

How to Define a Confidence Set for Functions: A New Justification of the Area Method

Vladik Kreinovich^{a*}, Gang Xiang^b, and Michael Oberguggenberger^c

^a*Department of Computer Science, University of Texas, El Paso, TX 79968, USA*

^b*Philips Healthcare, Business Line RIS, 4050 Rio Bravo, El Paso, TX 79912, USA*

^c*Unit for Engineering Mathematics, University of Innsbruck, 6020 Innsbruck, Austria*

(Received June 2010)

Due to uncertainty, in many problems, we only know the probability of different values. In such situations, we need to make decisions based on these probabilities: e.g., we must tell the user which values are possible and which are not. Often – e.g., for a normal distribution – the probability density is everywhere positive, so, theoretically, all real values are possible. In practice, it is usually safe to assume that values whose probability is very small are not possible. For a single variable, this idea is described by a confidence interval C , the interval for which the probability to be outside is smaller than a given threshold p_0 . In this way, if we know that a variable x is normally distributed with mean a and standard deviation σ , we can conclude that x is within the interval $C = [a - k \cdot \sigma, a + k \cdot \sigma]$, where k depends on p_0 (usually, $k = 2, 3$, or 6).

When a random object is a function $f(x)$, we similarly want to find a *confidence set* C of functions, i.e., the set for which the probability to be outside is smaller than p_0 . To find such a set, it is possible to use the following *area method*: they define the area $I(f)$ under the graph of f (i.e., in mathematical terms, an integral $I_0(f) = \int f(x) dx$), select a confidence interval for $I_0(f)$ and take, as C , the set of all the functions $f(x)$ for which $I_0(f)$ is within this interval.

At present, the area method is largely heuristic, with no justification explaining why exactly the integral functional $I(f)$ corresponding to the area should be used. In our paper, we provide a justification for the area method.

Keywords: confidence sets; area method; random functions

1. Confidence Sets and Area Method: Formulation of the Problem

Often, we have probabilistic uncertainty. In many practical situations, we must determine the values of certain parameters characterizing a given object. In some cases, these values can be determined directly, by measuring these values. In other situations, these values can only be determined indirectly, by measuring some auxiliary easier-to-measure quantities, and processing these measurement results to estimate the values of the desired quantities.

Measurement is never absolutely accurate, there is always measurement uncertainty. As a result, our estimates of the desired quantities are also uncertain.

Traditionally, in engineering practice, we assume that the sensors and measuring instruments used in the measurements are well calibrated and that, therefore, we know the probabilities of different values of measurement inaccuracy. In such situations, after the measurement and data processing, we know the probability of different values of the desired quantity; see, e.g., Rabinovich (2005).

*Corresponding author. Email: vladik@utep.edu

From probabilities to confidence intervals. In addition to knowing the probabilities of different values (and different combination of values), the user usually also wants to know which values (or combinations of possible) are possible.

For some probability distributions, the answer to this question is easy: e.g., if the probability distribution is located on an interval $[a, b]$, so that the probability to be outside this interval is 0, then we can conclude – with probability 1 – that the desired quantity x is located on this interval.

In other cases, the situation is not so easy. For many distributions, e.g., for a normal distribution, the probability density is everywhere positive. Theoretically, this means that all real values are possible – although the probability of some of the values is extremely small.

In practice, it is safe to assume that the occurrence of values whose probability is very small is impossible.

For a single variable, this idea is described by a *confidence interval* C , the interval for which the probability to be outside is smaller than a given threshold p_0 . For example, if we know that a variable x is normally distributed with mean a and standard deviation σ , we can conclude that x is within the interval

$$C = [a - k \cdot \sigma, a + k \cdot \sigma],$$

where k depends on p_0 ; usually, practitioners select $k = 2$, $k = 3$, or $k = 6$.

Comment. Note that we use the wording *confidence interval* in a broad statistical sense, going beyond its common usage in statistical practice; see, e.g., Kreinovich *et al.* (2002, 2006).

From confidence intervals to confidence sets. When we have several variables v_1, \dots, v_n , then we can also consider a confidence interval C_1, \dots, C_n for each of these variables, and then form a box

$$C_1 \times \dots \times C_n = \{v = (v_1, \dots, v_n) : v_1 \in C_1, \dots, v_n \in C_n\}$$

of all possible combinations of values $v_i \in C_i$. However, this box does not always provide a full picture of which combinations $v = (v_1, \dots, v_n)$ are possible and which are not. For example, if the random variables v_1 and v_2 are strongly correlated, then only the combinations for which $v_1 \approx v_2$ are possible. In this example, the set of all possible combinations is a proper subset of the above box.

We therefore need to describe a general confidence set. Moreover, we need to describe not just a single confidence set, but the family of confidence sets $C(p_0)$ corresponding to different values p_0 – probabilities to find the actual random vector outside the corresponding set $C(p_0)$.

Describing a family of confidence sets by a function. The value p_0 is the probability of a random vector not getting into the confidence set – i.e., equivalently, the probability of a random vector getting into the complement $-C(p_0)$. Thus, the smaller the value p_0 , the smaller the complement $-C(p_0)$ – and thus, the larger the confidence set $C(p_0)$.

Similarly to confidence intervals, it is therefore reasonable to require that if an element v belongs to the confidence set $C(p_0)$ for some $p_0 > 0$, then this element v also belongs to all the sets $C(p)$ with $p < p_0$. Under this requirement, for each element v , there is the smallest possible value $p(v)$ of the probability p_0 for which $v \in C(p_0)$. (To be more precise, there is an infimum, but if we consider closed

confidence sets that continually depend on p , then this infimum is indeed attained, i.e., we indeed have $v \in C(p(v))$.

Thus, instead of describing a family of sets, it is sufficient to describe a function $p(v)$ for which, for every p , the condition $v \in C(p)$ is equivalent to $p \leq p(v)$.

Sometimes, we have prior information about the probabilities. In some cases, we have some information about the probability distributions. For example, if the probability distribution on an n -dimensional space of points (v_1, \dots, v_n) is invariant with respect to arbitrary rotations in this space, then the probability density function $\rho(v_1, \dots, v_n)$ depends only on the distance $\|v\| = \sqrt{v_1^2 + \dots + v_n^2}$. In this case, it is reasonable to select confidence sets which have the same symmetry, i.e., sets which are also rotation-invariant. In terms of the function $p(v)$, this means that this function should also only depend on the distance $\|v\|$.

In other cases, instead of the distance, we may have other combinations of the original variables. Such cases can be summarized as follows:

- First, we select an appropriate function $I : \mathbb{R}^n \rightarrow \mathbb{R}$.
- Then, based on the given distribution, for each p_0 , we select confidence intervals $C_I(p_0)$ for the corresponding scalar random variable $I(v_1, \dots, v_n)$.
- Finally, as the desired confidence set $C(p_0)$, we take the set of all the values v for which $I(v) \in C_I(p_0)$:

$$C(p_0) \stackrel{\text{def}}{=} \{v = (v_1, \dots, v_n) : I(v) \in C_I(p_0)\}.$$

Comment. It should be noted that for this method, what is important are the sets corresponding to the function $I(v)$ and not the numerical values of the function itself. For example, if we *re-scale* the function $I(v)$ by taking $I'(v) = a + b \cdot I(v)$ or $I'(v) = (I(v))^3$, then the two functions $I(v)$ and $I'(v)$ generate the same family of “sub-level” sets $\{v : I(v) \leq i_0\}$ and thus, the same confidence sets $C(p_0)$.

It should be mentioned that this idea of constructing confidence sets is close to the concept of *potential clouds*, put forth in Fuchs and Neumaier (2009a,b).

Confidence sets for random functions: the area method. The same idea can be used when a random object is a function $f(x)$, e.g., a function from the class \mathcal{F} of all piece-wise continuous non-negative functions from real numbers to real numbers, with a compact support (i.e., equal to 0 outside a certain interval). In this case, we similarly want to find a *confidence set* C of functions, i.e., the set for which the probability to be outside is smaller than p_0 .

To find such a set, we:

- select an appropriate functional $I : \mathcal{F} \rightarrow \mathbb{R}$;
- based on the given distribution, for each p_0 , we select confidence intervals $C_I(p_0)$ for the corresponding scalar random variable $I(f)$; and
- as the desired confidence set $C(p_0)$, we take the set of all the functions f for which $I(f) \in C_I(p_0)$:

$$C(p_0) \stackrel{\text{def}}{=} \{f : I(f) \in C_I(p_0)\}.$$

In particular, Fellin and Oberguggenberger (2010) proposed to use the *area method* – the above method in which, as the functional $I(f)$, we take the area $I_0(f)$ under the graph of f – i.e., in mathematical terms, the integral $I_0(f) = \int f(x) dx$.

Example. The paper Fellin and Oberguggenberger (2010) considers the following *geotechnical* problem. Buildings are built on the soil. In addition to buildings, we also need to design dams, irrigation canals, bridges, roads, etc. All these structures cause stress in the surrounding soil. This stress can be divided into two components:

- *shear* stress τ – that is applied parallel or tangential to a face of the material, and
- *normal* stress σ – which is applied perpendicularly.

For each given value σ of the normal stress, there is a critical threshold τ_c of the shear stress that causes the soil to fail (i.e., to start developing fractures and/or other damage). In general, this critical threshold τ_c depends on the normal stress: $\tau_c = \tau_c(\sigma)$. In other words, how much shear stress a soil can sustain depends on how much normal stress is applied. Usually, the critical threshold τ_c increases with σ : indeed,

- If we apply a horizontal (shear) stress to a free-standing soil (with no vertical pressure), at some point, parts of the soil will start moving horizontally – i.e., the soil will fail.
- However, when the soil is weighed down by a normal stress, it is not as easy to move it horizontally, so the critical threshold for shear stress is larger.

When constructing structures, it is therefore important to know the dependence $\tau_c(\sigma)$ corresponding to the surrounding soil. This dependence has to be determined experimentally. For that, we select a family of functions $\tau_c = f(\sigma, c_1, \dots)$ which is sufficiently general to describe all possible observed dependencies. In the first approximation, we can assume that the dependence is linear: $\tau_c = c_1 + c_2 \cdot \sigma$; for more accurate estimations, we need to take into account that this dependence is usually non-linear, and so, more parameters c_i are needed.

To find the desired dependence, for several fixed values $\sigma_1, \dots, \sigma_n$ of the shear stress, we measure the corresponding values τ_1, \dots, τ_n of the critical shear stress. Then, we determine the values c_i from the condition that $\tau_j \approx f(\sigma_j, c_1, \dots)$ for all measurements $j = 1, \dots, n$. Substituting the resulting “nominal” values \tilde{c}_i into the general formula, we get the *nominal* dependence $\tau_c = f(\sigma, \tilde{c}_1, \dots)$.

Since measurements come with sizable measurement errors, the values c_i (and the resulting dependence $\tau_c = f(\sigma, c_1, \dots)$) are only determined approximately. In other words, the actual dependence $\tau_c(\sigma)$ may be somewhat different from the one that we obtained based on the statistical point estimates \tilde{c}_i for c_i . We want our structure to survive in the real soil, so we must make sure that it is stable not only for the nominal dependence, but also for all statistically possible dependencies – i.e., for all dependencies which are consistent with the our soil-testing measurement results.

Of course, in general, under a usual assumption that the measurement errors are normally distributed, arbitrary large measurement errors are possible – since the probability density of a normal distribution is always positive. As a result, in principle, huge deviations from the nominal dependence are possible – but they have a very small (negligibly small) probability, since for a normal distribution, probability of getting farther than, say, six standard deviations from the mean is negligibly small ($< 10^{-8}$). Thus, similarly to normal statistical situations, to describe realistically possible dependencies, we select a small threshold probability p_0 and dismiss the dependencies which are deviating too much from the nominal dependency – provided that the probability of such large deviations does not exceed p_0 .

In geotechnical engineering, the normal stress σ is usually reasonably well determined – by the weight of the structure and the area to which this weight is applied. Thus, we are interested in the dependence $\tau_c(\sigma)$ in a reasonably narrow interval $[\underline{\sigma}, \bar{\sigma}]$ of possible values of σ . Within this narrow interval, as a first approximation, it makes sense to approximate the *actual* value of the critical shear stress (corresponding to the actual value σ) by *average* value over the interval:

$\frac{1}{\bar{\sigma} - \underline{\sigma}} \cdot \int_{\underline{\sigma}}^{\bar{\sigma}} \tau_c(\sigma) d\sigma$. Thus, as a confidence set, we can select the set of all dependencies in which this average shear stress lies within some interval $[\underline{c}(p_0), \bar{c}(p_0)]$ (depending on the probability p_0):

$$\underline{c}(p_0) \leq \frac{1}{\bar{\sigma} - \underline{\sigma}} \cdot \int_{\underline{\sigma}}^{\bar{\sigma}} \tau_c(\sigma) d\sigma \leq \bar{c}(p_0).$$

For computational purposes, this condition can be simplified if we multiply all three sides of this double inequality by $\bar{\sigma} - \underline{\sigma}$; then, we get the equivalent condition

$$\underline{C}_I(p_0) \leq \int_{\underline{\sigma}}^{\bar{\sigma}} \tau_c(\sigma) d\sigma \leq \bar{C}_I(p_0),$$

where $\underline{C}_I(p_0) \stackrel{\text{def}}{=} (\bar{\sigma} - \underline{\sigma}) \cdot \underline{c}(p_0)$ and $\bar{C}_I(p_0) \stackrel{\text{def}}{=} (\bar{\sigma} - \underline{\sigma}) \cdot \bar{c}(p_0)$. This is exactly the area method.

What we do in this paper. The paper Fellin and Oberguggenberger (2010) addresses a geotechnical application, in which the area method arose from a specific, heuristic geotechnical motivation.

The goal of the present paper is to put this method in a wider context, and thus, to provide a justification for using the integral functional $I(f)$ as a general building block for confidence sets of random functions.

2. Towards a Justification of the Area Method

By \mathcal{F} , we will denote the set of all piece-wise continuous non-negative functions from real numbers to real numbers, with a compact support (i.e., which are equal to 0 outside some interval).

Definition 2.1: By a *confidence functional* $I(f)$, we mean a mapping from the set \mathcal{F} to real numbers.

First property: monotonicity. For a standard normal distribution, when we form a confidence interval, we exclude values which are larger than a certain threshold because these values have low probability. In other words, we assume that if a value x is too large to be included in the confidence interval, then all larger values are also too large to be included in this set.

It is reasonable to require a similar property for functions: if for some function $f(x)$, the value $I(f)$ is larger than or equal to a threshold t , and $g(x) \geq f(x)$ for all x , then we should have $I(g) \geq t$.

In particular, for $t = I(f)$, we should have $I(g) \geq I(f)$.

Thus, we arrive at the following definition.

Definition 2.2: We say that a confidence functional I is *monotonic* if for every two functions $f(x)$ and $g(x)$ for which $f(x) \leq g(x)$ for all x , we have

$$I(f) \leq I(g).$$

Second property: continuity. In practice, we get the values of the function $f(x)$ from measurements. Measurements are never absolutely accurate, they always provide only an approximation to the actual value of the corresponding quantity. To be more precise, instead of the exact values $f(x)$, we get the approximate values $\tilde{f}(x)$ for which $|\tilde{f}(x) - f(x)| \leq \varepsilon$, where $\varepsilon > 0$ is a known upper bound on the measurement inaccuracy.

It is reasonable to require that if we consider functions defined on the same interval $[a, b]$, and we use more and more accurate measurements ($\varepsilon \rightarrow 0$), then our estimates of the quality $I(f)$ should also become more and more accurate. In other words, we would like the functional $I(f)$ to be continuous in the following sense:

Definition 2.3: We say that a confidence functional is *continuous* if for any interval $[a, b]$ and for any functions f_n and f whose compact supports are included in this interval $[a, b]$, if $\sup |f_n(x) - f(x)| \rightarrow 0$, then $I(f_n) \rightarrow I(f)$.

Third property: shift-invariance in signal space. In some practical situations, we may have several different ways to describe the real-life process by a function $f(x)$.

For example, when we measure a seismic signal generated by a certain event, we can take, as the value $f(x)$, the actual signal at the moment x .

Alternatively, if we know that the signal always exceeds a certain minimal signal $f_0(x)$, then it is also reasonable to characterize a signal by the “excess” value, i.e., by the difference $f_{\text{alt}}(x) \stackrel{\text{def}}{=} f(x) - f_0(x)$ between the original signal $f(x)$ and the minimal signal $f_0(x)$.

It is reasonable to require that if the two signals $f(x)$ and $g(x)$ are “equivalent” in the sense that we have $I(f) = I(g)$, then the corresponding excess signals should also be equivalent, i.e., we should have $I(f - f_0) = I(g - f_0)$. In other words, if we “shift” two equivalent functions f and g by the same “shift function” $f_0(x)$, the resulting shifted functions remain equivalent.

Similarly, if the functions $f_{\text{alt}}(x) = f(x) - f_0(x)$ and $g_{\text{alt}}(x) = g(x) - f_0(x)$ are equivalent, then the functions $f(x) = f_{\text{alt}}(x) + f_0(x)$ and $g(x) = g_{\text{alt}}(x) + f_0(x)$ should also be equivalent. Thus, we arrive at the following definition:

Definition 2.4: We say that a confidence functional I is *shift-invariant in the signal space* if for all functions $f(x)$, $g(x)$, and $f_0(x)$, the equality $I(f) = I(g)$ implies that $I(f + f_0) = I(g + f_0)$.

Fourth property: shift-invariance in the input space. The value of the input x – e.g., for signal, of time – depends on the choice of the starting point for measuring time. In general, this choice is rather arbitrary, so the quality of a function $f(x)$ should not change if we simply change this starting point.

If we change the starting point to a new one which is x_0 seconds later, then we get a new function $f_{\text{new}}(x)$ describing the signal at the moment which is x second after the new starting point. Since the new starting point is x_0 seconds after the original one, this same moment of time corresponds to moment $x + x_0$ on the original time scale. At that moment of time, the value of the signal was $f(x + x_0)$. Thus, $f_{\text{new}}(x) = f(x + x_0)$. So, changing the starting point replaces the original signal $f(x)$ with a new signal $f_{\text{new}} = S_{x_0}(f)$ for which $f_{\text{new}}(x) = f(x + x_0)$. Thus,

the requirement that the quality of a signal does not change under this shift takes the following form:

Definition 2.5: We say that a confidence functional I is *shift-invariant in the input space* if for every function $f(x)$ and for every number x_0 , we have

$$I(f(x + x_0)) = I(f).$$

Definition 2.6: We say that a confidence functional I is *shift-invariant* if it is shift-invariant both in the signal space and in the input space.

Observation. One can easily check:

- that the area functional $I_0(f) = \int f(x) dx$ is monotonic, continuous, and shift-invariant, and
- that for every monotonic, continuous, and shift-invariant functional $I(f)$ and for every non-decreasing continuous function $F(z)$, the “re-scaled” functional $F(I(f))$ is also monotonic, continuous, and shift-invariant.

In particular, this means that all the functionals of the type $F(I_0(f))$ are monotonic, continuous and shift-invariant. It turns out that all monotonic shift-invariant functionals have this form:

Proposition 2.7: *Every monotonic continuous shift-invariant confidence functional has the form $I(f) = F(I_0(f))$ for some non-decreasing function $F(z)$.*

Comments.

- In the language of (Wang and Klir 2009, Section 4.4), $I = F^{-1}(I_0)$ is a quasi-measure with T -function F^{-1} and corresponding measure I_0 , the Lebesgue measure.
- As we have already mentioned, the actual value of the confidence functional $I(f)$ is irrelevant, so this functional is actually defined modulo an arbitrary monotonic transformation. More precisely, we can put the burden of scaling on the choice of confidence intervals $C_I(p_0)$ as defined in Section 1. Taking this into consideration, we can thus simply state that every confidence functional which is monotonic and shift-invariant *is* the area functional. Thus, we get the desired justification of the area functional.

3. Proof

1°. Let us first prove that if f is equivalent to g (in the sense of $I(f) = I(g)$), and f' is equivalent to g' (in the sense that $I(f') = I(g')$), then $f + f'$ is equivalent to $g + g'$, i.e., $I(f + f') = I(g + g')$.

Indeed, due to shift-invariance in the signal space, $I(f) = I(g)$ implies that $I(f + f') = I(g + f')$. Similarly, $I(f') = I(g')$ implies that $I(g + f') = I(g + g')$. Thus, $I(f + f') = I(g + f') = I(g + g')$, and $I(f + f') = I(g + g')$.

2°. We will follow the standard notation $\chi_{[a,b)}(x)$ for the characteristic function of the interval $[a, b)$, i.e., for the function that takes value 1 for $x \in [a, b)$ and 0 for all other x .

3°. The characteristic function notation enables us to explicitly describe the function $F(z)$ for which we will prove that $I(f) = F(I_0(f))$: namely, for every $z \geq 0$,

we take

$$F(z) \stackrel{\text{def}}{=} I(z \cdot \chi_{[0,1]}(x)).$$

When $z_1 \leq z_2$, we have $z_1 \cdot \chi_{[0,1]}(x) \leq z_2 \cdot \chi_{[0,1]}(x)$ for all x . Thus, since the functional I is monotonic, we conclude that

$$F(z_1) = I(z_1 \cdot \chi_{[0,1]}(x)) \leq I(z_2 \cdot \chi_{[0,1]}(x)) = F(z_2),$$

i.e., $F(z_1) \leq F(z_2)$. Thus, the above-defined function $F(z)$ is indeed non-decreasing.

Similarly, the continuity of the functional $I(f)$ implies that the above function $F(z)$ is continuous.

4°. Let us now prove that for every positive integer q and for every positive real number h , we have

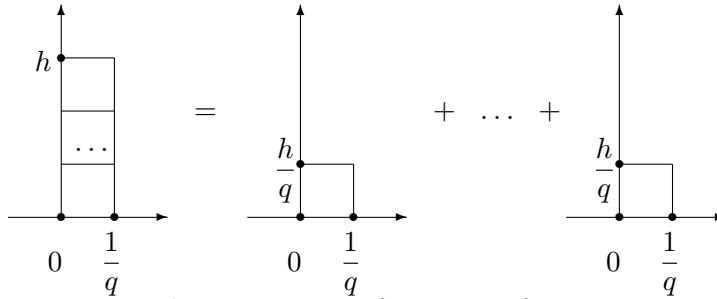
$$I(h \cdot \chi_{[0,1/q]}(x)) = F\left(\frac{h}{q}\right).$$

Indeed, the function $f(x) = h \cdot \chi_{[0,1/q]}(x)$ can be represented as the sum of q identical q times smaller functions located on the same interval $[0, 1/q]$:

$$f(x) = f_1(x) + \dots + f_q(x),$$

where

$$f_1(x) = \dots = f_q(x) = \frac{h}{q} \cdot \chi_{[0,1/q]}(x).$$



By shift-invariance in the input space, for every i from 1 to q , we have $I(f_i) = I(g_i)$, where

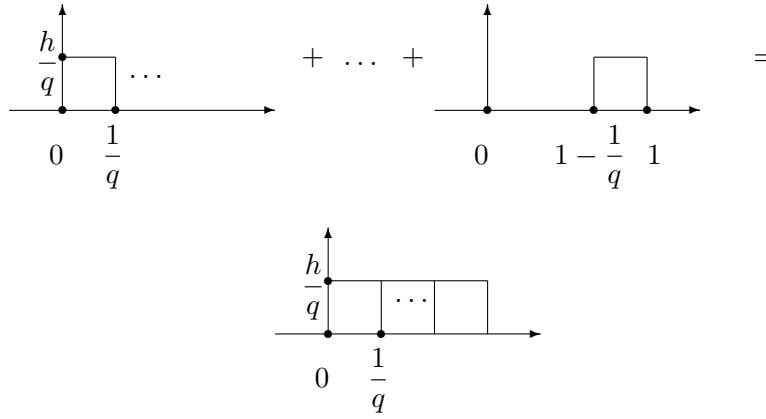
$$g_i(x) \stackrel{\text{def}}{=} f_i\left(x - \frac{i-1}{q}\right) = \chi_{[(i-1)/q, i/q]}(x).$$

Thus, due to Part 1 of this proof, we conclude that

$$I(f) = I(f_1 + \dots + f_q) = I(g_1 + \dots + g_q).$$

However, one can easily see that for every x , we have

$$g_1(x) + \dots + g_q(x) = \frac{h}{q} \cdot \chi_{[0,1]}(x) :$$



Thus, indeed,

$$I(f) = I\left(\frac{h}{q} \cdot \chi_{[0,1)}(x)\right) = F\left(\frac{h}{q}\right).$$

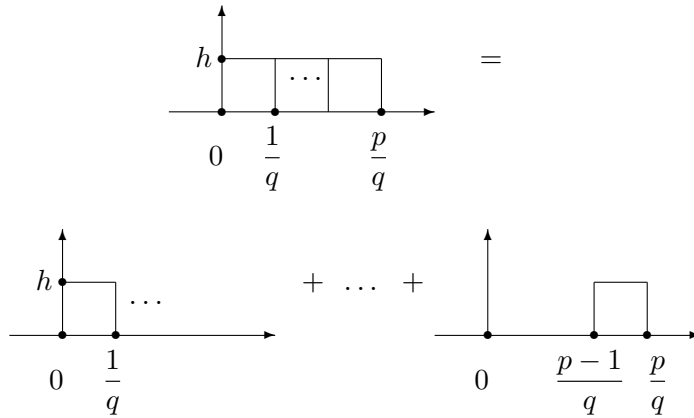
The statement is proven.

5°. Let us now prove that for every two positive integers p and q and for every positive real number h , we have

$$I(h \cdot \chi_{[0,p/q)}(x)) = F\left(h \cdot \frac{p}{q}\right).$$

Indeed, similar to Part 4, we can represent the function $f(x) \stackrel{\text{def}}{=} h \cdot \chi_{[0,p/q)}(x)$ as a sum $f(x) = f_1(x) + \dots + f_p(x)$, where

$$f_i(x) \stackrel{\text{def}}{=} h \cdot \chi_{[(i-1)/q, i/q)}(x) :$$



By shift-invariance in the input space, for every i from 1 to q , we have $I(f_i) = I(g_i)$, where

$$g_i(x) \stackrel{\text{def}}{=} f_i\left(x + \frac{i-1}{q}\right) = h \cdot \chi_{[0,1/q)}(x).$$

Thus, due to Part 1 of this proof, we conclude that

$$I(f) = I(f_1 + \dots + f_p) = I(g_1 + \dots + g_p).$$

However, one can easily see that for every x , we have

$$g_1(x) + \dots + g_p(x) = p \cdot h \cdot \chi_{[0,1/q)}(x).$$

We already know, from Part 4, that

$$I(p \cdot h \cdot \chi_{[0,1/q)}(x)) = F\left(\frac{p \cdot h}{q}\right).$$

Thus, indeed,

$$I(f) = I(g_1 + \dots + g_p) = F\left(h \cdot \frac{p}{q}\right).$$

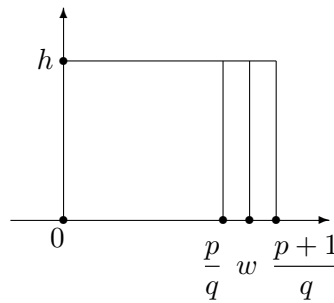
The statement is proven.

6°. Let us now prove that for every two positive real numbers $w > 0$ and $h > 0$, we have

$$I(h \cdot \chi_{[0,w)}(x)) = F(h \cdot w).$$

For rational values $w = \frac{p}{q}$, this statement was proven in Part 5. An arbitrary positive real number, for every integer q , lies between $\frac{p}{q}$ and $\frac{p+1}{q}$ for some integer $p = \lfloor q \cdot w \rfloor$. From $\frac{p}{q} \leq w \leq \frac{p+1}{q}$, one can easily conclude that for every x , we have

$$h \cdot \chi_{[0,p/q)}(x) \leq h \cdot \chi_{[0,w)}(x) \leq h \cdot \chi_{[0,(p+1)/q)}(x) :$$



Thus, due to monotonicity,

$$I(h \cdot \chi_{[0,p/q)}(x)) \leq I(h \cdot \chi_{[0,w)}(x)) \leq h \cdot I(\chi_{[0,(p+1)/q)}(x)).$$

For the left-most and right-most terms in this inequality, we already know the expressions, from Part 5. Substituting these expressions into the above inequality,

we conclude that

$$F\left(h \cdot \frac{p}{q}\right) \leq I\left(h \cdot \chi_{[0,w)}(x)\right) \leq F\left(h \cdot \frac{p+1}{q}\right).$$

When $q \rightarrow \infty$, we have $\frac{p}{q} \rightarrow w$ and $\frac{p+1}{q} \rightarrow w$, so $h \cdot \frac{p}{q} \rightarrow h \cdot w$ and $h \cdot \frac{p+1}{q} \rightarrow h \cdot w$. Since the function $F(z)$ is continuous, we therefore conclude, in the limit, that

$$I\left(h \cdot \chi_{[0,w)}(x)\right) = F(h \cdot w).$$

The statement is proven.

7°. Let us now prove that for every positive real number $h > 0$, and for every interval $\mathbf{x} = [\underline{x}, \bar{x})$, we have

$$I(h \cdot \chi_{\mathbf{x}}(x)) = F(h \cdot \Delta x),$$

where $\Delta x \stackrel{\text{def}}{=} \bar{x} - \underline{x}$ is the width of the interval \mathbf{x} .

This follows from Part 6 and from the shift-invariance in the input space. Indeed, if we shift the interval $[\underline{x}, \bar{x})$ to the left by \underline{x} , we get the interval $[0, \Delta x)$, so $I(h \cdot \chi_{\mathbf{x}}(x)) = I\left(h \cdot \chi_{[0, \Delta x)}(x)\right)$. From Part 5, we already know that $I\left(h \cdot \chi_{[0, \Delta x)}(x)\right) = F(h \cdot \Delta x)$. Thus, we have $I(h \cdot \chi_{\mathbf{x}}(x)) = F(h \cdot \Delta x)$. The statement is proven.

8°. Let us now show that for every piece-wise constant function that takes values $f(x_i)$ ($i = 1, \dots, n$) on intervals \mathbf{x}_i of width Δx_i , we have

$$I(f) = F\left(\sum_{i=1}^n f(x_i) \cdot \Delta x_i\right),$$

i.e., $I(f) = F\left(\int f(x) dx\right) = F(I_0(f))$.

Indeed, we can represent the function $f(x)$ as the sum

$$f(x) = f_1(x) + \dots + f_n(x)$$

of n functions $f_i(x)$ each of which attains the value $f(x_i)$ on the corresponding interval of width Δx_i and is equal to 0 outside the i -th interval. Due to Part 7, for each of these functions,

$$I(f_i) = F(f_i \cdot \Delta x_i) = I\left(f(x_i) \cdot \Delta x_i \cdot \chi_{[0,1]}(x)\right),$$

i.e., $I(f_i) = I(g_i)$, where we denoted

$$g_i(x) \stackrel{\text{def}}{=} f(x_i) \cdot \Delta x_i \cdot \chi_{[0,1]}(x).$$

Thus, due to Part 1, we have

$$I(f) = I(f_1 + \dots + f_n) = I(g),$$

where

$$g(x) \stackrel{\text{def}}{=} g_1(x) + \dots + g_n(x) =$$

$$f(x_1) \cdot \Delta x_1 \cdot \chi_{[0,1]}(x) + \dots + f(x_n) \cdot \Delta x_n \cdot \chi_{[0,1]}(x) =$$

$$\left(\sum_{i=1}^n f(x_i) \cdot \Delta x_i \right) \cdot \chi_{[0,1]}(x).$$

Therefore, by definition of the function $F(z)$, we conclude that

$$I(f) = I(g) = I \left(\left(\sum_{i=1}^n f(x_i) \cdot \Delta x_i \right) \cdot \chi_{[0,1]}(x) \right) = F \left(\sum_{i=1}^n f(x_i) \cdot \Delta x_i \right).$$

The statement is proven.

9°. Now, the integral of a piece-wise continuous function $f(x)$ can be obtained as a limit of integral sums for piece-wise constant approximations $\underline{f}_n(x)$ of $f(x)$ from below – and as a limit of similar approximation $\bar{f}_n(x)$ from above.

Due to monotonicity, for every n , we have $I(\underline{f}_n) \leq I(f) \leq I(\bar{f}_n)$. According to Part 8, we have $I(\underline{f}_n) = F(I_0(\underline{f}_n))$ and $I(\bar{f}_n) = F(I_0(\bar{f}_n))$, so

$$F(I_0(\underline{f}_n)) \leq I(f) \leq F(I_0(\bar{f}_n)).$$

When $n \rightarrow \infty$, we have $I_0(\underline{f}_n) \rightarrow I_0(f)$ and $I_0(\bar{f}_n) \rightarrow I_0(f)$. Thus, due to the continuity of the function $F(z)$, we get $F(I_0(\underline{f}_n)) \rightarrow F(I_0(f))$ and $F(I_0(\bar{f}_n)) \rightarrow F(I_0(f))$. Hence, in the limit, the above inequalities lead to the desired equality $I(f) = F(I_0(f))$.

The proposition is proven.

Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and DUE-0926721 and by Grant 1 T36 GM078000-01 from the National Institutes of Health. The authors are thankful to the organizers and participants of the 4th International Workshop on Reliable Engineering Computing REC'2010 (Singapore, March 3–5, 2010) for valuable discussions, and to the anonymous referees for their help.

References

Fellin, W., and Oberguggenberger, M., 2010. Robust assessment of shear parameters in geotechnics”, In: M. Beer, R.L. Muhamma, and R.L. Mullen, eds., *Proceedings of the 4th International Workshop on Reliable Engineering Computing REC'2010*, Singapore, March 3–5, 2010, 229–240.

- Fuchs, M. and Neumaier, A., 2009a. Autonomous robust design optimisation with potential clouds, *International Journal of Reliability and Safety*, 3, 23–34.
- Fuchs, M. and Neumaier, A., 2009b. Potential based clouds in robust design optimization, *Journal of Statistical Theory and Practice*, 3, 225–238.
- Kreinovich, V., Nguyen, H.T., Ferson, S., and Ginzburg, L., 2002. From computation with guaranteed intervals to computation with confidence intervals: A new application of fuzzy techniques. *Proceedings of the Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2002*, 418–423.
- Kreinovich, V., Xiang, G., Starks, S.A., Longpré, L., Ceberio, M., Araiza, R., Beck, J., Kandathi, R., Nayak, A., Torres, R., and Hajagos, J.G., 2006. Towards combining probabilistic and interval uncertainty in engineering calculations: algorithms for computing statistics under interval uncertainty, and their computational complexity, *Reliable Computing*, 12, 471–501.
- Rabinovich, S., 2005. *Measurement Errors and Uncertainties: Theory and Practice*. New York: American Institute of Physics.
- Sheskin, D.J., 2004. *Handbook of Parametric and Nonparametric Statistical Procedures*. Boca Raton, Florida: Chapman & Hall/CRC.
- Wang, Z., and Klir, G.J., 2009. *Generalized Measure Theory*. New York: Springer.