XXI IMEKO World Congress “Measurement in Research and Industry”
August 30 – September 4, 2015, Prague, Czech Republic

ISSUES IN FORENSIC MEASUREMENTS

Giovanni Battista Rossi and Francesco Crenna

Università degli Studi di Genova, Genova, Italy, gb.rossi@dimec.unige.it

Abstract – The role of measurement in forensic science is discussed, with focus on the impact of measurement uncertainty in forensic decision-making. The need of properly modelling the measurement process and the benefit of expressing results in probabilistic terms are pointed out. The interpretation of probabilistic statements in decision-making is also discussed and the use of scales of degree of certainty is suggested. The benefit of a multi- and interdisciplinary approach is outlined, as well as the need of future research in this area.

Keywords: forensic metrology, measurement science, uncertainty, measurement modelling, probability

1. EXPERT WITNESS TESTIMONY AND SCIENTIFIC EVIDENCE

Expert witness testimony often plays a key role in forensic proceedings. A relevant question in forensic science is the value to be attributed to them [1-2][1]. The attitude towards them has somewhat changed in (relatively) recent times, from an unquestioned acceptance, to a more critical attitude. According to Dondi [1-2], particular relevant in this regards was the so-called and quite famous “Daubert decision”, concerning a Supreme Court’s sentence issued in 1993, where precise and highly demanding criteria where given for the selection of expert testimonies. These criteria were also quite promptly adopted in the Federal Rules of Evidence, in the 2000 revision.

Expert testimony is tightly linked to the use of *scientific evidence*, based on measurement and testing procedures, such as chemical analysis, doping testing, fingerprints identification, DNA testing, and so on. Interestingly enough, in almost the same years the change of perspective in forensic science just mentioned took place, in the measurement community a critical analysis on the influence of measurement in decision-making was carried out, as a part of an investigation on the methodological aspects of measurement [4-7, 11]. In fact, in those studies, decision-making in manufacturing, quality assessment and legal metrology, was mainly investigated. It may be now worthwhile to extend such studies to the area of forensic science also. Interestingly enough, in this perspective, ethical [13-14] and epistemological aspects have also been recently pointed out [3,8, 10, 15-16].

[1] References are listed in chronological order, to give a sense of the historical development of the subject.

2. THE ROLE OF MEASUREMENT IN THE FORENSIC DECISIONAL PROCESS

Consider some factual hypothesis, H, such as “the subject assumed a doping substance”, “subject A is father of subject B”, “bread from this supply contains an amount of pesticides over threshold”, and some scientific evidence, E, that provides support to (or against) it. In many cases, E is related to some measurement or testing procedure. The way in which it is possible to gain information on H based on E may be *modelled*[2] through the Bayes-Laplace rule.

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E \mid H)P(H) + P(E \mid \bar{H})P(\bar{H})}, \quad (1)$$

where $\bar{H}$ is the complementary (negation) of H. Let us consider, for example, the simple case where the hypothesis to be tested, H, is: “the amount of a pollutant in a substance exceeds a threshold fixed by the law”.

Let us denote the amount of substance (for example pesticides in bread) by x, and the threshold value by a. Suppose we measure x by a proper instrument which yields an indication y, on the basis of which the final measurement value $\hat{x}$ is produced. We can now introduce the following events/statements[3]:

$$\begin{aligned} &H = x > a, \\ &\bar{H} = x \le a, \\ &E = \text{"the measurement value is } \hat{x}\text{"}. \end{aligned} \quad (2)$$

For applying formula (1) the corresponding probabilities need calculating. To this goal, it is very important to consider whence the needed information can be obtained. In some respects, considerations similar to those developed for conformance assessment apply [4, 7, 11], in others differences are substantial.

[2] Note that what we are doing here is a kind of modelling, although involving a relatively “high” level of discourse. This consideration will be useful in the discussion to follow.

[3] We are going to develop a probabilistic treatment of the subject. Probability may be seen either as a property of events (that may happen or not) or of statements (that may be true or false). We will keep an open attitude towards these two possibilities [12].

Let then $P(H)$ be the probability that may be assigned to the hypothesis under consideration, prior to making the measurement. In our example, if $p_{\mathcal{P}}(x)$ is a proper initial[4] distribution for the variable x,

$$P(H) = P(x > a) = \int_{x>a} p_{\mathcal{P}}(x)\mathrm{d}x\,. \tag{3}$$

It is essential to note that the information for assigning $P(H)$ comes from *outside* the measurement process, and usually it requires a different expertise: in this notation, this is highlighted by the subscript $\mathcal{P}$, for "process". In the case of conformity assessment for large-scale production, the origin of the distribution $p_{\mathcal{P}}(x)$ is straightforward: it comes from statistical data on the manufacturing process, based on historical production records. In that case, the involved expertise concerns mainly a knowledge of the production process, combined with statistical skills. In the case of forensic science, instead, the required expertise may concern the sociological area, and the related information may be obtained by sociological or epidemiological data, concerning the population to which the involved individual(s) belongs. The importance of this initial distribution should not be overlooked: in the literature on decision-making the cognitive bias related to overlooking its importance has been often emphasized [9]. Famous is the "taxi problem", that concerns the identification of the company to which a taxi car that has caused an incident, belongs, based on a testimony, which anyway is not fully reliable, due to the uncertain foggy environment where the fact occurred. Furthermore, in the town there are two taxi companies, with a quite different market share. Well, the moral of the story is that the identification should not be based solely on witness testimony, which can be modelled by a conditional probability, but also on the prior probability that a taxi in that town belongs to one of the two companies, which is related to the number of cars of each company on duty.

To some up, concerning the initial distribution, as a last chance, the possibility of assuming a uniform not informative shape remains, but other possibilities should be considered as well. In this regard, useful considerations, specific for the area of forensic science, are developed in Refs. [10, 16].

Let us now consider the conditional probability $P(E \mid H)$. In our simple example the required expertise, and the related information, comes from measurement science; in the general case, other kinds of expertise may be involved. For example, in the case of the identification of a person on the basis of a plurality of physical and/or biological features, the measurement of the individual features concerns measurement science (eventually in different disciplines), whilst their combination in a proper model for the identification may be of main pertinence to anthropological or biological sciences. Anyway, let us now remain on our example.

Consider a generic model of the measurement process, where x is the (unobservable) value of the measurand, y is the instrument indication and $\hat{x}$ is the measurement value [12]. Call *observation* the overall (physical) transformation that links x with y, and *restitution* the data-processing phase, where $\hat{x}$ is obtained on the basis of the acquisition of y [5]. Suppose we have made a measurement and obtained the instrument indication y. Let $p(y \mid x)$ be the conditional probability distribution (often a probability density function in the usual practice of physical measurement) that characterizes the measuring device. Then, for obtaining the final measurement result, we have to perform restitution [12][6], according to the Bayes-Laplace rule:

$$p(x \mid y) = \frac{p(y \mid x)}{\int_X p(y \mid x)\mathrm{d}x}\,, \tag{4}$$

and take the expected value:

$$\hat{x} = \int_X x p(x \mid y)\mathrm{d}x\,, \tag{5}$$

where X is the domain of x.

Note now that $p(E \mid H) = p(\hat{x} \mid x > a)$ [7]. For calculating it, we need the conditional distribution $p(\hat{x} \mid x)$ that describes *the overall mapping* from x to $\hat{x}$. It may be obtained by

$$p(\hat{x} \mid x) = \int_Y \delta(\hat{x} - \mathrm{E}(x \mid y))\, p(y \mid x)\mathrm{d}y\,, \tag{6}$$

where E denotes the expectation operator, and δ is the Dirac-delta operator. Once $p(\hat{x} \mid x)$ has been obtain, we can calculate $p(E \mid H)$, by

$$p(E \mid H) = p(\hat{x} \mid x > a) = \int_{x>a} p(\hat{x} \mid x) p(x)\mathrm{d}x\,, \tag{7}$$

and,

[4] For this distribution, the term "prior" is also used, which has a somewhat subjective flavour. Therefore we prefer the more neutral term "initial distribution".

[5] A more complete model can be devised, where the input is the state of the measurand, rather than its numerical representation *x*, and, consequently, observation is the mapping of this state into the instrument indication. Yet, for the purpose of the present discussion, the proposed framework is sufficient, and its presentation is probably simpler.

[6] Note that the approach here proposed is in agreement with the principles of the GUM and has the advantage, over other competing interpretations of the GUM, that the meaning of the different probability distributions is clearly stated, in the framework of a coherent overall probabilistic measurement theory [5, 12].

[7] We have moved from $P(E \mid H)$ to $p(E \mid H)$ since we prefer to work in terms of probability density functions (pdfs), in the following.

$$p\left(E \mid \bar{H}\right) = p\left(\hat{x} \mid x \le a\right) = \int_{x \le a} p\left(\hat{x} \mid x\right) \mathrm{d}x \,. \quad (8)$$

Note now two very important points.

- In (4) we have assumed a uniform, not informative, initial distribution for x: this is correct, since here we are *inside* the measurement process, and, in this context, *only information on x provided by the instrument should be used*, and not information coming from outside sources.
- Great care must be taken in using the correct conditional distributions, in particular in using $p\left(\hat{x} \mid x\right)$ in (7).

Overlooking these aspects, may lead to inconsistent results: see Refs [5-7] for a thorough discussion of this point. Lastly, the probability of the hypothesis H, based on scientific evidence E, may be obtained by combining formulae (1), (3), (7) and (8).

Note now that scientific evidence has ultimately led to a probabilistic statement on the (hypothesized) state of things. But, how can we use such a statement in forensic decision-making?

3. FORENSIC DECISION-MAKING

Once that information has been obtained, a decision has to be taken, based on that information, but, in the general case, not exclusively on it. Different approaches are possible. One possibility is to perform a *significance test* on the hypothesis H. Note that if H has to be proven, *beyond any reasonable doubt* [13, 14], we have rather to test the opposite assumption, i.e., $\bar{H} = x \le a$, and to try and falsify it, according to accepted epistemological criteria. To do this, an acceptance region, $B = \left[0, b\right]$, with $b \ge a$, such that the probability, p_0, that $\hat{x} \in B$, when $\bar{H}$ is true, be *very high*, has to be identified. The (apparent) advantage of this approach is that that a final yes/no decision is attained. But this only an apparent certainty, since the choice of p_0 is *conventional*.

In the case of conformance assessment in the manufacturing area, a relatively simple, at least in principle, strategy for identifying the acceptance region can be devised through *cost analysis* [11][8]. In fact, it is often possible to estimate a cost related to accepting a not conformal item and a cost related to rejecting a conformal one. Then, for any possible choice of the acceptance region, the associated costs can be estimated and an optimum strategy in term of cost reduction may be identified. But this may be hardly applicable, in the forensic domain, for the difficulty of associating "costs" to wrong decisions, especially in the case of criminal right. Therefore, alternative approaches deserve consideration.

Another possibility is offered by the *likelihood ratio* corresponding to the two competing hypotheses:

$$\Lambda = \frac{P\left(H \mid E\right)}{P\left(\bar{H} \mid E\right)} \,. \quad (9)$$

Yet, again the problem arises of fixing a value for Λ, for which the hypothesis under test has to be accepted.

If we accept to renounce to, strict, certainty, as a further alternative approach, we may try to evaluate the "degree of certainty", so to speak, that may be obtained from scientific evidence, of a given type. In this regard, a probability figure cannot be easy to interpret: what degree of certainty corresponds to a probability, of, say, 0.9?

Altough it may not be easy, nor even possible, to answer such a question in general terms, yet a criterion may be found, for homogeneous classes of problems. For example, in Ref. [3], for the case of DNA testing, a so-called Himmel's scale is reported, that links "plausibility of kinship", which is a probability, to "likelihood of kinship", which is expressed in linguistic terms, that are understandable by not specialists also, and therefore better suited for addressing decision-making. If we generalize this idea to "probability" and "degree of certainty" respectively, we obtain the relationship in Table 1[9].

Table 1. An interpretation of probability in terms of degree of certainty, inspired by the Himmel's scale [3].

Probability	**Degree of certainty**
$p > 0.998$	Practically proved
$0.991 \le p < 0.998$	Extremely likely
$0.950 \le p < 0.991$	Very likely
$0.90 \le p < 0.95$	Likely
$0.80 \le p < 0.90$	Undecided
$p < 0.80$	Not useful

Possibly similar interpretation scales could be devised for likelihood ratio also. Yet this latter approach would imply a noteworthy change of attitude. In fact, in this perspective, the result of scientific evidence cannot be anymore considered as providing a definitive indication for the decision to be taken, but rather as an important piece of information that can validly contribute, hopefully together with other pieces of evidence, to the formulation of a fair trial.

[8] Note that decision-making based on cost analysis is, conceptually, a different matter from significance testing. Although the mathematics may be similar, the epistemological status of these two approaches is completely different, since significance testing aims at ascertain, as far as possible, the state of things, whilst cost analysis mainly tends to obtain maximum benefit from the decision.

[9] We obviously recommend to read the original paper and the references there quoted, for a full understanding of the original specific topic.

4. A NUMERICAL EXAMPLE

Let us illustrate the above ideas on a simple numerical example. Consider the measurement of the total content of pesticides in some sample of bread, to determine whether it is within the limit established by law or not [4]. Let the threshold be $a = 2.0$ mg/kg, and let the measurement procedure be characterised by a normal (Gaussian) distribution[10],

$$p(\hat{x} \mid x) = u^{-1}\varphi\left(u^{-1}(\hat{x} - x)\right). \tag{10}$$

where $\varphi(\xi) = (2\pi)^{-1/2}\exp\left(-\xi^2/2\right)$ is the standard – zero-mean and unitary variance – normal distribution, and let $u = 0.34$ mg/kg.

We also need an initial probability distribution for x, $p(x)$. It we have no experimental information about that, we may proceed under the assumption that $P(H) = P(\bar{H}) = 0.5$. If we further assume that $x \in [0, 2a]$, a $p(x)$ consistent with the assumption $P(H) = P(\bar{H}) = 0.5$, is a constant distribution over the range of x, that is $p(x) = (2a)^{-1}$, for $x \in [0, 2a]$, and $p(x) = 0$, elsewhere.

Then we simply obtain:

$$p(x \mid \hat{x}) = u^{-1}\varphi\left(u^{-1}(x - \hat{x})\right). \tag{11}$$

Note the difference with formula (10): there the wright-side member is a function of $\hat{x}$, with x fixed, here instead x is the variable and $\hat{x}$, but otherwise there is a perfect symmetry.

The probability $P(H \mid E)$, can also be calculated as

$$\begin{aligned} P(H \mid E) &= P(x > a \mid \hat{x}) = \int_{x>a} p(x \mid \hat{x})\mathrm{d}x \\ &= \int_{x>a} u^{-1}\varphi\left(u^{-1}(x - \hat{x})\right)\mathrm{d}x, \end{aligned} \tag{12a}$$

and

$$P(\bar{H} \mid E) = P(x \le a \mid \hat{x}) = 1 - P(H \mid E). \tag{12b}$$

[10] In reality, this holds true only when we far enough from the origin of the x axis. Otherwise, since the quantity is non negative by definition, the probability of a value ≤ 0 collapses in the (finite) probability of the zero value.

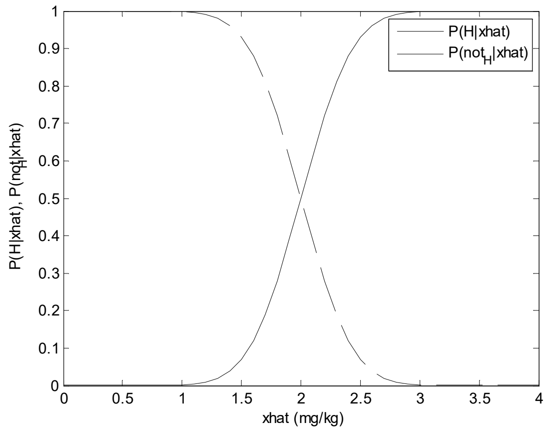

Fig. 1. The probabilities $P(H \mid E)$ and $P(\bar{H} \mid E)$, as a function of the measurement value $\hat{x}$.

A picture of these probabilities, as a function of the measurement value, $\hat{x}$, is shown in Figure 1.

After this premise, let us now consider different ways of expressing the result of measurement and the corresponding decisional criteria.

In the case the sole measurement result is communicated, we may only take

$$H \Leftrightarrow \hat{x} > a = 2.0 \text{ mg/kg}. \tag{13}$$

But, as we have seen, this strategy *does not account for all the information associated to the measurement process, which should include an uncertainty statement* [13, 14].

If the result of measurement is instead expressed as a coverage interval, based on the expanded uncertainty, $x = \hat{x} \pm U$, an *uncertainty region* appears. A decision can be taken only by considering to which part the *burden of proof* is to be applied. Consider that *H has to be proved*, we would take

$$H \Leftrightarrow \hat{x} > a + U = 2.7 \text{ mg/kg}, \tag{14}$$

having assumed $U = 2u$ and rounded off the result. Yet there still a possibility of this to be not true, and the probability associated to this event remains undefined.

So far we have considered cases in which the measurement result is provided according in the typical current formats. If instead a probabilistic approach is followed, an entire set of alternative possibilities opens.

One is offered by *statistical significance testing*. In this case also, we should consider on which side the burden of proof stands, since the procedure is highly asymmetrical. Assuming that we have to prove H, that is that the pesticide content is over the threshold, the null hypothesis, that is the hypothesis to be contrasted is $\bar{H}$, that is $x \le a$. We have therefore to prove that the assumption that the amount of pesticides is *under* the threshold is *not* supported by data. To perform the test we have to consider the probability (density) of the measurement value when $\bar{H}$ is true, that is

$$p(\hat{x} \mid x \le a) \propto \int_0^a p(\hat{x}, x)\mathrm{d}x = \int_0^a p(\hat{x} \mid x)\, p(x)\mathrm{d}x . \quad (15)$$

After assuming a probability density $p(x)$ as above specified, we obtain the result shown in Figure 2, and the related cumulative distribution in Figure 3.

Then, after choosing a "high" probability p_0, we can identify a second, safeguard, threshold b, such that,

$$P(\hat{x} \le b \mid x \le a) == \int_0^b \left[\int_0^a p(\hat{x} \mid x)\, p(x)\mathrm{d}x \right] \mathrm{d}\hat{x} = p_0 . \quad (16)$$

For example, for $p_0 = 0.99$, the corresponding decisional rule will be, in our example:

$$H \Leftrightarrow \hat{x} > b = 2.4 \text{ mg/kg} . \quad (17)$$

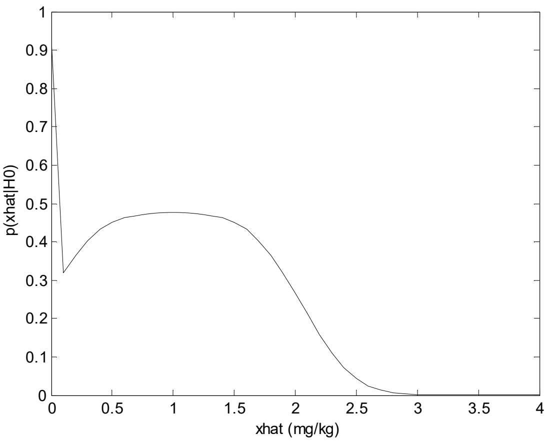

Fig. 2. The probability density $p(\hat{x} \mid x \le a)$, as a function of the measurement value $\hat{x}$.

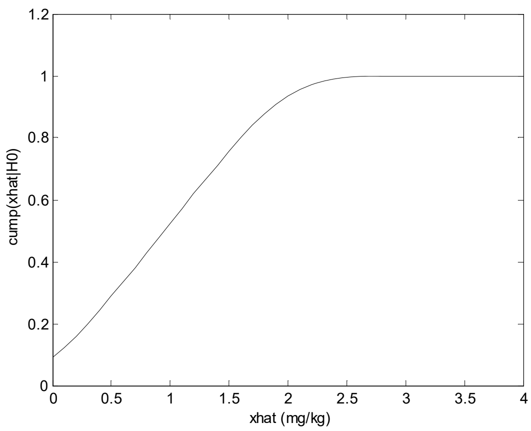

Fig. 3. The cumulative distribution for $\hat{x}$, given $\bar{H}$, as a function of the measurement value $\hat{x}$.

Note one major difference with the previous approaches that do not use explicitly probability: *here the scientific basis of what we have done is clearly identified: we affirm* H *, since the alternative hypothesis has been rejected at a* p_0 *significance level*. On the hand, note that this method strongly depends upon the choice of the probability distribution $p(x)$ [11].

Another possibility is offered by the *likelihood ratio*, which is sometimes envisaged as a possible alternative to significance testing, especially in the neo-Bayesian school of thought. The result in our example, is reported in Figure 4. It is apparent that there is a sudden increase between say, $\hat{x} = 2.3$ mg/kg and $\hat{x} = 2.5$ mg/kg, but it is hard to identify an appropriate threshold for the decision.

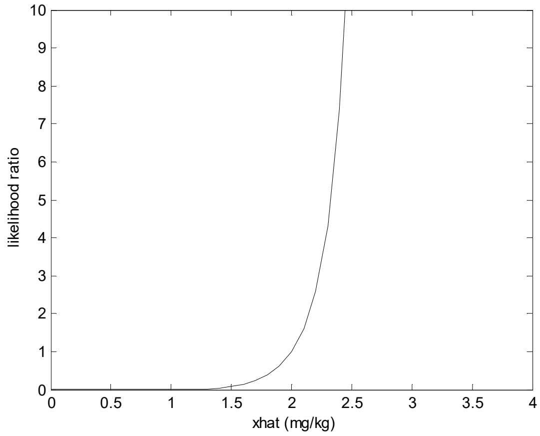

Fig. 4. The likelihood ratio as a function of the measurement value $\hat{x}$.

A further alternative consists in considering the probability $P(H \mid \hat{x})$ and to affirm H if such probability is "high", that is if it is greater that a previously selected value, p_1. If we take, e.g., $p_1 = 0.95$, considering the probabilities in Figure 1, we obtain

$$H \Leftrightarrow \hat{x} > 2.6 \text{ mg/kg} . \quad (18)$$

But is it p_1 appropriate?

Lastly, an alternative possibility is to consider, instead of a single probability value p_1, a "degree of certainty", based on the above mentioned Himmel's scale, which in our example, looks like in Table 2. Therefore, an appropriate "degree of certainty" can be selected, depending upon the gravity of the decision to be taken.

[11] The methods to follow also depend upon $p(x)$, but such dependence is less critical.

Table 2. An application of Himmel's scale to the example under consideration.

Measurement value $\hat{x}$	**Degree of certainty about** H
$\hat{x} > 3.0$	Practically proved
$2.8 \le \hat{x} < 3.0$	Extremely likely
$2.6 \le \hat{x} < 2.8$	Very likely
$2.4 \le \hat{x} < 2.6$	Likely
$2.3 \le \hat{x} < 2.4$	Undecided
$\hat{x} < 2.3$	Not useful

5. FINAL REMARKS

We have offered an introductory discussion of the role of measurement in forensic science. We have shown, in particular, how to included measurement information in the decisional process, in such a way as to properly account for measurement uncertainty. This requires a proper modelling of the measurement process and the calculation of the related probability distributions. We have outlined some critical aspects that require special attention, to avoid biased results.

We have also suggested the development and use of scales of degree of uncertainty for a proper interpretation of scientific evidence. From the analysis here proposed, it appears how decision-making requires the combination of different pieces of expertise. It is very important that the entire decisional process that involves scientific evidence is modelled, in such a way as to including information coming from different sources. In this perspective the (trite) question of the subjective or objective nature of probability becomes, in our opinion, totally irrelevant.

On the other hand, the importance of multi- and interdisciplinary research should be stressed, as well as of the development of a common measurement theory and science.

REFERENCES

[1] A. Dondi, Problemi di utilizzazione delle "conoscenze esperte" come "expert witness testimony" nell'ordinamento statunitense, in Rivista trimestrale di diritto e procedura civile, 2001, pp. 1133 ss.

[2] A. Dondi, Utilizzazione delle conoscenze esperte nel processo civile – Alcune ipotesi di caraattere generale, in Studi di diritto civile uin onore di Giuseppe Tarzia, Milano, 2005, pp. 843 ss.

[3] F. Taroni, S. Bozza., C. Aitken, Decision analysis in forensic science, Journal of Forensic Science, 5 (2005) 894-905

[4] G. B. Rossi, F. Crenna, A probabilistic approach to measurement-based decisions, Measurement 39 (2006) 101-119.

[5] MG Cox, Gb Rossi, PM Harris, A Forbes, A probabilistic approach to the analysis of measurement processes, Metrologia 46 (2009) L9-L10

[6] I. Lira, Comment on 'A probabilistic approach to the analysis of measurement processes', Metrologia 46 (2009) L8

[7] MG Cox, Gb Rossi, PM Harris, A Forbes, Reply to the comment on 'A probabilistic approach to the analysis of measurement processes', Metrologia 40 (2008) 493-502

[8] J. Wallace, Ten methods for calculating the uncertainty of measurement, Science and Justice, 50 (2010) 182-186.

[9] D. Kahneman, Thinking, fast and slow, Allen Lane, London, 2011

[10] M. C. Galavotti, Probability, statistics and law, in D. Dieks et al. (eds.) Probabilities, Laws, and Structures, Springer, 2012.

[11] L. Pendrill, Using measurement uncertainty in decision-making and conformity assessment, Metrologia 51 (2014) S2016-S218.

[12] G. B. Rossi, Measurement and Probability, Springer, 2014.

[13] A. Ferrero, V. Scotti, Forensic metrology: a new application field for measurement experts across techniques and ethics, IEEE Instrumentation and Measurement Magazine 16 (2013) 14-17

[14] A. Ferrero, V. Scotti, Forensic metrology: when measurement science meets ethics, Proc. IEEE Int. Symp. On Etics in Engineering, Science and Technology, ETHICS 2014, Chicago, IL, USA May 2014

[15] C. Aitken, P. Roberts, G. Jackson, Fundamentals of probability and statistical evidence in criminal proceedings, Royal Statistical Society, 2014.

[16] P. Garbolino, Probabilità e logica della prova, Giuffré, 2014.