982) On unbiased L_p regression estimators. J Am Stat Assoc 77:652–654. doi:10.2307/2287729
5. Nyquist H (1983) The optimal L_p norm estimator in linear regression models. Commun Statist Theor Methods 12:2511–2524. doi:10.1080/03610928308828618
6. Gonin R (1989) Money AH nonlinear Lp-norm estimation. Marcel Dekker, New York
7. Sposito VA *Some properties of L_p – estimators*, in K. D. Lawrence and J.L. Arthur, editors, Robust Regression, Analysis and Applications, chapter 2, pp. 23–58, Marcel Dekker, New York, 1990
8. Hwang SY (1975) On Monotonicity of L_p and l_p norms. IEEE Trans Acoust Speech Signal Process ASSP-f 3:593–594
9. Pennecchi F, Callegaro L (2006) Between the mean and the median: the L_p estimator. Metrologia 43(3):213–219. doi:10.1088/0026-1394/43/3/004