PERIODICA POLYTECHNICA SER. MECH. ENG. VOL. 37, NO. 4, PP. 223-235 (1993)

INCREASE OF INFORMATION IN THE COURSE OF MEASUREMENTS

PÉTER BÖLÖNI

National Office of Measures of Hungary
P.O.BOX 19, H-1531 Budapest, Hungary

Received: July 1, 1992

Abstract

Results of measurements are messages about the possible values of measurand. As measurands can be described as discrete sources of information Shannon's formula is applicable to describe the missing quantity of information both before and after the measurement. Quantitative description of the measurement uncertainty provides the chance to use the discrete version of the Shannon's formula. Fixing zero level of the necessary information different measurement results became comparable.

Keywords: measurement, information, quantity, uncertainty, measurand, source of information.

Introduction

An increasing part of human population is living in conditions of more and more technical nature. Some information of quantitative type are necessary to accommodate to and to improve these conditions. Primary sources of quantitative information are the rising number of measurements, results of which are obtained from measurement processes. Up to now the traditional definition of measurement emphasises its nature of process: Measurement is the set of operations having the objective to determine the value of a quantity [1]. In describing the performances of measuring instruments used for measuring quantities of more complex nature like spectra, distributions or integrated values over a period the definition of measurement by the theory of sets has some advantages: Measurement performs a transformation from the set of the possible values of the measurand to the set of possible results of measurements [2]. The growing need for a tool to compare measurements of different nature and the common use of computers to control measurements and to process results call for a third definition related to the generally accepted information obtaining character of measurements. This third definition might be as follows: Measurement is a process decreasing the insufficiency of information about the measurand. The aim of this paper is to show why measurement cannot eliminate the insufficiency of information completely and to show a way to characterize the quantity of information obtained in the course of measurements.

Information and Quantity of Information

Information in general sense is a certain concrete part of the human knowledge which can be described with words and phrases. From scientific point of view information is a basic term of the theory of information and cannot be defined within the frame of this theory itself (just like to the term of point which cannot be defined in the geometry, to the term of probability which cannnot be defined within the mathematical theory of probability and to the term of time which cannot be defined within the theory of physics). It is common to describe the method to measure the basic quantities of a science rather than to define the basic terms of them, e.g. time is the quantity measured with a clock in specified manner. Theory of information has a similar structure: instead of defining the term of information a method is given to measure the quantity of information. As all of the complex information can be expressed with a set of the elementary information (quantum of information equals the information content of an answer of YES or NO of equal probability to a question to be decided), the quantity of information of any complex message can be expressed as a function of information content of the answers YES and NO completely determining the message. Information content or information quantity of a message is a function of the number of questions to be answered with YES and NO (but the quantity of information is not equal to the number of questions itself).

Occurrence of one particular event in a complete set with N number of events discarding each other can be identified with $\log_2 N$ number of questions if the probability of occurrence of all the events is the same and the strategy of searching is optimal. (Optimal strategy is described in the theory of searching.) If the figures of 0, 1, 2, 3 ... 9 can occur on the second digit of a digital display of a frequency meter with equal probability, then $n = \log_2 10 = 3.32$ questions (to be decided or answered with YES-s and NO-s) are usually necessary to identify the figure really occurring on the second digit. Amount of information provided by the second digit equals the information content of 3.32 answers of YES and/or NO. As the possible number of events is not a whole power of 2, in most cases $\log_2 N$ is not a whole number. (But the average number of necessary questions is still characteristic of the system of events or of the message or of any other source of information, just like the average number of children in families belonging to specified social groups is not a whole number although the number of children is always whole in real families.)

Probability of occurrence of elementary events is not equal in most of the practical cases. There are in English more words beginning with L than with K, so probability of a single word message starting with L is higher than that of starting with the letter K. Probabilities of (elementary) events are not equal in this example. System of events with different probabilities of occurrence changes the optimal strategy of search. If, for example, the $P(B)$ probability of occurrence of the event marked with B is higher than 0.5 in a set of events comprising four events of A, B, C and D, then it is advisory to ask for the occurrence of event B first. If answer is YES, we have saved the rest of questions. Number of questions necessary to identify the i^{th} event equals $\log_2 \frac{1}{P_i}$ now denoting the probability of occurrence of the i^{th} event by P_i. This probability-based optimal strategy has a disadvantage of asking late for the less probable events. Identification of less probable events requires more questions in this way. Expectable number of questions changes from the case of equal probabilities

$$I = \log_2 N \tag{1}$$

to the case of different probabilities

$$I = \log_2 \frac{1}{P_i}\ , \tag{2}$$

where I is the expectable or average number of necessary questions,
N is the number of independent and mutually discarding events,
P_i is the probability of occurrence of the i^{th} event.

Unit of necessary questions seems to be identical to the unit of information which is 1 bit, the quantity of information of an answer to a question to be decided. (Most of information is located in the question and not in the answer in this type of questions and answers in common parlance.)

As a lot of messages have a random nature it is convenient to characterize the messages from the same source with the average quantity of information. Events of higher probability of occurrence share in the average with less number of necessary questions but with higher frequency whereas events of less probability of occurrence share in the average with higher number of necessary questions but less frequency of occurrence. Number of necessary questions is weighted with the respective P_i probabilities, this way, in the course of calculating the number of the average necessary questions or information. The in average necessary information equals

$$H = \sum_{i=1}^{N} P_i \cdot \log_2 \frac{1}{P_i} \text{ bit}\ . \tag{3}$$

Formula (3) has been introduced by dr. Claude Elwood Shannon. (It is to mention here that the author couldn't demonstrate the convergence of H average information characterising the discrete and the continuously variable sources when the discriminations of possible events converge to zero. As some model calculations have shown definite divergence, it is hard to compare the discrete and the continuous version of formulae for calculation of information obtained in the course of measurement.)

Prior Information about the Measurand

Both the increase of resolution of a measuring instrument and the accuracy of results are limited by instability of measurand and conditions, noise and the costs and time consumption of measurements. For this reason results of measurements always have a limited number of digits. As measurement results are messages of series of decimal digits with limited lengths (completed with decimal points, units, signs, etc.), the value of measurand can be always regarded as an element of a finite set of values. Consequently, the possible values of measurand can be described as a complete set of discrete values with P_i probability of occurrence of the i-th value. Specifying the respective probabilities at each possible consecutive value of measurand the probability distribution of measurand is achieved prior to the measurement. Description of measurand with a probability distribution prior to the measurements is unusual in metrology now but this is a way (if not the only way) to correspond the measurand and any other source of information theory to information it is dealing with.

Let us denote the resolution of measuring instrument or the smallest detectable difference of the measurand with Δx, and the limits of range within which the actual value of measurand x lays with $L \leq x \leq K$. In principle the $P(x_i)$ probabilities of the events that the actual value of the measurand fall in the range of

$$\left[x - \frac{\Delta x}{2}; \qquad x + \frac{\Delta x}{2}\right]$$

can be determined for all the x_i values. This set of events is complete because actual value of the measurand can fall in one of these intervals only. Two more intervals can be added, if necessary, one to describe the cases when the actual value of the measurand is smaller than the L lower limit of the measuring range of the instrument and one if the actual value is higher than the K upper limit but this slight refinement does not affect the principle. Actual type and parameters of $P(x)$ prior probability distribution describe our information about the measurand before measurements. Substituting the respective $P(x_i)$ and Δx values in formula (3) one can calculate the $H(1)$ information prior to the measurements. In fact, this quantity of information ought to be obtained to get the actual value of the measurand perfectly determined or known. This quantity of information is missing before the measurement.

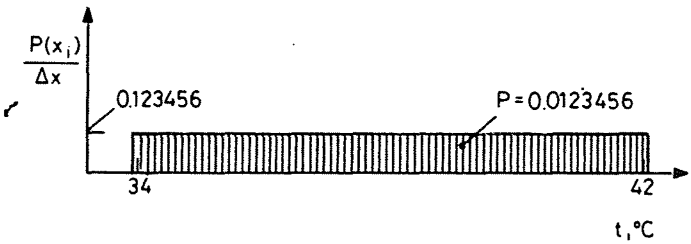

Fig. 1. Complete set of events and minimum level of information

Fig. 1 demonstrates minimum level of information about a measurand with a mercury in glass medical thermometer as an example. Let the resolution of a medical thermometer Δx correspond to the generally expected accuracy of 0.1 °C, the lower limit of the measurements range $L = 34$ °C and the upper limit $K = 42$ °C. Number of the possible events is

$$\frac{K-L}{\Delta x}+1=\frac{42-34}{0.1}+1=81 \ .$$

When we know that the actual temperature has to fall in one of these intervals but we do not know more then all the 81 intervals shall be assigned with the same probability of $P(i)$. As the actual temperature will fall in one of the intervals with $P = 1$ probability, sum of 81 equal $P(i)$ probabilities equals 1. Thus, the value of $P(i)$ is $1/81 = 0.01234$ now. Using formula (3) one can calculate the information necessary to identify the actual temperature:

$$H(0)=81\cdot 0.01234\cdot \log_2\frac{1}{\frac{1}{81}}=6.34 \text{ bit} \ .$$

That is to say fixing the resolution and the range, 6.34 bit information ought to be obtained to get the measurand perfectly described in this worst case (when nothing more can be supposed or expected about the measurand but the resolution and the range). Index 0 is to indicate the state of minimum level of information.

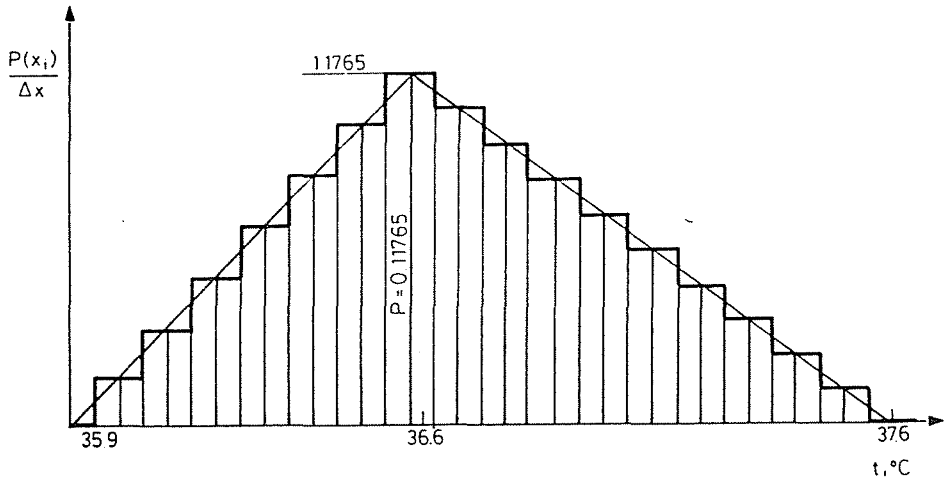

Fig. 2. Level of information before measurements

There are usually more information about the measurand than the resolution and the range before the measurements. The expectation of observer excludes some part of measurement range and gives different probabilities for the rest between maximum and zero. For a person looking healthy temperatures less than 36 °C and more than 37.5 °C are not realistic.[1] In this limited range a modus is of about 36.6 °C. Let us suppose in this example that probabilities from this maximum value decrease to zero on both sides of the modus with constant slope according to the observers' opinion. Table 1 lists the respective $P(x_i)$ probabilities.

Table 1

35.9 °C	0.0000	36.5 °C	0.1008	37.1 °C	0.0588
36.0 °C	0.0168	36.6 °C	0.1176	37.2 °C	0.0471
36.1 °C	0.0336	36.7 °C	0.1059	37.3 °C	0.0353
36.2 °C	0.0504	36.8 °C	0.0941	37.4 °C	0.0235
36.3 °C	0.0672	36.9 °C	0.0824	37.5 °C	0.0118
36.4 °C	0.0840	37.0 °C	0.0706	37.6 °C	0.0000

A more realistic level of prior information can be calculated now with formula (3). Index 1 is to indicate the realistic level of information before

[1] These values correspond to the practice of medical thermometry used in Central Europe.

measurement. In this example $H(1) = 3.79$ bit. That is to say fixing the resolution and the range and supposing certain preliminary distribution of measurand, 3.79 bit information ought to be obtained to get the measurand perfectly described. Difference in values of $H(0)$ and $H(1)$ indicates a difference in the prior knowledge about the measurand.

Uncertainty of Measurement Results and Residual Insufficiency of Information

Apart from the specification of measurand and conditions measurement results consist of

- the numerical value for estimation of the true value of the measurand,
- the unit or the reference to the conventional scale used,
- the uncertainty of the measurement result and
- the specification of the supposed type of probability density function of results (or errors). Most of metrologists agree that lack of specification on the type of probability density function corresponds to the supposition of the normal (Gaussian) distribution of the observations and errors.

Supposing a type of distribution function of observations (or errors) and using the parameters characterising the uncertainty of results, one can calculate the $Q(x_i)$ probabilities of events, the true value of measurand falling in the range of

$$\left[x - \frac{\Delta x}{2}; x + \frac{\Delta x}{2}\right] .$$

In the most part of the measuring range Q will not differ from zero practically. For the rest of few x values the $Q(x_i)$ probabilities are differing from zero and can be substituted in formula (3).

$$H(2) = \sum_{i=1}^{N} Q(x_i) \log_2 \frac{1}{Q(x_i)} \text{ bit} . \tag{4}$$

Index of 2 is to indicate the quantity of missing information characterising the level of information about the measurand after the measurement has been completed.

Let us suppose a result of (36.8 ± 0.1) °C. As this x_m result is the best possible estimation of the true value of measurand, the probability of $Q(x_m)$ is the highest, say 0.96. The uncertainty specification of 0.1 °C allows the measurand to be 36.7 °C and 36.9 °C as well. The supposed distribution function of results permits to calculate the probabilities of all particular results (not only the maximal one). Let both the 36.7 °C and 36.9 °C

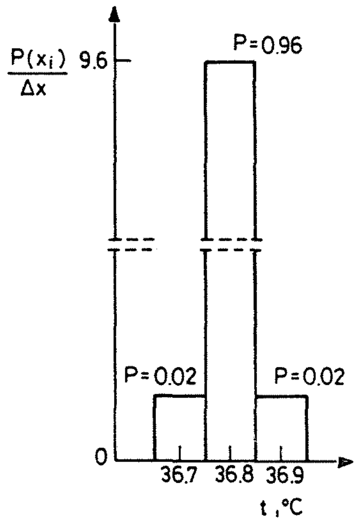

Fig. 3. Level of information after measurements

results have $Q = 0.02$ probability and the rest of the x values practically zero. Shannon's formula (3) allows again to calculate the missing part of information. It is

$$H(2) = 2 \cdot 0.02 \cdot \log_2 \frac{1}{0.02} + 0.96 \cdot \log_2 \frac{1}{0.96} = 0.25 \text{ bit.}$$

A much better level of information has been calculated now from the probability distribution of results (and not from the possible values of measurand) with formula (4). In this example $H(2) = 0.25$ bit. That is to say having the result and supposing some distribution of it, 0.25 bit information ought to be obtained to get the measurand perfectly described. Difference in values of $H(0)$ or $H(1)$ and that of $H(2)$ indicates the difference between the prior knowledge and the amount of knowledge after measurement.

It might be interesting to note that if the result were associated with uncertainty less than $\Delta x/2$, then the only $Q(x_m)$ probability would be equal to 1 (see *Fig. 4*). In this case Shannon's formula gives $H(p) = 0$, that is to say no information is needed to complete our knowledge about the measurand if the result does not have uncertainty more than $\Delta x/2$. This is an indication of the principle perhaps expressed by Hinchin for the first time: Shannon's formula measures the quantity of information necessary to eliminate the insufficiency of information about the state of a source rather than measuring the information content of the source itself (4)

$$H(P) = P \cdot \log 1/P = 1 \cdot \log 1/1 = 0 \text{ bit .} \tag{5}$$

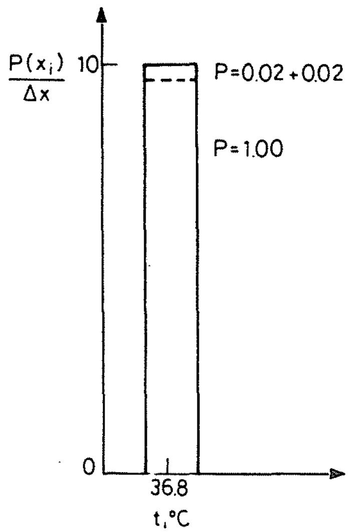

Fig. 4.

Quantity of Information Obtained in the Course of Measurement

In order to determine the information obtained by a measurement the previous considerations shall be summed up and completed first:

$H(0)$ or $H(1)$ information shall be obtained to eliminate the insufficiency of information prior to the measurement.

$H(2)$ information shall be obtained to eliminate the insufficiency of information rest after the measurement.

$H(p) = 0$, no information is needed when the measurand (or the state of information source) is perfectly determined.

Mathematical properties of equivalency and additivity are valid for two quantities of information both of them described by Shannon's formula.

Let us denote the level (or amount) of information associated with measurand prior to the measurement by $G(0)$ or $G(1)$ and by $G(2)$ after the measurement. Adding $H(0)$ or $H(1)$ information to the prior value of $G(0)$ or $G(1)$, respectively, one comes to the state of perfectly determined $H(p)$ when no information is needed. Similarly, adding $H(2)$ information to the $G(2)$ level of information after the measurement, one comes to the state of perfectly determined $H(p)$ when no information is needed.

$$G(0) + H(0) = H(p) \text{ or } G(1) + H(1) = H(p) \text{ or } G(2) + H(2) = H(p)\ . \quad (6)$$

As $H(p) = 0$ according to (5) and all the H values calculated with the Shannon's formula are not negative, we have got negative values for all the levels of information characterising the measurand (before and after the measurement, too):

$$G(0) = -H(0) \text{ or } G(1) = -H(1) \text{ and } G(2) = -H(2)\,. \tag{7}$$

The $H(m)$ amount of information obtained by the measurement can be calculated finally as a difference of levels of information after and before the measurement:

$$H(m) = G(2) - G(0) \text{ or } H(m) = G(2) - G(1)\,. \tag{8}$$

Substituting the calculable values from (7) we got:

$$H(m) = H(0) - H(2) \text{ or } H(m) = H(1) - H(2)\,. \tag{9}$$

Supposing equal probability of all temperatures in the range, the measurement resulted in $H(m) = 6.34 - 0.25 = 6.09$ bit of information. Having some experience in medical thermometry, that is using some previous information, $H(m) = 3.79 - 0.25 = 3.54$ bit information is gained. As the information quantity of any messages about the state of source of information or any result of measurement calculated with Shannon's formula is always more than zero, the level of information or quantity of information of the source or the measurand has to be less than zero (to meet the requirement of formula (5)).

Examples

E1. Let us have two pressure gauges of Bourdon type with $0\ldots 100$ bar measurement range each, with resolutions 0.5 bar and 0.1 bar and the error limits specified for gauges 1 bar and 0.1 bar, respectively. How much more information can the second gauge supply about the measurand? (For simplicity's sake we suppose that the measurand does not have any preferred range or value and the probability of results is equal within the ranges of error limits).

Number of the different possible values of measurand for the first gauge is 200 and 1000 for the second. The number of permissible results are 5 with the first gauge as there are 1 bar/0.5 bar = 2 more possible values within the error range in both sides of the result itself. For the second gauge the number of permissible results are 3. The difference of supplied information $H(D)$ equals:

$$H(D) = [H(0)_2 - H(2)_2] - [H(0)_1 - H(2)_1]\,,$$
$$H(D) = [\log_2 1000 - \log_2 3] - [\log_2 200 - \log_2 5]\,,$$
$$H(D) = 3.06 \text{ bit}\,.$$

E2. Let us have results of known standard deviation of S from a normal (Gaussian) distribution function. How much more information will the average of nine repeated measurements provide about the measurand than the average of two if the resolution is $0.1 \cdot S$? (For example $S =$ 1 mg, 1μV, 1 nm, etc.).

For normal distributions the $S(A)$ standard deviation of the average of results equals the standard deviation of results over the square root of the n number of results $S(A) = S/\sqrt{n}$. To use Shannon's formula the respective probabilities shall be substituted calculated with the formula for normal distribution:

$$P(x_i) = \frac{1}{\sqrt{2\pi} S(A)} \cdot e^{\frac{-(x_i-m)^2}{2S^2(A)}} \cdot \Delta x \,. \tag{10}$$

Performing the calculation

$$H(2)_2 = \sum_{x_i-m=-3S(A)}^{x_i-m=3S(A)} P(x_i) \cdot \log_2 \frac{1}{P(x_i)} = 2.43 \text{ bit} \,.$$

For the average of nine results the calculation gives the result $H(2)_9 = 1.89$ bit. Amount of a prior information $H(0)$ can be considered equal for the two cases as neither measurand nor range and resolution changed. The difference of information $H(D)$ is obtained from two cases:

$$\begin{aligned}
H(D) &= [H(0) - H(2)_9] - [H(0) - H(2)_1] \,, \\
H(D) &= [H(0) - 1.89] - [H(0) - 2.43] \,, \\
H(D) &= 0.54 \text{ bit} \,.
\end{aligned}$$

It is to note here that supposing the same level of a prior information, the difference of information obtained by two different measuring instruments and/or methods only depends on the instruments and methods as the $H(0)$ values are eliminated mathematically from the expression of difference. This approach allows to demonstrate that the sharper the distribution of results, the less the uncertainty after measurement.

Doubts and Further Considerations

As it was mentioned earlier, no bridge has been found up to now between the continuous and the discrete version of the theory of information for the interpretation of measurement results. The problem is that most of the metrologists use continuous functions as hypothesis because of the simplicity of use and the extensive practice of application. But description of the measurand with continuous values supposes exact – without uncertainty – results. Such a result could be only written as an infinite long message which ought to contain infinite amount of information. The continuous version of the theory gives similar answers to some questions but the results were not found convertible numerically.

It is easy to identify the source of information and the measurand at the first sight. But some principal difficulties must be overcome to use the above described approach for metrological purposes. The true value of the measurand is usually not known whereas the message is considered to be known completely (before coding). Measurements are often changing the value of the measurand whereas it is hard to find any cases when coding changes the original state of the source. Noise does not affect the coding but affects the measuring processes, and measurement results are messages usually not intended to decoding at all. But the chance to characterise the information quantity of measurements gives good reason for efforts of further improvement of the application of information theory in metrology.

Summary

Determining the prior distribution probabilities of the measurand's different values, Shannon's formula becomes applicable to calculate the missing quantity of information about the measurand. As all the measurement results have more or less uncertainties, some insufficiency of information will remain after each measurement. Fixing a zero level, the information providing property of different measurements can be compared.

Acknowledgement

The author wishes to express his grateful thanks to his colleagues Mr László Tihanyi and to Mr Béla Gyarmati for their inspiration, contributions to the development of the ideas and assistance in excavating the base concepts out of the great number of papers and references of confusing nature in some cases. A lot of thanks are due to Mr György Pataki for critiques to fit the ideas to more general aspects of engineering and for the assistance provided to the demonstration.

References

1. ISO IEC OIML BIPM: International Vocabulary of Basic and General Terms in Metrology.
2. Introduction to Metrology General. Textbook of the National Office of Measures of Hungary, 1988. Edited and partly written by dr Péter Bölöni (in Hungarian language).
3. CLAUDE E. SHANNON – WARREN WEAWER: The Mathematical Theory of Communication, University of Illinois, 1972.
4. A. JA. HINCSIN: Ponjatie entropii v teorii verojatnoszyej. Uszpehi matematicseszkih nauk, 1953.