PERIODICA POLYTECHNICA SER. MECH. ENG. VOL. 40, NO. 2, PP. 121–142 (1996)

QUALIFICATION OF MEASURING INSTRUMENTS ON THE BASIS OF INFORMATION THEORY

János ROSTÁS, Zoltán KAPOSVÁRI and Mihály SZALAI

Department of Precision Mechanics and Optics
Technical University of Budapest
H–1521 Budapest, Hungary

Received: Oct. 20, 1995

Abstract

Measuring instruments can be interpreted as tools for gaining information – since taking measurements is in fact gathering information. This way instruments are qualified according to the efficiency of the information gathering process.

For this purpose parameters to describe efficiency both in the context of gaining information compared to our *a priori* knowledge and the efficiency of the sensor to display channel are introduced. The measuring instrument is described as an information transmitting channel and efficiency is described using measures well accepted in information theory (entropy, mutual information, information gain, etc.).

The new method has an ability to describe measuring instruments solely on the basis of their potential for providing information and independently of their mechanical structure, working principle, etc.. The advantage of the method over conventional ones is that instrument qualification can be based on a principle related more closely to the core function of measuring instruments, that is, gathering information.

Keywords: measuring instrument, qualification, information theory.

1. Introduction

Since the publication of Shannon's fundamental works [1],[2] scientists have paid more and more attention to the exploration and examination of information theory's potentials in diverse fields (technical, economical sciences, biology, aesthetics) [5],[6],[7] in parallel with the development of the mathematical background [3], [4], [8].

Realising the growing potentials and universal applicability of information theory, we propose a method that applies it to the problem of qualifying measuring instruments.

It has been accepted that measuring instruments or their accuracy are described by error functions defined over the operating range [8]. An example is the concept of accuracy classes, which qualifies the instrument according to its allowed error expressed in percents of maximum capacity.

However, since the error is found as the reciprocal of accuracy the prescribed error margins can only be indirectly used to describe accuracy and they give no information regarding the parameter being measured and the instrument's information transmitting capability. Thus it seems to be necessary to introduce a new method or parameter that would have the advantage of being more objective than those used today. The method (parameter) developed by the authors qualifies measuring instruments by taking into account the statistical structure of the parameter measured and the instrument's information gathering efficiency.

2. The Possibilities Provided by Information Theory

Among the concepts of information transmitting

- Channel (more precisely channel transmittance) and
- Information Gain

seem to be well suited for the qualification of instruments.

From an Information Theory perspective any object that transmits information from one point to another in time or space may be called a *Channel* [1]. Let us suppose that information is represented by the value of some random variable U (input) which will appear at the other end of the Channel as another random variable X (output). This information transmitting process is illustrated in *Fig. 1*.

Fig. 1.

This so-called channel model can be interpreted as a model of taking measurements, as it is described below. The value of the measured quantity U (the focus of the information gaining process) is influenced by numerous external factors and the internal state of the object under measurement. The value X appearing on the instrument's display, beside the input U, is determined by a number of – known or unknown but, due to technical difficulties, not recognizable (internal and external) – factors. Random fluctuation of these factors results in similar fluctuations of U and X, rendering the probabilistic modelling of these latter very convenient. It is the inherent quality of the measuring instrument to be able to provide information on U through the readout X. This process of taking measurements can be modelled in fact with the channel model in *Fig. 1*, by substituting 'Instrument' for 'Channel'.

According to Information Theory, stochastically linked variables U and X provide information on each other's value. The information provided is quantified by the so-called *Mutual Information* $I(U, X)$. Communication theory uses the above quantity for describing Channel (transmitting) quality.

The other concept from Information Theory suitable for the description of instruments is Information Gain, since the measuring process can be regarded as an information gaining protocol. Our aim is to collect information on the yet unknown value of U with the help of the measuring instrument.

By Information Gain we mean any extra information obtained as compared to our a priori knowledge. A priori information may be represented by a distribution $Q(u)$ of the value of U. When the measuring process yields distribution $P(u)$ of U with a smaller uncertainty, we have gained information as compared to our a priori knowledge (see *Fig. 2* where $q(u)$ denotes the density function of distribution $Q(u)$ and $p(u)$ that of $P(u)$, the Information Gain is measured by the quantity $D(P||Q)$).

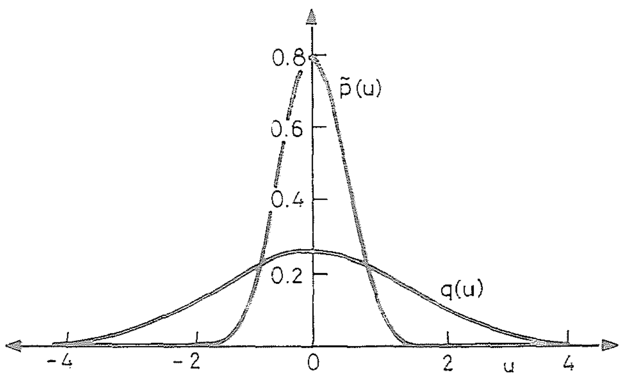

Fig. 2.

When taking measurements (for the first time) $Q(u)$ is best described by a Uniform distribution, containing the smallest possible amount of information. However, in practical situations there is always some a priori knowledge on U available. A simple example is an interval $[U^*_{\min}, U^*_{\max}]$ where we expect the value of U to fall in. Such intervals can be of different kind as illustrated in *Fig. 3*.

Due to a common reference – in case of a correctly selected instrument – this interval is the same as the measuring range of the instrument and will be refined later.

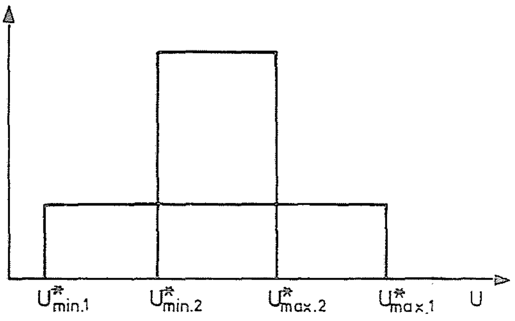

Fig. 3.

3. Discretization of the Random Variables

Although the quantities measured are often regarded as continuous, in practice there must be a finite difference $\Delta u (\Delta u \neq 0)$ between values to be discernible. In all practical situations there is a minimum Δu (a characteristic of the instrument called resolution) for which it is true that the instrument fails to discern two values differing by less than Δu. Within the measuring range of the instrument

$$N_u = \frac{U^*_{\text{max}} - U^*_{\text{min}}}{\Delta u} \tag{3.1}$$

exists. Let

$$\begin{aligned} U^*_0 &\equiv U^*_{\text{min}} , \\ U^*_j &= U^*_0 + j\Delta u , \quad j = 0, 1, 2, \ldots , N_u \end{aligned} \tag{3.2}$$

bounds for the intervals, let

$$q_j = P(U \in [U^*_{j-1}, U^*_j]) , \quad j = 1, 2, \ldots , N_u \tag{3.3}$$

be the probability of a measured value falling in the above interval. Further, let

$$P(U \in [U^*_{\text{min}}, U^*_{\text{max}}]) = 1 . \tag{3.4}$$

That is, the probability of the measured value being within the range of the instrument is 1. In other words, the set of all q_j is a certain event. That is

$$\sum_{j=1}^{N_u} q_j = 1 \,. \tag{3.5}$$

Following a similar path we conclude the same for read out values of X. That is

$$p_i = P(X \in [X_{i-1}^*, X_i^*]) \,, \quad i = 1, 2, \dots, N_x \,, \tag{3.6}$$

$$\sum_{i=1}^{N_x} p_i = 1 \tag{3.7}$$

and introducing p_{ij} product events we find

$$p_{ij} = P(\{X \in [X_{i-1}^*, X_i^*]\} \wedge \{u \in [U_{j-1}^*, U_j^*]\}) \,, \tag{3.8}$$

$$\sum_{j=1}^{N_u} \sum_{i=1}^{N_x} p_{ij} \,. \tag{3.9}$$

4. Applying the Channel Concept

In Information Theory the Mutual Information for U and X is given by [2]

$$I(X, U) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_u} p_{ij} \log \frac{p_{ij}}{p_i q_j} \,. \tag{4.1}$$

Further, the uncertainty of random variable X is described with

$$H(X) = -\sum_{i=1}^{N_x} p_i \log p_i \,, \tag{4.2}$$

the so-called Shannon's entropy [2]. A similar expression may be given for the uncertainty of the variable U.

In Information Theory, with the aid of

$$p_{i|j} = \frac{p_{ij}}{p_j} \,; \quad p_j \neq 0 \,, \tag{4.3}$$

conditional probabilities, the so-called conditional entropy

$$H(X|U) = \sum_{i=1}^{N_u} p_j \sum_{i=1}^{N_x} p_{i|j} \log p_{i|j} , \tag{4.4}$$

has also been introduced. This measures the uncertainty of the value X with condition U [2].

Using uncertainties the Mutual Information becomes

$$P(U \in [U^*_{\min}, U^*_{\max}]) = 1 \tag{4.5}$$

or

$$I(U, X) = H(U) - H(U|X) . \tag{4.6}$$

The Mutual Information above gives the amount of information the readout X has on U. If U and X were independent (that would indicate severe malfunction of the instrument!) then we would find

$$I(U, X) = H(U) - H(U|X) \tag{4.7}$$

and as a result the Mutual Information would be zero.

Considering that for $I(X, U)$ (according to *Eqs.* (4.5), (4.6))

$$0 \leq I(X, U) \leq H(X) \tag{4.8}$$

always holds, it is convenient to introduce

$$\mu_I = \frac{I(X, U)}{H(X)} = 1 - \frac{H(X|U)}{H(X)} ; \quad H(X) \neq 0 . \tag{4.9}$$

The value this variable can take is within [0,1] since $H(X|U) \leq H(X)$.

5. Applying the Concept of Information Gain

In Information Theory Information Gain is defined as follows

$$D(\tilde{P}||Q) = \sum_{j=1}^{N_u} \tilde{p}_j \log \frac{\tilde{p}_j}{q_j} . \tag{5.1}$$

This can also be expressed as

$$D(\tilde{P}||Q) = -H(U) - \sum_{j=1}^{N_u} \tilde{p}_j \log q_j \tag{5.2}$$

reduction of uncertainty, where p_j is a realised value of the distribution p with reduced tail area.

Let us denote the entropy representing the larger uncertainty observed before (a priori) the measurement made by

$$H_a(U) = -\sum_{j=1}^{N_u} q_j \log q_j \; . \tag{5.3}$$

Considering that $D(p\|Q) \geq 0$, the definitions of p_j, q_j and (5.2) we find that

$$-H(U) \geq \sum_{j=1}^{N_u} \tilde{p}_j \log q_j \; , \tag{5.4}$$

$$-\sum_{j=1}^{N_u} q_j \log q_j \geq -\sum_{j=1}^{N_u} \log q_j \geq 0 \; . \tag{5.5}$$

From the above results we get

$$D(\tilde{P}\|Q) \leq H_a(U) \; . \tag{5.6}$$

Observing *Eq.* (5.6) another quantity describing Information Gain may be defined as

$$\mu_D = \frac{D(\tilde{P}\|Q)}{H_a(U)} \; . \tag{5.7}$$

This quantity also will take a value in the interval [0, 1].

The combination of *Eqs.* (4.9) and (5.7) leads us to the introduction of

$$\mu = \mu_I \mu_D \; , \tag{5.8}$$

a general characteristic of measuring instruments that qualifies them on the basis of the quality of both the Information Gain and transmission.

6. The Case of Discrete Distributions

To find the general instrument characteristic $I(X, U)$ in *Eqs.* (4.9) and (5.7), $D(p\|Q)$ and $H_a(U)$ must be computed (estimated).

In the case of considering a Uniform distribution for $Q = (q_1, q_2, \dots, q_{N_u})$, that is in any $[U_{j-1}, U_j]$ interval of $[U^*_{\min}, U^*_{\max}]$ the value of U will fall with probability

$$q_j = \frac{1}{N_u} \; ; \quad j = 1, 2, \dots, N_u \; , \tag{6.1}$$

then $Eq.$ (5.3) becomes

$$H_a(U) = \log N_u \ . \tag{6.2}$$

Similarly, for the second term of $Eq.$ (5.2) we find

$$-\sum_{j=1}^{N_u} \tilde{p}_j \log q_j = \log N_u \ . \tag{6.3}$$

Using the above results the Information Gain

$$D(\tilde{P}||Q) = \log N_u - H(U) = H_a(U) - H(U) \tag{6.4}$$

is just the difference between the two (before and after taking the measurement) uncertainties (iff Q has Uniform distribution !).

Substituting $Eq.$ (6.4), (5.7) now becomes

$$\mu_D = 1 - \frac{H(U)}{H_a(U)} \ . \tag{6.5}$$

As it has been shown the quantities given by $Eqs.$ (4.9) and (5.7) are both smaller than unity and consequently their product in $Eq.$ (5.8) is smaller, too. Thus all three quantities are suitable for describing quality.

Substituting $Eq.$ (3.1), (6.4) becomes

$$D(\tilde{P}||Q) = \log \frac{(U^*_{\max} - U^*_{\min})}{\Delta u} - H(U) \ . \tag{6.6}$$

Next we need to find values for $I(X,U)$ and $H(u)$ in $Eq.$ (6.6). For this purpose let us consider some known continuous distribution.

7. The Problem of Continuous Distributions

In the majority of technical situations the results of measurement are best represented by some continuous – mostly Normal – distribution. Due to certain benefits – and a smaller number of minimum data points – it is worth taking advantage of the properties of the Normal distribution.

When X and U random variables are continuous, the Mutual Information in (4.1) becomes

$$I(X,U) = \int\int f(x,u) \log \left(\frac{f(x,u)}{h(x)\tilde{p}(u)}\right) dx du \tag{7.1}$$

and Information Gain can be expressed by the following integral

$$D(\tilde{P}\|Q) = \int \tilde{p}(u) \log \left(\frac{\tilde{p}(u)}{q(u)}\right) du \ . \qquad (7.2)$$

In the case of finite ($\Delta u \neq 0$, but sufficiently small) resolution

$$H(U) = -\sum_{j=1}^{N_u} p_j \log p_j \qquad (7.3)$$

entropy with the approximation

$$p_j \approx \tilde{p}(\overline{u}_j)\Delta u \ , \quad j = 1, 2, \ldots, N_u \ , \qquad (7.4)$$

may be expressed as follows

$$H(U) = -\sum_{j=1}^{N_u} \tilde{p}(\overline{u}_j) \log\tilde{p}\ (\overline{u}_j)\Delta u - \ \log \Delta u \sum_{j=1}^{N_u} \tilde{p}(\overline{u}_j)\Delta u \ . \qquad (7.5)$$

where u_j denotes the value at the middle of the interval $[U_{j-1}, U_j]$. Unlike *Eq.* (7.1) or (7.2), the above quantity cannot be expressed as an integral since in the case of $\Delta u \to 0$

$$\lim_{\Delta u \to 0} \sum_{j=1}^{N_u} \tilde{p}(\overline{u}_j) \log \tilde{p}(\overline{u}_j)\Delta u \to \int \tilde{p}(u) \log \tilde{p}(u) du \ . \qquad (7.6)$$

$$\lim_{\Delta u \to 0} \sum_{j=1}^{N_u} \tilde{p}(\overline{u}_j)\Delta u \to \int \tilde{p}(u) du \ . \qquad (7.7)$$

Thus *Eq.* (7.5) becomes

$$\lim_{\Delta u \to 0} H(U) = -\int \tilde{p}(u) \log \tilde{p}(u) du - \lim_{\Delta u \to 0} \ \log \Delta u \to \infty \ . \qquad (7.8)$$

This means that entropy $H(U)$ is not identical to the expression in *Eq.* (7.7) and consequently is not suitable for the purpose of measuring uncertainty. Although it is common to call the above integral the entropy of the continuous distribution (and it proves to be a useful quantity in many cases) it is not the same as the original concept.

Considering that instruments always have a resolution Δu that is larger than zero and with sufficiently small Δu the differences

$$\int \tilde{p}(u) \log \tilde{p}(u) du - \sum_{j=1}^{N_u} \tilde{p}(\overline{u}_j) \log \tilde{p}(\overline{u}_j) \Delta u \tag{7.9}$$

and

$$\int \tilde{p}(u) du - \sum_{j=1}^{N_u} \tilde{p}(\overline{u}_j) \Delta u \tag{7.10}$$

are small, the entropy may be approximated by the formula

$$H(U) \approx - \int \tilde{p}(u) \log \tilde{p}(u) du - \log \Delta u \ . \tag{7.11}$$

The useful nature of the above quantity is demonstrated in *Fig. 4*.

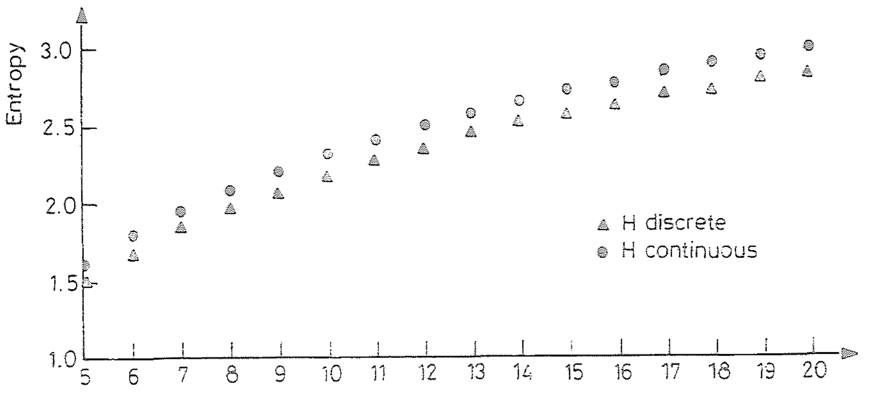

Fig. 4.

Beside being useful for numerical approximation, the formula (7.11) will also reveal the link between resolution and uncertainty. The first term in *Eq.* (7.11) is a constant value A. Thus the entropy may be expressed as

$$H(U) \approx A - \log \Delta u \tag{7.12}$$

or

$$H(U) \approx \log \frac{B}{\Delta u} \ , \tag{7.13}$$

where $A = \log B$.

8. An Example for the Choice of Continuous Distributions

Let $f(x,u)$ be the density function of a two-dimensional, $h(x)$ and $\tilde{p}(u)$ a one dimensional Normal and $q(u)$ that of a Uniform distribution (see Appendix A). Considering the above distributions and switching from an arbitrary (log) to – one that suits the nature of the application better – natural (ln) logarithm (which means that information is measured in 'nat') the Mutual Information in *Eq.* (7.1) becomes

$$I(X,U) = -\frac{1}{2}\ln(1 - r_{xu}^2) . \tag{8.1}$$

where r_{xu} is the correlation coefficient of random variables X and U.

Considering a Uniform distribution with the density function

$$q(u) = \begin{cases} \dfrac{1}{U^*_{\max} - U^*_{\min}} & \text{if} \quad U \in [U^*_{\min}; U^*_{\max}] \\ 0 & \text{otherwise} \end{cases} . \tag{8.2}$$

Information Gain in *Eq.* (7.2) becomes

$$D(\tilde{P}\|Q) = \ln(U^*_{\max} - U^*_{\min}) - \ln(\sqrt{2\pi e}\sigma_u) . \tag{8.3}$$

The above quantity is a function only of the scale parameters of the two distributions and independent of the instrument's properties.

According to *Eq.* (7.11) the formula

$$H(U) \approx \ln(\sqrt{2\pi e}\sigma_u) - \ln \Delta u \tag{8.4}$$

approximates entropy in the case of Uniform distribution with density function $q(u)$ (see Appendix B) and

$$H_a(U) \approx \ln(U^*_{\max} - U^*_{\min}) - \ln \Delta u \tag{8.5}$$

is the result when a Uniform distribution with Density function $q(u)$ is considered.

Substituting into *Eq.* (8.3) (and adding the term $(+\ln \Delta u - \ln \Delta u) = 0$) the Information Gain becomes

$$D(\tilde{P}\|Q) = \ln\left(\frac{U^*_{\max} - U^*_{\min}}{\Delta u}\right) - \ln\left(\frac{\sqrt{2\pi e}\sigma_u}{\Delta u}\right) . \tag{8.6}$$

Combining the above with *Eqs.* (8.4) and (8.5) we get

$$D(\tilde{P}\|Q) = H_a(U) - H(U) . \tag{8.7}$$

Examining *Eq.* (8.7) together with *Eq.* (6.4) we see that Information Gain (but only if a Uniform distribution is considered) is the difference between uncertainty before and after taking the measurement. This interpretation of Information Gain would be nice to use in more general cases. For this end (and for no other reason or purpose) we introduce arbitrarily another Information Gain, similarly to *Eq.* (8.7):

$$\tilde{D}(\tilde{P}\|Q) = H_a(U) - H(U) \ . \tag{8.8}$$

where, from *Eq.* (7.11)

$$H_a(u) \approx -\int p_a(u) \ln p_a(u) du - \ln \Delta u_a \ , \tag{8.9}$$

$$H(U) \approx -\int p(u) \ln p(u) du - \ln \Delta u \ . \tag{8.10}$$

Choosing $p_a(u) \equiv q(u)$ Uniform and $p(u) \equiv \tilde{p}(u)$ Normal distributions

$$\begin{aligned} \tilde{D}(\tilde{P}\|Q) &= \ln(U^*_{\max} - U^*_{\min}) - \ln \Delta u_a - \ln(\sqrt{2\pi e}\sigma_u) + \ln \Delta u \ , \\ \tilde{D}(\tilde{P}\|Q) &= \ln(U^*_{\max} - U^*_{\min}) - \ln(\sqrt{2\pi e}\sigma_u) + \ln \frac{\Delta u}{\Delta u_a} \ . \end{aligned} \tag{8.11}$$

This alternatively interpreted Information gain is sensitive to the resolutions and can even be increased by changing the ratio of the two resolutions.

In the rest of the paper we will use this new interpretation instead of that given in *Eq.* (5.7) to describe the quality aspect of Information Gain as follows

$$\tilde{\mu}_D = \frac{\tilde{D}(\tilde{P}\|Q)}{H_a(U)} = 1 - \frac{H(U)}{H_a(U)} \ . \tag{8.12}$$

Consequently the general characteristic in *Eq.* (5.8) changes to

$$\tilde{\mu} = \mu_I \tilde{\mu}_D \ . \tag{8.13}$$

9. The Efficiency of Information Gain

After substituting *Eqs.* (8.9), (8.10) and considering $\tilde{p}(u)$ Normal, $q(u)$ Uniform distributions the *Eq.* (8.12) quality characteristic becomes

$$\tilde{\mu}_D = 1 - \frac{\ln\left(\sqrt{2\pi e}\frac{\sigma_u}{\Delta u}\right)}{\ln\left(\frac{U^*_{\max} - U^*_{\min}}{\Delta u_a}\right)} \ . \tag{9.1}$$

It is our expectation for *Eq.* (9.1) to give a higher value for an instrument with better resolution. In the following we verify this expectation and examine the characteristic's response to changes in Δu_a.

From *Eq.* (8.11)

$$\tilde{D}(\tilde{P}\|Q) = \ln\left(\frac{U^*_{\max} - U^*_{\min}}{\sqrt{2\pi e}\sigma_u}\frac{\Delta u}{\Delta u_a}\right) > 0 \ . \tag{9.2}$$

From *Eqs.* (8.4) and (8.5)

$$H(U) \approx \ln\left(\sqrt{2\pi e}\frac{\sigma_u}{\Delta u}\right) > 0 \ , \tag{9.3}$$

$$H_a(U) \approx \ln\left(\frac{U^*_{\max} - U^*_{\min}}{\Delta u_a}\right) > 0 \ .$$

$$H_a(U) \approx \ln\left(\frac{U^*_{\max} - U^*_{\min}}{\Delta u_a}\right) > 0 \ . \tag{9.4}$$

Combining *Eqs.* (9.1), (9.2) and (9.3) we find that

$$\Delta u < \sqrt{2\pi e}\sigma_u \ , \tag{9.5}$$

$$\Delta u_a < U^*_{\max} - U^*_{\min} \ , \tag{9.6}$$

$$\Delta u_a < \kappa \Delta_u \ , \tag{9.7}$$

where

$$\kappa = \frac{U^*_{\max} - U^*_{\min}}{\sqrt{2\pi e}\sigma_u} > 1 \ . \tag{9.8}$$

Since even allowing

$$\Delta u = \sqrt{2\pi e}\sigma_u \tag{9.9}$$

it still holds that

$$\Delta u_a < \kappa\Delta u < \frac{U^*_{\max} - U^*_{\min}}{\sqrt{2\pi e}\sigma_u}\sqrt{2\pi e}\sigma_u = U^*_{\max} - U^*_{\min} \ . \tag{9.10}$$

Consequently if *Eq.* (9.7) is fulfilled *Eq.* (9.5) always holds.

Thus the requirements (9.5) through (9.7) are sensible only for values of Δu and Δu_a falling in the shaded area in *Fig. 5*. Examining the quality characteristic for values of Δu and Δu_a in the shaded area we find that:

a) Increasing (decreasing) Δu makes μ_D increase (decrease) in the interval [0,1].

b) Increasing (decreasing;) Δu_a makes μ_D decrease (increase) in the interval [0,1].

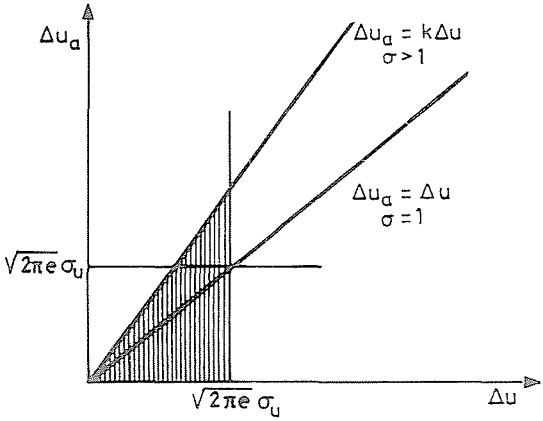

Fig. 5.

From this we learn that it is the quantity Δu_a whose effect on μ_D fulfils our expectations based on engineering common sense. Consequently, we will interpret Δu_a as the resolution of the instrument and Δu as some reference resolution. That is

$$\Delta u_m \equiv \Delta u_a \ , \tag{9.11}$$

$$\Delta u_r \equiv \Delta u \ . \tag{9.12}$$

It also follows from the above that the definition of *a priori* and *a posteriori* entropy – since they are dependent on the resolution – has to be completed by adding that any value is valid only at a given pair of resolutions Δu_m and Δu_r.

Another problem that we confront is with the interpretation of U^*_{min} and U^*_{max} in the denominator of *Eq.* (9.1). Contrary to what we observe in all practical situations, increasing $U^*_{\text{max}} - U^*_{\text{min}}$ in *Eq.* (9.1) will increase μ_D without limit. In reality, increasing the above quantity over a certain limit brings no further Information Gain. In fact, if $U^*_{\text{max}} - U^*_{\text{min}}$ is sufficiently large it may be judged without any instrument whether U lies in $[U^*_{\text{min}}, U^*_{\text{max}}]$ or not and from this point on the 'Information Gain' is independent of the instrument. This problem is addressed in the next section.

10. Introducing New Interval Bounds

In order to avoid the previous mentioned problems relating to the interval $[U^*_{\text{min}}, U_{\text{max}}]$ it is necessary to introduce it in a way that would serve the needs of instrument qualification better. Let us select the new $U_{\min}, U_{\max}$ so that the measured quantity U will fall in the selected interval with a prescribed high probability

$$P(\tilde{U}_{\min} \le U \le \tilde{U}_{\max}) = 1 - p \ , \tag{10.1}$$

where p is a very small quantity. Writing *Eq.* (10.1) somewhat differently, we have

$$P(U \in [\tilde{U}_{\min}, \tilde{U}_{\max}]) = P(U < \tilde{U}_{\min}) + P(U > \tilde{U}_{\max}) = p \ . \tag{10.2}$$

Let us write our condition on p in a symmetric form

$$P(U < \tilde{U}_{\min}) = \frac{p}{2} = 0.001 \tag{10.3}$$

and

$$P(U > \tilde{U}_{\max}) = \frac{p}{2} = 0.001 \ . \tag{10.4}$$

By introducing the variable

$$t = \frac{u - m_u}{s_u}\sqrt{n} \ , \tag{10.5}$$

where m_n is the expected value of random variable U, conditions (10.3) and (10.4) may be written

$$P(t < -t_p) = \frac{p}{2} = 0.001 \ , \tag{10.6}$$

$$P(t > t_p) = \frac{p}{2} = 0.001 \ . \tag{10.7}$$

With t_p we have

$$\tilde{U}_{\min} = m_u - t_p\frac{s_u}{\sqrt{n}} \ , \tag{10.8}$$

$$\tilde{U}_{\max} = m_u + t_p\frac{s_u}{\sqrt{n}} \ . \tag{10.9}$$

Considering the (worst) case of $n = 1$

$$\tilde{U}_{\max} - \tilde{U}_{\min} = 2t_p s_u \cong 2t_p\sigma_u \ . \tag{10.10}$$

Taking $n = 1$ degrees of freedom $p/2 = 0.001$ gives $t_p/2 = 636.610$. Substituting into *Eq.* (10.10) we get

$$\tilde{U}_{\max} - \tilde{U}_{\min} \cong 1273\sigma_u \ . \tag{10.11}$$

In the following we will use this result in *Eq.* (9.1).

11. Defining the Reference Resolution

After computing the constant term and considering *Eqs* (9.11), (9.12) and (10.11), (9.1) qualification parameter becomes

$$\tilde{\mu}_D = 1 - \frac{\ln\left(4.133\frac{\sigma_u}{\Delta u_r}\right)}{\ln\left(1273\frac{\sigma_u}{\Delta u_m}\right)} , \tag{11.1}$$

where σ_u is estimated from the sample, Δu_m is the instrument's resolution and Δu_r is the reference resolution. A few examples to demonstrate the effect of the reference resolution on $\tilde{\mu}_D$ are given in *Table 1*.

Table 1

	Δu_r				
Δu_m	σ_u	$\frac{1}{2}\sigma_u$	$\frac{1}{5}\sigma_u$	$\frac{1}{10}\sigma_u$	$\frac{1}{100}\sigma_u$
$\frac{1}{5}\sigma_u$	0.838	0.759	0.654	0.577	0.312
$\frac{1}{10}\sigma_u$	0.850	0.777	0.680	0.606	0.363
$\frac{1}{100}\sigma_u$	0.879	0.820	0.742	0.636	0.487

From the examples we learn that a suitable value of Δu_r will be in the interval $[\sigma_u, 1/5\sigma_u]$ but $1/10\sigma_u$ is still sensible.

Since the Information Gain in *Eq.* (8.11) is proportional to the ratio $\Delta u_r/\Delta u_m$, the values under the main diagonal of *Table 1* give good results. Considering this and *Eq.* (9.4) it holds that

$$\Delta u_m < \Delta u_r < \sqrt{2\pi e}\sigma_u \approx 4.33\sigma_u . \tag{11.2}$$

On the other hand, the sensitiveness of *Eq.* (11.1) to σ_u – when it does not include instrument error – indicates that this quantity is a characteristic of not only the instrument but of the Information gaining process as well. However, the sensitiveness to σ_u is quite weak. From a practical point of view, the above qualification parameter seems to be suitable for the purpose of qualifying instruments.

12. The Efficiency of the Information Transmitting Process

In order to find a value for I, the quantity measuring Information Transmittance efficiency, it is necessary to compute $I(X,U)$ Mutual Information and $H(X)$ Entropy (or $H(X|U)$ Conditional Entropy instead of $I(X,U)$). From *Eq.* (8.1) we see that it is sufficient to estimate the correlation coefficient r_{xu}. For this, measured pairs of values of U and X are sufficient. We measure

a) U with a high precision calibrating instrument (values u_i)
b) values of X with the instrument to be qualified (values x_i).

Measured values are substituted into

$$r_{xu} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(u_i - \overline{u})}{\left[\sum_{i=1}^{N}(x_i - \overline{x})^2 \sum_{i=1}^{N}(u_i - \overline{u})^2\right]^{\frac{1}{2}}} \tag{12.1}$$

to estimate the correlation coefficient r_{xu}. This variable takes a value close to unity for quality instruments and consequently its application may be difficult since the function

$$I(X,U) = -\frac{1}{2}\ln(1 - r_{xu}^2) \tag{12.2}$$

exhibits a very steep slope for $r_{xu} \geq 0.9$ (see *Table 2* and *Fig. 6*). Thus quality instruments need two or more '9' decimals to be properly described. However, to give an estimate of accurate to two or more decimals calls for 100 (or orders of magnitude more) data points that is practically unfeasible. This is explained by the fact that for $r_{xu} = 0.01$ and $r_{xu} \geq 0.99$ the loss of Mutual Information

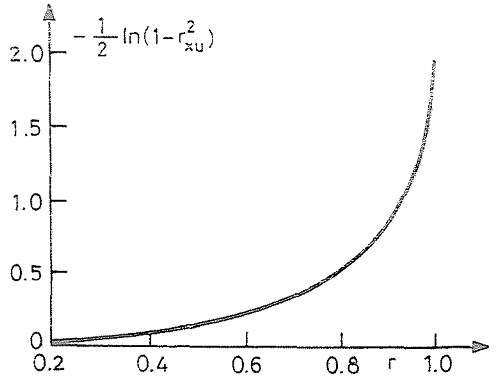

Fig. 6.

$$\Delta I = \frac{\delta I}{\delta r_{xu}} = \Delta r_{xu} = \frac{r_{xu}}{1 - r_{xu}^2}\Delta r_{xu}\ , \tag{12.3}$$

Table 2

	r_{xu}	$-\frac{1}{2}\ln(1-r_{xu}^2)$		r_{xu}	$-\frac{1}{2}\ln(1-r_{xu}^2)$		r_{xu}	$-\frac{1}{2}\ln(1-r_{xu}^2)$
1.	0.1	0.0005	7.	0.87	0.71	13.	0.995	2.30
2.	0.3	0.05	8.	0.9	0.83	14.	0.999	3.11
3.	0.5	0.14	9.	0.93	1.00	15.	0.9999	4.26
4.	0.75	0.41	10.	0.95	1.16	16.	0.99999	5.41
5.	0.78	0.47	11.	0.97	1.41	17.	0.999999	6.56
6.	0.8	0.61	12.	0.99	1.96	18.	0.9999999	7.71

Table 3

	r_{xu}	ΔI	I
1.	0.99	0.497	1.96
2.	0.999	4.997	3.11
3.	0.9999	49.997	4.26

is substantial (see *Table 3*).

This also means that the increased error is a problem, too. When we have three or four decimal '9' in the correlation coefficient the error is far greater than the information. This implies that Mutual Information cannot be estimated with this method. In order to avoid such problems it is necessary to classify instruments into two classes.

Class I absorbs instruments for which the correlation of X and U

$$r_{xu} \leq 0.9 \ . \tag{12.4}$$

To these instruments *Eq.* (12.2) is readily applied.

Class II absorbs instruments for which the correlation of X and U

$$r_{xu} > 0.9 \ . \tag{12.5}$$

In this case the estimate is based on the direct estimation of $H(X)$ and $H(X|U)$.

13. Estimating the Entropies Directly

To make a direct estimation of the entropies $H(X)$ and $H(X|U)$ the method described in [3] may be used. That is, the intervals $[U_{(1)}, U_{(N)}]$, $[X_{(1)}, X_{(N)}]$

(where $U_{(1)}, U_{(N)}$ and $X_{(1)}, X_{(N)}$ denote the first and Nth members of a series in increasing order) are divided into k sub-intervals where

$$N = k^2 \ln k^2 \ . \tag{13.1}$$

N is the number of data points. With positive roots of *Eq.* (13.1)

$$\Delta X = \frac{X_{(N)} - X_{(1)}}{k} \ , \tag{13.2}$$

$$\Delta U = \frac{U_{(N)} - U_{(1)}}{k} \tag{13.3}$$

the interval bounds are defined as follows

$$\tilde{U}_i = U_{(1)} + i\Delta U \ , \quad i = 0, 1, \ldots, k \ , \tag{13.4}$$

$$\tilde{X}_i = X_{(1)} + i\Delta X \ , \quad i = 0, 1, \ldots, k \ . \tag{13.5}$$

Then we count positive m_{il} $(i, l = 1, 2, \ldots, k)$ falling in the areas $[X_{i-1}, X_i] \times [U_{j-1}, U_j]$ – which is a function of the data. Having m_{il} we compute

$$m_{i\circ} = \sum_{l=1}^{k} m_{il} \ , \quad i = 1, 2, \ldots, k \ . \tag{13.6}$$

The next step is to estimate entropies

$$H(X) \approx \sum_{i=1}^{k} \frac{m_{i\circ}}{N} \ln \frac{m_{i\circ}}{N} \tag{13.7}$$

$$H(X|U) \approx \sum_{i=1} \frac{m_{i\circ}}{N} \sum_{l=1} \frac{m_{il}}{N} \ln \frac{m_{il}}{N} \ . \tag{13.8}$$

Substituting the above into *Eq.* (4.9) we estimate

$$\mu_I = 1 - \frac{(H(X|U)}{H(X)} \ . \tag{13.9}$$

14. Constructing a Reference Resolution

In order to avoid a fully arbitrary Reference Resolution (even in the already defined intervals) it is necessary to base it on an expression that is independent of our subjective judgement.

Table 4

	SIGMA-type mikator	ABBE-type vertical gauge	Digital readout micrometer	Mechanical micrometer
$\tilde{\mu}_D$	0.8448	0.8326	0.7734	0.7734
μ_I	0.7636	0.7239	0.7035	0.6828
μ	0.6451	0.6027	0.5441	0.5281

Let us derive this expression from the partitioning k used for the estimation of entropies (13.7) and (13.8). We select the intervals defined in *Eq.* (13.3) as reference resolution, that is

$$\Delta u_r \equiv \Delta u = \frac{U_{(N)} - U_{(1)}}{k} . \tag{14.1}$$

This also satisfies condition (9.4) since the probability of $U_{(N)} - U_{(1)} < 6\sigma_u$ is high. That is

$$P((U_{(N)} - U_{(1)}) < 6\sigma_u) \approx 1 . \tag{14.2}$$

By substituting *Eq.* (14.1) into *Eq.* (14.2) we find that the probability of

$$\Delta u_r = \frac{U_{(N)} - U_{(1)}}{k} < \frac{6}{k}\sigma_u \tag{14.3}$$

is high. Now even with $k = 2$ (9.4) is fulfilled:

$$\Delta u_r < 3\sigma_u < 4.133\sigma_u . \tag{14.4}$$

Considering the above, both *Eqs.* (11.1) and (13.9) can be computed and consequently

$$\mu = \mu_I \tilde{\mu}_D \tag{14.5}$$

is easily found.

15. Computing Instrument Characteristics from Measured Data

In order to find the Instrument Characteristics four different instruments were used to take 25 measurements on each of 10 different test objects. The data was used to compute the quantity μ_D describing Information Gain, μ_I describing the efficiency of Information Transmittance and the Instrument Characteristic μ, the product of the above two, which qualifies the instrument according to its accuracy as an information gaining tool. Results appear in *Table 4*.

Analysis of the results shows that the Instrument Characteristic μ developed by the authors is suitable for the qualification of Measuring Instruments on the basis of Information Theory.

Appendix A

The density function of a one-dimensional Normal distribution:

$$h(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-m_x}{\sigma_x}\right)^2} . \tag{A1}$$

Formally the $\tilde{p}(u)$ function just like $h(x)$ only has to change variable and parameters.

The density function of a two-dimensional Normal distribution:

$$f(x,u) = \frac{1}{2\pi\sigma_x\sigma_u\sqrt{1-r_{xu}^2}} e^{\frac{1}{2(1-r_{xu}^2)}\left[\left(\frac{x-m_x}{\sigma_x}\right)^2 - 2r_{xu}\left(\frac{x-m_x}{\sigma_x}\right)\left(\frac{u-m_u}{\sigma_u}\right)+\left(\frac{u-m_u}{\sigma_u}\right)^2\right]} . \tag{A2}$$

Appendix B

In the case of $\tilde{p}(u)$ Normal distribution by analogy Form (A1)

$$\tilde{p}(u) = \frac{1}{\sqrt{2\pi}\sigma_u} e^{-\frac{1}{2}\left(\frac{u-m_u}{\sigma_u}\right)^2} , \tag{B1}$$

$$\ln \tilde{p}(u) = -\ln(\sqrt{2\pi}\sigma_u) - \frac{1}{2}\left(\frac{u-m_u}{\sigma_u}\right)^2 , \tag{B2}$$

$$\int \tilde{p}(u)\ln\tilde{p}(u)du = -\ln(\sqrt{2\pi}\sigma_u)\int \tilde{p}(u)du - \frac{1}{2}\int \tilde{p}(u)\left(\frac{u-m_u}{\sigma_u}\right)^2 du ,$$

$$\int \tilde{p}(u)du = 1 \quad \text{and} \quad \int \tilde{p}(u)(u-m_u)^2 du = \sigma_u^2 , \quad \text{therefore}$$

$$\int \tilde{p}(u)\ln\tilde{p}(u)du = -\ln(\sqrt{2\pi}\sigma_u) - \frac{1}{2}\frac{\sigma_u^2}{\sigma_u^2} = -\ln(\sqrt{2\pi}\sigma_u) - \frac{1}{2}\ln e ,$$

$$\int \tilde{p}(u)\ln\tilde{p}(u)du = -\ln(\sqrt{2\pi e}\sigma_u) . \tag{B3}$$

In the case of

$$q(u) = \frac{1}{U^*_{\max} - U^*_{\min}} . \tag{B4}$$

Uniform distribution and $\tilde{p}(u)$ Normal distribution:

$$\int \tilde{p}(u) \ln q(u) du = -\ln(U^*_{\max} - U^*_{\min}) \int \tilde{p}(u) du,$$

$$\int \tilde{p}(u) \ln q(u) du = -\ln(U^*_{\max} - U^*_{\min}) \; . \tag{B5}$$

References

1. SHANNON, C. E.: A Mathematical Theory of Communication. *Bell System Tech. J.*, Vol. 27, pp. 379–423, 623–656, 1948.
2. SHANNON, C. E. – WEAVER, W.: The Mathematical Theory of Communication. Univ. of Illinois Press, Urbana, 1949.
3. MATHAI, A. M. – RATHIE, P. M.: Basic Concepts in Information Theory and Statistics. Wiley Eastern Ltd. New Delhi, 1975.
4. CSISZÁR, I. – KÖRNER, J.: Information Theory. Academic Press, Budapest, 1981.
5. REZA, F. M.: Bevezetés az információelméletbe. Műszaki Könyvkiadó, Budapest, 1966.
6. WILSON, A. G.: Entropy in Urban and Regional Modelling, Pion Ltd. London, 1970.
7. SNICKAR, F., – WEIBULL, J. W.: A Minimum Information Principle: Theory and Practice Regional Science and Urban Economics, Vol. 7, pp. 137–168, 1977.
8. BÖLÖNI, P.: Increase of Information in the Course of Measurements. *Periodica Polytechnica.* Budapest, Vol. 37, No. 4, pp. 223–235, 1993.
9. BECKWITH, T. G. – LEWIS BUCK, N.: Mechanical Measurements, Addison-Wesley Publishing Company, 1969.
10. FRITZ, J.: Információelmélet. A számítástechnika matematikai alapjai c. tanfolyam jegyzet, 1972.
11. REIMANN, J.: Valószínűségelmélet és matematikai statisztika mérnököknek. Tankönyvkiadó, Budapest, 1992.
12. PRÉKOPA, A.: Valószínűségelmélet. Műszaki Könyvkiadó, Budapest, 1972.