Bayesian Approach for Inconsistent Information

M. Stein[a], M. Beer[b], V. Kreinovich[c]

[a] *Department of Civil and Environmental Engineering, National University of Singapore, Singapore*
[b] *Institute for Risk and Uncertainty, School of Engineering, University of Liverpool, UK*
[c] *Department of Computer Science, University of Texas at El Paso, USA*

Abstract

In engineering situations, we usually have a large amount of prior knowledge that needs to be taken into account when processing data. Traditionally, the Bayesian approach is used to process data in the presence of prior knowledge. Sometimes, when we apply the traditional Bayesian techniques to engineering data, we get inconsistencies between the data and prior knowledge. These inconsistencies are usually caused by the fact that in the traditional approach, we assume that we know the *exact* sample values, that the prior distribution is *exactly* known, etc. In reality, the data is imprecise due to measurement errors, the prior knowledge is only approximately known, etc. So, a natural way to deal with the seemingly inconsistent information is to take this imprecision into account in the Bayesian approach – e.g., by using fuzzy techniques. In this paper, we describe several possible scenarios for fuzzifying the Bayesian approach.

We also show that in engineering problems, it is important to use the full Bayesian approach, by showing that the frequently used computational simplifications – e.g., assuming a fixed value for one of the parameters and assuming that parameters are independent – may lead to inaccurate (or even erroneous) results.

In this paper, to implement the corresponding fuzzy versions of the Bayesian formulas, we use straightforward computations of the related expression – which makes our computations reasonably time-consuming. Computations in the traditional (non-fuzzy) Bayesian approach are much faster – because they use algorithmically efficient reformulations of the Bayesian formulas. We expect that

similar reformulations of the fuzzy Bayesian formulas will also drastically decrease the computation time and thus, enhance the practical use of the proposed methods.

Keywords: Fuzzy-Bayes, imprecise data, uncertainty quantification, imprecise probabilities, fuzzy random variables

---

1. Inconsistent Knowledge and Data: Formulation of the Problem

Inconsistency in uncertainty quantification. One of the main objectives of uncertainty quantification is the specification of an appropriate mathematical uncertainty model which is the best reflection of reality. The available information to build such a model may appear in various forms including data, knowledge, guestimates, etc.; this information may be of a quantitative or of a qualitative nature. In general, different pieces of information corresponding to the same problem can be of different nature, different precision, different quality, etc. For example, a random sample may consist of set-valued sample elements; so, we have to deal with a mix of probabilistic and non-probabilistic information. In many practical cases, this leads to inconsistencies.

For example, different statistical tests applied to the same data may lead to inconsistent results. In cases when both data and subjective assessments are available, these may not agree with one another. Alternatively, different experts may have conflicting opinions about the same problem. Also, observations of the same problem, done in a direct and indirect manner, may lead to conflicting conclusions. In all these cases, when we formulate an uncertainty model, we have to deal with the problem of *inconsistencies*. Various cases of such inconsistencies are described and discussed in the special issue [5] and in the monograph [6] (see also references therein). In many such situation, a significant issue in dealing with inconsistencies is the analysis of *imprecise* data [11].

In the present study, we focus on two important types of inconsistency, when there is a moderate mismatch either between data and expert estimates, or between different parts of data. We show that both types of inconsistency

have a common solution – through a "marriage" of imprecise probabilities with Bayesian statistics.

Inconsistent knowledge: a brief description of the practical problem. In engineering practice, we sometimes encounter the practical problem of inconsistent knowledge. Specifically, in engineering, we usually have a large amount of prior knowledge. So, when we process the results of measurements and/or observations, we need to take the prior knowledge into account.

Traditionally, probabilistic and statistical methods are used to process measurement and observation results. In the probabilistic and statistical approach, prior knowledge is usually described by a prior distribution, and well-known Bayesian techniques can be used to process data in the presence of this prior knowledge.

The problem is that sometimes, the observations are (slightly) inconsistent with the prior knowledge. Let us give an example.

Inconsistent information: a simple example. When designing a bridge, we may assume, based on the past observation, that the wind speed w is always between 0 and 50 km/h, i.e., that the possible values of w belong to the interval $[0, 50]$. Since we have no reason to belief that some of these values are more probable and some are less probable, it is reasonable to assume that all the values within this interval are equally probable, i.e., that the prior distribution of w is uniform on this interval.

Suppose now that during a recent storm, we have measured the speed $w = 50.1$. From the *practical* viewpoint, this is not a problem. Indeed, first, the difference between 50.1 and 50 is small, so the actual wind speed – which, due to measurement error, can be slightly different from the measured value – can as well be below 50. Second, the systems are usually designed with an extra reliability, so the bridge should be able to withstand winds slightly stronger than 50 km/h. However, from the *purely mathematical* viewpoint, we have an inconsistency: according to the prior knowledge, the wind speed should be smaller than or equal to 50, while we have observed a value larger than 50.

How this problem is resolved now. Such "slight inconsistency" situations frequently occur in engineering practice. At present, there is no general recipe for dealing with this problem, practitioners deal with these problems on a case-by-case basis. For example, in the above case, a reasonable strategy for a practitioner is to somewhat widen the prior interval $[0, 50]$. How much to increase depends on the person.

The existing empirical approach to solving the inconsistency problem is not very satisfactory. A change of the range would make sense if we observe a *drastic* inconsistency, e.g., if we observed a storm with $w = 100$ km/h. In this case, the prior information is wrong, and we indeed need to update it.

However, when we observe a slight inconsistency (like $w = 50.1$), then, as we have mentioned, there is no intuitive contradiction with the prior knowledge. The mathematical inconsistency comes from the fact that we erroneously treat imprecise values – like 50 or 50.1 – as absolutely precise ones.

It is desirable to come up with a better approach for dealing with inconsistent information. In view of the above, it is desirable to come up with a more intuitively acceptable approach to dealing with the inconsistent information – ideally, by explicitly taking into account that the parameters which describe the data and the prior distribution are imprecise.

What we do in this paper. In this paper, we describe a natural way to deal with the imprecise values that enables us to process seemingly inconsistent data.

The structure of this paper. The structure of this paper is as follows. In order to explain how to naturally modify the existing Bayesian approach to engineering data processing, we first recall, in Section 2, the main motivations for this approach and the resulting algorithms. In Section 3, we use these motivations to come up with a natural way of taking imprecision into account. From the purely *theoretical* viewpoint, the problem is thus solved; however, from the practical *computational* viewpoint, performing a full Bayes update is often too computationally intensive to be practical – even in the simplified situations, when we

do not take imprecision into account. In such situations, practitioners often use simplified approaches: e.g., select a fixed value of one of the parameters or assume that different parameters are independent.

In Section 4, we show that one needs to be very cautious with such simplifications, since they can lead to inaccurate (and sometimes erroneous) results. This leaves us with an important open problem: to come up with a computationally efficient implementation of the corresponding Bayesian techniques.

Another important open problem is related to the fact that there are other approached for dealing with inconsistent data, such as the Dempster-Shafer approach. It would be nice to compare different approaches.

2. Traditional Bayesian Approach to Engineering Data Processing: Main Motivations and Resulting Algorithms

The structure of this section. In order to explain how to naturally modify the existing Bayesian approach to engineering data processing, let us first recall the main motivations for this approach and the resulting algorithms. We start by explaining why we need to consider prior information in the first place, then we explain how to describe prior knowledge, and finally, we explain how to update the knowledge based on the new observations and measurements.

Need to consider prior knowledge when processing engineering data. Both in science and in engineering, we often face the problem of processing data: we have observations $x_1, \ldots, x_n$, and we need to make conclusions based on these observations. In principle, the same standard statistical techniques can be used (and are used) in science and in engineering: e.g., we can use the usual statistical algorithms to estimate the mean, the variance, the correlation between different quantities, etc.

However, in spite of this similarity, there is an important difference between data processing in science (especially in state-of-the-art science) and in engineering. Indeed, one of the main objectives of science is to acquire new knowledge.

As a result, in science, the whole reason for making observations and measurements, the whole reason for collecting the data $x_1, \ldots, x_n$, is to find this new knowledge. Sometimes, we have *some* prior knowledge about the studied effect, but in many cases – especially in state-of-the-art science – we do not have much knowledge about the situation, we are doing exploratory research. In such situations, prior knowledge can be safely ignored, and the only information that is worth analyzing is the data itself.

As an example, let us consider a recent discovery of extraterrestrial planets of new type: Jupiter-type planets which are located very close to their stars. We have observed several such planets, and we want to make conclusions based on the observed data $x_1, \ldots, x_n$. This is a completely new phenomenon, for which there is no prior knowledge – except maybe some vague guesses. In such situations, we want to be sure that the conclusions that we make are objective – in the sense that they are based on the data and not on the vague guesses that form prior knowledge.

In contrast, in most engineering problems, we have a large amount of prior knowledge. The main reason for this large amount of prior knowledge is that, in contrast to science whose nature is mostly exploratory, engineering is mostly about *applying* knowledge to design systems, structures, processes, etc. Of course, engineering also has elements of exploration and discovery; however, when we, e.g., design a new bridge, we do not simply experiment with different designs in the hope that one of them works; we use the prior knowledge about bridges, about materials, about winds and possible earthquakes in this area – to make sure that the newly designed bridge lasts for the desired period of time. As a result, in engineering data processing, it is extremely important to take into account not only the data itself, but also the prior knowledge about the data.

How can we describe this prior knowledge?

Ideal description of prior information: a probability distribution. One of the main fundamental ideas behind applications of science and engineering is the

repeatability of results: if one laboratory measures the speed of light, another laboratory is expected to get the same result; if one bridge was successfully tested, it is expected that at any other similar location, a similar bridge will also withstand the desired load. The main difference between repeatability in fundamental science and engineering is that in fundamental science, the repeatability is (potentially) exact: fundamental constants like the speed of light are the same no matter where we measure them. In contrast, engineering parameters slightly change: e.g., even two similarly manufactured steel beams have slightly different mechanical properties, and for two similarly designed pieces of concrete, properties such as compressive strength can differ by as much as 30%.

So, while in fundamental science, we have a *single* value of a quantity, in engineering, we have a *population* of different values. To fully describe the corresponding population, we need to know which values are possible, and how frequent are different possible values. In other words, to fully describe the corresponding population, we need to know the probability distribution on the set of possible values.

Need for a finite-parametric family of probability distributions. One way to describe a probability distribution is to describe the probability density $f(x)$ for different values x. In principle, there are infinitely many possible values of the quantity x, and for each of these values, we can set a different value $f(x)$; the only constraint on all these parameters $f(x)$ is that the total probability should be equal to one: $\int f(x)\,dx = 1$. Thus, to fully describe a probability distribution, we need to describe infinitely many parameters. In practice, in the computer, we can only store finitely many numbers. So, we need to limit ourselves to finite-parametric probability distributions.

How to describe distributions from a finite-parametric family. Let $\theta = (\theta_1, \ldots, \theta_m)$ be the vector formed by the corresponding parameters. For each value of θ, we have a probability distribution $f(x \,|\, \theta)$ corresponding to these parameter values. In other words, we have a function that maps each pair (x, θ) into a non-negative

value $f(x \mid \theta)$, the probability density at value x under the condition that the actual parameters are θ. This function is usually called a *likelihood function*.

What do we know about the parameters θ: the notion of a prior distribution. The probability distribution of the sample values is uniquely determined by the parameter vector θ. Usually, we do not know the exact distribution, which means that we do not know the exact value θ. Instead, we have a *partial* information about θ: we know which values θ are possible and how probable are different possible values θ.

Usually, statisticians assume that we know the probabilities of different values θ, i.e., we know the *prior* probability distribution which can be described, e.g., by the corresponding probability density function $g(\theta)$.

Which prior distributions are used. Many different probability distributions are used as priors.

Sometimes, all we know is that the parameters belong to a certain area, and we have no reason to believe that some values from this area are more probable than others. In this case, as we have mentioned earlier, it makes sense to assume that all the possible values θ are equally probable, i.e., that the prior distribution is a *uniform* distribution on a given area.

In other cases, there are many independent small factors which contribute to our uncertainty in θ. It is known – this fact is known as the Central Limit Theorem – that, under reasonable conditions, the distribution of the sum of many such small random variables is close to normal. In such cases, it is reasonable to consider *normal* prior distribution.

We can also consider lognormal distributions, beta- and gamma-distributions, etc. In many cases, practitioners use prior distributions that make updates computationally easier; such prior distributions are known as *conjugate priors*.

How to combine prior knowledge with observations: Bayes theorem. As we have mentioned, prior distributions describe pour prior knowledge. After we perform measurements and observations, we gain additional knowledge, so we need to

update the corresponding distribution – by combining prior knowledge with the results x of measurement and observation.

Bayes theorem is a known probabilistic way of combining the prior knowledge with observations. This theorem is based on the following situation. Suppose that we have N possible mutually exclusive hypotheses $H_1, \ldots, H_N$ that cover all possible situations. Suppose also that we know the prior probabilities $P_0(H_i)$ of each of these hypotheses; then $\sum\limits_{i=1}^{N} P_0(H_i) = 1$. Suppose that for some event E, we know the conditional probability $P(E \,|\, H_i)$ that E occurs under hypothesis H_i. Our objective is to describe how the probabilities of different hypotheses will be updated if we actually observe the event E, i.e., to find the conditional probabilities $P(H_i \,|\, E)$.

The solution proposed by Bayes is based on the fact that the conditional probability $P(A \,|\, B)$ of A under the condition B can be defined as a fraction of those cases in which A holds as a proportion of the events in which B hold, i.e., as a ratio

$$P(A \,|\, B) = \frac{P(A \,\&\, B)}{P(B)}. \tag{1}$$

Thus, for the desired probability $P(H_i \,|\, E)$, we have

$$P(H_i \,|\, E) = \frac{P(E \,\&\, H_i)}{P(E)}. \tag{2}$$

Since the hypotheses H_i are mutually exclusive and cover all possible situations, we conclude that $P(E) = \sum\limits_{j=1}^{N} P(E \,\&\, H_j)$. Thus, we have

$$P(H_i \,|\, E) = \frac{P(E \,\&\, H_i)}{\sum\limits_{j=1}^{N} P(E \,\&\, H_j)}. \tag{3}$$

For the known probability $P(E \,|\, H_j)$, we similarly have

$$P(E \,|\, H_j) = \frac{P(E \,\&\, H_j)}{P_0(H_j)}, \tag{4}$$

so $P(E \,\&\, H_j) = P(E \,|\, H_j) \cdot P_0(H_j)$. Substituting this expression into the formula (3), we get the Bayes theorem

$$P(H_i \,|\, E) = \frac{P(E \,|\, H_i) \cdot P_0(H_i)}{\sum\limits_{j=1}^{N} P(E \,|\, H_j) \cdot P_0(H_j)}. \tag{5}$$

In our case, different mutually exclusive hypotheses are different values of the parameter vector θ, and the events are the observed values x; hence, e.g., the conditional probability $P(E \,|\, H_i)$ is proportional to $f(x \,|\, \theta)$. In this case, we have infinitely many possible values θ, so instead of the sum, we will have an integral:

$$g(\theta \,|\, x) = \frac{f(x \,|\, \theta) \cdot g(\theta)}{\int f(x \,|\, \nu) \cdot g(\nu)\, d\nu}. \tag{6}$$

The resulting distribution for θ is known as a *posterior* distribution.

How to get numerical estimates from the posterior distribution. The posterior distribution provides us with a full information about the probability of different possible parameter vectors θ. In many engineering situations, practitioners do not need that much information, all they need is an approximate value of each of the parameters – and maybe an indication of how accurate are these approximate values.

For a random variable Θ with the posterior probability density $g(\theta \,|\, x)$, we need to find a single approximating vector $\theta_{\rm p}$ for which, e.g., the mean square difference between the actual value θ and this approximating vector is the smallest possible: $\int (\theta - \theta_{\rm p})^2 \cdot g(\theta \,|\, x)\, d\theta \to \min$. Differentiating this expression with respect to $\theta_{\rm p}$ and equating this derivative to 0, we conclude that $\theta_{\rm p} = \int \theta \cdot g(\theta \,|\, x)\, d\theta$. In other words, we conclude that the optimal approximation $\theta_{\rm p}$ is the *expected value* $E[\theta]$ of the posterior distribution.

Comment. To gauge the accuracy of this approximation, we can use the corresponding mean square difference $\int (\theta - \theta_{\rm p})^2 \cdot g(\theta \,|\, x)\, d\theta$ which, as one can easily check, is the *variance* of the posterior distribution. Alternatively, we can find the confidence areas that contain the actual values θ with the given certainty.

Convergence to the true values depends on how accurate is the prior knowledge. The prior distribution is based on our prior knowledge, i.e., in effect, on prior measurements and observation of this phenomenon and of similar phenomena. When we add more observations and measurements, the resulting posterior distribution takes into account both the prior observations and the new ones.

The more new observations we add, the larger their relative contribution to the posterior distribution, and thus, the closer the resulting posterior estimates $E[\Theta]$ to the actual value of the parameters θ. In other words, when we add more and more measurement, this estimate converges to the actual value θ.

The speed of this convergence depends on how probable the actual value was in the prior distribution. If in the original distribution, the actual vector θ was very probable, the process converges fast, and we get a good approximation for the actual parameters θ after a few measurements. On the other hand, if in the original prior distribution, the actual vector θ was not very probable – i.e., if we did not have much prior knowledge about θ – then we need many more observations and measurements to get a good approximation to θ.

3. A Natural Way to Take Imprecision Into Account

Inconsistent knowledge: reminder. The Bayesian approach works well in many practical applications, but sometimes, it cannot be applied because we have inconsistency.

Inconsistency occurs, e.g., in the following case. Let x be equal to θ with probability 1, and the prior distribution of the parameter θ_1 be a uniform distribution on an interval $[a_1, b_1]$. In this case, only values x_1 from this interval $[a_1, b_1]$ are possible. Thus, if the observed value x_1 is outside this interval – e.g., is slightly larger than b_1 – then we get an inconsistency between the observation result and the prior information.

In this case, the Bayes formula cannot be applied. Indeed, the fact that $x = \theta$ with probability 1 means that $f(x \,|\, \theta) = 0$ for $\theta \neq x$. However, for $\theta = x$, we have $g(\theta) = 0$, since the value $x = \theta$ is outside the interval $[a_1, b_1]$ where the uniform probability density is different form 0. Thus, the numerator $f(x \,|\, \theta) \cdot g(\theta)$ is equal to 0 both for $x \neq \theta$ and for $x = \theta$ and is, hence, always equal to 0. The denominator is also equal to 0, and so the whole formula cannot be applied.

The main reason for inconsistency is that we do not take imprecision into account: reminder. As we have mentioned, the seeming inconsistency is caused by the fact that we do not take imprecision into account.

For example, the likelihood function $f(x \,|\, \theta)$ describes the distribution of the *actual* values x, but what we observe is, due to the measurement errors, slightly different from the actual values. Similarly, as we have mentioned, the choice of a prior distribution describing our prior knowledge is a somewhat subjective task; instead of the given prior distribution, we could as well select a closer one. Finally, our selection of the finite-parametric family of distributions $f(x \,|\, \theta)$ is somewhat subjective, we could have selected a close family.

Towards a natural way to take imprecision into account: from intervals to fuzzy. How can we take imprecision into account? Let us first consider the imprecision with which we know each observed value x_i. This imprecision comes from *measurements*, in which the observed value x_i is, in general, somewhat different from the actual (unknown) value $x_i^{\rm act}$.

Sometimes, we know the exact probability of each possible value of the measurement error $\Delta x_i \stackrel{\rm def}{=} x_i - x_i^{\rm act}$; however, often, the only information that we have is the upper bound Δ_i on the measurement error (to be more precise, on the absolute value of the measurement error). In this case, after the measurements, the only information that we have about the actual (unknown) value $x_i^{\rm act}$ is that this value belongs to the *interval* $[x_i - \Delta_i, x_i + \Delta_i]$.

Usually, the guaranteed upper bound Δ_i is an overestimation – caused by the need to provide an absolute guarantee. Often, in addition to this absolutely guaranteed bound, designers and manufacturers can also provide us with smaller bounds – but these smaller bounds come with a certain degree of uncertainty. The smaller uncertainty α we want, the larger the bounds. For each uncertainty level α, we have a corresponding bound $\Delta_i(\alpha)$ and thus, we can conclude that with this uncertainty, the actual value $x_i^{\rm act}$ belongs to the interval $\mathbf{x}_i(\alpha) = [x_i - \Delta_i(\alpha), x_i + \Delta_i(\alpha)]$. If we have two degrees $\alpha < \alpha'$, then the interval corresponding to the larger degree of uncertainty α' is a subinterval of the

interval corresponding to the smaller degree α: $\mathbf{x}_i(\alpha') \subseteq \mathbf{x}_i(\alpha)$. Families of intervals that satisfy this property are called *nested* since they are all contained in each other like nested dolls.

We can gauge the expert's degree of uncertainty by asking the expert to mark this degree by a number on a scale from 0 to 1. Thus, the degree of uncertainty becomes a number from the interval $[0,1]$: 0 means there is no uncertainty, we are absolutely certain, while 1 means that we are not certain at all. So, we have a nested family of intervals labeled by numbers from the interval $[0,1]$.

Each such nested family corresponds to a *fuzzy number*, whose α-cuts are exactly these intervals $\mathbf{x}_i(\alpha)$; see, e.g., [13]. For that, we can define the membership function $\mu_i(x_i)$ as $\max\{\alpha : x_i \in \mathbf{x}_i(\alpha)\}$. Thus, a natural way to describe each imprecise value x_i is to describe it as a *fuzzy number*; see, e.g., [6, 8, 9, 13].

Comment. There are many ways to quantify the expert's uncertainty level α. In particular, we can base this quantification on probabilistic arguments, by describing the degree of uncertainty in terms of confidence levels or credibility levels. A practically useful heuristic procedure of this type is described in the Appendix.

Historical comment. The need to extend statistical techniques to set-valued and fuzzy samples is well-understood, and many methods have been proposed for such an extension; see, e.g., [2, 4, 7, 9].

How to describe imprecision in prior distribution. Just like a distribution for x is selected from a finite-parametric family of distributions, the prior distribution $g(\theta)$ is also selected from a finite-parametric family of distributions, i.e., it has the form $g(\theta) = f_p(\theta \,|\, b^{(0)})$ for some family of distributions $f_p(\theta \,|\, b)$ for some specific vector $b^{(0)}$.

Similar to imprecision in x, here imprecision means that we do not know the exact vector $b^{(0)}$; instead, we have a fuzzy vector, for which we know the alpha-cuts $\mathbf{b}_k(\alpha)$ for each of the components b_k.

How to describe imprecision in the likelihood function. To describe imprecision in the likelihood function, we can use a similar idea: instead of considering a *single* function $f(x\,|\,\theta)$, let us explicitly take into account that this function is usually only a subclass of some more general class of probability distributions, i.e., that $f(x\,|\,\theta) = f_\ell(x\,|\,\theta, d^{(0)})$ for some family of functions $f_\ell(x\,|\,\theta, d)$ and for some specific vector $d^{(0)}$. Here, imprecision means that we do not know the exact vector $d^{(0)}$; instead, we have a fuzzy vector, for which we know the alpha-cuts $\mathbf{d}_\ell(\alpha)$ for each of the components d_ℓ.

Need to generalize Bayes formula to the fuzzy case. Now that, in general, the sample, the prior distribution, and the likelihood function are fuzzy, we need to extend the Bayes formula to this case.

Historical comment. In our description, we follow the main ideas summarized by R. Viertl and his co-authors in [10, 11, 12]; in the following text, we explain in what aspects what we are doing is different from the approaches presented in [10, 11, 12].

How to generalize Bayes formula to the interval case. Let us first consider the simplest case when instead of an exact value, we have *intervals* of possible values: we have intervals $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$, $\mathbf{b}_k = [\underline{b}_k, \overline{b}_k]$, and $\mathbf{d}_\ell = [\underline{d}_\ell, \overline{d}_\ell]$. For each combination of values $x_i \in \mathbf{x}_i$, $b_k \in \mathbf{b}_k$, and $d_\ell \in \mathbf{d}_\ell$, we can use the Bayes theorem and compute the values

$$g(\theta\,|\,x) = \frac{f(x\,|\,\theta, d)\cdot g(\theta, b)}{\int f(x\,|\,\nu, d)\cdot g(\nu, b)\,d\nu}. \tag{7}$$

Based on these values, we can then estimate, for each component θ_j, the expected value $\theta_{\mathrm{p},j} \stackrel{\mathrm{def}}{=} E[\theta_j] = \int \theta_j \cdot g(\theta\,|\,x)\,d\theta$. Thus, for each combination of vectors $x = (x_1, x_2, \ldots)$, $b = (b_1, \ldots)$, and $d = (d_1, \ldots)$, we defined the vector $\theta_{\mathrm{p}}(x, b, d)$ with coordinates

$$\theta_{\mathrm{p},j}(x, b, d) \stackrel{\mathrm{def}}{=} \int \theta_j \cdot \frac{f(x\,|\,\theta, d)\cdot g(\theta, b)}{\int f(x\,|\,\nu, d)\cdot g(\nu, b)\,d\nu}\,d\theta. \tag{8}$$

Different values $x_i \in \mathbf{x}_i$, $b_k \in \mathbf{b}_k$, and $d_\ell \in \mathbf{d}_\ell$ lead, in general, to different vectors $\theta_{\mathrm{p}}(x, b, d) = (\theta_{\mathrm{p},1}(x, b, d), \ldots, \theta_{\mathrm{p},m}(x, b, d))$. It is therefore reasonable

to describe the "posterior" range $R_{\rm p}(\mathbf{x},\mathbf{b},\mathbf{d})$ of all "posterior" vectors $\theta_{\rm p}$ which can be obtained from different combinations of values $x_i \in \mathbf{x}_i$, $b_k \in \mathbf{b}_k$, and $d_\ell \in \mathbf{d}_\ell$. In other words, for each combination of three interval-valued vectors $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots)$, $\mathbf{b} = (\mathbf{b}_1, \ldots)$, and $\mathbf{d} = (\mathbf{d}_1, \ldots)$, we define the range as

$$R_{\rm p}(\mathbf{x},\mathbf{b},\mathbf{d}) \stackrel{\rm def}{=} \{\theta_{\rm p}(x,b,d) : x \in \mathbf{x}, b \in \mathbf{b}, d \in \mathbf{d}\}, \tag{9}$$

where $x = (x_1, x_2, \ldots) \in \mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots)$ means that $x_1 \in \mathbf{x}_1$, $x_2 \in \mathbf{x}_2$, etc.; $b \in \mathbf{b}$ and $d \in \mathbf{d}$ are defined similarly.

Comment. In a similar manner, we can define the ranges of all other characteristics of the posterior distribution.

How to generalize Bayes formula: from interval case to the fuzzy case. Processing fuzzy data is usually described by Zadeh's Extension Principle; see, e.g., [13]. It is known that under reasonable assumptions, this principle reduces to the following natural idea: for each α, to find the α-cut of the result, we process the α-cuts of all the inputs.

For example, to get the fuzzy set describing the vector $\theta_{\rm p}$, we can do the following. For each α, we find the α-cuts $\mathbf{x}_i(\alpha)$, $\mathbf{b}_k(\alpha)$, and $\mathbf{d}_\ell(\alpha)$ of the corresponding fuzzy sets. Then, for each combination of values $x_i \in \mathbf{x}_i(\alpha)$, $b_k \in \mathbf{b}_k(\alpha)$, and $d_\ell \in \mathbf{d}_\ell(\alpha)$, we compute the values

$$g(\theta \,|\, x) = \frac{f(x \,|\, \theta, d) \cdot g(\theta, b)}{\int f(x \,|\, \nu, d) \cdot g(\nu, b)\, d\nu} \tag{10}$$

and

$$\theta_{{\rm p},j} = \int \theta_j \cdot g(\theta \,|\, x)\, d\theta. \tag{11}$$

The range of all resulting vectors $\theta_{\rm p} = (\theta_{{\rm p},1}, \ldots, \theta_{{\rm p},m})$ is the desired α-cut for the fuzzy vector $\theta_{\rm p}$. In other words, an m-dimensional fuzzy set for $\theta_{\rm p}$ is defined as a fuzzy set for which each α-cut is defined as $R_{\rm p}(\mathbf{x}(\alpha), \mathbf{b}(\alpha), \mathbf{d}(\alpha))$, where $\mathbf{x}(\alpha) \stackrel{\rm def}{=} (\mathbf{x}_1(\alpha), \mathbf{x}_2(\alpha), \ldots)$, $\mathbf{b}(\alpha) \stackrel{\rm def}{=} (\mathbf{b}_1(\alpha), \ldots)$, and $\mathbf{d}(\alpha) \stackrel{\rm def}{=} (\mathbf{d}_1(\alpha), \ldots)$.

How this is different from Viertl's approach. Our formulas follow the main *ideas* described in the first half of [12]; our only change is that we take into account

not only possible imprecision of data and prior distribution, but also possible imprecision of the likelihood function.

The main difference is that we follow these original *ideas*, while the actual *algorithms* described by Viertl and others [10, 11, 12] modify the resulting formulas, so that they will be able to simplify computations by retaining the sequential character of Bayesian update. One can see that their modification results in a *narrower* interval for the resulting probability density. In many engineering applications, we want to provide guaranteed estimates; in this case, we should not artificially narrow down the resulting imprecision intervals – even if it leads to faster computations, we should return the exact ranges produced by the above Bayes formulas.

Important comment. Our main motivation for taking imprecision into account came from the need to process inconsistent knowledge. Inconsistent knowledge is where imprecision *has* to be taken into account. The existence of seemingly imprecise knowledge highlights the fact that there is imprecision in many practical situations. Of course, in many cases, we have imprecision without inconsistency. In such cases, to get the most adequate description of our knowledge, it is also necessary to take this imprecision into account.

4. How to Transform the Above *Theoretical* Solution into a *Practical* Solution: Examples and Challenges

Need to speed up computations. From the purely *theoretical* viewpoint, the problem of processing inconsistent knowledge is solved: we have explicit formulas describing how to produce the fuzzy estimates for the desired parameters.

However, from the practical *computational* viewpoint, the above formulas are not very efficient: they require that we consider all possible combinations of values, and there can be many different combinations. Even in the simplified situations, when we do not take imprecision into account, performing a full Bayes update is often too computationally intensive to be practical – because we need to compute the corresponding multi-D integral.

How Bayesian computations are sped up now. To speed up computations, practitioners often use simplified approaches: e.g., select a fixed value of one of the parameters or assume that different parameters are independent.

What we show in this section. In this section, we show, on numerical examples, that one needs to be very cautious with such simplifications, since they can lead to inaccurate (and sometimes erroneous) results.

Moreover, we will show that this simplification-caused inaccuracy occurs not only for the (relatively rare) cases of inconsistent knowledge, it occurs in the general situation in which we take imprecision into account. To prove this generality, we try to make our examples as generic as possible – in particular, we use examples without inconsistency.

How we perform computations. Since the main objective of our numerical examples is to compare *results* of different formulations of the problem, in all our computations, we use straightforward numerical computations without worrying about these computations requiring too much computation time. Specifically, to compute the values of the corresponding integrals, we discretized the range of values of each variables and computed the corresponding integral sum. To find the range of possible values of different variables and combinations of variables, i.e., to solve the corresponding optimization problems, we used component-wise optimization.

Case study: general description. In all our examples, we use the same numerical data: 20 results of measuring the compressive strength of concrete; see middle column of Table 1.

We chose this data set because this same data set is used in our previous paper [2] as an illustration of alternative (non-Bayesian) fuzzy techniques; the fact that we perform our Bayesian computations on the same sample enables us to compare these two different fuzzy approaches.

We assume that the values f_c are normally distributed, with some mean μ_X and standard deviation σ_X: $f_c \sim N(\mu_X, \sigma_X)$. In this case, the observed data is

Number i of realizations	Compressive strength	Fuzzy compressive strength
	$x_i = f_{ci}\ [N/mm^2]$	$\tilde{x}_i = f_{ci}\ [N/mm^2]$
1	28.3	$\langle 26.3, 28.3, 30.3 \rangle$
2	31.5	$\langle 29.5, 31.5, 33.5 \rangle$
3	35.2	$\langle 33.2, 35.2, 37.2 \rangle$
4	29.8	$\langle 27.8, 29.8, 31.8 \rangle$
5	27.6	$\langle 25.6, 27.6, 29.6 \rangle$
6	30.7	$\langle 28.7, 30.7, 32.7 \rangle$
7	25.2	$\langle 23.2, 25.2, 27.2 \rangle$
8	34.6	$\langle 32.6, 34.6, 36.6 \rangle$
9	28.9	$\langle 26.9, 28.9, 30.9 \rangle$
10	19.2	$\langle 17.2, 19.2, 21.2 \rangle$
11	26.8	$\langle 24.8, 26.8, 28.8 \rangle$
12	35.3	$\langle 33.3, 35.3, 37.3 \rangle$
13	26.3	$\langle 24.3, 26.3, 28.3 \rangle$
14	23.1	$\langle 21.1, 23.1, 25.1 \rangle$
15	20.2	$\langle 18.2, 20.2, 22.2 \rangle$
16	29.2	$\langle 27.2, 29.2, 31.2 \rangle$
17	25.7	$\langle 23.7, 25.7, 27.7 \rangle$
18	34.2	$\langle 32.2, 34.2, 36.2 \rangle$
19	24.8	$\langle 22.8, 24.8, 26.8 \rangle$
20	22.8	$\langle 20.8, 22.8, 24.8 \rangle$

Table 1: Sample of the cylinder compressive strength f_c

just f_c, the parameter vector $\theta = (\theta_1, \theta_2)$ consists of two components $\theta_1 = \mu_X$ and $\theta_2 = \sigma_X$, and the likelihood function has the well-known form

$$f(x \,|\, \theta) = \frac{1}{\sqrt{2 \cdot \pi} \cdot \theta_2} \cdot \exp\left(-\frac{(x - \theta_1)^2}{2\theta_2^2}\right). \tag{12}$$

Prior distributions: non-fuzzy case. In this paper, we consider two cases corresponding to the simplest prior distributions: uniform and normal. For each of these cases, we consider two subcases:

- the 1-D subcase, when the value of the standard deviation $\theta_2 = \sigma_X$ is fixed, and we only consider a non-degenerate distribution for $\theta_1 = \mu_X$, and
- the 2-D subcase, when we consider a non-degenerate joint distribution of both parameters.

Each 1-D subcase corresponds to the 1-D prior distribution, i.e., to the above practical idea of fixing the value of one of the parameters (in this case, of $\theta_2 = \sigma_X$).

In all four subcases, we take 28.5 as the center for the distribution of $\theta_1 = \mu_X$ and 4.5 as the center for the distribution of $\theta_2 = \sigma_X$. (The value 28.5 is close to the sample mean, and 4.5 is close to the sample standard deviation.)

In both uniform subcases, we assume that $\theta_1 = \mu_X$ is uniformly distributed on the interval $[27.0, 30.0]$ centered at the value 28.5. The difference between these subcases is that:

- in the 1-D subcase, we assume that $\theta_2 = \sigma_X$ is equal to the selected value 4.5, while
- in the 2-D subcase, we assume that θ_2 is uniformly distributed on the interval $[3.0, 6.0]$ centered around this value.

In both normal subcases, we assume that μ_X is normally distributed with mean 28.5 and standard deviation 2.0. The difference between these subcases is that:

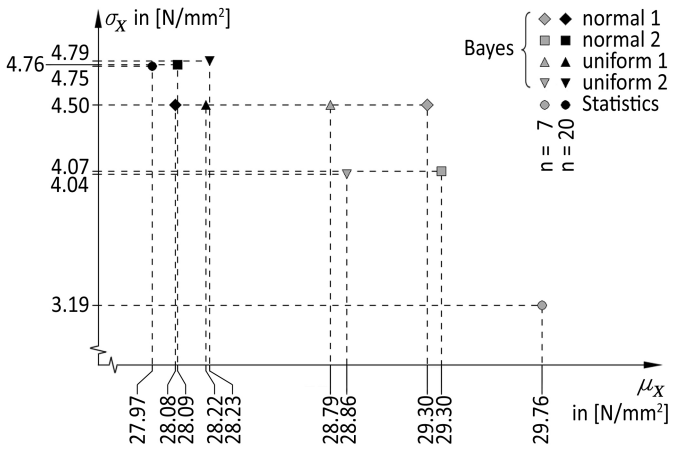

Figure 1: Estimates for the distribution parameters of X

- in the 1-D subcase, we assume that $\theta_2 = \sigma_X$ is equal to the selected value 4.5, while
- in the 2-D subcase, we assume that θ_2 is also normally distributed, with mean 4.5 and standard deviation 1.0.

Comment. In both 2-D cases, for simplicity, we consider an example in which the prior distributions for the mean $\theta_1 = \mu_X$ and for the variance $\theta_2 = \sigma_X$ are independent. This is only done to make the example simpler. In a generic 2-D distribution of the vector $\theta = (\theta_1, \theta_2)$, components θ_1 and θ_2 are, of course, not necessarily independent.

Numerical results show the limitations of the method of fixing the values of some parameters. Figure 1 shows the results of applying the Bayesian update for all four subcases; we separately show the results of using only the first 7 measurements and of using all 20 measurements.

We see that in the 2-D subcases, as we increase the sample size, the resulting estimates for μ_X get closer to the sample mean. In contrast, in the 1-D subcases, when we prematurely fix σ_X at the value 4.5, not only this value does not get close to the actual value, but the resulting estimates for μ_X do not get as close

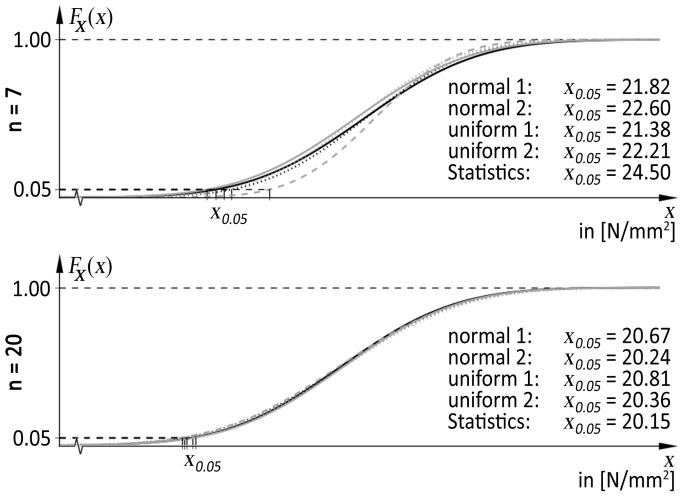

Figure 2: Distribution functions for X, based on different parameter estimators and varying sample size; 5% quantiles.

to the sample mean as in the 2-D subcases – if the fixed value is not close to the statistic as in the case of the first 7 sample elements.

We can also compare:

- the distribution $f(x \mid \theta_{\mathrm{p}})$ for x corresponding to the posterior mean values θ_{p} of the parameters with
- the distribution $f(x \mid \theta_{\mathrm{s}})$ corresponding to the sample mean and sample standard deviation θ_{s}.

The x-distributions $f(x \mid \theta_{\mathrm{p}})$ corresponding to 4 different prior distributions – as well as the 5%-quantiles of these x-distributions – are shown on Figure 2, both for the size $n = 7$ and for the size $n = 20$. The 5%-quantile is a practically important characteristic: it is the concrete strength which is guaranteed in 95% of the cases.

From Figure 2, we can see that for the smaller sample $n = 7$, the 5%-quantiles obtained from the 1-D (simplified) computations are reasonably conservative: we conclude that 95% of the cases will have strength ≥ 21.28 or ≥ 21.82, while in reality, they are even ≥ 24.50. However, for the larger sample ($n = 20$):

- our conclusion based on the 1-D simplification is that the strength is $\geq$ 20.81 in 95% of the cases, while
- in reality, the only thing that we can guarantee about the concrete strength in 95% of the cases is that this strength is ≥ 20.15.

Such over-estimation of the concrete strength can potentially lead to disastrous structural failures.

How we can take into account measurement imprecision. When we apply the traditional (non-fuzzy) Bayes approach, we use the actual measurement results. When we apply the fuzzy approach, we take into account the measurement uncertainty by replacing each value x_i with a triangular membership function with a center at x_i and endpoints at $x_i - 2$ and $x_i + 2$. The resulting fuzzy numbers are described in the last column of Table 1.

The triangular membership function is selected for illustrative purposes only – because it is computationally the simplest.

Two possible approaches: full Bayesian update and a simplified version assuming parameter independence. To explain the possibility of two different approaches, let us start with the simplest case of interval uncertainty, when both the prior distribution $g(\theta)$ and the likelihood function $f(x \,|\, \theta)$ are known exactly, but the values $x = (x_1, \ldots, x_n)$ are known with interval uncertainty. In other words, for each i, we only know the interval $[\underline{x}_i, \overline{x}_i]$ of possible values. For each combination of values $x_i \in [\underline{x}_i, \overline{x}_i]$, we can perform the Bayes update, find the posterior distribution $g(\theta \,|\, x)$, and then compute the means $\theta_{\rm p} = (\theta_{{\rm p},1}, \ldots)$ of the parameters over this posterior distribution.

In the *full* Bayes update, we try all possible combinations of $x_i \in [\underline{x}_i, \overline{x}_i]$, and come up with the multi-D range of possible values of $\theta_{\rm p}$, which reflects interaction between the parameters (known as dependability problem in interval analysis). For example, in the 2-D case, when we have two parameters $\theta = (\theta_1, \theta_2)$, we get a 2-D range. A general 2-D range is difficult to describe and difficult to compute. So, to simplify computations, we can instead, for each

parameter θ_j, compute the interval $[\underline{\theta}_{\mathrm{p},j}, \overline{\theta}_{\mathrm{p},j}]$ of possible values of this parameter. In this case, for each parameter, we only need to compute two numbers $\underline{\theta}_{\mathrm{p},j}$ and $\overline{\theta}_{\mathrm{p},j}$, so these computations are much faster.

Once we computed the intervals, then – instead of returning the actual multi-D range – we return the box

$$[\underline{\theta}_{\mathrm{p},1}, \overline{\theta}_{\mathrm{p},1}] \times \ldots \times [\underline{\theta}_{\mathrm{p},m}, \overline{\theta}_{\mathrm{p},m}] \tag{13}$$

which is guaranteed to contain the multi-D range. (Actually, one can easily see that this is the smallest box that contains the actual range.)

For example, in the 2-D case, we return a 2-D box $[\underline{\theta}_{\mathrm{p},1}, \overline{\theta}_{\mathrm{p},1}] \times [\underline{\theta}_{\mathrm{p},2}, \overline{\theta}_{\mathrm{p},2}]$.

The box is easier to compute than the actual range, but, as we see later, it contains additional points. In other words, we give up some accuracy in describing the results – and we gain a drastic decrease in computation time.

In the fuzzy case, when all the inputs are fuzzy numbers, instead of a *crisp* set of possible values of θ, we get a *fuzzy*. Namely, for each level α, we apply the above interval procedure to the α-cuts $\mathbf{x}_i(\alpha)$ of the input fuzzy numbers. The resulting multi-D range – or the multi-D box approximating this range – serve as the α-cut for the desired fuzzy set.

Computational results: case of imprecise data. Let us start by considering the above 2-D normal distribution as the prior. On Figure 3, we show the results of applying the above algorithm to this prior and to the fuzzy data from the last column of Table 1. The 2-D areas correspond to the level $\alpha = 0$, when we allow all possible values from the intervals $[x_i - 2, x_i + 2]$.

We perform the full Bayesian update – when we take into account interaction between the parameters $\theta_1 = \mu_X$ and $\theta_2 = \sigma_X$ – twice: for the first $n = 7$ measurement results and for the whole sample ($n = 20$). The range corresponding to $n = 7$ is described by the dashed like, the range corresponding to $n = 20$ by a solid line. For $n = 20$, we also used a simplified approach – in which we ignore the interaction between the parameters. We can say that the resulting 2-D box indeed contains many more points (a much bigger area) than the actual range.

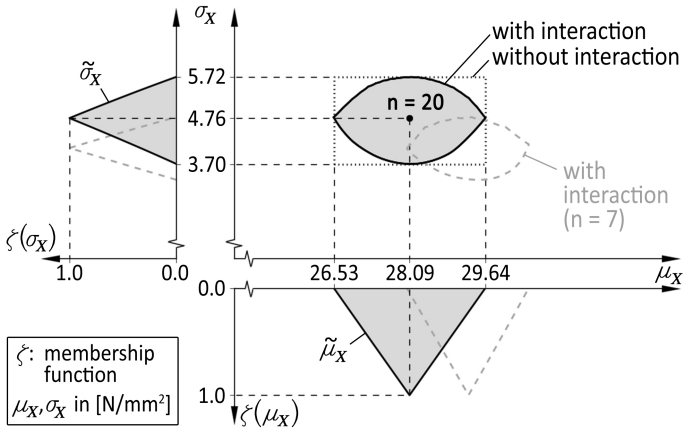

Figure 3: $\mu_X - \sigma_X$ interaction after Bayesian update with imprecise data for the 2-D normal prior.

In addition to the 2-D ranges corresponding to $\alpha = 0$, Figure 3 also depicts membership functions for $\theta_1 = \mu_X$ and $\theta_2 = \sigma_X$ corresponding to individual parameters. For each of these membership functions and for each α, the corresponding α-cut is the range of the values of the j-th parameter when we consider values $x_i \in \mathbf{x}_i(\alpha)$. In particular, for $\alpha = 1$, each input α-cut degenerates into a single point x_i, so the resulting value θ_j becomes exactly the value that we obtained earlier in the case of exact inputs for this prior distribution.

Observation: imprecision does not decrease with sample size. It is important to observe that the size of the 2-D region does not decrease when we increase the sample size – in contrast to the traditional Bayesian approach, as the sample size increases, the estimates become more and more accurate, and the size of the confidence region decreases.

Imprecision does not decrease with sample size: an intuitive explanation. This observed phenomenon is in good accordance with the usual sample estimates, e.g., with the sample arithmetic average

$$x_{\text{av}} = \frac{x_1 + \ldots + x_n}{n} \tag{14}$$

as an estimate for the mean μ_X. Indeed, when we take an average of n independent random variables with standard deviation σ_X, then the standard deviation of this average is equal to $\frac{\sigma_X}{\sqrt{n}}$ and thus, decreases when the sample size increases.

In the case of interval uncertainty, when the only information that we have about the actual (unknown) values $x_i^{\rm act}$ is that they belong to the intervals $[x_i - \Delta, x_i + \Delta]$, we return the set of all possible values of the average

$$x_{\rm av}^{\rm act} = \frac{x_1^{\rm act} + \ldots + x_n^{\rm act}}{n} \tag{15}$$

when $x_i^{\rm act} \in [x_i - \Delta, x_i + \Delta]$. The average is an increasing function of all its n variables $x_i^{\rm act}$, so its smallest possible value $\underline{x}_{\rm av}^{\rm act}$ is attained when all n variables takes their smallest possible values $x_i^{\rm act} = x_i - \Delta$. In this case,

$$\underline{x}_{\rm av}^{\rm act} = \frac{X_1 + \ldots + X_n}{n} = \frac{(x_1 - \Delta) + \ldots + (x_n - \Delta)}{n} = x_{\rm av} - \Delta. \tag{16}$$

Similarly, the largest possible value $\overline{x}_{\rm av}^{\rm act}$ of $x_{\rm av}^{\rm act}$ is attained when all n variables takes their smallest possible values $x_i^{\rm act} = x_i + \Delta$. In this case,

$$\overline{x}_{\rm av}^{\rm act} = \frac{x_1^{\rm act} + \ldots + x_n^{\rm act}}{n} = \frac{(x_1 + \Delta) + \ldots + (x_n + \Delta)}{n} = x_{\rm av} + \Delta. \tag{17}$$

Thus, the range of possible values of the mean takes the form $[x_{\rm av} - \Delta, x_{\rm av} + \Delta]$. The width of this interval does not decrease when we increase the sample size.

Comparison with the results of traditional (non-Bayesian) fuzzy data processing. It is interesting to compare the above results with the sample mean and standard deviation that can be computed – by using the same Zadeh's Extension Principle – from the fuzzy inputs. This comparison is presented on Figure 4. The difference between the two estimates is that in the Bayes approach, in addition to the data, we also use prior information. Since we use additional information in the Bayes approach, we expect the results of this approach to be more precise – and indeed, the ranges corresponding to the Bayes approach are smaller than the original statistical one. It can also be seen that, with increasing sample size, the Bayesian result converges towards the result from fuzzy statistics.

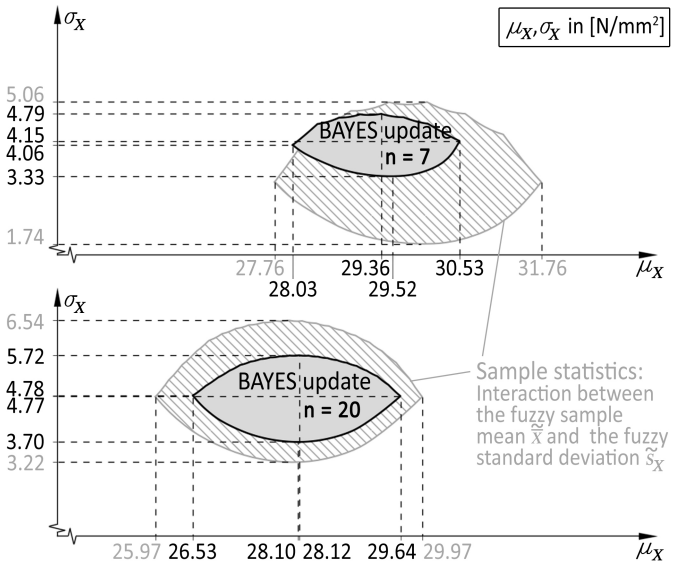

Figure 4: Comparison of the estimation results from Bayesian update with imprecise data with the results from sample statistics with imprecise data.

Thus, by the way, we show why it is beneficial to use Bayes approach even when we do not have inconsistent knowledge – because this approach enables us to use additional knowledge and thus, get more precise estimates.

Towards fuzzy probability distributions. Different values of the parameters θ correspond to different x-distributions $f(x\,|\,\theta)$. In particular, different possible values θ lead to different cumulative distribution functions

$$F(x) \stackrel{\text{def}}{=} \text{Prob}(X \le x). \tag{18}$$

Thus, for each x, instead of the single value $F(x)$, we have a range (interval) of possible values corresponding to different values θ.

For each level α, when we consider α-cuts of the inputs, we get the corresponding range of values of θ and thus, the interval of possible values of $F(x)$. So, for each x, we have a nested family of intervals corresponding to different α. In other words, for each x, $F(x)$ is now a fuzzy number. We can thus say that we now have a fuzzy-valued cumulative distribution function (CDF); see,

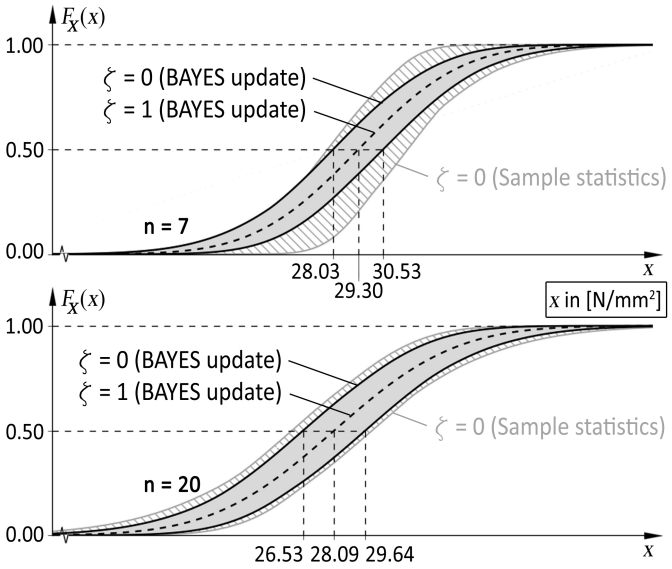

Figure 5: Comparison of the resulting CDF's

e.g., [1]. In Figure 5, we depict the corresponding fuzzy CDF for $\alpha = 0$ and for $\alpha = 1$. We can see that the range for the cdf $F(x)$ obtained by using statistical estimates (without taking into account prior knowledge) is much wider than the actual range – and is, thus, a not very accurate description of the actual range.

Figure 6 depicts what happens if we perform similar computations for the 2-D *uniform* prior distribution. In this case also, if we ignore the interaction between the parameters, we increase the fuzziness of the results.

Similarly, for different values θ, we get different 5%-quantiles. Thus, for each α, instead of a single quantile, we have the whole interval of possible quantile values and overall, we have a nested family of intervals – i.e., a fuzzy number. In Figure 7, we depict the corresponding membership functions for the 5%-quantile. We can see that the membership functions obtained by the simplified method – without taking into account interaction between parameters – are much wider and are, thus, less accurate than the ones obtained by using the full Bayes approach.

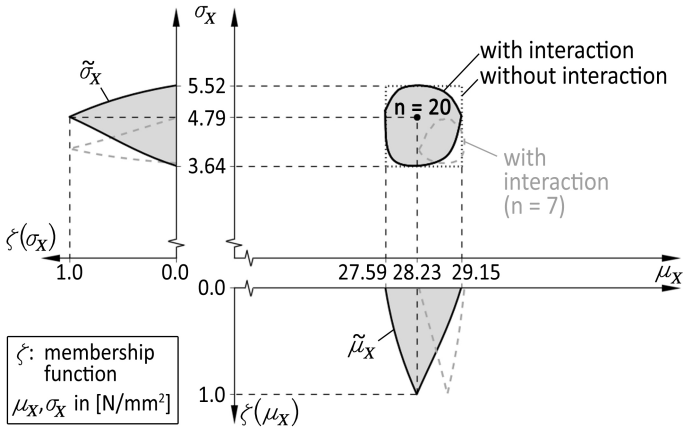

Figure 6: $\mu_X - \sigma_X$ interaction after Bayesian update with imprecise data for the 2-D uniform prior.

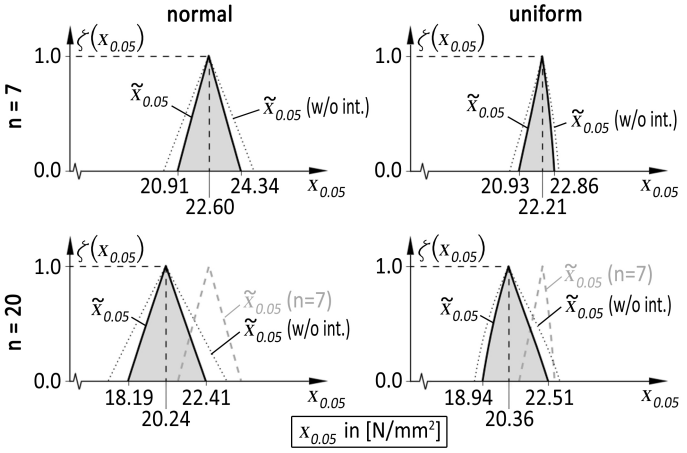

Figure 7: Fuzzy quantile $\tilde{x}_{0.05}$ of the fuzzy random variable $\tilde{X}$ based on imprecise data.

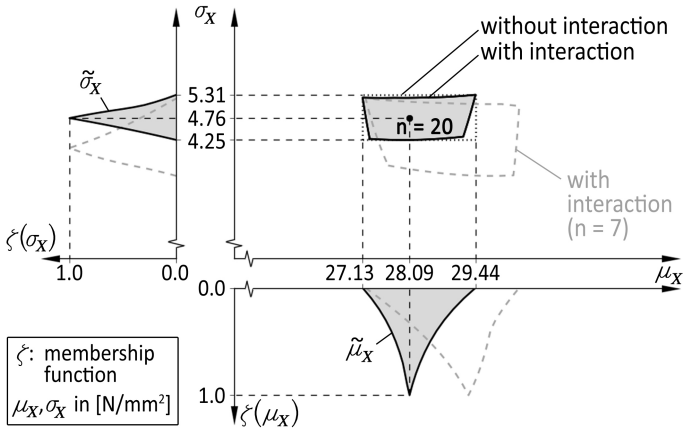

Figure 8: $\mu_X - \sigma_X$ interaction after Bayesian update with imprecise 2-D normal prior distribution function.

What if the prior distribution $g(\theta)$ is fuzzy. In the above computations, we assumed that the data are fuzzy, but the prior distribution is known exactly. Let us describe what happens if the data are exact but the prior distribution is fuzzy, i.e., has the form $g(\theta, b)$ for some fuzzy vector $b = (b_1, b_2)$. We will consider two cases: normal and uniform distributions. In both cases, the prior distribution for $\theta_1 = \mu_X$ is crisp, only the prior distribution for $\theta_2 = \sigma_X$ is fuzzy.

For the normal distribution, we assume that $g(\theta_2, b)$ is a normal distribution with mean b_1 and standard deviation b_2. In the crisp cases, we took $b_1 = 4.5$ and $b_2 = 1.0$. Here, we assume that b_1 is a triangular fuzzy number $b_1 = \langle 3.5, 4.5, 5.5\rangle$ and b_2 is a triangular fuzzy number $b_2 = \langle 0.75, 1.0, 1.5\rangle$.

For the uniform distribution, we assume that $g(\theta_2, b)$ is a uniform distribution on the interval $[b_1, b_2]$. In the crisp cases, we took $b_1 = 3.0$ and $b_2 = 6.0$. Here, we assume that b_1 is a triangular fuzzy number $b_1 = \langle 2.0, 3.0, 4.0\rangle$ and b_2 is a triangular fuzzy number $b_2 = \langle 5.0, 6.0, 7.0\rangle$.

For each α, different values $b_k \in \mathbf{b}_k(\alpha)$ lead, in general, to different values of θ_{p}. The corresponding ranges are described on Figure 8 for the normal case and on Figure 9 for the uniform case. Figure 10 describes the fuzzy quantiles for both cases.

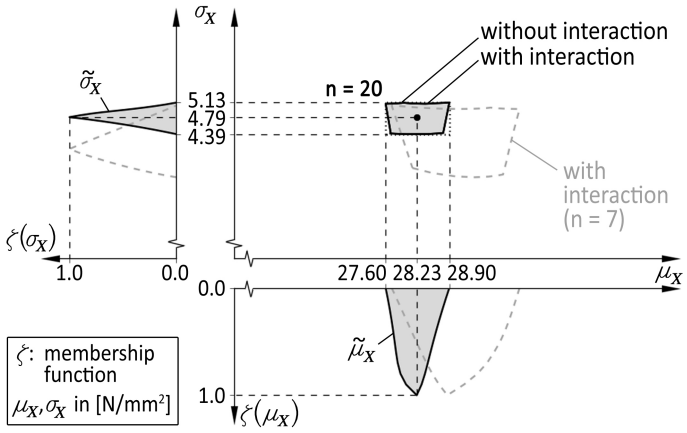

Figure 9: μ_X–σ_X interaction after Bayesian update with imprecise 2-D uniform prior distribution function.

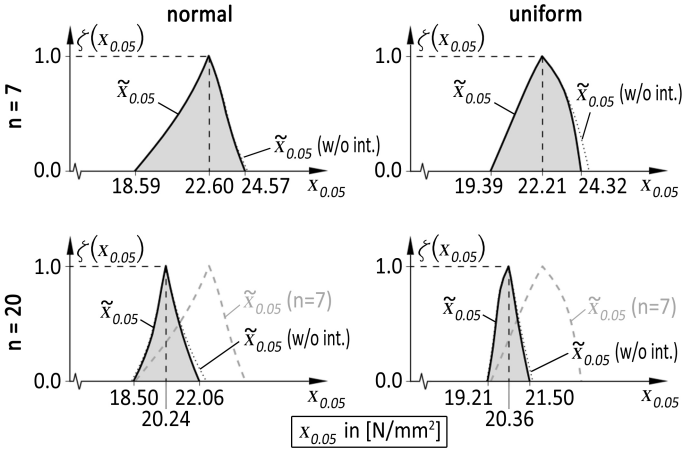

Figure 10: Fuzzy quantile $\tilde{x}_{0.05}$ of the fuzzy random variable $\tilde{X}$ based on imprecise prior distribution functions.

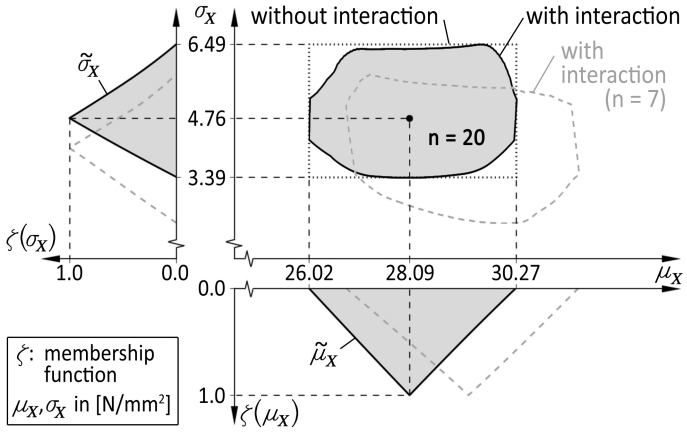

Figure 11: $\mu_X - \sigma_X$ interaction after Bayesian update with imprecise prior distribution function and imprecise data for the case of 2-D normal prior.

In general, the larger the sample, the smaller should be the influence of the prior distribution, i.e., the closer the estimates to the actual values of the corresponding parameters. And indeed, we see that as the sample size increases, the range of possible values of θ_{p} shrinks.

It is interesting to observe the difference between the shapes of the corresponding membership functions. In the uniform case, the membership functions are concave, and a small reduction of the imprecision of the prior distribution leads to a larger reduction in the imprecision of the results. In contrast, in the normal case, the membership functions are convex, so a reduction in imprecision in the prior $g(x)$ is less effective to reduce the imprecision of the estimation results.

What if both data and prior distributions are fuzzy. The results for this case are shown in Figures 11, 12, and 13. We observe a combination of the effects discussed for the individual cases. Whilst the effect of the prior imprecision decreases with increases sample size, the imprecision of the data becomes dominating. The result also converges towards the solution according to fuzzy statistics.

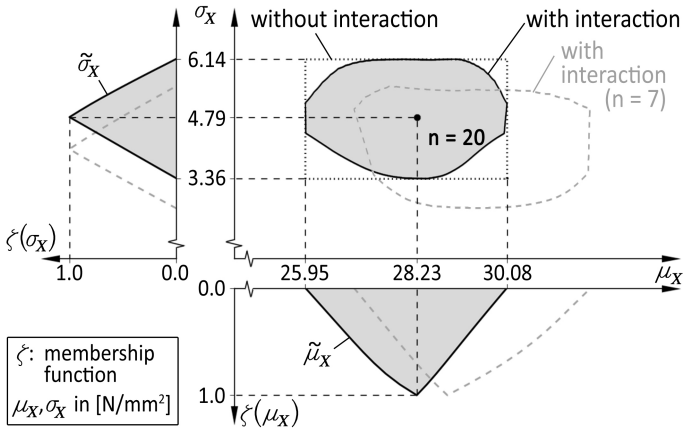

Figure 12: $\mu_X - \sigma_X$ interaction after Bayesian update with imprecise prior distribution function and imprecise data for case of 2-D uniform prior.

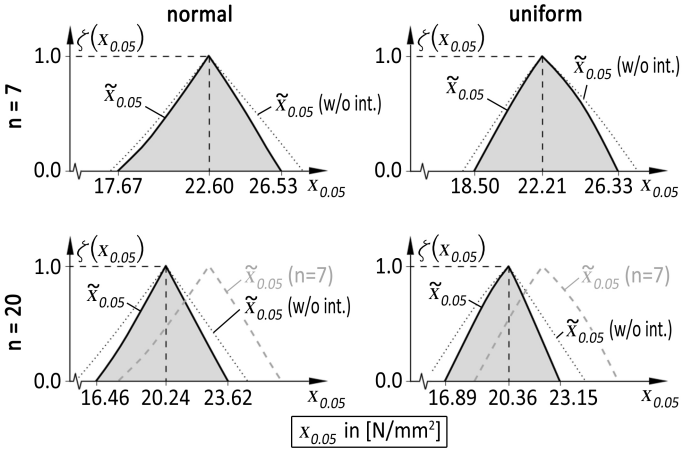

Figure 13: Fuzzy quantile $\tilde{x}_{0.05}$ of the fuzzy random variable $\tilde{X}$ based on imprecise prior distribution and imprecise data.

5. Conclusions and Remaining Open Problems

Conclusions. Inconsistent information represents a common problem in engineering practice. In case of inconsistent information, traditional Bayesian statistical methods – that only take into account probabilistic uncertainty – cannot be applied. In such situations, we also need to take into account imprecision with which we know the data, the prior distribution, and the likelihood functions. Once we take this imprecision into account – by considering fuzzy data, fuzzy prior distributions, and fuzzy likelihood functions – we can use fuzzy Bayesian updates and reasonable results.

In general, the Bayesian approach combines the prior information with the information contained in the data. The more data we add, i.e., the larger the data sample, the larger the role of the data and the smallest the influence of the prior. As a result, both the influence of the prior and the uncertainty of the resulting estimates decreases as the sample size increases. In the fuzzy Bayesian case, the uncertainty similarly decreases, as well as the imprecision caused by the imprecise prior. In contrast, the imprecision caused by the imprecise data does not decrease with the sample size.

First open problem: towards efficient computations. As we have mentioned earlier, the existing algorithms for the full Bayes update are often very computationally intensive. To speed up data processing, practitioners make simplifying assumptions. In the previous section, we have shown that one needs to be very cautious with such simplifications, since they can lead to inaccurate (and sometimes erroneous) results. This leaves us with the first important open problem: to come up with a computationally efficient implementation of the corresponding Bayesian techniques.

Second open problem: comparison with other approaches. Another important open problem is related to the fact that there are other approaches for dealing with inconsistent data (including the Dempster-Shafer approach); see, e.g., [2, 4, 7, 9]. It is desirable to compare different approaches.

Acknowledgements

The authors gratefully acknowledge the financial support by National University of Singapore through the Ministry of Education Academic Research Fund, Grant No. R246000234133, by the US National Science Foundation grants HRD-0734825 (Cyber-ShARE Center of Excellence) and DUE-0926721, by the US Grant 1 T36 GM078000-01 from the National Institutes of Health, and by a grant on from the US Office of Naval Research.

The authors are very thankful to the anonymous referees for valuable suggestions.

References

[1] M. Beer, Fuzzy probability theory, In: Encyclopedia of Complexity and Systems Science, Vol. 6, Springer, New York, 2009, pp. 4047–4059.

[2] M. Beer, Engineering quantification of inconsistent information, International Journal of Reliability and Safety 3 (1–3) (2009) 174–200.

[3] M. Beer, S. Ferson, Fuzzy probability in engineering analyses, in: B. Ayyub (Ed.), Proceedings of the First International Conference on Vulnerability and Risk Analysis and Management (ICVRAM 2011) and the Fifth International Symposium on Uncertainty Modeling and Analysis (ISUMA 2011), 11.-13. April 2011, University of Maryland, ASCE, Reston, VA, USA, 2011, pp. 53–61.

[4] M. A. Gil and O. Hryniewicz, Statistics with imprecise data. In: Encyclopedia of Complexity and Systems Science, Springer, Springer, New York, 2009, pp. 8679–8690.

[5] J. C. Helton and W. L. Oberkampf (eds) Special Issue on Alternative Representations of Epistemic Uncertainty, Reliability Engineering & System Safety, 85 (1–3) (2004) 1–369.

[6] G. J. Klir, Uncertainty and Information: Foundations of Generalized Information Theory, Wiley-Interscience, Hoboken, 2006.

[7] R. Kruse, K. Meyer, Statistics with Vague Data, Reidel, Dordrecht, 1987.

[8] B. Möller, M. Beer, Fuzzy Randomness – Uncertainty in Civil Engineering and Computational Mechanics, Springer, Berlin, 2004.

[9] R. Viertl, Statistical Methods for Non-Precise Data, CRC Press, Boca Raton New York London Tokyo, 1996.

[10] R. Viertl, Foundations of fuzzy bayesian inference, Journal of Uncertain Systems 2 (3) (2008) 187–191.

[11] R. Viertl, Statistical Methods for Fuzzy Data, Wiley, Chichester, 2011.

[12] R. Viertl, D. Hareter, Generalized Bayes' theorem for non-precise a-priori distribution, Metrika 59 (2004) 263–273.

[13] H.-J. Zimmermann, Fuzzy Set Theory and Its Applications, Kluwer Academic Publishers, Boston London, 1992.

AppendixA. A Heuristic Relation Between Fuzzy Degrees and Probabilities

As we have mentioned, fuzzy estimates for the data come from the fact that in addition to the guaranteed bound Δ on the possible value of measurement error, we also have smaller bounds $\Delta(\alpha) < \Delta$ which only hold with a certain degree of uncertainty α. If we have a large number of such measuring instruments, then, for each level α, we can ask the following natural question: how frequent are the cases for which the measurement error exceeds the corresponding bound $\Delta(\alpha)$.

To get an approximate heuristic answer to this question, following [3], let us assume that:

- the corresponding fuzzy number n has the simplest symmetric triangular shape, and
- the measurement error is – as is the case of many measuring instruments – normally distributed.

Theoretically, in a normal distribution with mean μ and standard deviation σ, all real values have a non-zero probability density and are, thus, possible. However, in practice, the probability of large deviations is so small that it can safely be neglected. Usually, some number $k_0 = 2$, 3, or 6 is selected, and it is assumed that all deviations larger than $k_0 \cdot \sigma$ are impossible. For example, the probability of deviations exceeding 3σ is $\approx 0.1\%$, the probability of deviations exceeding 6σ is $\approx 10^{-8}$, etc. Once we select the corresponding small probability η_0 – or, equivalently, the *credibility* $c_0 = 1 - \eta_0$, i.e., the probability with which the random variable is contained in the interval – we then cut off the tails with probability $\eta_0/2$ on both sides of the normal distribution and claim that the random variable is guaranteed to be within the resulting interval. The usual interpretation of the symmetric triangular number $n = \langle \ell, m, r \rangle$ is that the corresponding value is guaranteed to be within the interval $[\ell, r]$. Thus, it makes sense to identify this interval with the corresponding interval coming from the normal distribution; see Figure A.14.

For each uncertainty level α, we can now compute the probability c that the corresponding normally distributed random variable will be inside the corresponding α-cut $\mathbf{n}(\alpha)$. Then:

- the probability c is a good estimate of the frequency with which the actual measurement error is limited by $\Delta(\alpha)$, and
- the complement $\eta = 1 - c$ of this probability c is a good estimate for the frequency with which the actual measurement error exceeds this bound.

For $c_0 = 0.99$, the corresponding relation between α and c is described in Figure A.15.

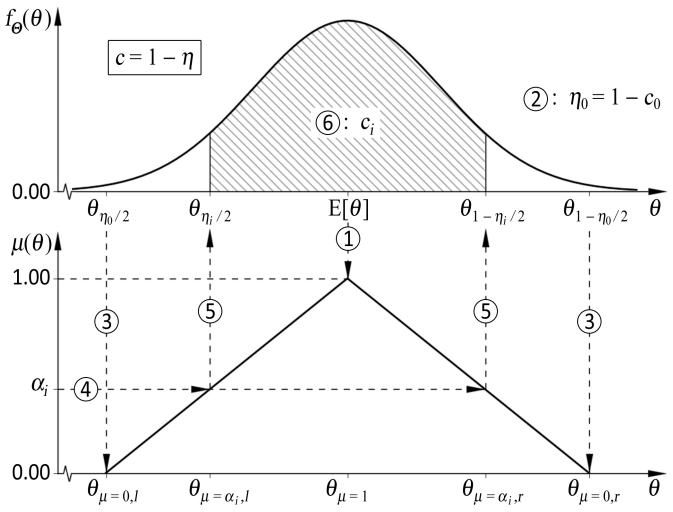

Figure A.14: Derivation of heuristic relationship between α-level and credibility level

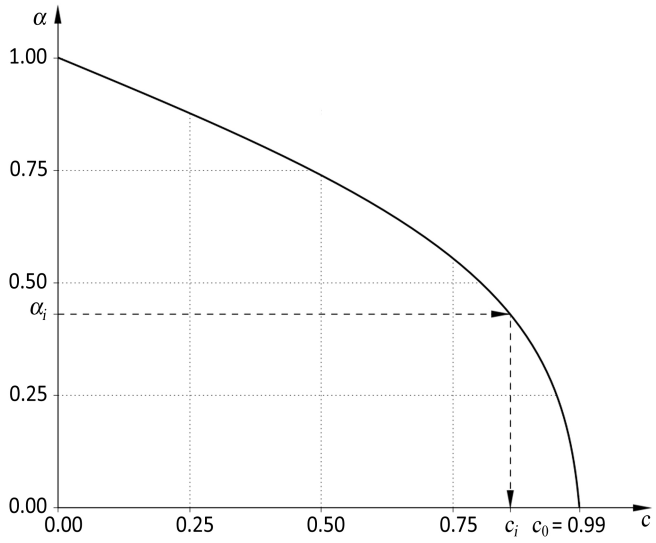

Figure A.15: Heuristic relationship between α-level and credibility level for $c_0 = 0.99$