

## Enhancement of Natural Extension

**Igor Kozine**  
Risø National Laboratory  
Denmark  
igor.kozine@risoe.dk

**Victor Krymsky**  
Ufa State Aviation Technical University  
Russia  
kvg@mail.rb.ru

### Abstract

The theory of imprecise previsions admits the use of a wide variety of statistical evidence. Nevertheless, some existing evidence, for example, in reliability applications, cannot be utilized by models developed within its framework. In the pursuit of reducing imprecision, any available evidence should become an input to modeling. It is suggested to take a different look at the natural extension, the basic constructive step in the theory. It is shown that natural extension can be viewed as a problem belonging to the realm of variational calculus, which opens up new perspectives for obtaining tighter intervals.

**Keywords.** Imprecise probability, statistical reasoning, natural extension, variational calculus, reliability analysis

### 1 Introduction

In spite of the existence of a number of risk/reliability and other applied models built on imprecise statistical reasoning, only a few of them have ever been used in practice – and then only hesitantly –, the rest remaining firmly in the academic realm. Do they lack adequate promotion by their practitioners, or are there other primary obstacles that prevent them from being widely applied? We believe that the main obstacle to the practical application of imprecise statistical models is thoroughly familiar to the group of experts who practise interval computations: it is namely the rapid growth in imprecision that occurs when intervals are propagated through mathematical models.

Should this state of affairs be regarded as unalterable, or can this weakness in the model be remedied? If the growth in imprecision is due to a deficiency in the model, what is its basic cause in mathematical terms, and how can we attempt to develop a more adequate model?

A cause of the large imprecision in computed previsions should be sought in the mechanism producing the previsions. It is called natural extension and it may be seen as the basic constructive step in statistical reasoning;

it enables us to construct new coherent previsions from old ones [1].

Natural extension can appear in different forms. Four forms of it were described in [2]. Each of them has pros and cons in the context of a specific application. The use of a proper form can substantially facilitate inference and computation of the probability measures of interest.

We suggest taking a different look at the natural extension, an approach which opens up new perspectives for obtaining tighter intervals.

It is shown that natural extension can be viewed as the problem of finding an extremal of a functional, a problem which belongs to the realm of variational calculus. If this path is followed, the modeller can utilise more versatile information than is possible with the natural extension suggested by Walley [1] and Kuznetsov [3]. For example, as demonstrated in this paper, bounds on probability density functions and their derivatives can be utilised by the new form of natural extension, which is an effective way of obtaining tighter bounds of statistical measures.

### 2 Different Forms of Natural Extension

Suppose there is a continuous random variable, for example, a lifetime  $X$  of a component or system defined on the sample space  $[0, T]$  and information about this variable is represented as a set of  $n$  interval-valued expectations of functions  $f_1(X), \dots, f_n(X)$ . Denote these expectations  $\bar{a}_i = \overline{M}(f_i)$  and  $\underline{a}_i = \underline{M}(f_i)$ ,  $i = 1, \dots, n$ , where  $\bar{a}_i$  and  $\underline{a}_i$  upper and lower bounds for the expectations, correspondingly. For computing new expected values  $\overline{M}(g)$  and  $\underline{M}(g)$  of a function  $g(X)$  from the available information, natural extension can be used in the following primal form:

$$\left. \begin{aligned} \underline{M}(g) &= \inf_P \int_0^T g(x) \rho(x) dx \\ \overline{M}(g) &= \sup_P \int_0^T g(x) \rho(x) dx \end{aligned} \right\} \quad (1)$$

subject to

$$\left. \begin{aligned} \underline{a}_i &\leq \int_0^T f_i(x) \rho(x) dx \leq \overline{a}_i, i \leq n \\ \int_0^T \rho(x) dx &= 1, \rho(x) \geq 0 \end{aligned} \right\} \quad (2)$$

Here the infimum and supremum are taken over the set  $P$  of all admissible (matching the constraints) probability density functions  $\rho(x)$  satisfying conditions (2). Solutions (1) exist if all the constraints (2) form a non-empty subset  $P_0 \subseteq P$ . If the subset  $P_0$  is empty, this means that the set of evidence is conflicting. If all the evidence is interval-valued (this is a particular case of imprecise evidence), then two interval-valued judgements on the same prevision are called conflicting if they do not intersect.

It should be noted that problems (1)-(2) are linear and the dual optimization problems can be written for them. For  $\underline{M}(g)$ , for example, the dual problem is the following [2], [3]:

$$\underline{M}(g) = \sup_{c_0, c_i, d_i} \left( c_0 + \sum_{i=1}^n (c_i \underline{a}_i - d_i \overline{a}_i) \right)$$

subject to  $c_0 \in \mathbf{R}$ ,  $c_i, d_i \in \mathbf{R}_+$ ,  $i = 1, \dots, n$ , and for any  $0 \leq x \leq T$ ,  $c_0 + \sum (c_i - d_i) f_i(x) \geq g(x)$

Values  $\overline{M}(g)$  and  $\underline{M}(g)$  are often called upper and lower previsions and functions  $f_i(X)$  and  $g(X)$  are called gambles. Note that the lower and upper previsions  $\overline{M}(g)$  and  $\underline{M}(g)$  can be regarded as the bounds for an unknown precise prevision  $M(g)$  which is called a linear prevision.

Natural extension is a general mathematical procedure for calculating new previsions from initial judgements. It produces a coherent overall model from a certain collection of imprecise probability judgements and may be seen as the basic constructive step in interval-valued statistical reasoning.

The crux of optimisation problems (1)-(2) is that their solutions obtained as a result of solving linear programs are defined on the family of degenerate probability

distributions<sup>1</sup>, which are included on equal footing in the set of all admissible probability distributions over which the solution is sought. As proven in [2], solving these optimisation problems on the set of all admissible probability distributions gives the same solution as that obtained on only the set of degenerate distributions:

$$\rho^*(x) = \sum_{k=1}^{n+1} c_k \delta(x, x_k), \quad (3)$$

where  $c_k \in \mathbf{R}_+$ ,  $\sum_{k=1}^{n+1} c_k = 1$ , and  $\delta(x, x_k)$  is the Dirac function which has unit area concentrated in the immediate vicinity of point  $x_k$ .

By substituting the degenerate class of densities (3) into objective function (1) for  $\underline{M}(g)$  and constraints (2) we obtain

$$\underline{M}(g) = \inf_{c_k, x_k} \sum_{k=1}^{n+1} c_k g(x_k) \quad (4)$$

subject to

$$\left. \begin{aligned} \sum_{k=1}^{n+1} c_k &= 1, \quad c_k \geq 0, k = 1, \dots, n+1, \\ \underline{a}_i &\leq \sum_{k=1}^{n+1} c_k f_i(x_k) \leq \overline{a}_i, i \leq n \end{aligned} \right\} \quad (5)$$

We refer to the natural extension (4)-(5) as the degenerate form.

All this would simply be mathematical subtlety – that is, of little interest to practitioners – if it did not give us a clue to deriving more precise previsions of interest for continuous random variables. For some variables it is often not realistic to assume that the probability masses are concentrated in a few points as opposed to being continuously distributed over the set of possible outcomes. In reliability applications, probability masses of time to failure cannot (except in very special cases) concentrate in a very few points of the positive real line. Ignoring this evidence is one of the causes (we hold it to be the root cause) of high imprecision in reliability applications as well as in other applications.

### Example 1.

The sample set of a continuous random variable  $X$  is an interval  $[0, T]$ . The only available information about  $X$  is point-valued probability  $b$  of finding its value within an

<sup>1</sup> The probability distribution of a continuous random variable is referred to as degenerate if the probability masses are concentrated in a finite number of points belonging to the continuous set of possible states

interval  $[\underline{q}, \bar{q}] \subseteq [0, T]$ . That is,  $\Pr(x \in [\underline{q}, \bar{q}]) = b$ . What are the lower and upper bounds for the expected value of  $X$ ?

Natural extension in its primal form appears as follows:

$$\underline{M}(X) = \min_p \int_0^T x \rho(x) dx \quad \text{subject to}$$

$$\int_0^T \rho(x) dx = 1, \quad \rho(x) \geq 0, \quad \text{and} \quad \int_0^T I_{[\underline{q}, \bar{q}]}(x) \rho(x) dx = b,$$

where  $I_{[\underline{q}, \bar{q}]}(x) = 1$  if  $x \in [\underline{q}, \bar{q}]$  and  $I_{[\underline{q}, \bar{q}]}(x) = 0$  otherwise.

Its counterpart in the degenerate form, as follows from (4)-(5), is the optimization problem

$$\underline{M}(X) = \inf_{c_1, c_2} (c_1 x_1 + c_2 x_2) \quad \text{subject to}$$

$$c_1 + c_2 = 1, c_i \geq 0 \quad \text{and} \quad c_1 I_{[\underline{q}, \bar{q}]}(x_1) + c_2 I_{[\underline{q}, \bar{q}]}(x_2) = b.$$

From the constraints it can be concluded that  $c_1 I_{[\underline{q}, \bar{q}]}(x_1) + c_2 I_{[\underline{q}, \bar{q}]}(x_2) = b$  holds only if  $x_1 \notin [\underline{q}, \bar{q}]$ ,  $x_2 \in [\underline{q}, \bar{q}]$  and  $c_2 = b$ , which entails  $c_1 = 1 - b$ . Plugging  $c_1$  and  $c_2$  into the objective function brings us to the simple optimisation problem

$$\underline{M}(X) = \inf_{x_1} ((1-b)x_1 + bx_2).$$

The infimum is attained with  $x_1 = 0$  and  $x_2 = \underline{q}$ , that is,  $\underline{M}(X) = b\underline{q}$ .

Thus, the probability distribution function delivering the infimum to the objective function degenerates into the one with probability masses concentrated in two points  $x_1 = 0$  and  $x_2 = \underline{q}$  with masses  $(1-b)$  and  $b$ , correspondingly. This case is presented in Fig. 1.

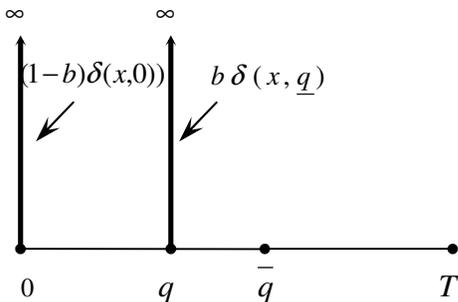


Figure 1: Degenerate probability distribution function providing the infimum to the objective function

### 3 An example Where Natural Extension Fails to Utilise Evidence

An attempt to mitigate the influence of degenerate probability distributions on the solutions and to obtain tighter bounds was undertaken in [4]. No significant effect was obtained through the introduction of judgements on the skewness and unimodality of the distributions as, in this case, the peaks of degenerate distributions simply become repositioned and probability masses become redistributed among the peaks. The nature of the distributions defining the solutions remained unchanged.

In the attempt to achieve tighter bounds, it seems natural to try to constrain the underlying probability distributions such that they rule out the degenerate distributions. This can be done through imposing a restriction on the upper bound of the probability density function. (This device is not new and was used, for example, in [5] and [6].) In some practical situations, such bounds can be elicited from experts. For example, in reliability applications, the expert could be asked: ‘‘What is the largest possible percentage of failures per year for a given component with a definite age?’’ In other cases, such bounds can be obtained from the statistical data or from a physical model of the corresponding phenomenon.

Once an upper bound to the probability density function is known, it can be used to restrict the set of feasible probability distributions and rule out the degenerate ones. Let us introduce such an upper bound  $K \in \mathbf{R}_+$  on the values of the probability density function, i.e.,

$$0 \leq \rho(x) \leq K = \text{const} \quad \text{for } \forall x. \quad (6)$$

Since the overall probability over the interval  $[0, T]$  is equal to 1, the upper bound  $K$  satisfies the inequality  $KT \geq 1$ .

By bounding the density function, the set of constraints to optimisation problem (1)-(2) is complemented by inequality (6) which, as it turns out, complicates the optimisation problem drastically.

It is chiefly through duality theory that a linear program can be viewed in its proper perspective and solved. For primal problem (1)-(2) complemented by constraint (6), the dual optimisation problem has the infinite number of dual variables. This is because there are as many dual variables as primal constraints, and in our case the inequality  $\rho(x) \leq K$  is to be regarded as denoting an infinite set of constraints  $\rho(x_i) \leq K \quad i=1, \dots, n, n \rightarrow \infty$ . Thus, not being able to employ the dual form of natural extension nor its degenerate form, we become devoid of the key

mechanism for the construction of coherent imprecise models, natural extension.

One would anyway arrive at this stopping point in case of trying to use non-linear constraints, as real-life statistical evidence in many cases cannot be confined to linear constraints.

In the section below we suggest taking a different look at the primal form of natural extension (1)-(2), an approach which opens up new perspectives for obtaining tighter intervals.

#### 4 Natural Extension as a Problem of the Calculus of Variations

The mathematical program (1)-(2) can be modified slightly to make it amenable to the calculus of variations. The calculus is based on the statement that we can always apply a small change  $\pm\delta\rho(x)$  to a function  $\rho(x)$ . (Here  $\delta\rho(x)$  denotes a variation of  $\rho(x)$ , and the symbol  $\delta$  should not be confused with the Dirac function). Applying variation  $\pm\delta\rho(x)$  to a function  $\rho(x)$  has the consequence that  $\rho(x)$  can become negative, which is in contradiction with the inequality  $\rho(x)\geq 0$ .

The requirement  $\rho(x)\geq 0$  can be satisfied differently by introducing another function  $z(x)$  for which

$$\rho(x) = z^2(x). \quad (7)$$

We then have to replace  $\rho(x)$  by  $z^2(x)$  in the expressions for the objective functions and constraints.

The other inequalities in constraints (2) are turned into equalities by introducing yet other unknown functions  $u_{(i,1)}(x)$  and  $u_{(i,2)}(x)$ ,  $i=1, \dots, n$ , such that

$$\int_0^T f_i(x) \cdot (z^2(x) - u_{(i,1)}^2(x)) dx = \underline{a}_i, \quad (8)$$

$$\int_0^T f_i(x) \cdot (z^2(x) + u_{(i,2)}^2(x)) dx = \overline{a}_i, \quad (9)$$

More information on this technique can be found, for example, in [7].

After having made the above changes, the problem of finding the lower and upper bounds for  $M(g)$  now has  $z(x)$ ,  $u_{(i,1)}(x)$  and  $u_{(i,2)}(x)$ ,  $i=1, 2, \dots, n$ , as decision variables. Thus the original problem (1)-(2) turns into the following:

$$\inf_{z(x)} \int_0^T g(x) z^2(x) dx \quad \text{and} \quad \sup_{z(x)} \int_0^T g(x) z^2(x) dx \quad (10)$$

subject to (8), (9) and

$$\int_0^T z^2(x) dx = 1. \quad (11)$$

Optimization problem (10) subject to (8), (9) and (11) is another form of natural extension amenable to variational calculus. Constraints like (8), (9) and (11), which are integrals of some unknown functions, are called *isoperimetric constraints* [8].

The conventional way of solving problem (10) subject to (8), (9) and (11) is to replace it with an unconstrained optimization problem. In this case the integrand of the objective function

$$F(z, x) = g(x) z^2(x) \quad (12)$$

is replaced by

$$F^*(z, u_{(i,1)}, u_{(i,2)}, x) = g(x) z^2(x) + \lambda_0 z^2(x) + \sum_{i=1}^n \lambda_i f_i(x) (z^2(x) - u_{(i,1)}^2(x)) + \sum_{i=1}^n \lambda_{i+n} f_i(x) (z^2(x) + u_{(i,2)}^2(x)) \quad (13)$$

where  $\lambda_i \in \mathbf{R}$ ,  $i=0, \dots, 2n$ , are (unknown) Lagrange multipliers that could be derived from a system of the Euler-Lagrange equations (see below) complemented by equations-constraints (8), (9) and (11).

The unconstrained optimization problem, which is to be solved now, appears as follows:

$$\inf_{z, u_{(i,1)}, u_{(i,2)}} \int_0^T F^*(z, u_{(i,1)}, u_{(i,2)}, x) dx. \quad (14)$$

For an unconstrained optimization problem the solutions satisfying the necessary condition of optimality can be derived from the Euler-Lagrange equations [8]. For problem (14) these equations take the following form:

$$\left. \begin{aligned} \frac{\partial F^*(z, u_{(i,1)}, u_{(i,2)})}{\partial z} &= \frac{d}{dx} \left( \frac{\partial F^*(z, u_{(i,1)}, u_{(i,2)})}{\partial \dot{z}} \right) \\ \frac{\partial F^*(z, u_{(i,1)}, u_{(i,2)})}{\partial u_{(i,1)}} &= \frac{d}{dx} \left( \frac{\partial F^*(z, u_{(i,1)}, u_{(i,2)})}{\partial \dot{u}_{(i,1)}} \right) \\ \frac{\partial F^*(z, u_{(i,1)}, u_{(i,2)})}{\partial u_{(i,2)}} &= \frac{d}{dx} \left( \frac{\partial F^*(z, u_{(i,1)}, u_{(i,2)})}{\partial \dot{u}_{(i,2)}} \right) \end{aligned} \right\} \quad (15)$$

where  $\dot{z} = dz/dx$ ;  $\dot{u}_{(i,1)} = du_{(i,1)}/dx$ ,  $\dot{u}_{(i,2)} = du_{(i,2)}/dx$ .

By plugging (13) into (15) we obtain

$$z(x) \cdot \left( g(x) + \lambda_0 + \sum_{i=1}^n (\lambda_i + \lambda_{i+n}) f_i(x) \right) = 0 \quad (16)$$

$$\lambda_i \cdot u_{(i,1)}(x) = 0, \quad i=1, \dots, n, \quad (17)$$

$$\lambda_{i+n} \cdot u_{(i,2)}(x) = 0, \quad i=1, 2, \dots, n. \quad (18)$$

Let us examine equation (16). It holds if  $z(x)=0$  for all  $x \in [0, T]$ . But this would be in contradiction with constraints (8), (9) and (11). Thus  $z(x) \neq 0$ , at least in some points or possibly inside some subintervals of  $[0, T]$ . From (16) for those points where  $z(x) \neq 0$  it holds that

$$g(x) + \sum_{i=1}^n (\lambda_i + \lambda_{i+n}) f_i(x) = -\lambda_0. \quad (19)$$

Let us now denote  $\xi(x) = g(x) + \sum_{i=1}^n (\lambda_i + \lambda_{i+n}) f_i(x)$ , and consider as an example  $g(x) = I_{[0, x_1]}(x)$  and all the other gambles  $f_i(x)$ ,  $i = 1, 2, \dots, n$  as linear functions. This case is depicted in Fig. 2.

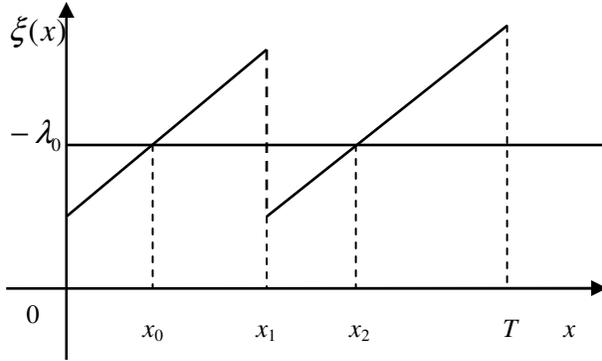


Figure 2: An example of  $\xi(x)$  satisfying the necessary condition of optimality

In order to satisfy constraint (11) and to hold equation (16) true the probability density function delivering an extremum to the objective function  $\int_0^T F^*(z, u_{(i,1)}, u_{(i,2)}, x) dx$  can only be degenerate, i.e., concentrated in the three points  $x_0, x_1$  and  $x_2$ . This is formalised as  $\rho(x) = z^2(x) = \sum_{k=0}^2 c_k \cdot \delta(x, x_k)$ , where

$$\sum_{k=0}^2 c_k = 1. \quad \text{Thus, we have arrived at the case where}$$

optimal solutions belong to the family of degenerate distributions.

## 5 Utilising Boundary Constraints with the Variational Form of Natural Extension

Let us now turn back to the case where a boundary to the probability density function is known and we would like to utilise this knowledge to reduce, as we expect, imprecision in the probabilistic measures of interest. That is, we will seek inf and sup of the objective function (1) subject to constraints (2), (6). To solve this new problem, an approach based on the following theorem is proposed.

**Theorem 1.** *If for any interval  $\alpha \leq x \leq \beta$ ,  $0 \leq \alpha < \beta \leq T$  and for any  $h_0, h_1, \dots, h_n \in \mathbf{R}$  it holds that*

$$g(x) \neq h_0 + \sum_{i=1}^n h_i f_i(x),$$

*then probability density function  $\rho(x)$ , on which inf and sup are attained in problems (1) subject to (2) and (6), is a step-wise function whose values are either 0 or K.*

*Proof.* In this problem we have two direct constraints on the density function:  $\rho(x) \geq 0$  and  $\rho(x) \leq K$ . To adjust the constraints to the calculus of variations, we introduce some new functions  $z(x)$  and  $v(x)$  such that  $\rho(x) = z^2(x)$  and

$$z^2(x) + v^2(x) = K \quad (20)$$

Thus, we have a new optimisation problem with objective function (10) subject to (8), (9), (11) and (20).

With respect to noted above, newly introduced equality (20) should be referred to as holonomic constraint.

As we did it earlier, the primal problem with holonomic and isoperimetric constraints is replaced by a new unconstrained optimization problem

$$\inf_{z, v, u_{(i,1)}, u_{(i,2)}} \int_0^T F^{**}(z, v, u_{(i,1)}, u_{(i,2)}, x) dx, \quad (21)$$

where

$$\begin{aligned} F^{**}(z, v, u_{(i,1)}, u_{(i,2)}) &= g(x)z^2(x) + \lambda^*(x) \cdot (z^2(x) + v^2(x) - K) \\ &+ \lambda_0 z^2(x) + \sum_{i=1}^n \lambda_i f_i(x) (z^2(x) - u_{(i,1)}^2(x)) \\ &+ \sum_{i=1}^n \lambda_{i+n} f_i(x) (z^2(x) - u_{(i,2)}^2(x)), \end{aligned} \quad (22)$$

and  $\lambda^*(x), \lambda_0, \lambda_1, \dots, \lambda_{2n}$  are (unknown) Lagrange multipliers. Note that  $\lambda^*(x)$  is to be a function of  $x$  because it is multiplied by a holonomic constraint, while  $\lambda_0, \lambda_1, \dots, \lambda_{2n}$  are constants because they correspond to isoperimetric constraints [7].

For an unconstrained optimization problem the solutions satisfying the necessary condition of optimality can be derived from the Euler-Lagrange equations [8]. By applying the Euler-Lagrange equations, as we did for (14), we arrive at the following set of equalities:

$$z(x) \cdot \left( g(x) + \lambda^*(x) + \lambda_0 + \sum_{i=1}^n (\lambda_i + \lambda_{i+n}) f_i(x) \right) = 0, \quad (23)$$

$$\lambda^*(x) v(x) = 0, \quad (24)$$

$$\lambda_i \cdot u_{(i,1)}(x) = 0, \quad i=1,2,\dots,n, \quad (25)$$

$$\lambda_{i+n} \cdot u_{(i,2)}(x) = 0, \quad i=1,2,\dots,n. \quad (26)$$

Let us an interval  $[\alpha, \beta] \subseteq [0, T]$  is that on which  $z(x) \neq 0$ . How would  $z(x)$  behave on this interval and what values would it take?

According to (23), in those points  $x$  where  $z(x) \neq 0$  it holds that

$$\lambda^*(x) = -g(x) - \lambda_0 - \sum_{i=1}^n (\lambda_i + \lambda_{i+n}) f_i(x). \quad (27)$$

And according to (23), in those points  $x$  where  $\lambda^*(x) \neq 0$  it holds that  $v(x) = 0$ , which in turn, according to (20), results in  $\rho(x) = z^2(x) = K$ .

From (27) it follows that if

$$g(x) \neq -\lambda_0 - \sum_{i=1}^n (\lambda_i + \lambda_{i+n}) f_i(x),$$

then  $\lambda^*(x) \neq 0$  and  $\rho(x) = K$ .

By denoting  $h_0 = -\lambda_0$  and  $h_i = -(\lambda_i + \lambda_{i+n})$  we can rewrite

$$g(x) \neq h_0 + \sum_{i=1}^n h_i f_i(x),$$

which was to be proven.

The theorem enables us to reduce the original variational optimization problem to an easier problem of optimizing a multivariate function under algebraic constraints.

Indeed,

$[x_0, x_1), [x_2, x_3), [x_4, x_5), \dots, [x_{2m}, x_{2m+1})$  be intervals on which  $\rho(x) = K \neq 0$ , and  $[x_1, x_2), [x_3, x_4), [x_5, x_6), \dots, [x_{2m+1}, T)$  be intervals on which  $\rho(x) = 0$ . Let us denote

let  
the  
let  
the

$$G(x_j, x_{j+1}) = \int_{x_j}^{x_{j+1}} g(x) dx \quad (28)$$

$$\Phi_i(x_j, x_{j+1}) = \int_{x_j}^{x_{j+1}} f_i(x) dx, \quad i = 1, 2, \dots, n. \quad (29)$$

Then, problem (1) subject to constraints (2), (6) takes the following form:

$$\begin{aligned} \underline{M}(g) &= \min_{x_0, x_1, \dots} \left\{ K \cdot \sum_{j=0}^m G(x_{2j}, x_{2j+1}) \right\} \\ \overline{M}(g) &= \max_{x_0, x_1, \dots} \left\{ K \cdot \sum_{j=0}^m G(x_{2j}, x_{2j+1}) \right\} \end{aligned} \quad (30)$$

subject to

$$K \cdot \sum_{j=0}^m (x_{2j+1} - x_{2j}) = 1 \quad (31)$$

$$\underline{a}_i \leq K \cdot \sum_{j=0}^m \Phi_i(x_{2j}, x_{2j+1}) \leq \overline{a}_i, \quad i = 1, 2, \dots, n. \quad (32)$$

If the number of intervals  $m$  is known, this optimization problem can be solved by using standard numerical techniques such as gradient methods, simplex-based search methods, genetic algorithms, etc. In simple cases, the solution can be found in an analytical form.

How can we find  $m$ ? One idea is to start with the smallest value  $m$ , corresponding to having one interval with nonzero density, and to solve the optimization problem with this  $m$ . Then, increase  $m$  by 1 and solve the problem again, etc. Repeat the process until when for a new  $m$  you get exactly the same optimising function  $\rho(x)$  as for the previous  $m$  – this will mean that a further subdivision of intervals will probably not change the value of objective function (1).

**Example 2.** Utilising knowledge on the boundary of the density function

In this example, the statistical evidence about a random value  $X$  we have at hand is a boundary  $K$  on the probability density function and, as in Example 1,  $\Pr(x \in [\underline{q}, \overline{q}]) = b$ . What are the lower,  $\underline{M}(X)$ , and upper bounds,  $\overline{M}(X)$ , for the expected value of  $X$ ?

It is found that increasing  $m$  step by step by 1 starting from 0 does not change the optimising density function  $\rho(x)$  after  $m$  exceeds 1. That is, the solution of problem (30)-(32) must be sought for  $m=1$ . (Note that  $m=1$  corresponds to having two intervals on which the probability density function is different from 0.)

Depending on the disposition of  $\underline{q}$  within the interval  $[0, T]$ , the probability density function delivering the minimum to the objective function is calculated differently.

If  $\underline{q} \geq \frac{1-b}{K}$ ,

$$\rho(x) = \begin{cases} K & \text{for } 0 \leq x \leq \frac{1-b}{K} \\ 0 & \text{for } \frac{1-b}{K} < x < \underline{q} \\ K & \text{for } \underline{q} \leq x \leq \underline{q} + \frac{b}{K} \\ 0 & \text{for } \underline{q} + \frac{b}{K} < x \leq T \end{cases},$$

and for  $\underline{q} < \frac{1-b}{K}$ :

$$\rho(x) = \begin{cases} K & \text{for } 0 \leq x \leq \underline{q} + \frac{b}{K} \\ 0 & \text{for } \underline{q} + \frac{b}{K} < x < \bar{q} \\ K & \text{for } \bar{q} \leq x \leq \bar{q} + \frac{1-qK-b}{K} \\ 0 & \text{for } \bar{q} + \frac{1-qK-b}{K} < x \leq T \end{cases}$$

Let us assume that  $\underline{q} \geq \frac{1-b}{K}$ . Then it can be concluded that optimization problem (30)-(32) for the lower bound becomes as follows:

$$\underline{M}(X) = \min_{x_i} K \frac{x_1^2 - x_2^2 + x_3^2 - x_2^2}{2} \text{ subject to}$$

$$K(x_1 - x_0) + K(x_3 - x_2) = 1$$

$$K(x_3 - x_2) = b.$$

The next step is to plug the constraints into the objective function and observe that minimum is attained if  $x_2 = \underline{q}$ .

After doing this, we obtain

$$\underline{M}(X) = \min_{x_0} \frac{1}{2K} + x_0(1-b) + b \left( \underline{q} - \frac{1-b}{K} \right).$$

It is not difficult to see that the minimum is attained if  $x_0 = 0$ . Thus

$$\underline{M}(X) = \frac{1}{2K} + \frac{b(qK - (1-b))}{K}.$$

The probability density function delivering the minimum is shown in Fig. 3.

For the case when  $\underline{q} < \frac{1-b}{K}$  the solution is

$$\underline{M}(X) = \frac{1}{2K} + \frac{\bar{q}K - (qK + b) \cdot (1 + (\bar{q} - q)K - b)}{K}.$$

The solution to  $\bar{M}(X)$  can be obtained in a similar way to that for the lower bound.

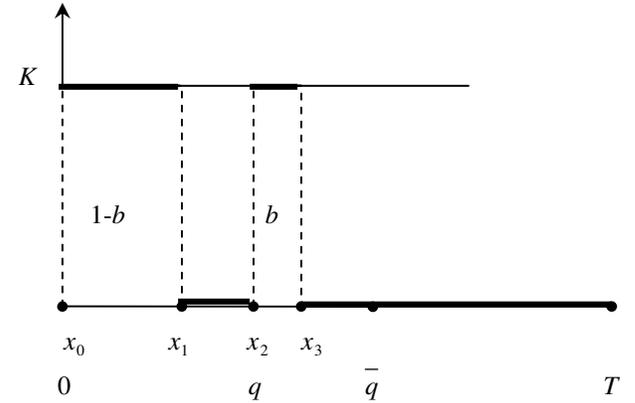


Figure 3: Bounded probability density function providing the infimum to the objective function

## 6 Bounded densities and their derivatives

In attempting to achieve tighter bounds, one can impose constraints on the derivatives (or their absolute values) of probability density functions. So, now we suppose that one has at hand an upper bound on the value of the probability density function and an upper bound on its derivative absolute value. Any other assumptions concerning the actual shape of the distribution are not introduced.

Once the additional upper bound is known, it can be used to restrict the set of admissible probability distributions and rule out the functions which derivatives take excessively high values.

Let us denote  $M \in \mathbf{R}_+$  an upper bound on the value of the probability density absolute value, i.e., for  $\forall x$

$$|dp(x)/dx| \leq M = \text{const.} \quad (33)$$

In the variational calculus set-up, now we seek inf and sup of the objective function (1) subject to constraints (2), (6) and (33). To solve this new problem, an approach based on the following theorem is proposed.

**Theorem 2.** If for any interval  $a \leq x \leq \beta$ ,  $0 \leq a < \beta \leq T$  and for any  $h_0, h_1, \dots, h_n \in \mathbf{R}$  it holds that

$$g(x) \neq h_0 + \sum_{i=1}^n h_i f_i(x),$$

then the probability density function  $\rho(x)$ , on which inf and sup are attained in problems (1) subject to (2), (6) and (33), is a stepwise linear function  $\rho(x) = \pm Mx + C$  whose values are bounded by  $K$  from above.

*Proof.* The logic of the proof is similar to that used to prove Theorem 1. The proof can be found in [9] which has been submitted for publication.

An example of the density function satisfying Theorem 2 is depicted in Fig. 4.

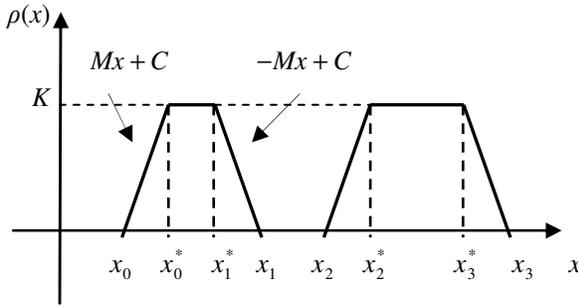


Figure 4: An example of the density function satisfying Theorem 2

The points in Fig. 5 marked with asterisks have the following values:  $x_0^* = x_0 + K/M$ ,  $x_1^* = x_1 - K/M$ ,  $x_2^* = x_2 + K/M$ ,  $x_3^* = x_3 - K/M$

Theorem 2 enables to reduce the original variational optimization problem to an easier one. This can be done because the shape of the density function is now known. Unknown are the points  $x_0, x_1, \dots, x_0^*, x_1^*, \dots$ , which become the parameters of the density function  $\rho(x) = \rho(x_0, x_1, \dots, x_0^*, x_1^*, \dots, x)$  and the decision variables in the new optimisation problem.

Let  $[x_0, x_1), [x_2, x_3), [x_4, x_5), \dots, [x_{2m}, x_{2m+1})$  be the intervals on which  $\rho(x) \neq 0$ . They can be interpreted as lower bases of the trapezoids (see Fig. 5). The upper bases of the trapezoids are the intervals  $[x_0 + K/M, x_1 - K/M), [x_2 + K/M, x_3 - K/M), \dots, [x_{2m} + K/M, x_{2m+1} - K/M)$ . And  $[x_1, x_2), [x_3, x_4), \dots, [x_{2m+1}, x_{2m+2})$  are the intervals on which  $\rho(x) = 0$ .

Now the optimisation problem appears as follows:

$$\begin{aligned} \underline{M}(g) &= \min_{x_0, x_1, \dots} \left\{ \sum_{j=0}^m G^*(x_{2j}, x_{2j+1}) \right\} \\ \overline{M}(g) &= \max_{x_0, x_1, \dots} \left\{ \sum_{j=0}^m G^*(x_{2j}, x_{2j+1}) \right\} \end{aligned} \quad (34)$$

subject to

$$K \cdot \sum_{j=0}^m (x_{2j+1} - x_{2j} - K/M) = 1, \quad (35)$$

$$\underline{a}_i \leq \sum_{j=0}^m \Phi_i^*(x_{2j}, x_{2j+1}) \leq \overline{a}_i, \quad i = 1, 2, \dots, \quad (36)$$

where

$$\begin{aligned} G^*(x_j, x_{j+1}) &= M \int_{x_j}^{x_j + K/M} g(x)(x - x_j) dx - \\ &M \int_{x_{j+1} - K/M}^{x_{j+1}} g(x)(x - x_{j+1}) dx + K \int_{x_j + K/M}^{x_{j+1} - K/M} g(x) dx \\ \Phi_i^*(x_j, x_{j+1}) &= M \int_{x_j}^{x_j + K/M} f_i(x)(x - x_j) dx - \\ &M \int_{x_{j+1} - K/M}^{x_{j+1}} f_i(x)(x - x_{j+1}) dx + K \int_{x_j + K/M}^{x_{j+1} - K/M} f_i(x) dx \end{aligned}$$

**Example 3.** Unbounded probability density function and bounded absolute values of its derivative

Let us consider an example in which  $|d\rho(x)/dx| \leq M$  is the only restriction on  $\rho(x)$ . What are the bounds on the expected value  $M(X)$  of the corresponding random variable?

In this example  $g(x)=x$ , which implies that everywhere  $g(x) \neq h_0$  meaning that theorem 2 can be applied.

Note first that as the condition  $\rho(x) \leq K$  is not imposed on the density function, the trapezoidal shape of the density is changed to the triangular one.

Let us start with  $m=0$  corresponding to having one interval  $[x_0, x_1)$  on which the probability density function is different from 0 and denote  $y = x_1 - x_0$ .

Here we have only one isoperimetric constraint:  $\int_0^T \rho(x) dx = M \frac{y^2}{4} = 1$ , where  $M \frac{y^2}{4}$  is the area the

triangle. Hence  $y = \frac{2}{\sqrt{M}}$ , or  $x_1 = x_0 + \frac{2}{\sqrt{M}}$ .

The formula for the expected value  $M(X)$  takes the form:

$$M(X) = \int_0^T x\rho(x)dx = \int_{x_0}^{x_0+1/\sqrt{M}} x(Mx - Mx_0)dx + \int_{x_0+1/\sqrt{M}}^{x_0+2/\sqrt{M}} x(-Mx + M(x_0 + 2/\sqrt{M}))dx.$$

And further

$$M(X) = \frac{M}{3} \left( (x_0 + 1/\sqrt{M})^3 - x_0^3 \right) - \frac{Mx_0}{2} \left( (x_0 + 1/\sqrt{M})^2 - x_0^2 \right) - \frac{M}{3} \left( (x_0 + 2/\sqrt{M})^3 - (x_0 + 1/\sqrt{M})^3 \right) + \frac{M(x_0 + 2/\sqrt{M})}{2} \left( (x_0 + 2/\sqrt{M})^2 - (x_0 + 1/\sqrt{M})^2 \right).$$

In the following we will keep in mind that  $x_0 \geq 0$ ,  $x_1 \leq T$ , and hence  $x_0 \leq T - \frac{2}{\sqrt{M}}$ .

It is easy to see that the smallest value of  $M(X)$  is attained when  $x_0 = 0$ , so  $\underline{M}(X) = \frac{1}{\sqrt{M}}$ .

Similarly, to obtain  $\overline{M}(X)$ , we take the largest possible value of  $x_0$ , i.e.  $x_0 = T - \frac{2}{\sqrt{M}}$ , which brings us to  $\overline{M}(X) = T - \frac{1}{\sqrt{M}}$ .

If we take  $m=1$  and do manipulations similar to the above, we find that the solutions do not change.

**Example 4.** *Bounded probability density function and bounded absolute values of its derivative*

Now we have two constraints (6) and (33), i.e.,  $0 \leq \rho(x) \leq K$  and  $|d\rho(x)/dx| \leq M$ . The question to answer is still the same: What are the bounds on the expected value  $M(X)$ ?

As we keep the function  $g(x)=x$  introduced in Example 3, theorem 2 can be also applied for this case.

Start with  $m=0$ . Here we have only one isoperimetric constraint (the area of the trapezoid equalised to 1):

$$\int_0^T \rho(x)dx = \frac{x_1 - x_0 + x_1 - x_0 - 2K/M}{2} K = 1, \quad \text{hence} \\ x_1 = 1/K + K/M + x_0.$$

The formula for the expected value  $M(X)$  takes the form:

$$M(X) = \int_0^T x\rho(x)dx = \int_{x_0}^{x_0+K/M} x(Mx - Mx_0)dx + K \int_{x_0+K/M}^{x_0+1/K} xdx + \int_{x_0+1/K}^{x_0+1/K+K/M} x(-Mx + M/K + K + Mx_0)dx.$$

And finally,

$$M(X) = \frac{M}{3} \left( (x_0 + K/M)^3 - x_0^3 \right) - \frac{Mx_0}{2} \left( (x_0 + K/M)^2 - x_0^2 \right) + \frac{K}{2} \left( (x_0 + 1/K)^2 - (x_0 + K/M)^2 \right) - \frac{M}{3} \left( (x_0 + 1/K + K/M)^3 - (x_0 + 1/K)^3 \right) + \frac{M(x_0 + 1/K) + K}{2} \left( (x_0 + 1/K + K/M)^2 - (x_0 + 1/K)^2 \right)$$

As  $x_0 \geq 0$  and  $x_1 \leq T$ , hence  $x_0 \leq T - \frac{1}{K} - \frac{K}{M}$ .

The smallest value of  $M(X)$  is attained when  $x_0 = 0$ , hence  $\underline{M}(X) = \frac{1}{2K} + \frac{K}{2M}$ .

Similarly, to obtain  $\overline{M}(X)$ , we take the largest possible value of  $x_0$ , i.e.,  $x_0 = T - \frac{1}{K} - \frac{K}{M}$ , hence  $\overline{M}(X) = (T - 1/K) + \frac{1}{2K} = T - \frac{1}{2K} - \frac{K}{2M}$ .

If we take  $m=1$  and do manipulations similar to the above, we find that the solutions do not change.

## 7 What is Still Dissatisfying?

There are at least two remaining problems with applying imprecise statistical reasoning to reliability analysis.

In reliability analysis, the pivotal characteristic is time to failure (or time between failures if a system is repairable), and a failure in a system can occur at any point of the lifetime. In contrast, the model presupposes that failures can take place only within some specific intervals but not at any point. This is because probability masses are not continuously distributed during the lifetime. In spite of bringing in more statistical evidence about time to failure, the situation does not seem to be remedied.

The other principle obstacle to reliability applications is the bounding condition on gambles, which in practice means dealing with bounded random values. That is, applying the reasoning to reliability implies that time to

failure is a bounded random value. Let us say, one must know the maximum time a system can survive in order to apply the theory. This is that what can hardly be known for certainty. Furthermore, as technical systems undergo preventive maintenance and are put out of operation based on volitional decisions rather than after observing their full inoperability, knowing the point behind which they become irrecoverable, and even defining what it means, make the bound on time to failure meaningless.

## 8 Summary and Conclusions

The usefulness of interval-valued statistical characteristics depends both on how tight the bounds are and on how easy they are to compute. The tightness of the bounds depends in turn on the amount of information available and that which can be utilised by the method, and on the method itself. The more relevant information the modeller has at hand and the greater the amounts of it that can be utilised by the model, the tighter the bounds are. We have been aiming at enhancing natural extension so that it could utilise a wider variety of statistical evidence, some of which is easy to acquire but not easy to utilise.

As has been demonstrated, natural extension can be viewed as the problem of finding an extremal of a functional, a problem which belongs to the realm of variational calculus. If this path is followed, the modeller can utilise more versatile information than is possible with the natural extension suggested by Walley [1] and Kuznetsov [3]. The present paper has demonstrated that imposing a restriction on the upper bound of the probability density function of a random value is an effective way of obtaining tighter bounds of statistical measures.

In some cases, common sense and intuition may suggest that the underlying distribution is for instance differentiable in any point or symmetrical without specifying a particular shape. Utilising this kind of evidence may drastically reduce imprecision in the resultant interval-valued statistical characteristics, and, it is clear, this evidence is acquired at a low cost; in some cases it can be gained at no effort.

We have been attempting to demonstrate in relation to the approach based on variational calculus that there is room for improvement without having to use unreliable data and introduce debatable assumptions as a means of obtaining reasonably precise results.

In the pursuit of robust reliability assessments, the next facing challenge is to update the existing reliability models so that they can take account of additional evidence, evidence that until now has not been requested owing to the models' incapacity to utilise it. The fact that there is currently a substantial amount of alternative evidence at our disposal presents other challenges. For

example, what kind of evidence is worth using in order to facilitate computations and make substantial headway in terms of tighter bounds? What constraints are most beneficial for what models? These are directions in which, we suggest, further work with the calculus of variations ought to proceed.

## Acknowledgement

Participation of V. Krymsky in the work described was partially supported by NATO grant CBP.NR.NREV 982410.

## References

- [1] Walley P. (1991) *Statistical reasoning with imprecise probabilities*. Chapman and Hall. New York.
- [2] Utkin L. and Kozine I. (2001) *Different faces of the natural extension*. In: Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications, ISIPTA '01, pp. 316-323.
- [3] Kuznetsov V. (1991) *Interval statistical models*. Radio and Sviaz. Moscow. (In Russian).
- [4] Utkin L. (2002) *Imprecise calculation with the qualitative information about probability distributions*. In: Proceedings of the conference on Soft Methods in Probability and Statistics. Eds. P. Grzegorzewski, O. Hryniewicz and M.A. Gil, Physica-Verlag, Heidelberg, New York, pp. 164-169.
- [5] Smith E.J. (1995) *Generalized Chebychev inequalities: theory and applications in decision analysis*. Operations Research, Vol. 43(5), pp. 807-825.
- [6] Kozine I.O., Krymsky V.G. (2003) *Reducing uncertainty by imprecise judgements on probability distributions: Application to system reliability*. In: Proceedings. 3rd International symposium on imprecise probabilities and their applications (ISIPTA '03), Lugano (CH), 14-17 July 2003. Bernard, J.-M.; Seidenfeld, T.; Zaffalon, M. (eds.), (Carleton Scientific, Waterloo (CA), 2003) (Proceedings in Informatics, 18), pp. 335-344.
- [7] Ivanov V.A., Faldin N.V. (1981) *Theory of optimal control systems*. Nauka Publ. Moscow. (In Russian).
- [8] Gelfand I.M., Fomin S.V. (2000) *Calculus of variations*. Dover Publ. New York, 240 p.
- [9] Kozine I., and Krymsky V. (2007) Computing interval-valued statistical characteristics: What is the stumbling block for reliability applications? Submitted to the Journal of Reliable Computing