XVI IMEKO World Congress
Measurement - Supports Science - Improves Technology - Protects Environment ... and Provides Employment - Now and in the Future
Vienna, AUSTRIA, 2000, September 25-28

DISTRIBUTION-FREE POSSIBILITY MODELLING OF POOR SENSOR INFORMATION

G. Mauris, V. Lasserre and L. Foulloy
Laboratoire d'Automatique et de Micro-Informatique Industrielle LAMII/CESALP
Université de Savoie, 41, Avenue de la plaine, 74016 Annecy Cedex, France

Abstract: At the application level, it is important to be able to define around the measurement result an interval which will contain an important part of the distribution of the measured values, that is, a confidence interval. When the sensor uncertainty is represented by a probability distribution, the confidence intervals can be easily deduced from it. But when the probability distribution cannot be identified due to poor sensor information, a more generalised representation must be used. To obtain confidence intervals in such a situation, available probabilistic methods are essentially the Bienayme-Chebychev and the Camp-Meidel inequalities. In this paper, after having recalled these methods, alternative approaches based on the possibility theory are considered. Distribution-free possibility distribution building based on the sets of all confidence intervals is proposed. According to the knowledge of uncertainty that is available, i.e. the range or the standard deviation of the measures, triangular and truncated triangular possibility distributions are respectively considered. These different possibility distributions which are fuzzy sets with uncertainty semantics are then compared in terms of the information provided.
Keywords: measurement uncertainty, possibility theory, probability theory

1 INTRODUCTION

Uncertainty is a key concept for the measurement expression [1][2][3]. Indeed, in many application domains, it is important to take the measurement uncertainties into account, especially in order to define around the measurement result an interval which will contain an important part of the distribution of the measured values, that is, a confidence interval [4]. Such an interval allows to define later decision risks, as for example the risk to accept a defective lot, the risk to exceed an alarm threshold, etc. Moreover, when the measurement uncertainty is defined, it is also sometimes necessary to propagate it. For example, the uncertainty associated to an environment temperature measurement will have to be propagated to define the uncertainty associated to the sound speed in this environment.

A main tool to deal with sensor measurement uncertainty is statistics [5]. This tool needs a mathematical support to be used, especially to propagate uncertainties. Two theories are mainly considered : the interval calculus [6] and the probability theory [5]. Although the interval calculus allows simple calculations, the resulting model is very imprecise. Moreover, it only supplies the confidence interval of the 100% confidence degree. Thus, the use of the probability theory seems to be necessary. But, the use of the probability theory to supply confidence intervals requires the identification of the probability law. This identification is not always possible, especially when too poor information is available - for instance, when this information is supplied by an expert or a poor sensor. In such cases, a distribution-free representation is required, that is, a way to supply confidence intervals without knowing the probability law. The purpose of this paper is to address this problem with the possibility theory by making bridges with the probability theory.

The paper will be organised as follows. The second section deals with the probabilistic solutions to build a distribution free representation of measurement uncertainty, i.e. the Bienayme-Chebychev and Camp-Meidel inequalities. We show how their results lead to a possibility distribution [7][8] (which is a normalised fuzzy set [9]), thus making a bridge between the probability theory and the possibility theory via the notion of confidence intervals. In the third section, we propose two other distribution-free possibility representations, having triangular or pseudo-triangular shapes, which can be used when a preferred value and a parameter characterising the uncertainty are available. The nature of this uncertainty parameter determines the possibility distribution that can be used:

- the triangular one when the range of the measurements is supplied,
- the pseudo-triangular one when the standard deviation of the measurements is supplied.

We show that, in both cases, the required confidence intervals can be very easily deduced by cutting the distribution horizontally at the considered levels. Finally, in the fourth section, a comparison of the different distribution-free possibility representations is proposed.

2 DISTRIBUTION-FREE PROBABILITY UNCERTAINTY REPRESENTATION

2.1 Bienayme-Chebychev and Camp-Meidel inequalities

When the probability law of considered measurements is unknown, the confidence intervals can be supplied by the Bienayme-Chebychev inequality that can be written as (let X be the quantity to be estimated):

$$P(x_m - ks \le X \le x_m + ks) \ge 1 - 1/k^2 \tag{1}$$

It allows to define confidence intervals of confidence value k centered on the mean-value x_m of an unknown probability law p, whose associated probability measure is noted P. Note that p only needs to be characterised by its mean-value x_m and its standard deviation σ, it does not necessarily need to be unimodal and symmetrical.

If the probability law is known as being unimodal and symmetrical, the Camp-Meidel inequality can be applied :

$$P(x_m - ks \le X \le x_m + ks) \ge 1 - 1/2.25k^2 \tag{2}$$

The confidence intervals for the Camp-Meidel inequality have a smaller length than those obtained by the Bienayme-Chebychev inequality. This difference is due to the richer knowledge available for the probability distribution in the former case.

2.2 Bridges with the fuzzy/possibility theories

In fact, the stacking of the whole set of confidence intervals (obtained with k varying between 0 and 1) defines a possibility distribution π having a value of 1 for x_m (on which the confidence intervals are centered)(see figure 1).

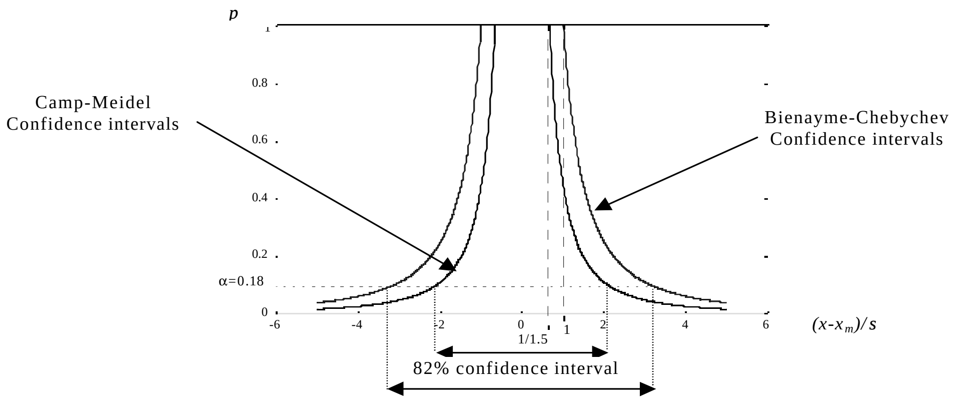

Figure 1. Bienayme-Chebychev and Camp-Meidel confidence intervals

In fact, the α–cuts of this possibility distribution (i.e. intervals I_α defined such as $I_\alpha=\{x/\pi(x)>\alpha\}$) are the confidence intervals of confidence degree $1-\alpha$. It is the possibility distribution π, which is the most specific possibility distribution consistent with all the probability distributions p, such that the inequality $\Pi(A)$♠$P(A)$ holds, with Π and P being respectively the possibility and probability measures associated to the distribution. The possibility distribution thus built can be seen as an upper bound of a set of probability distributions. Horizontal and vertical bridges are thus established between the probability and possibility theories.

3 DISTRIBUTION-FREE POSSIBILITY UNCERTAINTY REPRESENTATION

3.1 How to build a distribution-free possibility representation?

In practical cases, the physical measurement is characterised by a preferred value and an uncertainty parameter. The nature of this uncertainty parameter will determine the proposed representation that will be used:

- the triangular possibility representation when the range of the measurements is known,
- the truncated triangular possibility representation when the standard deviation associated to the measurements is known.

The question of the evaluation of the preferred value will not be tackled in this paper. Anyway, for unimodal and symmetrical probability distributions that are considered in this paper, the preferred value is generally given by the mean value x_m of the measures acquired.

3.2 The triangular possibilistic representation

In this paragraph, we consider the case where the range r of the measures is known (it is often the case when the uncertainty is estimated by an expert). In this case, due to the symmetry and the monotony property of the probability distribution on each side of x_m, the stacking of the confidence intervals has a curved shape on the left and on the right of x_m, whatever the probability law shape is. Though all the probability laws are bounded by the interval $[x_m - r/2, x_m + r/2]$, the triangular possibility distribution, having this interval as support and x_m as modal value, is an upper bound of the confidence intervals of any unimodal and symmetric probability law defined on $[x_m - r/2, x_m + r/2]$ (see figure 2).

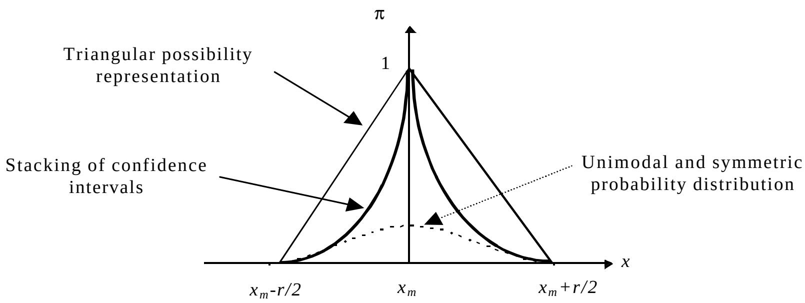

Figure 2. Distribution-free triangular possibility representation

Therefore, the triangular possibility (t.p.) representation provides a consistent distribution free possibility representation of measurement uncertainty.

3.3 The truncated triangular possibilistic representation

In this paragraph, we build a distribution-free possibility representation which can be used when the known uncertainty parameter is the standard deviation (which often happens when the uncertainty is estimated statistically from a set of measures). Indeed, it is a parameter which is much less unstable than the range [5].

For the sake of practical considerations, we will only consider the most common unimodal and symmetrical probability distributions of given mean value x_m and standard deviation σ, i.e. Gaussian, double-exponential, triangular laws, and also the uniform law. The stackings of their associated confidence intervals under the form of possibility distributions are plotted in figure 3.

If the upper bound of these possibility distributions is considered, a distribution-free possibility representation is obtained. According to the characteristics of these distributions, a truncated triangular possibility distribution can be a simple upper bound of them, consistent with the considered probability laws. The truncated triangular possibility (t.t.p.) representation is thus defined by the following four parameters x_m, x_e, e, x_n (see figure 3).

As the possibility distributions associated cross each other in the same point : $x_e = x_m + 1.73s$ (see figure 3), this point will be used to define the truncation point. The value ε=0.086 is deduced by taking the maximum of the possibility distributions at the point x_e,. Finally, in order to have an easy to handle possibility representation, this latter has also been bounded by $x_n = x_m + 3.2s$ which corresponds to the largest confidence interval of *99%* confidence for the considered four probability laws. The t.t.p. representation thus defined provides a simple distribution-free possibility representation of measures whose α-cuts bound the confidence intervals of the four probability laws.

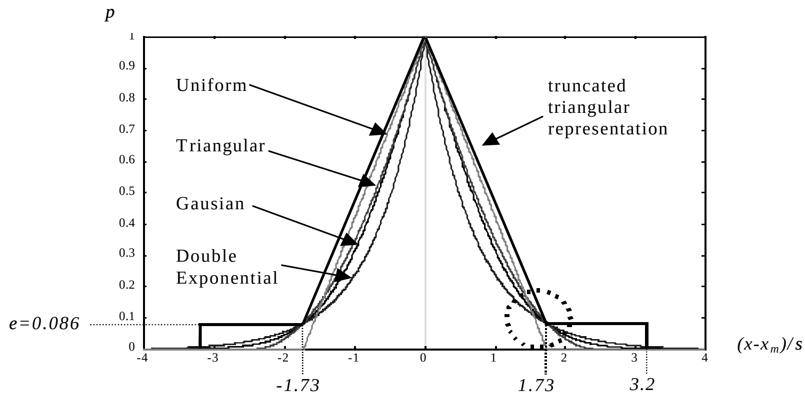

Figure 3. Possibility distributions associated to the four considered probability laws

4 COMPARISON

4.1 Problematics of comparison

To compare the different possibility representations of measurement uncertainty, criteria must be defined. Let us recall that though the possibility measure must be superior to the probability measure (i.e. to satisfy the consistency between the two theories), distribution-free possibility representations have been built based on this property. This possibility distribution is then used in further treatments in order to take decisions. Therefore, it is important for the different values possible for describing the entity under measurement to be clearly distinguished. This leads to the concept of specificity of a possibility distribution. The specificity value is defined by computing the area under the possibility distribution [8]. The smaller its specificity value is, the more informative the considered possibility distribution is.

4.2 Comparison with the Bienayme-Chebychev and Camp-Meidel inequalities

The Bienayme-Chebychev inequality supplies a possibility distribution which is far from the t.t.p. representation (see figure 4) in terms of specificity. It is understandable because the t.t.p. representation is an approximation of the set of confidence intervals for the four considered probability laws. On the contrary, the Bienayme-Chebychev inequality gives a set of confidence intervals for whatever the probability there are. Note that the Camp-Meidel inequality supplies a possibility distribution that it is neither far from what a superior bound of the four optimal possibility distributions could be, and nor far from the t.t.p. representation (see figure 4). Anyway, the t.t.p. representation is more interesting in terms of specificity. It is understandable because the t.t.p. representation can only be used when the probability modelling of the measurement result can be assimilated to one of these four probability laws studied.

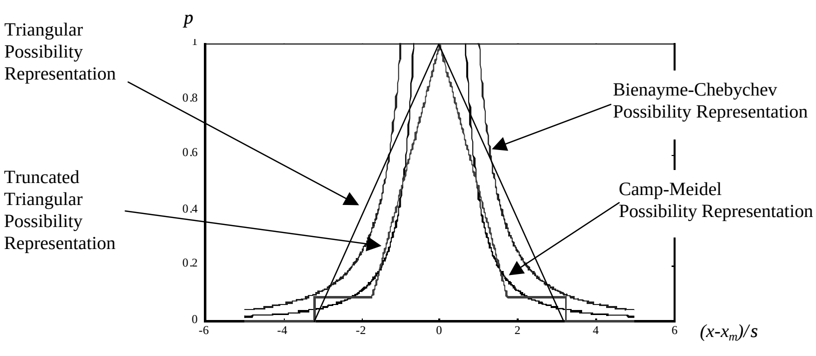

Figure 4. Comparison of possibility distributions

4.3 Comparison of the two distribution free possibility representations

To make the comparison, the two representations had to be built from the same uncertainty parameter for the considered measures. Let us first build the representation associated to the preferred value and the standard deviation of a measure set for which the range is not known. The support used for the t.p. representation will be $[x_m - x_n, x_m + x_n]$. The resulting specificity degree is $2.13s$ for the t.t.p. representation and it is $3.2s$ for the t.p. one. This t.p. representation is then less specific than the t.t.p. representation.

Then, let us build the t.t.p. representation associated to the preferred value and the range of a measure set whose standard deviation is not known. An upper value of the standard deviation to be associated to any probability distribution, whose support is known, can be deduced from this support, that is, from the range of the measure set. It is the standard deviation associated to the uniform law having the same support [5]. In the considered situation, the specificity degree associated to the t.t.p. representation is $2.13s$, and the specificity degree associated to the t.p. representation is $1.73s$. So, by using the t.t.p. representation instead of the t.p. one, some specificity is lost.

In fact, with the t.t.p. representation, a compromise between consistency and specificity is made. If less than these four probability laws are considered in order to build the distribution-free representation (for example, if the double exponential law is not taken into account), then some consistency can be relaxed and a better specificity will result. On the contrary, if more than these four laws are considered, then some specificity will be lost because the consistency condition will be more demanding. When the range and the standard deviation are both supplied, then the higher the standard-deviation/range ratio will be, the more interesting the t.p. representation will be.

5 CONCLUSION

In this paper, we have recalled the probabilistic methods to build a distribution-free representation of a measurement result, that is, to build the associated confidence intervals. Then, two possibility representations have also been proposed, according to the knowledge of uncertainty :

- when the uncertainty is expressed by the range of the measures, the triangular possibility (t.p.) representation is proposed.

- when the uncertainty is expressed by the standard deviation of the measures, the truncated triangular (t.t.p.) representation is proposed.

For both cases, only the measurement results which can be modelised by a unimodal and symmetrical probability law have been considered. We have shown that these possibility representations give interesting results comparable with the ones issued from probabilistic methods. When the uncertainty is expressed by a standard deviation, then a better specificity is obtained by using the t.p. representation. And when the uncertainty is expressed by the range of the measure set, then a better specificity is obtained by using the t.t.p. representation. So, it will not be useful to deduce a standard deviation from the supplied range of the measures, or vice versa, to improve the specificity of the possibility representation. And finally, when the range and the standard deviation are both supplied, then the higher the standard-deviation/range ratio will be, the more interesting the t.p.g. representation will be.

Our next studiess will concern the measurements results which can be modelised by a unimodal and dissymmetrical probability distribution. An important point is then to choose the preferred value. Must it be the modal, the mean or the median value ? It seems that the choice must be made before building a distribution-free possibility representation.

REFERENCES

[1] L. Mari, G. Zingales, Uncertainty in measurement science, in Proceedings of the Int. Workshop on Advances of measurement science, Kyoto, Japan, 1999, p. 177-189.
[2] L. Mari, Notes on the uncertainty of measurement, in Proceedings of the XV IMEKO World Congress, Osaka, Japan, 1999, Vol. II, p. 89-92.
[3] L. Reznik, K.P. Dabke, Evaluation of uncertainty in measurement: a proposal for application of intelligent methods, in Proceedings of the XV IMEKO World Congress, Osaka, Japan, 1999, Vol. II, p. 93-100.
[4] *Guide for the expression of uncertainty in measurement*, ISO 1993, 99 pages, 1993.
[5] Kendall M., Stuart A., *The advanced theory of statistics*, Ed. Griffin and Co., 1977.
[6] Moore R., *Interval Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1966.
[7] Zadeh L.A., Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems, Vol.1, no1 (1978) p. 3-28.

[8] Dubois D., Prade H., Fuzzy sets, probability and measurement, European Journal of Operational Research, no 40 (1989) p. 135-154.
[9] Zadeh L.A., Fuzzy sets, Information and Control 8 (1965), p. 338-353.

AUTHOR: Gilles MAURIS, Virginie LASSERRE and Laurent FOULLOY, Laboratoire d'Automatique et de Micro-Informatique Industrielle LAMII/CESALP, Université de Savoie, BP 806, 74016 Annecy Cedex, France, Tel: +33 450-09-65-52, Fax: +33 450-09-65-59, E-mail: mauris@esia.univ-savoie.fr