How to Estimate Amount of Useful Information, in Particular Under Imprecise Probability

Luc Longpré[1], Olga Kosheleva[2] and Vladik Kreinovich[1]
[1]*Department of Computer Science,* [2]*Department of Teacher Education*
University of Texas at El Paso, El Paso, TX 79968, USA
longpre@utep.edu, olgak@utep.edu, vladik@utep.edu

Abstract. Traditional Shannon's information theory describes the overall amount of information, without distinguishing between useful and unimportant information. Such a distinction is needed, e.g., in privacy protection, where it is crucial to protect important information while it is not that crucial to protect unimportant information. In this paper, we show how Shannon's definition can be modified so that it will describe only the amount of useful information.

Keywords: amount of information, entropy, utility, imprecise probabilities, interval uncertainty, useful information.

1. Formulation of the Problem: Traditional Information Theory Does Not Distinguish Between Useful and Unimportant Information

To gauge the amount of information, we need to be able to gauge uncertainty. The more we learn, the more information we have about the world. Our ultimate goal is to gain a complete knowledge of the world, i.e., to reach a situation when we should be able, based on this knowledge, to answer any question that anyone may have about the future or past state of the world.

For example:

- one of the main objectives of meteorology is to be able to predict future weather;
- one of the main objectives of celestial mechanics is to predict future locations of celestial bodies and man-made satellites,
- etc.

In practice, we rarely have the *complete* information, we usually have only *partial* information, based on which we cannot uniquely reconstruct the state of the world and thus, cannot always find an answer to a question about the world. In other words, in practice, we have *uncertainty.*

Additional information allows us to decrease this uncertainty. It is therefore reasonable to gauge the amount of information in the new knowledge by how much this information decreases the original uncertainty. Thus, a natural way to gauge the amount of information is to estimate how much this information decreases uncertainty. Hence, to gauge the amount of information, we need to be able to gauge the amount of uncertainty.

How to gauge the amount of uncertainty: idea. As we have mentioned earlier, uncertainty means that for some questions, we do not have a definite answer. Once we learn the answers to these questions, we thus decrease the original uncertainty. It is therefore reasonable to estimate the amount of uncertainty by the number of questions needed to eliminate this uncertainty.

Of course, not all questions are created equal. Some questions can have a simple binary "yes"-"no" question, some questions look for a more detailed answer – e.g., we can ask what is the value of a certain quantity. No matter what is the answer, we can describe this answer inside the computer and thus – since everything in the computer is represented as 0s and 1s – represent this answer as a sequence of 0s and 1s. Such a several-bits question can be represented as a sequence of on-bit questions:

- we can first ask what is the first bit of the answer,
- we can then ask what is the second bit of the answer,
- etc.

Since every question can thus be represented as a sequence of one-bit ("yes"-"no") questions, it is reasonable to measure uncertainty by the smaller number of such "yes"-"no" questions which are needed to eliminate this uncertainty – i.e., to uniquely determine the state of the corresponding system.

Let us recall how this idea can be transformed into exact formulas.

How to gauge uncertainty: finite case. Let us first consider the situation when we have finitely many N alternatives. In this case:

- if we ask one binary question, then we can get two possible answers (0 and 1) and thus, uniquely determine one of the two different states;
- if we ask two binary questions, then we can get four possible combinations of answers (00, 01, 10, and 11), and thus, uniquely determine one of the four different states;
- in general, if we ask q binary questions, then we can get 2^q possible combinations of answers, and thus, uniquely determine one of 2^q states.

So, to identify one of n states, the smallest number of binary questions q needed for this identification is the smallest values q for which $2^q \geq N$. This inequality is equivalent to $q \geq \log_2(N)$, and thus, the smallest number of binary questions is equal to the smallest integer which is greater than or equal to this logarithm, i.e., equal to $\lceil \log_2(q) \rceil$.

How to gauge uncertainty: finite case with known probabilities. In the previous text, we considered the situation when we have n alternatives about whose frequency we know nothing. In practice, we often know the probabilities $p_1, \ldots, p_n$ of different alternatives. In this case, instead of considering the *worst-case* number of binary questions needed to eliminate uncertainty, it is reasonable to consider the *average* number of questions.

This value can be described as follows. We have a large number N of similar situations with n-uncertainty.

– In $N \cdot p_1$ of these situations, the actual state is State 1.

– In $N \cdot p_2$ of these situations, the actual state is State 2,

– etc.

The average number of binary questions can be obtained if we divide the overall number of questions needed to determine the states in all N situations, by N.

Let us describe the uncertainty of this N-repetitions arrangement. Out of N situations, there are

$$\binom{N}{N \cdot p_1} = \frac{N!}{(N \cdot p_1)! \cdot (N - N \cdot p_1)!}$$

ways to select the situations in which the actual state is State 1. For each of these situations, there are

$$\binom{}{N \cdot p_1} = \frac{(N - N \cdot p_1)!}{(N \cdot p_2)! \cdot (N - N \cdot p_1 - N \cdot p_2)!}$$

ways to select, among the remaining $N - N \cdot p_1$ situations, $N \cdot p_2$ ones for which the actual state is State 2, etc. Thus, the overall number of possible arrangements is equal to the product

$$A = \frac{N!}{(N \cdot p_1)! \cdot (N - N \cdot p_1)!} \cdot \frac{(N - N \cdot p_1)!}{(N \cdot p_2)! \cdot (N - N \cdot p_1 - N \cdot p_2)!} \cdot \ldots =$$

$$\frac{N!}{(N \cdot p_1)! \cdot (N \cdot p_2)! \cdot \ldots \cdot (N \cdot p_n)!}.$$

To identify one of these A arrangements, we need to ask

$$Q = \log_2(A)$$

binary questions. For the above formula for A, we get

$$Q = \log_2(N!) - \sum_{i=1}^{n} \log_2((N \cdot p_i)!).$$

Here,

$$m! \sim \left(\frac{m}{e}\right)^m,$$

so

$$\log_2(m!) \sim m \cdot (\log_2(m) - \log_2(e)).$$

Thus,

$$Q = N \cdot \log_2(N) - N \cdot \log_2(e) - \sum_{i=1}^{n} (N \cdot p_i) \cdot \log_2(N \cdot p_i) + \sum_{i=1}^{n} (N \cdot p_i) \cdot \log_2(e),$$

and the average number $\overline{q}$ of "yes"-"no" questions is equal to

$$\overline{q} = \frac{Q}{N} = \log_2(N) - \log_2(e) - \sum_{i=1}^{n} p_i \cdot \log_2(N \cdot p_i) + \sum_{i=1}^{n} p_i \cdot \log_2(e).$$

Taking into account that $\log_2(N \cdot p_i) = \ln(N) + \ln(p_i)$, we get

$$\overline{q} = \log_2(N) - \log_2(e) - \sum_{i=1}^{n} p_i \cdot \log_2(N) - \sum_{i=1}^{n} p_i \cdot \log_2(p_i) + \left(\sum_{i=1}^{n} p_i\right) \cdot \log_2(e) =$$

$$\log_2(N) - \log_2(e) - \left(\sum_{i=1}^{n} p_i\right) \cdot \log_2(N) - \sum_{i=1}^{n} p_i \cdot \log_2(p_i) + \left(\sum_{i=1}^{n} p_i\right) \cdot \log_2(e).$$

Since $\sum\limits_{i=1}^{n} p_i = 1$, we get

$$\overline{q} = -\sum_{i=1}^{n} p_i \cdot \log_2(p_i).$$

This is the classical Shannon's formula for the amount of uncertainty (Shannon and Weaver, 1971).

How to gauge uncertainty: continuous case. In the continuous case, when the unknown(s) can take any of the infinitely many values from some interval, we need infinitely many binary questions to uniquely determine the exact value.

To estimate uncertainty, it then makes sense to consider the average number of questions needed to determine each value with a given accuracy $\varepsilon > 0$.

In other words, instead of determining the exact value x, we divide the real line into intervals $[x_i - \varepsilon, x_i + \varepsilon]$, where $x_{i+1} = x_i + 2\varepsilon$, and we want to find out to which of these intervals the actual value x belongs. For small ε, the probability p_i of belonging to the i-th interval is equal to $p_i \approx \rho(x_i) \cdot (2\varepsilon)$. Substituting this expression for p_i into the classical Shannon's formula, we get

$$\overline{q} = -\sum_{i=1}^{n} p_i \cdot \log_2(p_i) = -\sum_{i=1}^{n} \rho(x_i) \cdot (2\varepsilon) \cdot \log_2(\rho(x_i) \cdot (2\varepsilon)).$$

Using the fact that $\log(a \cdot b) = \log(a) + \log(b)$, we conclude that

$$\overline{q} = -\sum_{i=1}^{n} \rho(x_i) \cdot (2\varepsilon) \cdot \log_2(\rho(x_i)) - \sum_{i=1}^{n} \rho(x_i) \cdot (2\varepsilon) \cdot \log_2(2\varepsilon).$$

The first term in this sum has the form

$$-\sum_{i=1}^{n} \rho(x_i) \cdot \log_2(\rho(x_i)) \cdot (2\varepsilon) = -\sum_{i=1}^{n} \rho(x_i) \cdot \log_2(\rho(x_i)) \cdot \Delta x_i,$$

where we denoted $\Delta x_i \stackrel{\text{def}}{=} x_{i+1} - x_i$. One can easily see that this term is an integral sum for the interval $-\int \rho(x) \cdot \log_2(\rho(x))\, dx$ and thus, for small ε, is practically equal to this interval.

Similarly, the second term has the form

$$-\sum_{i=1}^{n} \rho(x_i) \cdot (2\varepsilon) \cdot \log_2(2\varepsilon) = -\log_2(2\varepsilon) \cdot \sum_{i=1}^{n} \rho(x_i) \cdot (2\varepsilon) = -\log_2(2\varepsilon) \cdot \sum_{i=1}^{n} \rho(x_i) \cdot \Delta x_i,$$

and is, thus, an integral sum for the integral $-\log_2(2\varepsilon) \cdot \int \rho(x)\, dx$. By the formula of complete probability, we have $\int \rho(x)\, dx = 1$, so the second term is simply equal to $-\log_2(2\varepsilon)$ and thus, the average number of binary questions $\overline{q}$ which is needed to determine x with accuracy ε is equal to

$$\overline{q} = -\int \rho(x) \cdot \log_2(\rho(x))\, dx - \log_2(2\varepsilon).$$

The first term in this expression does not depend on ε and thus, provides a good measure of how much uncertainty we have. This term

$$S \stackrel{\text{def}}{=} -\int \rho(x) \cdot \log_2(\rho(x))\, dx$$

was also introduced originally by Shannon and is known as *Shannon's entropy* (or simply *entropy*, for short) (Shannon and Weaver, 1971).

A similar formula holds in the multi-D case, when instead of a single variable x, we have a tuple of variables $\vec{x} = (x_1, \ldots, x_m)$:

$$\overline{q} = S - m \cdot \log_2(2\varepsilon),$$

where

$$S = -\int \rho(\vec{x}) \cdot \log_2(\rho(\vec{x}))\, d\vec{x}.$$

Need to distinguish between useful and unimportant information. Not all information is created equal:

– some pieces of information are useful, while

– other pieces of information are unimportant.

Whether the information is useful or not depends on what we plan to do with this information. For example:

– if we are interested in predicting weather in a given geographic area based on meteorological observations, then the specific colors at sunset or the colors and the smell of the fog are probably unimportant, while

– if we are analyzing polluting level, all this would be a very useful information.

Such distinction is important in privacy protection. A distinction between useful and irrelevant information is needed in privacy protection. Ideally, we would like to maintain full privacy, so that no one can gain any information about a person without his or her explicit permission. However, realistically, some information may be leaked. It is therefore important to distinguish between the cases when an important information was leaked and when an unimportant piece of information was leaked.

For example, if we want to keep the salaries private, then disclosing the higher bits of the salaries – i.e., in effect, the approximate values of the salaries – would be a major violation of privacy, while disclosing the lowest bits, i.e., number of cents in the annual salary, would be reasonably harmless.

It is therefore desirable to estimate the amount of *useful* information, i.e., information that affects the utility of different alternatives.

Such a distinction is also important in education. The distinction between relevant and irrelevant information underlies a successful training system designed by A. M. Zimichev et al. in the 1980s; see, e.g., (Zimichev, Kreinovich, et al., 1982; Zimichev, Kreinovich, et al., 1982a; Aló and Kosheleva, 2006). This system started with an attempt to resolve the following seeming contradiction.

On the one hand, the vast majority of pedagogues strongly supports the humanitarian belief that (almost) students are equal in their ability to learn. Some students may require different learning styles, different support environments, but the history of education has consistently shown that any group originally perceived as inferior in learning – be it by gender or by race of by country of origin – turns out to be as successful as everyone else. Standardized tests consistently prove that kids from different groups have, on average, the same ability to learn. There are big discussions on whether 10% differences between scores are meaningful or not, but the very size of these disputed differences confirms the big pictures: we are all born equally skilled in learning.

Based on the fact that we are all equally skilled in learning, one should expect that with advanced learning strategies, all kids should succeed equally in school. Alas, no matter how advanced is the pedagogical technique that we use in the actual classroom, usually, after each class, there is a big difference between how much different kids learned, difference often in the orders of magnitude.

Since this difference cannot be explained by the difference in the kid's ability to learn, then why is it there? and, most importantly, how can we overcome this difference and enable each kid to learn according to his or her full potential?

To understand the empirically observed difference, researchers decided to quantify how much the students recalled. They asked each student, after the class, to spend an hour or so writing down what this student remembered after the class, be it the material from the class, what was happening outside, etc. Then, the researchers tried their best to quantify this amount of information by counting it in bits.

When they counted the number of bits *relevant* to the material presented in the class (e.g., math), they got exactly the same order of magnitude difference as was expected. Surprisingly, however, when they counted the *total* amount of information, including both the information presented in the class and the irrelevant information (what was the teacher dressed in, what birds were signing, what kids at neighboring desks were saying, etc.), then the total amounts of information recalled by all the kids became approximately the same.

In other words, both the best learners and the worst learners remembered approximately the same amount of information, but the best learners remembered the material presented in the class, while the worst learners remembered a lot of irrelevant details but not what was taught.

2. How to Estimate the Amount of Useful Information: A Suggestion

Main idea. According to decision theory, the usefulness of each situation to a user can be described by a *utility function* $u(x)$; see, e.g., (Fishburn, 1969; Raiffa, 1970; Luce and Raiffa, 1989; Nguyen, Kosheleva, and Kreinovich, 2009).

Therefore, we propose to count the number of binary questions that are needed to determine each of the unknown variables with accuracy sufficient to determine the utility u with a given accuracy $\varepsilon > 0$.

From this viewpoint, if some variable is irrelevant, then it does not affect the utility at all, so we should not waste binary questions trying to find the value of this variable. If some variable is slightly relevant, then a very crude estimate of this variable will provide us with ε-accuracy in $u(x)$ – and therefore, few questions will be needed. On the other hand, if a variable is highly relevant, then we need exactly as many questions as before.

Let us transform this idea into a precise definition. Let us describe how this idea can be transformed into a precise definition.

We will start with the case when we have only one variable x, and then continue with the multi-D case.

Towards a precise definition: 1-D case. In the 1-D case, if we know x with uncertainty Δx, then the resulting values of the utility is known with the uncertainty

$$u(x + \Delta x) - u(x) \approx u'(x) \cdot \Delta x,$$

where $u'(x)$ denotes the derivative of the utility function. Thus, to get the desired accuracy ε in describing utility $u(x)$, we need to determine x with accuracy $\Delta x = \dfrac{\varepsilon}{|u'(x)|}$.

In this case, we divide the real line into intervals

$$\left[x_i - \frac{\varepsilon}{|u'(x_i)|}, x_i + \frac{\varepsilon}{|u'(x_i)|}\right],$$

where

$$x_{i+1} = x_i + \frac{2\varepsilon}{|u'(x_i)|},$$

and we want to find out to which of these intervals the actual value x belongs. For small ε, the probability p_i of belonging to the i-th interval is equal to

$$p_i \approx \rho(x_i) \cdot \Delta x_i = \rho(x_i) \cdot \frac{2\varepsilon}{|u'(x_i)|},$$

where we denote $\Delta x_i \stackrel{\text{def}}{=} x_{i+1} - x_i$.

Substituting this expression for p_i into the classical Shannon's formula, we get the following formula for the average number of questions $\overline{q}$:

$$\overline{q} = -\sum_{i=1}^{n} p_i \cdot \log_2(p_i) = -\sum_{i=1}^{n} \rho(x_i) \cdot \Delta x_i \cdot \log_2 \left(\rho(x_i) \cdot \frac{2\varepsilon}{|u'(x_i)|} \right).$$

Using the fact that $\log(a \cdot b) = \log(a) + \log(b)$, we conclude that

$$\overline{q} = -\sum_{i=1}^{n} \rho(x_i) \cdot \Delta x_i \cdot \log_2 \left(\frac{\rho(x_i)}{|u'(x_i)|} \right) - \sum_{i=1}^{n} \rho(x_i) \cdot \Delta x_i \cdot \log_2(2\varepsilon).$$

The first term in this sum has the form

$$-\sum_{i=1}^{n} \rho(x_i) \cdot \log_2 \left(\frac{\rho(x_i)}{|u'(x_i|} \right) \cdot \Delta x_i,$$

i.e., it is an integral sum for the interval

$$-\int \rho(x) \cdot \log_2 \left(\frac{\rho(x)}{|u'(x)|} \right) dx$$

and thus, for small ε, is practically equal to this interval.

Similarly to the traditional Shannon's entropy case, the second term is equal to $-\log_2(2\varepsilon)$.

Thus, the average number of binary questions $\overline{q}$ which is needed to determine the utility $u(x)$ with accuracy ε is equal to

$$\overline{q} = -\int \rho(x) \cdot \log_2 \left(\frac{\rho(x)}{|u'(x)|} \right) dx - \log_2(2\varepsilon).$$

The first term in this expression does not depend on ε and thus, provides a good measure of how much uncertainty we have. Thus, we arrive at the following conclusion.

1-D case: conclusion. The average number of binary questions needed to the determine the utility $u(x)$ with given accuracy $\varepsilon > 0$ is equal to $\overline{q} = S_u - \log_2(2\varepsilon)$, where

$$S_u \stackrel{\text{def}}{=} -\int \rho(x) \cdot \log_2 \left(\frac{\rho(x)}{|u'(x)|} \right) dx.$$

Thus, this quantity S_u can be viewed as the amount of useful information.

By using the fact that $\log(a/b) = \log(a) - \log(b)$, we conclude that

$$S_u = S + \int \rho(x) \cdot \log_2(|u'(x)|)\, dx,$$

where S is the traditional Shannon's entropy. The additional integral term is the mathematical expectation of $\log_2(|u'(x)|)$.

Discussion. In the particular case when $u(x) = x$, determining the utility with accuracy ε is equivalent to finding x with accuracy ε, so the new expression coincides with the traditional Shannon's entropy formula.

The smaller the derivative $|u'(x)|$, the less relevant the variable x – and the smaller the additional term and, thus, the amount S_u of useful information.

Multi-D case. In the multi-D case, for each of the variables x_j $(1 \le j \le m)$, the interval that guarantees accuracy ε in the utility has the width

$$\Delta x_j = \frac{2\varepsilon}{|u_{,j}|},$$

where we denoted

$$u_{,j} \stackrel{\text{def}}{=} \frac{\partial u}{\partial x_j}.$$

Thus, we divide the whole m-dimensional space into zones of volume

$$\Delta V = \frac{(2\varepsilon)^m}{\prod\limits_{j=1}^{m} |u_{,j}|}$$

and probability

$$p_i = \rho(\vec{x}_i) \cdot \frac{(2\varepsilon)^m}{\prod\limits_{j=1}^{m} |u_{,j}(\vec{x}_i)|}.$$

Substituting these probabilities into the Shannon's formula, we conclude that

$$\overline{q} = -\int \rho(\vec{x}) \cdot \log_2\left(\frac{\rho(\vec{x})}{\prod\limits_{j=1}^{m} |u_{,j}(\vec{x})|}\right) dx - \log_2(2\varepsilon).$$

Thus, as an amount of useful information, we can take the value

$$S_u = -\int \rho(\vec{x}) \cdot \log_2\left(\frac{\rho(\vec{x})}{\prod\limits_{j=1}^{m} |u_{,j}(\vec{x})|}\right) d\vec{x}.$$

By using the formula $\log(a/b) = \log(a) - \log(b)$, we can reformulate this formula in the following form:

$$S_u = S + \sum_{i=1}^{m} \int \rho(\vec{x}) \cdot \log_2(|u_{,j}(\vec{x})|)\, d\vec{x},$$

where S is the traditional Shannon's entropy and each additional term is the expected value of $\log_2(|u_{,j}(\vec{x})|)$, where $u_{,j}(\vec{x})$ is the j-th partial derivative of the utility function $u(\vec{x})$.

What if only have partial information about the probabilities. The above formulas assume that we have a complete knowledge of the corresponding probabilities. Specifically, these formulas assume that for every point $\vec{x}$, we know the corresponding value $\rho(\vec{x})$ of the probability density function.

In practice, however, we only have partial information about the probabilities. Specifically, instead of the exact value $\rho(\vec{x})$, we only know an lower bounds $\underline{\rho}(\vec{x})$ and upper bound $\overline{\rho}(\vec{x})$ on the actual (unknown) value $\rho(\vec{x})$, i.e., we only know that the the value $\rho(\vec{x})$ belongs to the interval $[\underline{\rho}(\vec{x}), \overline{\rho}(\vec{x})]$.

Many different probability distributions are consistent with this interval information. For different such distributions, in general, we get different values for the amount S_u of useful information.

We do not know which of the distributions are more probable and which are less probable. Thus, we do not know which values of S_u are more probable and which are less probable. In such situations, it makes sense to characterize the uncertainty by the worst-case scenario, i.e., by the largest of the corresponding values S_u:

$$\overline{S}_u \stackrel{\text{def}}{=} \max\left\{S_u : \underline{\rho}(\vec{x}) \le \rho(\vec{x}) \le \overline{\rho}(\vec{x}) \text{ for all } x \text{ and } \int \rho(\vec{x})\, d\vec{x} = 1\right\}.$$

To compute this largest value, we can take into account that the objective function S_u is concave and the corresponding domain

$$\left\{\rho(\vec{x}) : \underline{\rho}(\vec{x}) \le \rho(\vec{x}) \le \overline{\rho}(\vec{x}) \text{ for all } x \text{ and } \int \rho(\vec{x})\, d\vec{x} = 1\right\}$$

is convex. Thus, to compute the maximum $\overline{S}_u$, we can use one of the efficient convex optimization algorithms; see, e.g., (Bertsekas, 2015).

Comment. Algorithms for computing a similar bound $\overline{S}$ for the traditional Shannon's entropy S under such intervl uncertainty are presented in (Kreinovich, 1996; Xiang, Ceberio, and Kreinovich, 2007; Nguyen et al., 2012).

Acknowledgements

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, and by an award "UTEP and Prudential Actuarial Science Academy and Pipeline Initiative" from Prudential Foundation.

The authors are thankful to the anonymous referees.

References

Aló, R. and O. Kosheleva, Optimization Techniques under Uncertain Criteria, and Their Possible Use in Computerized Education, *Proceedings of the 25th International Conference of the North American Fuzzy Information Processing Society NAFIPS'2006*, Montreal, Quebec, Canada, June 3–6, 2006, pp. 593–598.

Bertsekas, D. P., *Convex Optimization Algorithms*, Atthena Scientific, Belmont, Massachusetts, 2015.

Chirayath, V., L. Longpré, and V. Kreinovich, Measuring privacy loss in statistical databases, In: H. Leung and G. Pighizzini (Eds.), *Proceedings of the Workshop on Descriptional Complexity of Formal Systems DCFS'2006*, Las Cruces, New Mexico, June 21–23, 2006, pp. 16–25.

Fishburn, P. C., *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.

Kreinovich, V., Maximum entropy and interval computations, *Reliable Computing*, 1996, Vol. 2, No. 1, pp. 63–79.

Luce, R. D., and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.

Nguyen, H. T., O. Kosheleva, and V. Kreinovich, Decision making beyond Arrow's 'impossibility theorem', with the analysis of effects of collusion and mutual attraction, *International Journal of Intelligent Systems*, 2009, Vol. 24, No. 1, pp. 27–47.

Nguyen, H. T., V. Kreinovich, B. Wu, and G. Xiang, *Computing Statistics under Interval and Fuzzy Uncertainty*, Springer Verlag, Berlin, Heidelberg, 2012.

Raiffa, H. *Decision Analysis*, Addison-Wesley, Reading, Massachusetts, 1970.

Shannon, C. E., and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Chicago, Illinois, 1971.

Xiang, G., M. Ceberio, and V. Kreinovich, Computing Population Variance and Entropy under Interval Uncertainty: Linear-Time Algorithms, *Reliable Computing*, 2007, Vol. 13, No. 6, pp. 467–488.

Zimichev, A. M., V. Kreinovich, et al. *Automatic control system for psychological experiments*, Final Report on Grant No. 0471-2213-20 from the Soviet Science Foundation, 1982 (in Russian).

Zimichev, A. M., V. Kreinovich, et al. *Automatic system for quick education of computer operators*, Final Report on Grant No. 0471-0212-60 from the Soviet Science Foundation, 1980–82 (in Russian).