Origninally Published by
Sandia National Laboratories as
SAND2002-4015, January 2003

SAND2015-4166J

Constructing probability boxes and Dempster-Shafer structures*

Scott Ferson, Vladik Kreinovich, Lev Ginzburg, Davis S. Myers
Applied Biomathematics
100 North Country Road
Setauket, New York 111733

Kari Sentz
Systems Science and Industrial Engineering Department
Thomas J. Watson School of Engineering and Applied Science
Binghamton University
P.O. Box 6000
Binghamton, NY 13902-6000

Abstract

This report summarizes a variety of the most useful and commonly applied methods for obtaining Dempster-Shafer structures, and their mathematical kin probability boxes, from empirical information or theoretical knowledge. The report includes a review of the aggregation methods for handling agreement and conflict when multiple such objects are obtained from different sources.

* The work described in this report was performed for Sandia National Laboratories under Contract No. 19094

Acknowledgments

We thank Vicki Bier of the University of Wisconsin at Madison and Robert Clemen of Duke University for helpful discussions. The report benefited immensely from reviews by William Oberkampf, Jon Helton, J. Arlin Cooper and Marty Pilch of Sandia National Laboratories and and W. Troy Tucker of Applied Biomathematics. We thank these readers for their many careful and thoughtful comments, although we did not adopt all of their suggestions. Opinions expressed herein are those of the authors.

Symbols

$\sim$	is distributed as
$\in$	is an element of
$\subseteq$	is a subset of, also enclosure of uncertain numbers (Section 4.1)
$\cup$	union
$\cap$	intersection
$\pm$	plus or minus
$\varnothing$	the empty set, i.e., the set having no members
$[\overline{F}, \underline{F}]$	probability-box specified by a left side $F(x)$ and a right side $\underline{F}(x)$ where $\underline{F}(x) \leq \overline{F}(x)$ for all $x \in \Re$, consisting of all nondecreasing functions F from the reals into [0,1] such that $\underline{F}(x) \leq F(x) \leq \overline{F}(x)$. (Section 2.1)
$\{(s_1,m_1),\ldots,(s_n,m_n)\}$	an enumeration of the elements of a Dempster-Shafer structure in terms of its focal elements s_i and their nonzero masses m_i
$f:A\rightarrow B$	a function f whose range is the set A and whose domain is the set B. In other words, for any element in A, the function f assigns a value that is in the set B
$H_c(x)$	the step function that is zero for all values of $x<c$ and one for all $x\geq c$
inf	infimum (for a finite set, simply the minimum)
normal(μ, σ)	a normal distribution with mean μ and standard deviation σ
$\Re$	the set of all real numbers
sup	supremum (for a finite set, simply the maximum)
uniform(a, b)	a uniform distribution ranging between a and b, where $a\leq b$
weibull(d, c)	a Weibull distribution with scale parameter (or characteristic life) d and shape parameter c, where $0\leq d$, $0\leq c$

1 Introduction

Risk analysts recognize two fundamentally distinct forms of uncertainty. The first is variability that arises from environmental stochasticity, inhomogeneity of materials, fluctuations in time, variation in space, or heterogeneity or other differences among components or individuals. Variability is sometimes called Type I uncertainty, or less cryptically, *aleatory* uncertainty to emphasize its relation to the randomness in gambling and games of chance. It is also sometimes called irreducible uncertainty because, in principle, it cannot be reduced by further empirical study (although it may be better characterized). The second kind of uncertainty is the incertitude that comes from scientific ignorance, measurement uncertainty, inobservability, censoring, or other lack of knowledge. This is sometimes called Type II uncertainty, or simply *epistemic* uncertainty. In contrast with aleatory uncertainty, epistemic uncertainty is sometimes called reducible uncertainty because it can generally be reduced by additional empirical effort at least in principle.

Several disparate theories have been proposed to represent what is known about real-valued but uncertain quantities. For instance, intervals (Moore 1966) have been used to represent purely epistemic uncertainty. For situations in which the uncertainty about quantities is purely aleatory in character, probability theory is usually preferred. When the gaps in our knowledge involve both aleatory and epistemic uncertainty, several competing approaches have been suggested. Some probabilists assert, for instance, that probability theory is sufficiently general to serve as the uncertainty calculus in this case as well. However, many disagree with this assertion. Walley (1991), inter alia, argued that a theory embracing imprecise probabilities would be needed. Shafer (1976) argued that an approach that takes account of the indistinguishability of underlying states within bodies of evidence would be required. We discuss these ideas below.

Several researchers have addressed the issue of representing incertitude or epistemic uncertainty within the context of probability theory. Several treatments of the problem have converged on essentially the same idea: that one can work with *bounds on probability* for this purpose. For example, this idea has been developed independently by Walley and Fine (1982), Williamson (1989), and Berleant (1993). There are suggestions that the idea has its roots in Boole (1854). Hailperin (1986) extensively developed the idea of interval probability in logical expressions. Hyman (1982) developed similar ideas for probabilistic arithmetic expressions in the density domain. Williamson and Downs (1990) introduced interval-type bounds on cumulative distribution functions, which we call "probability boxes", or "p-boxes" for short. They also described algorithms to compute arithmetic operations (addition, subtraction, multiplication and division) on pairs of p-boxes. These operations generalize the notion of convolution between probability distributions. Berleant (1993; 1996; Berleant and Goodman-Strauss 1998) described similar algorithms in the context of automatically verified computations.

The Dempster-Shafer approach to representing uncertainty was articulated by Dempster (1967) and Shafer (1976). The approach has been somewhat controversial, particularly with respect to the appropriateness of Dempster's rule for combining evidence. Yager (1986) described how to compute arithmetic operations on pairs of

Dempster-Shafer structures. Oberkampf et al. (2001) demonstrated how the theory and these convolutions could be used to account for epistemic uncertainty in engineering applications of risk analysis.

The use of p-boxes and Dempster-Shafer structures in risk analyses offers many significant advantages over a traditional probabilistic approach (Ferson and Long 1995; Ferson 2002). They provide convenient and comprehensive ways to handle several of the most practical serious problems faced by analysts, including

1. Imprecisely specified distributions,
2. Poorly known or even unknown dependencies,
3. Non-negligible measurement uncertainty,
4. Non-detects or other censoring in measurements,
5. Small sample size,
6. Inconsistency in the quality of input data,
7. Model uncertainty, and
8. Non-stationarity (non-constant distributions).

This document will illustrate how these problems can be handled with p-boxes and Dempster-Shafer structures.

Section 2 explains the connection between p-boxes and Dempster-Shafer structures on the real line. Section 3 reviews several methods for obtaining p-boxes and Dempster-Shafer structures from empirical information or theoretical knowledge. Section 4 reviews the aggregation methods available for synthesizing uncertainty from multiple information sources that has already been expressed in p-boxes or Dempster-Shafer structures. Section 5 introduces the problem of model uncertainty and discusses possible strategies to account for it in risk assessments.

2 Probability boxes and Dempster-Shafer structures

In the context of the real line, there is an intimate relationship between probability boxes and Dempster-Shafer structures which we will spell out in this section.

2.1 Probability boxes (p-boxes)

Suppose $\overline{F}$ and $\underline{F}$ are nondecreasing functions from the real line $\Re$ into [0,1] and $\underline{F}(x) \le \overline{F}(x)$ for all $x \in \Re$. Let $[\overline{F}, \underline{F}]$ denote the set of all nondecreasing functions F from the reals into [0,1] such that $\underline{F}(x) \le F(x) \le \overline{F}(x)$. When the functions $\overline{F}$ and $\underline{F}$ circumscribe an imprecisely known probability distribution, we call $[\overline{F}, \underline{F}]$, specified by the pair of functions, a "probability box" or "p-box" (Ferson 2002) for that distribution. This means that, if $[\overline{F}, \underline{F}]$ is a p-box for a random variable X whose distribution F is unknown except that it is within the p-box, then $\underline{F}(x)$ is a lower bound on $F(x)$ which is the (imprecisely known) probability that the random variable X is smaller than x. Likewise, $\overline{F}(x)$ is an upper bound on the same probability. From a lower probability measure* $\underline{P}$ for a random variable X, one can compute upper and lower bounds on distribution functions using (Walley 1991, page 203ff)

$$\overline{F}_X(x) = 1 - \underline{P}(X > x),$$
$$\underline{F}_X(x) = \underline{P}(X \le x).$$

As shown on the figure below, the left bound $\overline{F}$ is an upper bound on probabilities and a lower bound on quantiles (that is, the x-values). The right bound $\underline{F}$ is a lower bound on probabilities and an upper bound on quantiles.

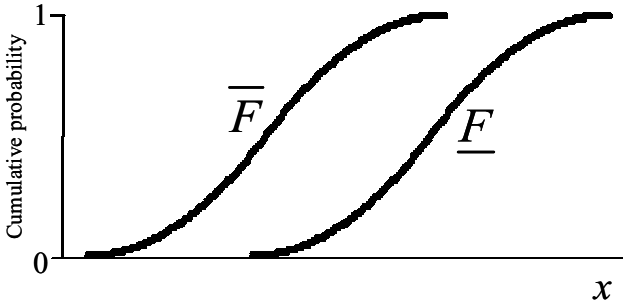


Figure 1: A probability box, or p-box, consisting of a left and right bound.

*Walley (1991) describes the lower probability for an event A as the maximum rate one would be willing to pay for the gamble that pays 1 unit of utility if A occurs and nothing otherwise. The upper probability of A can then be defined as the one minus the lower probability of the complement of A. It is not irrational for someone to assess a lower probability for an event that is strictly less than its upper probability. When this happens, the event is said to have "imprecise probability" for that person.

Williamson (1989; Williamson and and Downs 1990) gave practical algorithms for computing bounds on the result of addition, subtraction, multiplication and division of random variables when only bounds on their input distributions are given. The algorithms employ upper and lower discretizations of the quantile function (i.e., the inverse of the distribution function) in a manner that rigorously contains all discretization error that arises from using a finite number of levels to represent a continuously changing function. These algorithms form the practical basis for a computational risk analysis that need not rely on unwarranted distribution assumptions or over-specification of information.

Walley (1991) emphasized the idea that the use of imprecise probabilities, including distribution function bounds, does not require one to assume the actual existence of any underlying distribution function. One way that such a function might fail to exist is if the random process governing X is not stationary in the sense that its distribution function is changing over time. Although not embracing this idea on its own merits, Williamson and Downs (1990, page 110) did assert, "It is possible to work only with the…bounds themselves and not worry about the distributions within these bounds." This suggests that this approach could be useful in risk analyses even when the underlying stochastic processes are nonstationary or could never, even in principle, be identified to precise distribution functions.

2.2 Dempster-Shafer structures

In an ordinary discrete probability distribution on the real line, a nonzero probability mass is associated with each of the possible points of the distribution. All other values have a probability mass of zero and the probabilities for all the points in the discrete distribution add up to one. A Dempster-Shafer structure (Shafer 1976; Klir and Yuan 1995) on the real line is similar to a discrete distribution except that the locations at which the probability mass resides are *sets of real values*, rather than precise points. These sets associated with nonzero mass are called focal elements. The correspondence of probability masses associated with the focal elements is called the basic probability assignment. This is analogous to the probability mass function for an ordinary discrete probability distribution. Unlike a discrete probability distribution on the real line, where the mass is concentrated at distinct points, the focal elements of a Dempster-Shafer structure may *overlap* one another, and this is the fundamental difference that distinguishes Dempster-Shafer theory from traditional probability theory. Dempster-Shafer theory has been widely studied in computer science and artificial intelligence, but has never achieved complete acceptance among probabilists and traditional statisticians, even though it can be rigorously interpreted as classical probability theory in a topologically coarser space. (In the coarser space, each focal element is identified as a point.)

A finite Dempster-Shafer structure on the real line $\Re$ can be identified with its basic probability assignment, which is a map

$$m\colon 2^{\Re} \rightarrow [0,1]$$

where $m(\varnothing)=0$, $m(a_i)=p_i$ for focal elements $a_i \subseteq \Re$, $i=1,2,...,n$, and $m(D)=0$ whenever $D \neq a_i$ for all i, such that $0 < p_i$ and $\sum p_i=1$. To simplify matters, we'll also assume that the focal elements are closed intervals, rather than more complicated sets. Implementation on a computer of such a Dempster-Shafer structure would thus require storage for $3n$ floating-point numbers, one for each p_i and two for each corresponding interval.

The plausibility function $\text{Pls}:2^{\Re} \rightarrow [0,1]$ corresponding to a Dempster-Shafer structure with basic probability assignment m is the sum of all masses associated with sets that overlap with or merely touch the set $b \subseteq \Re$. Thus,

$$\text{Pls}(b) = \sum_{\substack{a \\ a \cap b \neq \phi}} m(a) = \sum_{\substack{i \\ a_i \cap b \neq \phi}} m(a_i) .$$

The belief function $\text{Bel}:2^{\Re} \rightarrow [0,1]$ is the sum of all masses associated with sets that are subsets of $b \subseteq \Re$ so that

$$\text{Bel}(b) = \sum_{\substack{a \\ a \subseteq b}} m(a) = \sum_{\substack{i \\ a_i \subseteq b}} m(a_i) .$$

Clearly, $\text{Bel}(b) \leq \text{Pls}(b)$. In fact, a Dempster-Shafer structure could also be identified with either of these functions. Given only one of m, Pls or Bel, one can compute the other two.

Yager (1986) defined arithmetic operations between Dempster-Shafer structures that generalize the notion of convolution between distribution functions. Yager (1986) also considered bounds on the distribution function of a random real-valued quantity characterized by a finite Dempster-Shafer structure. For a finite Dempster-Shafer structure with basic probability assignment m and n focal elements a_i having masses p_i, the upper bound for its distribution function is $\text{Pls}(g(z))$, where $g(z)$ is the set of all real numbers less than or equal to z, $g(z)=\{x: x \in \Re, x \leq z\}$. Thus, the function is

$$\text{Pls}(g(z)) = \sum_{\substack{i \\ a_i \cap g(z) \neq \phi}} m(a_i) = \sum_{\substack{i \\ \inf(a_i) \leq z}} p_i .$$

This is a right-continuous nondecreasing step function from the reals into [0,1] with at most n discontinuities located at the points $\inf(a_i)$. At these points the jumps in the function are p_i high. The associated lower bound on the distribution function is

$$\text{Bel}(g(z)) = \sum_{\substack{i \\ a_i \subset g(z)}} m(a_i) = \sum_{\substack{i \\ \sup(a_i) \leq z}} p_i .$$

This is a right-continuous nondecreasing step function from the reals into [0,1] with at most n discontinuities located at the points $\sup(a_i)$. At these points the jumps in the function are p_i high.

In this report, we are interested primarily in Dempster-Shafer structures whose focal elements are *closed intervals*. As we have seen, this restriction permits several convenient simplifications in the formulas above. It allows us to define a Dempster-Shafer structure as a collection of pairs consisting of an interval and a mass $\{([x_1, y_1], m_1), ([x_2, y_2], m_2), \ldots, ([x_n, y_n], m_n)\}$, where $x_i \leq y_i$ for all i, $\Sigma m_i = 1$, and $y_i \neq y_j$ whenever $x_i = x_j$.

2.3 Connection between the two structures

Dempster-Shafer theory was created to account for the fact that measurements taken from the real world are often imperfect. Unless it is a point, a focal element represents a *set* of possible x-values that available evidence or measurement do not distinguish. This indistinguishability among the values in a focal element expresses the epistemic limitations of the evidence. On the other hand, the mass assigned to any particular focal element is a precise number. This amounts to having uncertainty about the x-value and certainty about the p-value. In contrast, the probability bounding approach represented by p-boxes addresses the uncertainty about probabilities and generally presumes that the underlying events can be specified precisely as points z on the real line or half lines $g(z)$. This approach seems to consider the imprecision present in such problems to be primarily about the probabilities, rather than the x-values (contra Williamson 1989). Thus probability boxes express uncertainty about p but a kind of certainty about x.

One might think that it would be natural to combine these two approaches so as to treat uncertainty arising in both x and in p simultaneously, but such an approach is not necessary. There is a duality between the two perspectives (Walley 1991), and, moreover, each can be converted to the other. The relationship between the Dempster-Shafer structures and p-boxes is not one-to-one; there are many Dempster-Shafer structures corresponding to a single p-box. Consequently, translating a Dempster-Shafer structure to a p-box is not an information-preserving operation.* However, the duality between the two kinds of objects appears to be quite useful for applications in risk analysis. The rest of this section demonstrates the interconvertibility between these objects.

If we have a finite Dempster-Shafer structure composed of focal elements that are closed intervals of the real line, then we can specify it by explicitly listing the focal elements and their associated probability masses $\{([x_1, y_1], m_1), ([x_2, y_2], m_2), \ldots, ([x_n, y_n], m_n)\}$. For such a structure, one can always obtain from it an associated p-box. The left bound of the p-box will be defined by the function

---

*For instance, the Dempster-Shafer structure with a single focal element $\{x : 0 \leq x < 0.1, \text{ or } 0.9 < x \leq 1\}$ leads to the same p-box as another one with a single focal element $\{x : 0 \leq x \leq 1\}$, yet the first clearly contains much more information (i.e., that the value cannot be in the interval [0.1, 0.9]). Even when we restrict our attention to Dempster-Shafer structures with interval focal elements, there is some information lost in going to a p-box. For instance, the Dempster-Shafer structures {([1, 3], 0.5), ([2,4], 0.5)} and {([1,4], 0.5), ([2,3], 0.5)} both lead to the same p-box.

$$\overline{F}(z) = \sum_{x_i \le z} m_i .$$

This is the cumulative plausibility function for the Dempster-Shafer structure. The right bound of the p-box will be defined by

$$\underline{F}(z) = \sum_{y_i < z} m_i .$$

This is the cumulative belief function for the Dempster-Shafer structure. Yager (1986) recognized this pair of functions as bounds that generalize a distribution function for a random variable. The functions $\overline{F}(z)$ and $\underline{F}(z)$ define a p-box because they are each nondecreasing functions from the reals into the interval [0,1] and the function $\underline{F}(z)$ is always less than or equal to $\overline{F}(z)$ for every value of z.

Conversely, if we have a p-box, it is always possible to obtain from it a Dempster-Shafer structure that approximates the p-box. For example, consider the p-box depicted below. The left and right bounds of the p-box are shown as thick, black step functions from zero to one along some unspecified real horizontal axis. To find a Dempster-Shafer structure corresponding to this p-box, simply draw a series of horizontal lines, one from each corner of the step function to the other bound. These horizontal lines are shown below as thin gray lines. Sometimes there is a corner on one bound that is at the same height as a corner on the other bound (as in the second gray line from the bottom). In this case the horizontal lines will be coincident. This process describes a collection of rectangles of various sizes and locations. The location of a rectangle along the horizontal axis defines a focal element of the Dempster-Shafer structure. The height of each rectangle is the mass associated with that interval.

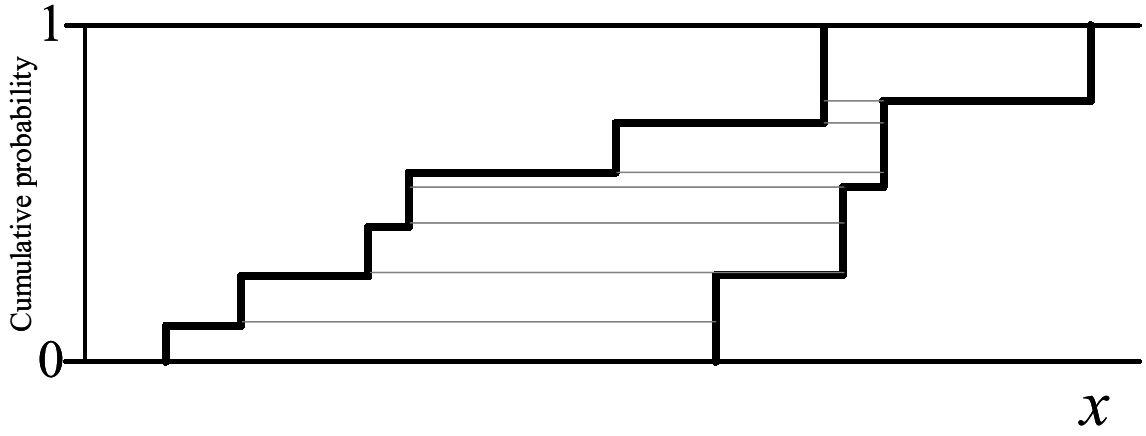


Figure 2: Discretization of a p-box.

When the p-box has curves rather than step functions for its bounds, a discretization is necessary to produce the associated Dempster-Shafer structure, which will therefore be an approximation to the p-box. This idea is illustrated in the figure below. The rectangles have now become thin slivers. These slivers can, by construction, have equal height. In this case, the probability masses of the focal elements in the Dempster-Shafer structure will all be the same. Naturally, the more numerous these slivers, the better the approximation will be. See Section 3.5.6.3 for a prescription about how the endpoints of these approximating slivers should be defined.

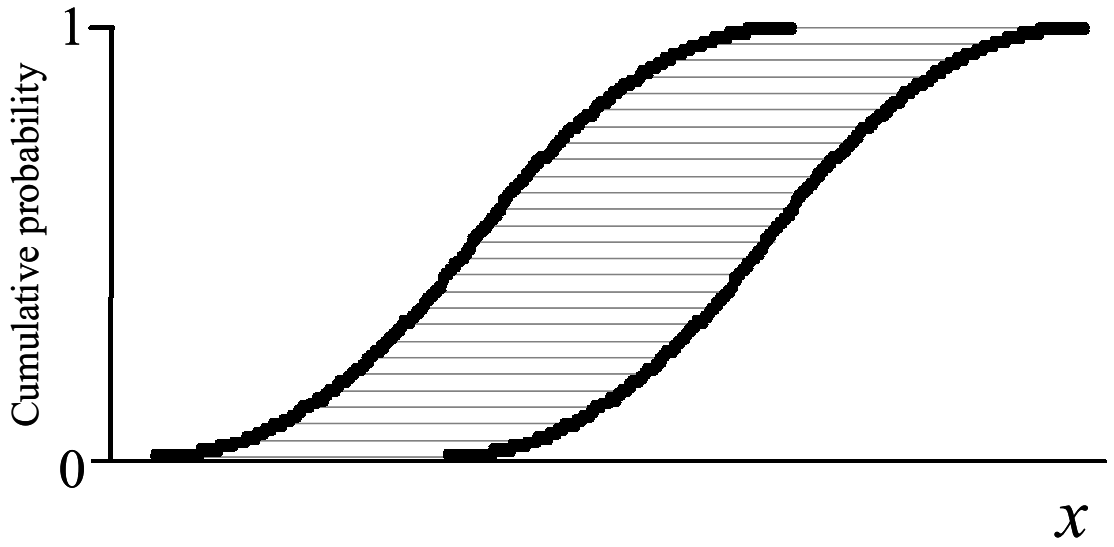


Figure 3: Equiprobable discretization of a p-box with continuous bounds.

For the purposes of discussion in this document, we can define the "canonical" Dempster-Shafer structure associated with a p-box to be that approximation produced by using a discretization with 100 equiprobable thin rectangles or slivers. Formally speaking, the definition of a Dempster-Shafer structure calls for any slivers that are identical in breadth and horizontal location to be condensed into a single focal element. The mass associated with the focal element would be the sum of the all the masses associated with slivers identical to that focal element. However, there would be no mathematical difference if we didn't bother with this condensation step. We could allow redundant elements (the same focal element listed more than one time), and there would be no fundamental difference, except in computational efficiency, between a Dempster-Shafer structure that has been condensed from a discretization and the original discretization that had redundant elements.

Most of the methods described in Sections 3 and 4 of this document for characterization and aggregation were originally conceived for application either to p-boxes or to Dempster-Shafer structures. However, because of the very close connection between these two objects, it turns out that any characterization or aggregation method useful for one of these objects can also be applied to the other. Furthermore, we

conjecture here—but defer the proof for future work—that p-boxes and Dempster-Shafer structures are essentially equivalent for the purposes of risk analysis (Regan et al. 2002).

2.4 Uncertain numbers

Finally, it is worth noting that intervals and (precise) discrete probability distributions are each special cases of both Dempster-Shafer structures and p-boxes. An interval [*a*, *b*] can be identified with a Dempster-Shafer structure having a single focal element consisting of that interval with an associated mass of one. An interval can be identified with a p-box whose left bound is the unit step function at *a* and whose right bound is the unit step function at *b*. A precise distribution function can be identified with a p-box whose bounds are coincident with the distribution function. The Dempster-Shafer structure obtained by canonical discretization of this p-box would have focal elements that are point values. It would just be a discrete probability distribution that approximated the precise distribution function.

In this report, we will use the phrase "uncertain numbers" to refer to a general class of objects that include real numbers, intervals, probability distributions, p-boxes, and finite Dempster-Shafer structures whose focal elements are closed intervals on the real line.

3 Characterization: where do they come from?

Oberkampf et al. (2001) has argued that Dempster-Shafer theory might be a useful formalism for risk assessments. Ferson and Ginzburg (1996; Ferson 2002) have made a similar claim for probability bounds analysis. Once one believes that Dempster-Shafer structures and p-boxes could be useful in the practice of risk analysis, a fundamental question is how to obtain them for use as inputs in calculations. In particular, how can we represent empirical and theoretical information in Dempster-Shafer structures and p-boxes? It turns out that there are basically five ways to do this:

1. Direct assumption,
2. Modeling,
3. Appeal to robust Bayes methods,
4. Constraint propagation, and
5. Observation of measurements.

Each of these approaches will be reviewed and illustrated with numerical examples in this section. Although the most commonly used approaches may well be the first two listed above, the last two, propagation of constraints and observation of measurements, are perhaps the most objective and may merit the most attention. Throughout the rest of Section 3, we shall occasionally note whether the method being reviewed has the following properties:

- **Rigor-preserving**: the resultant Dempster-Shafer structure or p-box is sure to completely bound the uncertainty so long as its specifications are sure bounds,

- **Best Possible**: the resultant Dempster-Shafer structure or p-box could not be any tighter without more information, and

- **Sample Uncertainty**: the resultant Dempster-Shafer structure or p-box represents a statistical confidence claim (such as "95% of the time the uncertain number is constructed it will completely enclose the true value or distribution").

Readers may elect to skip directly to Section 3.6, which outlines suggestions for how an analyst should go about picking a characterization method for use in a particular case. Section 4 on aggregation methods addresses the issue of how Dempster-Shafer structure or p-box estimates about a single quantity obtained from multiple sources can be combined.

3.1 Direct assumption

In this approach, the analyst, or perhaps a consulted expert, simply makes up the input based on what is known about the underlying quantity. This is probably the easiest way

to get an input, and—for the same reason—it is also probably the hardest to defend to others.

3.1.1 Basic idea

The central justifying idea behind this approach is that, sometimes, scientists and engineers may have knowledge about a system including at least a partial mechanistic understanding of how the variation in a particular quantity arises. This knowledge enables the risk analyst to deduce from first principles what statistical distribution that variable should be modeled with. Various examples are given in the subsections below. When the information is detailed enough to specify a distribution completely, the p-box will degenerate to a precise cumulative distribution function (CDF).

In other cases, physics-based or other mechanistic knowledge may be limited in a way that allows the analyst to specify the *shape* or family of the distribution, but not to say precisely what the parameters of the distribution are. In this situation, it is straightforward to construct a p-box that encloses all possible CDFs. For some distribution families, only two CDFs need to be computed to enclose the p-box. For instance, the expression "uniform([a, b], [c, d])" might be used to specify a p-box for a variable that is sure to be uniformly distributed with a minimum value somewhere between a and b, and a maximum value somewhere between c and d. This collection of distributions can be circumscribed by a pair of distribution functions. The left bound of the p-box is the cumulative distribution function for a uniform distribution between a and c. We can use the expression "uniform(a, c)" to denote this. It is the upper bound on probabilities and the lower bound on quantiles. The right bound of the p-box is uniform(b, d), which is the lower bound on probabilities and the upper bound on quantiles. For most distribution families, however, four or more crossing CDFs need to be computed to define the p-box. For instance, if normal([a, b], [c, d]) specifies a p-box for a variable sure to be normally distributed whose mean is in the interval [a, b] and whose standard deviation is in the interval [c, d], then the p-box is illustrated below. The upper part of the left side of the p-box is determined by the CDF of the normal distribution with mean a and standard deviation c. The lower part of the left side is determined by the distribution with mean a and standard deviation d. The right side of the p-box similarly involves the b value for the mean but both values for the standard deviation.

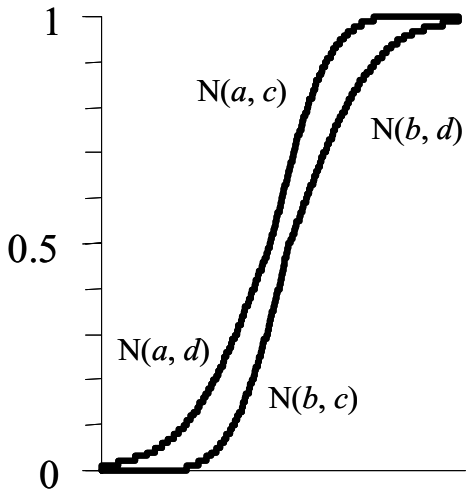


Figure 4: A p-box whose bounds arise from varying parameters of a normal distribution within intervals.

For the roughly two-dozen distribution families commonly used in risk analysis, this simple enveloping strategy is rigor-preserving. That is, if the intervals used are themselves rigorous bounds on the parameters of the distribution, this strategy yields structures that are rigorous bounds on the distribution. Moreover, they are also best possible bounds in the sense that they could not be any tighter and yet still enclose all possible distributions. If, on the other hand, the intervals used to specify parameters arose as confidence intervals (which are statistical rather than rigorous), then the p-box appears to contain sample uncertainty and represent a statistical claim about the underlying distribution, although the exact nature of this claim has not been articulated.

Before we consider some numerical examples, we briefly review some of the circumstances and arguments that can be used to justify some particular distribution families. See also Section 3.4 for other ways to use partial information about a random variable to construct a p-box or Dempster-Shafer structure based on first principles.

3.1.1.1 Normal distribution

According to the central limit theorem, under some reasonable conditions, the sum of many small random variables tends to a normal distribution (also known in some quarters as the Gaussian distribution). The central limit theorem holds so long as the individual variates are roughly independent and have bounded variance, so that none dominates the sum. Thus, when the error in the quantity x can be represented as the sum $x = x_1 + x_2 + \ldots + x_n$ of a large number of small independent error components x_i, the probability distribution for x is close to normal. This is the reason why the normal distribution is so frequently observed and used in practice (see, for example,Wadsworth, 1990). The normal distribution, like all other precise probability distributions, is a degenerate p-box. In some situations, we may only have intervals for the possible values for the parameters of the normal distribution.

3.1.1.2 Cauchy distribution

There are many similar situations that lead to different distribution families. For example, in a setting similar to the central limit theorem, if we keep the variables x_i small in some reasonable sense but allow them to have large (even infinite) standard deviations by allowing thick tails, we get distributions from the class of infinitely divisible distributions. This class includes not only normal distributions, but also Cauchy distributions, with probability density function $\rho(x)=\Delta/\pi(\Delta^2+(x-a)^2)$.

3.1.1.3 Lognormal distribution

Other examples of distributions that can be derived from first principles come from situations in which the error is caused by many small components x_i, but these components are combined with an operation other than addition. For instance, one situation commonly encountered is multiplicative noise, in which a value s is multiplied by some value so that s becomes $s\,k$. For example, when a communication signal passes through the atmosphere, its amplitude changes depending on the specific properties of the medium. Suppose that we have several layers with independent noise values n_i and, correspondingly, independent multiplicative coefficients $k_i=1+n_i$. When a signal passes through each layer, it is multiplied by $1+n_i$. By the time the signal passes through all the layers, it is multiplied by the product of many independent coefficients $k_i=1+n_i$. When we apply logarithms, the product turns into the sum, and the central limit theorem follows. Thus, the distribution for the product of many independent factors tends to be lognormal in shape.

3.1.1.4 Extreme value distributions

Another situation which is very useful in risk analysis applications is the situation of the "weakest link", when a certain event happens if at least one of the numerous quantities x_1, …, x_n exceeds a certain threshold x_0. Thus, the event occurs if the largest $x=\max(x_1, \ldots, x_n)$ of these quantities exceeds x_0. To analyze such events, we therefore need to analyze the distribution of such maxima. It is known that under reasonable conditions, when $n\to\infty$, the distribution of the maximum typically tends to one of three standard distributions: Gumbel or extreme value distribution ($F_{\mathrm{I}}(x) = 1-\exp(-\exp(ax+b))$), the Fréchet ($F_{\mathrm{II}}(x) = 1-\exp(1-(-x)^{-\alpha})$) or the Weibull ($F_{\mathrm{III}}(x) = 1-\exp(1-x^{\alpha})$) distribution, collectively known as Gumbel-type distributions (Galambos, 1978; Wadsworth, 1990). Thus, for large n, we may assume that the distribution of $x=\max(x_1, \ldots, x_n)$ can be described by one of these distributions. Extreme value distributions are widely used in reliability analysis to model a variety of phenomena, including failures under stress, temperature extremes, flood data, telephone data, and electrical insulator lifetimes. Gumbel (1958) and Castillo (1988) review the theory and describe many applications.

3.1.1.5 Uniform distribution

Symmetry or invariance considerations can sometimes justify an assumption about the shape of a distribution. For example, it may sometimes be the case that no values are more probable than any others. This implies of course that all values are equally probable. In such situations, it is reasonable to select a distribution for which the value of

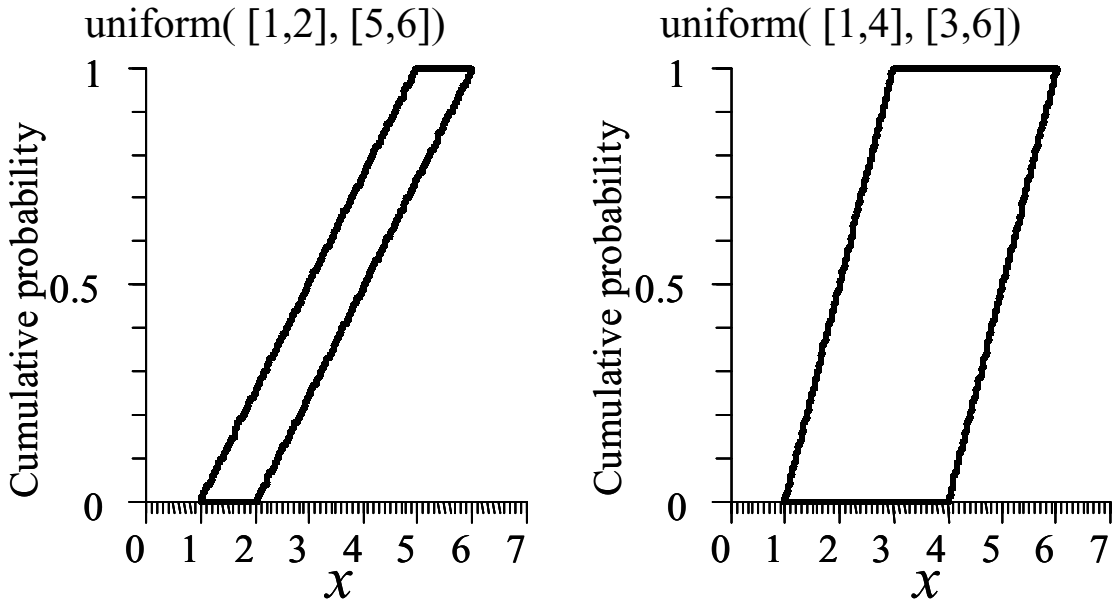
the probability density $\rho(x)$ is the same for all the points x. In this case, we get a uniform distribution on an interval $[a, b]$, for which $\rho(x)=1/(b-a)$ for all $x \in [a, b]$. Probability values, for instance, are often modeled with uniform distributions.

3.1.1.6 Exponential distribution

Some components exhibit neither aging nor "burn-in" improvement, and their risk of failure is constant over time. If the probability of a correctly functioning system becoming faulty between times t and $t+\Delta$ does not depend on t, it may be reasonable to assume that this probability is indeed constant. This assumption leads to the exponential distribution, in which $F(t)=1-\exp(-\lambda t)$ for some constant $\lambda>0$. This distribution might be appropriate, for instance, when component lifetimes are determined by accidental insults rather than as a result of gradual degradation of materials or wearing out of mechanical parts.

3.1.2 Numerical examples

The graphs below depict p-boxes for four cases in which the shape of the distribution is assumed to be known, but the relevant parameters of these distributions are known only to within intervals. For instance, the upper, left-hand graph depicts the case in which the distribution is known to be uniform with a minimum value somewhere in the interval [1, 2] and a maximum value somewhere in the interval [5, 6]. As these intervals grow narrower, the p-box approaches a precise probability distribution in a natural way.



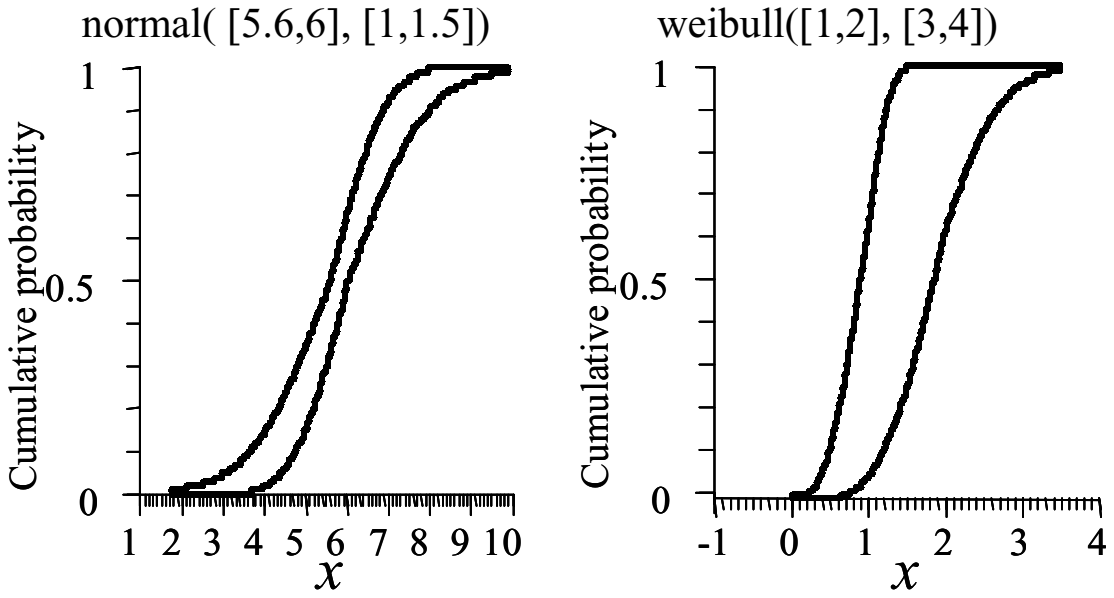


Figure 5: Example p-boxes associated with random variables whose distribution shape is known and whose parameters are known within intervals.

Note that the upper, right-hand graph depicts a similar situation, except that the two intervals for the endpoints of the distribution actually overlap each other. This situation does not, in itself, seem very remarkable; it is certainly plausible that such a case could arise in practice. But it is worth noting that representing this case in a two-dimensional Monte Carlo simulation could be problematic. What would happen, for instance, if in one of the replicates the randomly selected minimum happened to be larger than the randomly selected maximum? A special strategy would be required to handle this anomalous case in the simulation. The reason, of course, is that the minimum and maximum values are not independent of one another. They are related by the constraint that the former must be less than or equal to the latter. There are, in fact, many such constraints among probability distributions. The parameters of a triangular distribution are another obvious example. Frey and Rhodes (1998) have documented some of these dependencies in the beta and other distributions and shown how they can seriously complicate two-dimensional (second-order) Monte Carlo simulations. It is important to note that no evasive maneuver is necessary to represent such a case with p-boxes, however. By representing only the *hull* of the uncertainty about the underlying variable, the p-box avoids entanglement in this particular kind of complication. Ferson (2002) allows the computation of best possible p-boxes for several dozen named distribution families ranging from Bernoulli, beta and binomial to uniform, Wakeby and Weibull.

The lower, left-hand p-box circumscribes a normal distribution whose mean is sure to lie within the interval [5.6, 6] and whose standard deviation is within the interval [1, 1.5]. The lower, right-hand p-box describes a Weibull distribution whose scale parameter (or characteristic life) is between 1 and 2, and whose shape parameter is between 3 and 4. Note that, although a normal distribution theoretically has infinite tails, the tails here are truncated to a finite range. The Weibull distribution is bounded on the left by zero, but is unbounded on the right. It too is truncated at some upper quantile. Where these distributions are truncated is a choice for the analyst, and is part of the assumption used

to select these structures. See Section 3.4.4 for an argument why it is reasonable to truncate infinite tails.

It is straightforward to compute the approximating Dempster-Shafer structure associated with these p-boxes in the canonical way (see Section 2.3). For instance, the canonical Dempster-Shafer structure associated with the upper, left-hand p-box uniform([1,2], [5,6]) would be the collection of 100 interval-mass pairs{ ([1, 2.04], 0.01), ([1.04, 2.08], 0.01), ([1.08, 2.12], 0.01), ([1.12, 2.16], 0.01), …, ([4.88, 5.92], 0.01), ([4.92, 5.96], 0.01), ([4.96, 6], 0.01) }.

3.1.3 Caveats

Bertrand Russell expressed the fundamental caveat about making assumptions in a famous quip*: "Assumption has many advantages. Chiefly these are the same as those of theft over honest toil." The essence of the problem is that an assumption merely *says* something is so, which of course is rather different from really *knowing* it to be so from empirical observation. Any conclusions that come from risk analyses based on inputs selected by assumption are necessarily contingent on these assumptions being true. This limitation, although obviously profound, nevertheless does not seem to have had much of a chilling effect on the propensity of analysts to employ this method for selecting inputs.

3.1.3.1 Human tendencies to underestimate uncertainty

It is well known in the psychometric literature (Kahneman et al. 1982; Henrion and Fischhoff 1986; inter alia) that humans, including both experts and lay persons, routinely and substantially underestimate uncertainty. This occurs, for instance, in almost every scientific and engineering discipline, ranging from predictions about the weather, to stock market forecasting and voter preference polling, from calculations about the speed of light to projections about how fish will be biting for anglers. Whenever humans rely on their perceptions to characterize uncertainty rather than studying it systematically and quantitatively, they will tend to make predictable errors. In general, subjectively assessed interquartile ranges are typically estimated to be much narrower than they actually are and therefore to include less than the nominal 50% of the actual variation, sometimes as little as half of what is expected. Underestimation of the tail risks is even worse. Subjectively assessed 98% confidence regions fail to include the true values 20 to 45% of the time, rather than the mere 2% nominally expected. This overconfidence is ubiquitous among humans, and pervasive even in scientific and engineering disciplines. Its consequences for risk analysis are obvious and substantial, especially when inputs are selected by assumption or, more generally, whenever uncertainty is assessed subjectively (see Gigirenzer 2002; Morgan and Henrion 1990, pages 102ff).

3.1.3.2 Overzealous use of the central limit theorem

The central limit theorem is widely abused in risk analysis as a justification for a normal or lognormal distribution shape. The mere fact that a variable can be partitioned into a sum of several quantities, or factored into a product of several quantities, will generally not ensure that the variable's distribution will be normal or lognormal. The ancillary

*The exact wording is variously quoted, perhaps because Russell found the need to restate it on many occasions.

conditions of the theorem, including independence and similarity of variance of the parts, are also needed for the application to be justified.

3.1.3.3 Tails wagging the distribution

Exactly where the tails of an uncertain number are truncated can, in some unusual situations, have a non-negligible consequence for the rest of the p-box or Dempster-Shafer structure. The graph below illustrates the effect. The graph is a detail of the left tails of the results of two convolutions*. Depicted are the cumulative plausibility and belief functions resulting from adding (under independence) two uncertain numbers. The first uncertain number is *A*, which is represented as a canonical Dempster-Shafer structure approximating a normal distribution with mean in the interval [10, 11] and variance in [1, 4]. The second uncertain number, *B*, is represented as a canonical Dempster-Shafer structure approximating a uniform distribution over the range [10, 20]. Because a normal distribution in theory has infinite tails, it will be necessary to truncate the range of *A* to some finite range. (This is necessary for the computer to be able to handle the computation. But, because there are no physical variables with infinite ranges, the truncation would be desirable from a modeling perspective even if the computer could handle calculations with unbounded quantities.) But where should the tails of *A* be truncated? Will our choice make a difference to the results of calculations involving *A*? When *A* is truncated to the range [−10, 30], the convolution with *B* yields the result shown with the gray lines. When the first addend is truncated outside [4.85, 16.15] instead, the convolution yields the result shown with black lines. The figure shows that the choice about where the tails are truncated makes a difference not only in the leftmost focal element but also in the other focal elements comprising the tail of the structure. The effect is most pronounced in convolutions under independence and is largest near the tails, but is almost always small in magnitude even there. In most cases, except for the extreme focal element, the effect is so small that it is typically obscured by the width of the lines graphing the result. In rare situations, however, it may be important to assess the effect on the final results of the assumption about where the tails are truncated.

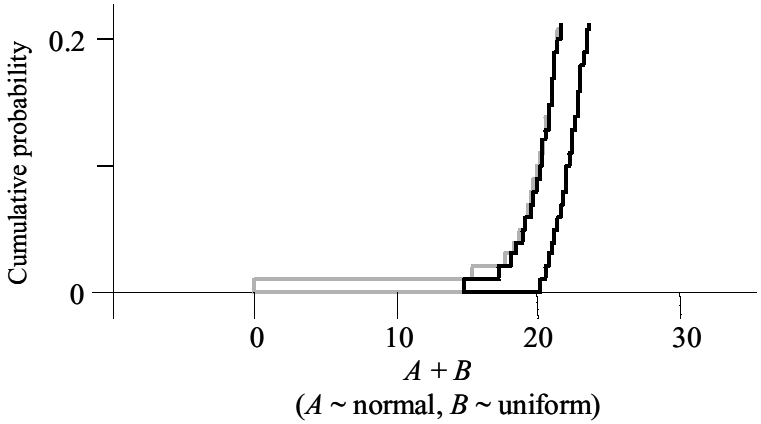


Figure 6: Left tails of p-boxes arising from convolution of *A* and *B* (see text) with different truncation limits for *A*.

*Convolution is the generalization of addition and other arithmetic operations for distributions. See section 3.2.1.1 for a discussion of convolution.

3.2 Modeling

3.2.1 Basic idea

The most important uncertain quantities characterized in risk analyses are estimated by modeling. For instance, we do not estimate the risk of system failure of the space shuttle by building many shuttles and deploying them in field tests to estimate the risk empirically. Nor would we trust the method of direct assumption (discussed in Section 3.1) in which an analyst's intuition or principled argument is used to say what the distribution of risks looks like. In the case of the space shuttle, as indeed in most risk analyses, the estimation is obtained by *modeling* the quantity or distribution of interest. In this approach, the estimation problem is broken into subproblems, which, it is hoped, are easier to solve. The desired estimate is then reconstructed from these pieces. Morgan and Henrion (1992, page 116) call this approach disaggregation, and recount the reasons that it is in common use in quantitative assessments. Of course, this approach begs the question in the sense that it merely transfers the chore of getting the inputs to a different level. Nevertheless, the approach is so important and so pervasive that it merits consideration here. It is, after all, a fundamental approach used throughout risk analysis and scientific modeling in general. We typically do not have enough direct empirical information or expert knowledge about a quantity of concern and we resort to analysis, i.e., breaking it down into its component parts, to study it more easily. It is difficult, for instance, to produce a direct empirical estimate of the chance that a nuclear reactor core melts down; the sample size is, thankfully, too small. But we can make estimates that switches or pumps malfunction, that containments are breached, that fires or electrical surges occur. Using these estimates, together a physics-based understanding of how the reactor functions, we can synthesize a substantially better estimate of the risk of a core melt.

The modeling approach to be employed in any particular case is determined by the modeler who understands something about the underlying physics or engineering. For instance, it is the modeler who decides whether the decomposition will be in terms of breaking a sum into its addends, or a product into its factors, or a quotient into its numerator and divisor, or some other kind of decomposition. There are many ways to decompose a quantity in question into other quantities. We outline here an assortment of the elementary ways. In practice, modeling often consists of several or even many of these steps, which are sequentially applied to build up the desired estimate out of the subproblems. These operations are convolution, transformations, enveloping, intersection, mixture, composition and deconvolution. Most of these operations are also described later in Section 4 because they are aggregation operators.

3.2.1.1 Convolution

In probability theory, convolution is the operation between distribution functions F and G, corresponding respectively to independent random variables X and Y, that yields the distribution of the sum $X+Y$. The notion can be generalized in several ways. First, we can speak of convolutions for functions other than sum, such as difference, product,

quotient, minimum, maximum, power, etc. Second, the random variables might be dependent rather than independent. And third, we might only have Dempster-Shafer structures or p-boxes rather than precise distribution functions. We will use the word convolution to denote any of several operations that generalize ordinary binary arithmetic operations on the reals, such as addition, subtraction, etc., to uncertain numbers such as p-boxes and Dempster-Shafer structures.

Yager (1986) explained how to compute convolutions for Dempster-Shafer structures under the assumption of independence. The convolution is a Cartesian product of the focal elements of the two operands. The requisite calculations are illustrated with a numerical example in Section 3.2.3. The algorithm for convolving p-boxes under independence (Williamson and Downs 1990; Berleant 1993; 1996) is essentially the same as that described by Yager.

Using a convolution to model a desired quantity depends on the analyst being able to assert knowledge about the form of a model that relates the desired quantity to other quantities. Aside from knowing which operation (addition, subtraction, etc.) is involved, it is also important to know how the variables might be dependent on one another. Any assumption about this dependence needs justification, just as does the form of the model itself. Assuming independence for the sake of mathematical convenience may seriously underestimate or overestimate the uncertainty in the convolution. Fortunately, it is never necessary to make this assumption if it is not justified. Williamson and Downs (1990) give an algorithm for computing pointwise best possible bounds on the distribution of a sum (or difference, product, quotient, etc.) using a mathematical result due to Frank et al. (1987) based on copulas* (Nelsen 1999). The result is the best possible over all possible dependency or correlation structures. Berleant and Goodman-Strauss (1998) describe an alternative strategy that is based on linear programming.

These convolution algorithms for the general case are rigor-preserving in the sense that if the inputs are sure to enclose their respective quantities, the result of the convolution will also enclose the desired quantity, so long as the model is correct (Williamson and Downs 1990). Moreover, based on a theorem due to Frank et al. (1987), the algorithms will produce pointwise best possible results for any single convolution.

When convolutions are chained together in a sequence to represent complex models, the results can be best possible so long as there are not multiple occurrences of a single uncertain quantity among the operands. This sensitivity to repeated variables seems to be a common feature of uncertainty calculi (see Manes 1982; Moore 1966). The Cartesian-product algorithms for the independent case are not even rigor-preserving in the case of repeated variables. The problem arises because of the inconsistency of using a formulation that assumes quantities are independent on variables (the repeated ones) that are manifestly not independent. This complication has been observed in a variety of schemes to perform arithmetic with probability information, including discrete probability distributions (Kaplan 1981) and stepwise Monte Carlo simulation (Ferson 1996). Hyman's (1982) significance arithmetic arose as a strategy to control this problem. A very simple strategy that is often applicable is to use algebra rearrangements to re-express the model in a form that does not contain repeated variables. Another approach that is often appropriate when the model cannot be re-expressed in this way is to model the repeated variables as *perfectly correlated* rather than independent.

*A copula is the function that joins together marginal distributions to form a joint distribution function.

3.2.1.2 Transformation

A model used to estimate the quantity of interest may be more complicated than a string of convolutions. For instance, it may also include transformations such as logarithm, sine or absolute value. Unlike a convolution, which takes two inputs, a transformation may take only one. Monotone transformations such as square root, logarithm, exponential, etc. are easy to compute for Dempster-Shafer structures and, via canonical discretization, p-boxes. The transformation is applied to each focal element according to the rules of interval arithmetic (Moore 1966). For instance, if $D=\{([x_1, y_1], m_1), ([x_2, y_2], m_2), \ldots, ([x_n, y_n], m_n)\}$ is a Dempster-Shafer structure, then $\ln(D)$ can be computed as $\{([\ln x_1, \ln y_1], m_1), ([\ln x_2, \ln y_2], m_2), \ldots, ([\ln x_n, \ln y_n], m_n)\}$, so long as every focal element is strictly greater than zero. The probability masses are not changed in the transformation. Such transformations are generally rigor-preserving and best possible.

Non-monotone transformations are handled essentially the same way, although they can be trickier to implement, because different focal elements in the input can sometimes map to the same focal element in the output. Any transformation that is defined on intervals can be extended to Dempster-Shafer structures and therefore p-boxes by a canonical discretization.

3.2.1.3 Envelope

A model may also include other kinds of operations such as, for instance, an envelope operation. This operation is used when the analyst knows that *at least one* of multiple p-boxes encloses the distribution of the quantity in question. Enveloping can be applied to Dempster-Shafer structures by first converting them to p-boxes. If there are n p-boxes $[\overline{F}_1, \underline{F}_1]$, $[\overline{F}_2, \underline{F}_2]$, …, $[\overline{F}_n, \underline{F}_n]$, then their envelope is defined to be $[\overline{F}^*, \underline{F}^*]$, where $\overline{F}^*(x) = \max(\overline{F}_1(x), \overline{F}_2(x), \ldots, \overline{F}_n(x))$ and $\underline{F}^*(x) = \min(\underline{F}_1(x), \underline{F}_2(x), \ldots, \underline{F}_n(x))$. This operation is clearly rigor-preserving and yields best possible results. Enveloping is discussed more extensively in Section 4.4.

3.2.1.4 Intersection

This operation is used when the analyst knows that *each* of multiple p-boxes encloses the distribution of the quantity in question. This kind of intersection can be applied to Dempster-Shafer structures by first converting them to p-boxes. If there are n p-boxes $[\overline{F}_1, \underline{F}_1]$, $[\overline{F}_2, \underline{F}_2]$, …, $[\overline{F}_n, \underline{F}_n]$, then their intersection is defined to be $[\overline{F}^*, \underline{F}^*]$, where $\overline{F}^*(x) = \min(\overline{F}_1(x), \overline{F}_2(x), \ldots, \overline{F}_n(x))$ and $\underline{F}^*(x) = \max(\underline{F}_1(x), \underline{F}_2(x), \ldots, \underline{F}_n(x))$. The operation is undefined if the resulting bounds cross each other so that strictly $\overline{F}^*(x) < \underline{F}^*(x)$ for any x. This operation is both rigor-preserving (Rowe 1988) and yields best possible results. This intersection is discussed further in Section 4.3.

3.2.1.5 Mixture

A model used to estimate an uncertain quantity may also include a mixture operation. Mixture models are appropriate for a quantity that is selected from one of several different values, with known probabilities. In the context of probability distributions, mixing has also been called “averaging” because the values of the distribution functions

are averaged together for every x-value. The result of mixing n p-boxes $[\overline{F_1}, \underline{F_1}]$, $[\overline{F_2}, \underline{F_2}]$, …, $[\overline{F_n}, \underline{F_n}]$, with respective weights w_1, w_2, …, w_n, is $[\overline{F}^*, \underline{F}^*]$, where $\overline{F}^*(x) = (w_1\ \overline{F}_1(x) + w_2 \overline{F}_2(x) + \ldots + w_n\ \overline{F}_n(x)) / \Sigma\ w_i$ and $\underline{F}^*(x) = (w_1\ \underline{F}_1(x) + w_2\ \underline{F}_2(x) + \ldots + w_n\ \underline{F}_n(x)) / \Sigma\ w_i$. The weights must be positive.

The weighted mixture of n finite Dempster-Shafer structures with basic probability assignments m_1, m_2, …, m_n has the basic probability assignment

$$m^*(A) = \frac{1}{\sum_i w_i} \sum_i w_i m_i(A) .$$

Again, the weights must be positive. When the focal elements are intervals, an equally weighted mixture of two finite Dempster-Shafer structures $\{([a_1, b_1], m_1), ([a_2, b_2], m_2), \ldots, ([a_n, b_n], m_n)\}$ and $\{([c_1, d_1], k_1), ([c_2, d_2], k_2), \ldots, ([c_r, d_r], k_r)\}$ is just $\{([a_1, b_1], m_1/2), ([a_2, b_2], m_2/2), \ldots, ([a_n, b_n], m_n/2), ([c_1, d_1], k_1/2), ([c_2, d_2], k_2/2), \ldots, ([c_r, d_r], k_r/2)\}$, assuming the two structures have no focal elements in common. If they have common elements, a condensation step that sums the masses might be needed.

When the mixture weights are known precisely, this operation is rigor-preserving and best possible. It is also possible to compute mixtures when the weights are known only to within intervals, but this is complicated by the constraint that weights must sum to one. When bounds for the weights are obtained from statistical confidence procedures, the resulting p-box or Dempster-Shafer structure will contain sample uncertainty.

Mixing is also discussed in Section 4.7.

3.2.1.6 Composition

A model can also contain a composition, in which a Dempster-Shafer structure or p-box is used as a *parameter* in specifying another Dempster-Shafer structure or p-box. For instance, suppose that we know an uncertain quantity has an exponential distribution, but the only available estimate of its parameter is itself a Dempster-Shafer structure. A strategy to effect such a composition depends on the fact that a Dempster-Shafer structure (or a p-box once discretized) is but a collection of intervals with associated probability masses. This collection asserts that the uncertain quantity is within each of these intervals with the given probability. This interpretation allows us to compute the composition as the weighted mixture of the uncertain numbers generated by using the interval focal elements as the parameters. The weights used in the mixture are the masses associated with those focal elements. This strategy should work in all situations where interval parameters can be used as inputs. (A numerical example of composition is given below in Section 3.2.3.)

If, like a transformation, a composition uses a single input, there is no need to consider any question about dependence. However, if there are multiple inputs in a composition, dependencies can become important. This can happen, for instance, if both parameters of Weibull distribution are estimated by Dempster-Shafer structures or p-boxes. In such a case, a Cartesian product such as that employed to compute convolutions will be necessary to realize the uncertain number. Research is needed to understand whether and how distributional constraints such as described by Frey and

Rhodes (1998) should be taken into account when p-boxes and Dempster-Shafer structures are composed.

3.2.1.7 Deconvolution

Deconvolution is an operation that untangles a convolution. For instance, if it is known that the convolution of X and Y yields Z, deconvolution is an operation that can obtain an estimate of, say, X based on empirical estimates of both Y and Z. Deconvolutions are also called "backcalculations" in the risk analysis literature (Burmaster et al. 1995; Burmaster and Thompson 1995). Typically, deconvolutions are defined to maintain some critical property in the system. For instance, one might ask what constraints on X will guarantee that the Z that results from the forward convolution will not exceed its constraints. Algorithms to compute deconvolutions are under development, but it is already clear that they cannot generally be formulated in a way that yields best possible results.

3.2.2 Caveats

Breaking a problem into subproblems won't be beneficial unless there is more information available to solve the subproblems than there is to solve the primary problem. It is not always altogether clear when one should break an estimation problem into subproblems and when it would be better to simply use whatever information is available to estimate a quantity directly. For most quantities, there is a choice between modeling and other approaches to estimation. Sometimes the level of decomposition needs to be taken further to tackle the subproblems by the same strategy. But, clearly, sometimes decomposition can be carried too far. Mosleh and Bier (1992) reviewed the problem of finding the optimal level to which to decompose.

This approach introduces the additional wrinkle of whatever uncertainty there may be about the model used for the reconstruction. This can be especially problematic when independence assumptions are used indiscriminately or are used without specific justification. Without a strategy to address this model uncertainty, any results would be contingent of course on the analyst having gotten the model correct.

3.2.3 Numerical examples

We will consider three numerical examples in this section. In each of the graphs shown, the ordinate will be cumulative probability. The abscissa will be the axis of the quantity in question.

The three graphs below depict the modeling of a sum. The primary problem was to estimate a quantity Z which is known to be a sum $X + Y$ where the addends are independent. First, reliable estimates of X and Y must be obtained, using any available method (including modeling). The quantity X is modeled as a lognormal distribution whose mean is the interval [20, 23] and whose standard deviation is in the interval [3.5, 4.5]. The distribution is truncated at the 0.5^{th} and 99.5^{th} percentiles. The canonical Dempster-Shafer structure for X is {([11.0, 15.0], 0.01), ([11.6, 16.7], 0.01), ([12.4, 17.1], 0.01), ..., ([27.9, 35.7], 0.01), ([29.2, 37.5], 0.01)}. This is just the collection of intervals that when assigned masses of 0.01 best approximates the imprecisely specified probability distribution X. The quantity Y has a symmetric triangular distribution, with minimum value 10, mode 20 and maximum 30. The focal elements of this object are

very narrow intervals. Indeed, they could have been points, but are just wide enough to contain the representation error introduced by the canonical discretization. The Dempster-Shafer structure is therefore {([10, 11.4], 0.01), ([11.4 12.0], 0.01), ([12.0 12.4], 0.01), ..., ([28.0 28.6], 0.01), ([28.6 30.0], 0.01)}. Once these estimates are in hand, we can use convolution to compute the estimate of their sum. The resulting sum is shown in the graph on the right.

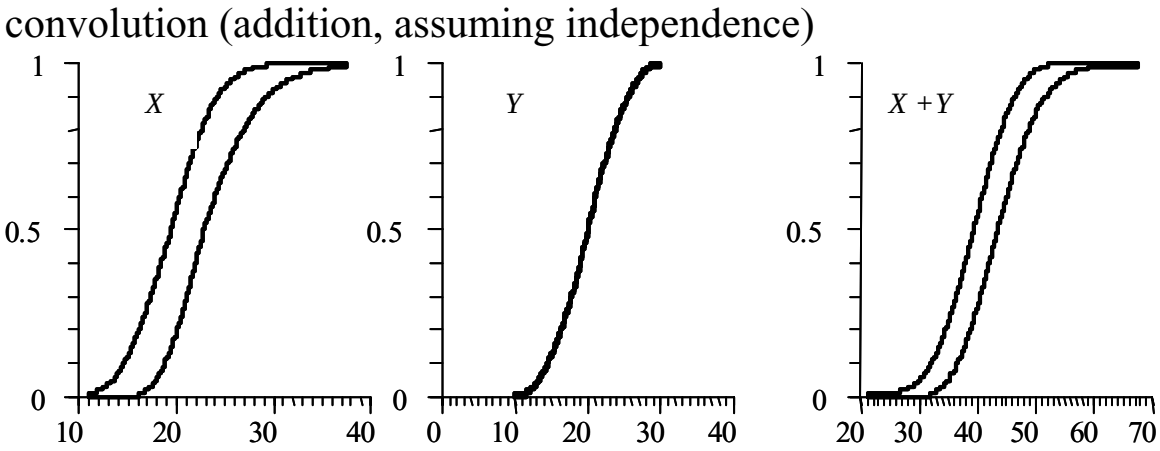


Figure 7: Convolution (right) of *x* (left) and *y* (middle).

How exactly is this addition computed? The matrix below shows a few of the calculations. The first line in each cell is an interval focal element and the second line is the probability mass associated with that focal element. The elements of *X* are arrayed along the top row of the matrix. The elements of *Y* are in the first column. The cells inside the matrix form the Cartesian product, crossing each element from *X* with every element from *Y*. The first line of a cell inside the matrix is determined by interval arithmetic on the corresponding focal elements from *X* and *Y*. Because the model asserts that the quantity is the sum of *X* and *Y*, each of these interval operations is addition. The second line in each cell is the probability mass associated with the interval on the first line. Note that the probability masses in the top row and first column are each 0.01; these are the mass that arose from the canonical discretizations. The masses inside the matrix are all 0.0001, which is the product (under independence) of 0.01 and 0.01. Because there are 100 focal elements in both *X* and *Y*, there will be 10,000 focal elements in their sum. Williamson (1989) describes a condensation strategy that can reduce this number back to 100 in a way that conservatively captures uncertainty.

addition, independent	**[11.0, 15.0] 0.01**	**[11.6, 16.7] 0.01**	**[12.4, 17.1] 0.01**	…	**[27.9, 35.7] 0.01**	**[29.2, 37.5] 0.01**
[10, 11.4] 0.01	[21, 26.4] 0.0001	[21.6, 28.1] 0.0001	[22.4, 28.5] 0.0001	…	[37.9, 47.1] 0.0001	[39.2, 48.9] 0.0001
[11.4 12.0] 0.01	[22.4, 27] 0.0001	[23, 28.7] 0.0001	[23.8, 29.1] 0.0001	…	[39.3, 47.7] 0.0001	[40.6, 49.5] 0.0001
[12.0 12.4] 0.01	[23, 27.4] 0.0001	[23.6, 29.1] 0.0001	[24.4, 29.5] 0.0001	…	[39.9, 48.1] 0.0001	[41.2, 49.9] 0.0001
⋮	⋮	⋮	⋮		⋮	⋮
[28.0 28.6] 0.01	[39, 43.6] 0.0001	[39.6, 45.3] 0.0001	[40.4, 45.7] 0.0001	…	[55.9, 64.3] 0.0001	[57.2, 66.1] 0.0001
[28.6 30.0] 0.01	[39.6, 45] 0.0001	[40.2, 46.7] 0.0001	[41, 47.1] 0.0001	…	[56.5, 65.7] 0.0001	[57.8, 67.5] 0.0001

Other convolutions, such as subtraction, multiplication, division, minimization, maximization, powers, etc., can also be used. Algorithms to handle these cases are more-or-less straightforward generalizations of that for addition (Yager 1986; Williamson and Downs 1990; Berleant 1996; Ferson 2002). It is also possible to compute a convolution with a dependence assumption other than independence.

The second example is illustrated in the next set of three graphs below. In this case, the analyst is confident that the desired quantity is an equal stochastic mixture of two other quantities. In other words, she is sure that the desired quantity is selected with equal probability from one of two other quantities, *X* and *Y*. In this example, these quantities have been estimated, respectively, by an interval [1, 3] and a quantity known to have a normal distribution with a mean in the interval [5.6, 6] and a standard deviation in the interval [1, 1.5]. The mixture operation involves (vertically) averaging the probability values of the respective bounds of the p-boxes *X* and *Y*. This is equivalent to forming a Dempster-Shafer structure by pooling the focal elements from both *X* and *Y* and simultaneously halving their masses (to keep the total of all masses equal to one). The nature of this operation is perhaps obvious from the picture of the result below and we need not give more detail about the particular algorithm used. The parts of the resulting mixture are so distinguishable because the inputs *X* and *Y* did not overlap much.

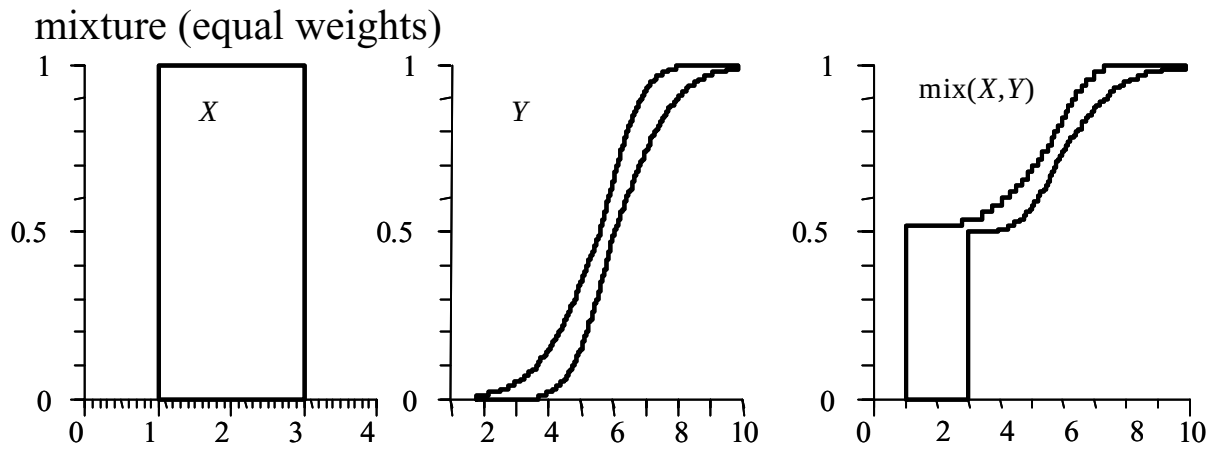


Figure 8: Stochastic mixture (right) of *x* (left) and *y* (middle) assuming even 50:50 weights.

The last example in this section on estimation by modeling is composition. It is illustrated in the two graphs below. The desired estimate, shown on the right, was modeled as a normal distribution with variance equal to one and mean equal to another uncertain quantity X. The estimate for X (on which the desired estimate was based) is shown on the left. It is the Dempster-Shafer structure {([1,2], 1/3), ([4,6], 1/3), ([9,10], 1/3)}. The composition was computed as the equal mixture (see Sections 3.2.1.5 and 4.7) of three p-boxes: normal([1,2], 1), normal([4,6], 1) and normal([9,10], 1), that is, three normal distributions with means in the intervals [1,2], [4,6] and [9,10] and variances all equal to one.

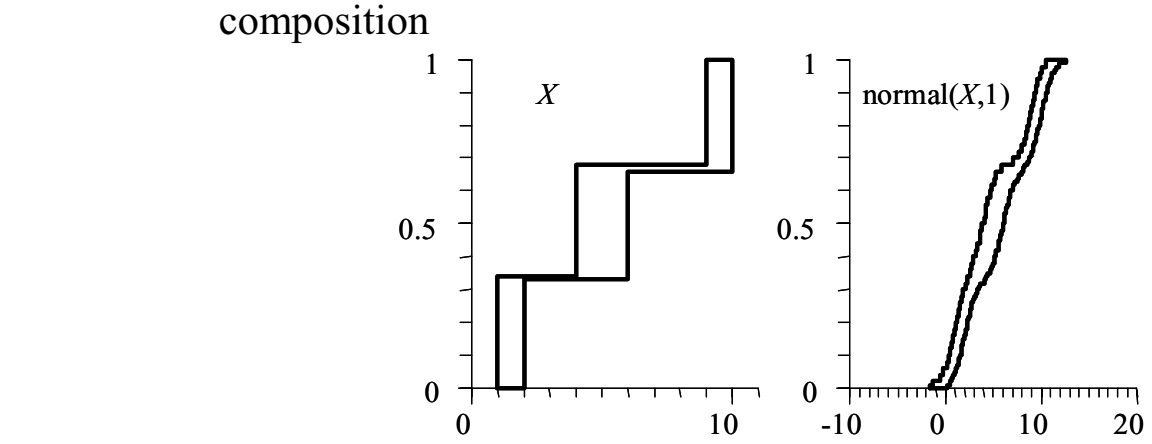


Figure 9: Composition (right) of the trimodal distribution *x* as the mean of a normal distribution.

*3.3 Robust Bayes methods**

In this section, we explain how to obtain a p-box or Dempster-Shafer structure from the objects already developed by analysts using robust Bayes methods. Bayesian methods are an important way—some would say an essential way—by which inputs for a risk analysis are selected. In a regular application of Bayes' rule, a prior distribution and a likelihood function are combined to produce a posterior distribution, which may then be used as an input in a risk analysis. Bayes' rule is

$$p(\theta \mid E) = p(\theta)\, p(E \mid \theta) \,/\, p(E)$$

where p denotes probability mass (or density), θ is a value of the quantity in question, E denotes the evidence being considered, $p(\theta)$ is the prior probability for a value θ, $p(E \mid \theta)$ is the conditional likelihood function that expresses the probability of the evidence given a particular value of θ, and $p(E)$ is the probability of having obtained the observed

*Bayes' rule is also a method of aggregation and is reviewed in that context in Section 4.6.

evidence. This divisor, sometimes called the normalizing factor, is the sum (or integral) with respect to θ of the product of the prior and the probability of observing a value if the value were actually θ. This rule is applied for all values of θ to obtain $p(\theta \mid E)$, which is the distribution of θ given the evidence. For most Bayesians, the prior distribution is obtained from the opinion or belief of the analyst. It is intended to represent, at least initially, the analyst's *subjective* knowledge before any specific evidence is considered. It may be the result of amorphous preconceptions or physics-based reasoning or a combination of the two. The likelihood function represents a model, also perhaps taken from the subjective knowledge of the analyst, of what data implies about the variable in question. It is part of the Bayesian dogma of ideal precision (Walley 1991) that both of these functions are precise.

The connection of Dempster-Shafer structures and p-boxes to Bayesian methods is through the approach developed by Berger (1985) and others known as robust Bayes. In robust Bayesian analysis, the insistence on having a single, precise prior distribution and a single, specific likelihood function is relaxed. In their places, entire classes of distributions and functions are used. In this section, we explain how to obtain a p-box or Dempster-Shafer structure from the objects already developed by analysts using robust Bayes methods.

3.3.1 Basic idea

Robust Bayes methods acknowledge that it is sometimes very difficult to come up with precise distributions to be used as priors (Insua and Ruggeri 2000). Likewise the appropriate likelihood function that should be used for a particular problem may be in doubt. In robust Bayes, standard Bayesian analysis is applied to all possible combinations of prior distributions and likelihood functions selected from *classes* of priors and likelihoods considered empirically plausible by the analyst. This approach has also been called "Bayesian sensitivity analysis". In this approach, a class of priors and a class of likelihoods together imply a class of posteriors by pairwise combination through Bayes' rule. Robust Bayes also uses a similar strategy to combine a class of probability models with a class of utility functions to infer a class of decisions, any of which might be the answer given the uncertainty about best probability model and utility function. In both cases, the result is said to be robust if it's approximately the same for each such pair. If the answers differ substantially, then their range is taken as an expression of how much (or how little) can be confidently inferred from the analysis. Although robust Bayes is clearly inconsistent with Bayesian idea that uncertainty should be measured by a single additive probability measure and that personal attitudes and values should always be measured by a precise utility function, the approach is often accepted as a matter of convenience (e.g., because the cost or schedule do not allow the more painstaking effort needed to get a precise measure and function). Some analysts also suggest that robust methods extend the traditional Bayesian approach by recognizing incertitude as of a different kind of uncertainty.

The p-box associated with the resulting class of posteriors can be found simply by cumulating each of the posterior distributions and forming the envelope or convex hull of these cumulated posterior distributions. The associated Dempster-Shafer structure can be obtained from the p-box in the canonical way.

3.3.2 Numerical example

Suppose that the prior distribution is within the class of all normal distributions having a mean within [−1, +1] and a variance within [1, 2.5]. Suppose further that the likelihood function is also characterized by a normal shape, with a mean in the interval [14, 16] and variance in the interval [1.7, 3]. In the illustration below, a few prior distributions and likelihood functions from their respective classes are drawn on a θ-axis in terms of density (the vertical axis not shown). Also shown on the same axis are several representatives of the (infinitely large) class of posterior distributions that are obtained by applying Bayes rule to every possible pair of prior distribution and likelihood function. Because the priors and likelihoods in this example are conjugate pairs, it is easy in this example to compute the posteriors. When a prior is normal with a known variance v_1 and a likelihood is normal with a known variance v_2, the posterior distribution is normal with variance $1/(1/v_1 + 1/v_2)$. Its mean is $(m_1/v_1 + m_2/v_2)/(1/v_1 + 1/v_2)$, where m_1 and m_2 are the means of the prior and likelihood respectively. These expressions allow us to compute the posterior that would arise from each combination of a prior and a likelihood.

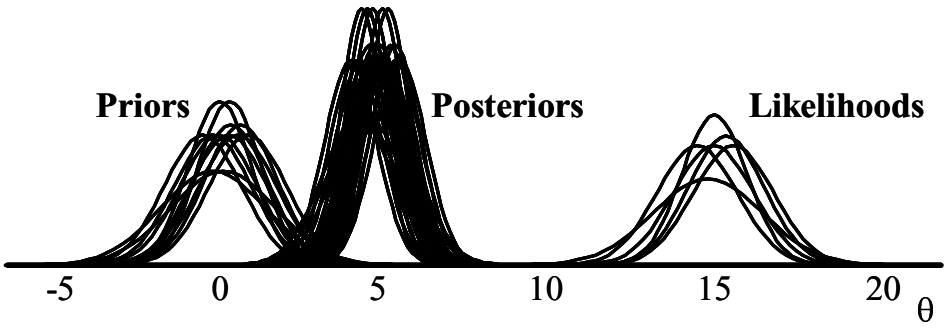


Figure 10: Example of robust Bayes combination of several prior distributions and likelihood functions to obtain many possible posterior distributions.

To get the associated p-box, we just have to cumulate each of the posterior distributions and find their convex hull. In this example, the resulting p-box is defined by the class of normal distributions having a mean in the interval [2.7, 10] and a variance in the interval [0.6, 1.4]. This p-box is depicted below (truncated at the one-half and 99.5th percentiles). The abscissa is the θ value, and the ordinate is cumulative probability. The associated Dempster-Shafer structure can be obtained from this p-box in the canonical way, simply by discretizing it into equiprobable horizontal slivers. In this case, the Dempster-Shafer structure would be the collection of interval-mass pairs { ([−0.0165897, 8.37037], 0.01), ([0.301737, 8.50761], 0.01), ([0.503706, 8.61084], 0.01), …, ([4.32963, 12.7166], 0.01), ([4.54594, 13.0079], 0.01) }. The left bound of the first focal element and the right bound of the last are determined by the truncation used in this example.

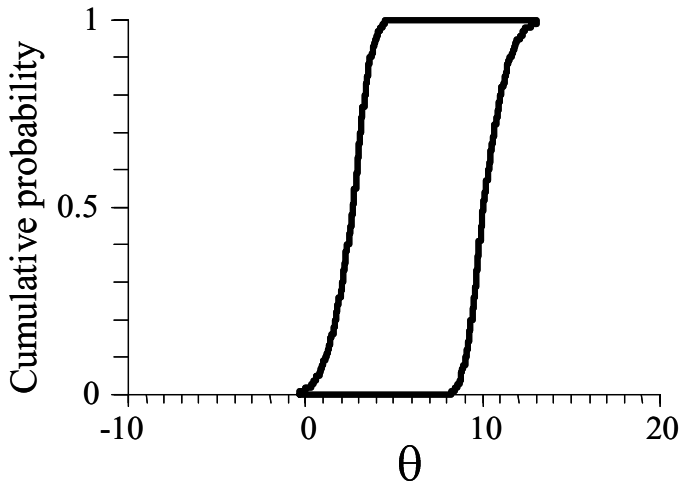


Figure 11: P-box resulting from enveloping all the posterior distributions (see text).

3.3.3 Caveats

The concerns that attend the use of Bayesian methods about specifying a single prior and likelihood are relaxed by the use of classes in the robust approach, but this approach is still subject to some of the remaining caveats associated with Bayesian methods in general. In particular, this approach exhibits what can be called the "zero preservation" problem. In this problem, any values on the real line for which the prior distribution is surely zero will remain with zero probability in the posterior, no matter what the likelihood is and no matter what new data may arrive. In the case of robust Bayesian analysis, the zeros are those regions of the real line where *all* the prior distributions in the class are identically zero. This is a mathematical form of intransigence, in that a previous conception is immutable in the face of new evidence or argument. This is the extreme case of a more general problem of Bayesian methods having to do with their possible insensitivity to surprise in the form of unanticipated data (Hammitt 1995). For instance, consider the numerical example above. This example illustrates the not uncommon result that the posterior can substantially disagree with both the prior and the new data (Clemen and Winkler 1999). Analysts might prefer fidelity to one or the other when they are in conflict. Another somewhat troublesome result is that, despite the apparently surprising nature of the evidence, the posterior distributions are generally tighter (that is, have smaller variance) than the prior distributions. In the case of extreme surprise such as this, one might hope that a methodology would always yield a result that represented more uncertainty, not less. To be fair, we should point out that the wide discrepancy between the priors and the likelihoods was introduced in this example purely for graphical reasons to ensure the reader could visually distinguish the three classes of curves. One might typically expect the priors and the likelihoods to overlap much more broadly. The details of this example depend in part on the use of normality assumptions and could differ if other shapes were used. Nevertheless, this example does illustrate some of the characteristics of Bayesian methods that may be troublesome to analysts.

Although the calculation in the numerical example above could be done almost by inspection, this is not generally possible. In fact, the computational difficulties associated with robust Bayes methods can be severe. Depending on how the class of priors and the class of likelihoods are specified, the actual calculations required may be rather burdensome. Berger (1985) gives a review of this method and describes its application in practice.

3.4 Constraint specification

Constraint propagation uses available information about a distribution or its underlying random variable as constraints to derive p-boxes that encapsulate this information. This information comes from theoretical argument, analyst judgment, expert opinion or empirical data about surrogate cases. Although the methods described in this section have their most natural application to p-boxes, the associated Dempster-Shafer structure can always be obtained (except for any infinite tails) by canonical discretization from any p-box produced by one of these methods.

In general, the problem of translating what is known about a quantity into rigorous and best possible bounds is a difficult problem. In this section, we review a variety of classical and recent results that can be used to circumscribe the distribution of a random variable given some limited information about it. The information that can be used includes limits on quantiles, information about summary statistics such as mean, mode or variance, and qualitative information about distribution shape, such as whether it is symmetric or unimodal. In many cases considered in this section, the calculation of the bounds is a relatively simple matter even though it formally implies consideration of an infinite family of distributions. However, we emphasize that *an analyst need not follow or even understand the details of these derivations in order to use them effectively in practice*. Once established, they serve as a library of algorithms by which to construct p-boxes from given sets of information.

Smith (1995) reviews several techniques for deriving distribution bounds, including limitations on entropy, which we do not discuss. It is worth mentioning, however, that the methods described in this section are very similar in spirit to those used in applications of the maximum entropy criterion (Jaynes 1957; Levine and Tribus 1978; Grandy and Schick 1991; Lee and Wright 1994), in which the distribution that has the largest entropy from a class of distributions obeying certain constraints is selected to use as the input. The difference is that we use the envelope about the class of distributions as the model of the information rather than picking a single exemplary distribution from the class to use for that purpose.

An issue that recurs several times in this section is how the limits on the distribution can be controlled so they are not vacuously large. For instance, if a p-box is the vacuous interval $[-\infty, \infty]$, it is likely to be of very little use computationally. To be practical, the range must be finite and the left bound of the p-box must be zero for sufficiently small x-values, and the right bound must reach one for large values. Three different strategies may be employed to ensure this convergence: range limits, variance limits, density limits. These are discussed in the subsections below.

3.4.1 Basic idea of range limits

Sometimes an analyst can specify the possible *range* of a quantity. This involves specifying what its smallest and largest possible values could be, its minimum and maximum. In some cases, the range may be deduced from theoretical limits, such as zero to one for a proportion. Sometimes analysts or domain experts may feel comfortable asserting a putative range from their specific or general knowledge about the quantity in question. In this case subsequent analyses will be contingent on the hypothesis that it actually encloses the true range of the quantity.

If an analyst can specify the range, then the tails are simply truncated to this interval. In principle, if there is other information available, one may only require *one* endpoint of the range to construct a useful p-box.

In the following subsections, we consider a handful of the most commonly encountered situations in which the range and ancillary information is available. There are a few other special cases that are not explicitly reviewed here for which p-boxes have been worked out (see Ferson 2002). Moreover, there are, no doubt, many situations analysts will encounter for which no ready solutions have yet been developed. Such cases will require ad hoc analyses to derive rigorous and best possible procedures.

3.4.1.1 Minimum, maximum

If it is know that the quantity cannot be smaller than a nor larger than b, then the interval $[a, b]$ is used as the representation of this fact. This interval is mathematically equivalent to the p-box $[H_a(x), H_b(x)]$, where $H_c(x)$ is the unit step function at c. The associated Dempster-Shafer structure for this case is $\{([a, b], 1)\}$.

3.4.1.2 Minimum, maximum, mean

If, in addition to the range, the mean (i.e., the mathematical expectation) of a random variable is also known, the p-box can be tighter. Despite its simplicity, Rowe (1988) was apparently the first to publish the best possible inequality for this case. Let m, M and μ be the minimum, maximum and mean values respectively. First consider the x-values between the minimum and the mean. The upper bound on probability over this range can be found by determining the largest possible values attained by a distribution function under the specified constraints. Consider an arbitrary value $x \in [m, \mu]$. The value p of a distribution function at x represents probability mass at and to the left of x. However much mass there is, it must be balanced by mass on the right of the mean. The greatest possible mass would be balanced by assuming that the rest of the probability, $1-p$, is concentrated at M. Likewise, the arrangement of mass on the left side requires the least balance when it is all concentrated at the point x. These considerations lead to the expression $px + (1 - p) M = \mu$ which can be solved to yield $p = (M - \mu) / (M - x)$, specifying the largest value of the distribution function for the value x. If there were any more probability mass at values less than or equal to x, the constraint of the mean could not be satisfied by any arrangement of mass at values less than or equal to M. Clearly then, the spike distributions defined by this expression describe the bounding distribution over the range $[m, \mu]$, subject to the fundamental constraint $0 \le p \le 1$.

The position of the lower bound is determined by the degenerate distribution which has all its mass at the mean. Its distribution function is zero from m to μ. Lower and

upper bounds for values larger than the mean can be derived by similar (but inverted) arguments. The resulting p-box is then $[U(x), L(x)]$, where

$$\Pr(X < x) \le U(x) = \begin{cases} (M-\mu) / (M-x), & \text{if } x < \mu \\ 1, & \text{if } \mu \le x, \end{cases}$$

$$\Pr(X < x) \ge L(x) = \begin{cases} 0, & \text{if } x < \mu \\ (x-\mu) / (x-m), & \text{if } \mu \le x. \end{cases}$$

Note that this formulation can handle interval uncertainty about the estimate of the mean. The implementation would simply use interval arithmetic to compute the necessary values and enforce the constraint that probability must be between zero and one.

These bounds are optimal in the sense that they could not be any tighter without excluding at least some portion of a distribution function satisfying the specified constraints. It is important to understand, however, that this does not mean that any distribution whose distribution function is inscribed within this bounded probability region would necessarily satisfy the constraints.

The associated Dempster-Shafer structure can be obtained by canonical discretization.

3.4.1.3 Minimum, maximum, median

Even more potent information is knowledge about the median (the 0.5 quantile), which pinches the uncertainty about the distribution to a definite point at the 50% probability level. Having reliable knowledge of other percentiles would correspond to similar points at other probability levels through which we can be sure the true distribution, whatever it is, must pass. Of course, if the estimate of the median or some percentile is not a point but an interval, then this pinching is less severe. When the information about a quantity is limited to knowing the minimum, maximum and an interval estimate of the median, we obtain bounds on probability such as those depicted in the graph below.

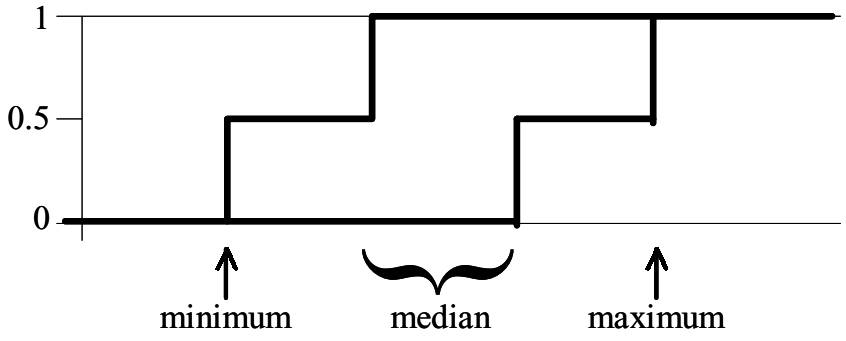


Figure 12: Best possible p-box based on knowledge of the minimum, median, and maximum values of a random variable.

The Dempster-Shafer structure for this object is just {([*m*, right(*med*)], 0.5), ([left(*med*), *M*], 0.5)}, where *m*, *M*, and *med* are the minimum, maximum and median values, and the right( ) and left( ) functions yield the right and left endpoints of an interval.

3.4.1.4 Minimum, maximum, mode

If one can assume that the underlying distribution is unimodal and that reliable estimates are available for the minimum, maximum and modal* values of the quantity, then the probability box like that shown in the graph below can be obtained by transformation of the interval p-box of Section 3.4.1.1. The justification for this transformation is beyond the scope of this report, but it is described by Smith (1995; see also Dharmadhikari and Joag-Dev 1986; Young et al. 1988).

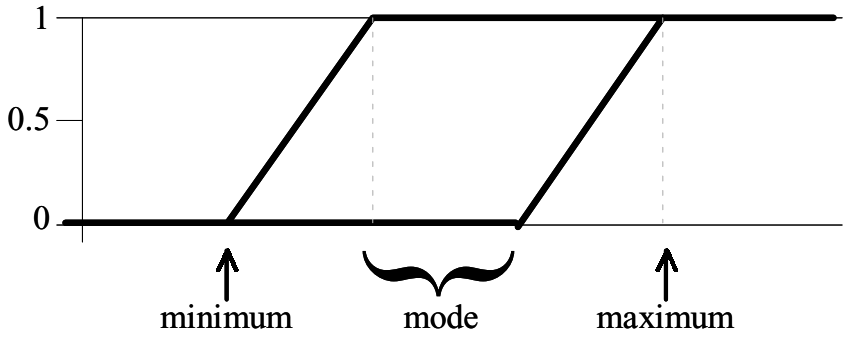


Figure 13: Best possible p-box based on knowledge of the minimum median and maximum values of a random variable.

Again, we emphasize that not every distribution contained in this region satisfies the given constraints. However, the bounds are optimal in the sense that they could not be tighter without excluding some distribution that does satisfy the specified constraints.

3.4.2 Caveats

The fundamental admonition about specifying the possible range of a quantity is not to underestimate it. The range is intended to represent the possible extent of the variable, not merely its likely extent or its observed extent. 'Possible' is usually bigger than 'probable' or 'actual'. The most extreme value ever observed may be far less extreme than the most extreme possible. In recognition of this fact, analysts should be careful not to be deceived by their data.

It is always possible to use intersection (Sections 3.2.1.4 and 4.3) to combine p-boxes obtained by these methods, but doing so may be a suboptimal strategy. For instance, if one knew the range, mean and mode, one might try to use this information by computing the minimum-maximum-mean p-box and then intersecting it with the minimum-maximum-mode p-box. This strategy is rigor-preserving; so long as both p-boxes are sure to enclose the underlying distribution, their intersection will too. However, the strategy does not always yield best possible results, even if both p-boxes

*The mode of a distribution is its most common value.

are themselves best possible. There may be some constraint or interaction that is not recognized by crude intersection.

3.4.3 When mean and variance are known from sampling

The classical Chebyshev inequality (Feller 1968, page 152; Allen 1990, page 79; Chebyshev 1874; Markov 1886) can be used to compute bounds on the distribution function of a random variable given that we know the mean and variance of the random variable. It is rare in practice, however, that an analyst would know the mean and variance of a random variable and yet not be able to say anything else about its distribution. Saw et al. (1984; 1988) derived bounds on distribution functions based on *sample* estimates for the mean and variance that generalize the classical Chebyshev inequality. The bounds are for exceedance past an absolute value, so the inequality is double-sided. Shown on the graph below are the Chebyshev limit and the Saw et al. limits for various sample sizes for the right side of the p-box having zero mean and unit variance. (The p-box is symmetric.) The Chebyshev limit is the smooth curve that, generally, has the largest value. The Saw et al. limits are step functions in various shades of gray. In general, the higher the sample size *n*, the higher the curve. The Chebyshev limit corresponds to a Saw limit with a sample size of infinity. The right tail of each limit becomes unbounded at some probability higher than a critical level, which depends on the sample size.

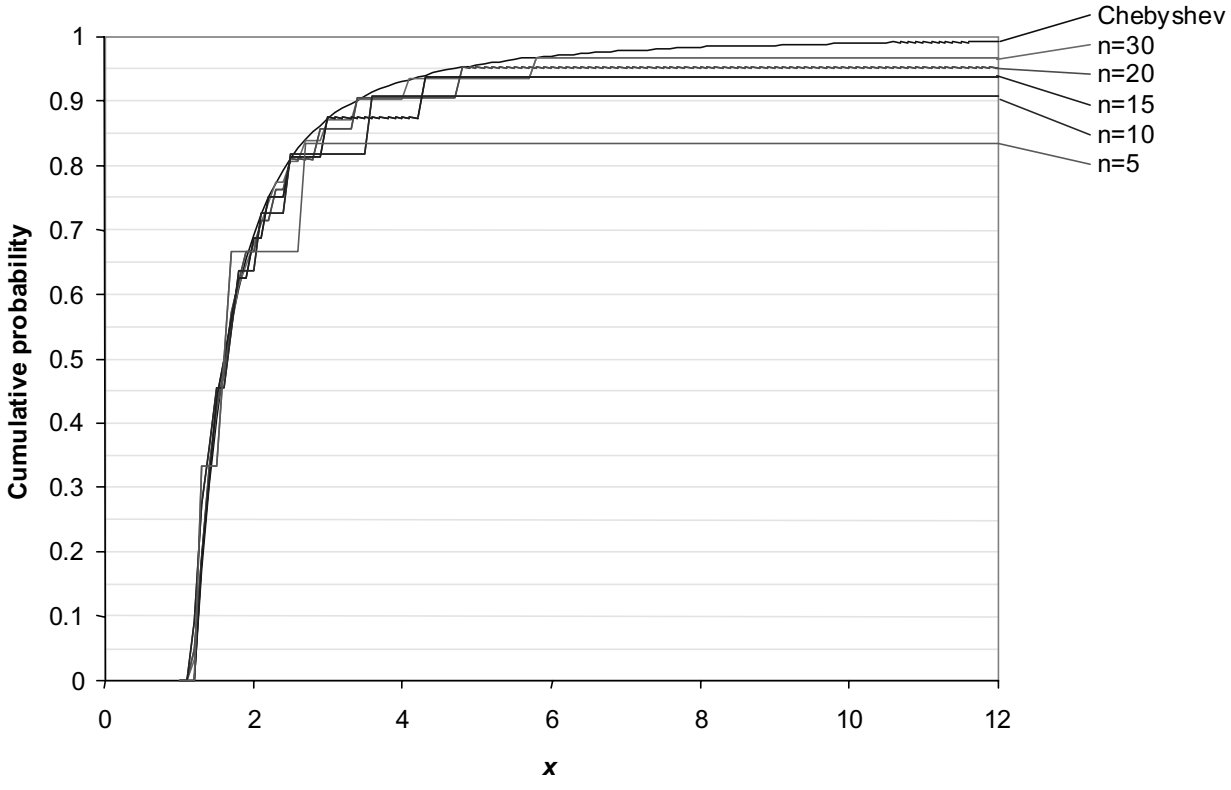


Figure 14: Right sides of p-boxes based on sample means and variances (equal to zero and one respectively) for several sample sizes.

There are several surprising things about this plot. The first is that the decrease from the Chebyshev limit is as small as it is. The decrease is really very modest for small sample sizes, except at the right tail. What is both surprising and counterintuitive is that the tightening of the bound with increasing sample size is not monotonic. Indeed the bounds cross each other! As explained by Saw et al. (1984), this can be understood as a result of the discreteness of the underlying distributions determining the limits. It is also surprising, as the authors point out, that some of the limits are slightly better than the original Chebyshev inequality at certain values of x.

3.4.4 Means and variances always 'exist'

Mathematically, the distribution of a random variable may fail to have a mean or variance. For instance, Student's t distribution with two degrees of freedom theoretically has no variance because its formula does not converge to a finite value. Similarly, the quotient of independent unit normals, which follows a Cauchy distribution, has neither a variance nor mean. As a practical matter, however, we do not consider nonexistence of moments to be of any real significance for risk analysts. Infinite means and variances are merely mathematical bugaboo that need not concern the practically minded. All random variables relevant to real-world risk analyses come from bounded distributions. As a practical matter, even a very comprehensive risk analysis need never include a mathematically infinite distribution for any variable. Analysts concerned with infinite tails of distributions are addressing *mathematical* problems, not risk analysis problems. All the moments of any bounded distribution are finite and therefore 'exist' in the mathematical sense.

On the other hand, just because the moments are finite, does not imply they are determinate. In fact, it may usually be the case that only an indeterminate estimate of a mean or variance is available. In such situations, we can use intervals to represent the value, whatever it is, in some range. We can then use elementary interval arithmetic (Moore 1966) and the methods described in this section to infer the implications of such moment estimates for p-boxes and Dempster-Shafer structures and propagate them through calculations even though we cannot specify their values precisely.

3.4.5 Basic idea of density limits

In some instances, analysts or the experts they consult may be able to describe upper or lower bounds on probability *densities*. Mallows (1958) and Smith (1990; 1995) discuss the use of bounds on probability density to constrain the cumulative probability function. In fact, this topic is classical, originating perhaps with Markov (1898). Such bounds are shown in the display below. The black trapezoid marks a hypothetical pointwise upper bound on the probability density function. This bound can be called a "cap". The shape of the cap need not be a trapezoid of course, and the only restriction on the cap is that the area under it must be greater than one. If it equals one, then the cap will completely specify a (precise) distribution. So long as the cap has a finite support* itself, the tails of the p-box or Dempster-Shafer structure must have (the same) finite support. The gray line marks a lower bound on the density, which can be called a "bubble". The area of bubble must be less than one. If it is equal to one, then, again, the probability density

*The support of such a function is the set of x-values over which it is not zero.

function will be thus specified. Specifying a bubble does not reduce the extent of the tails, but it can substantially tighten the resulting p-box and Dempster-Shafer structure.

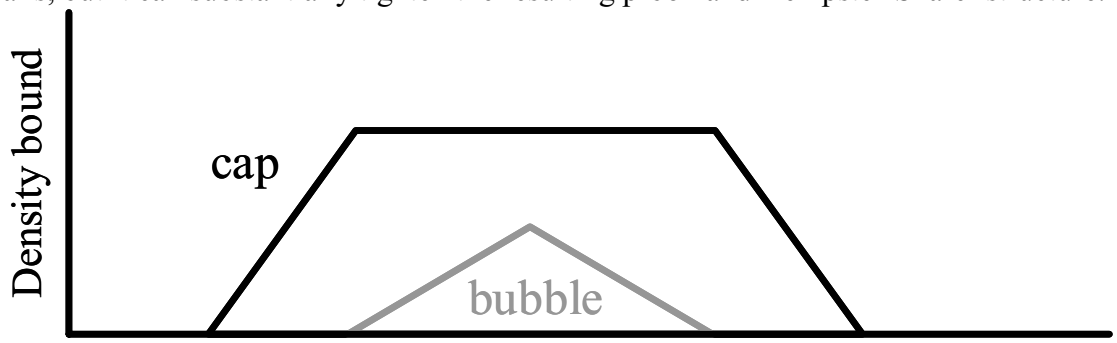


Figure 15: Hypothetical upper bound (cap, in black) and lower bound (bubble, in gray) on an unknown probability density function.

Given that the underlying probability density function must lie inside the area circumscribed by the cap and the bubble, one can immediately deduce strict bounds on the cumulative distribution function. In addition to being rigorous, these bounds will also generally be best possible.

This approach may be especially useful in interactive graphical elicitations, which can be helpful in group settings and for informants who are uncomfortable with specifying values numerically.

3.4.6 Caveats

The difference between knowing bounds on the density function and knowing bounds on the cumulative distribution function seems rather artificial. Some analysts may be hard to imagine how in practice one could know one but not the other. Perhaps the real practical advantage of this strategy is that it expresses the bounds in terms with which some experts or informants may feel more comfortable.

3.4.7 Graphical example

The gray lines in the graph below indicate the cap and bubble. These constraints don't seem to say much directly or specifically about the distribution function, nor do they create any particular specifications about the moments or any order statistics. But they do give a general picture of the shape of the density. We see that values less than 1 and greater than 5 are impossible. We see that the values around 3 should be frequent in the distribution, although 3 may not be the mode. The resulting probability box implied by the cap and bubble is shown in black. Like all p-boxes, its vertical scale is cumulative probability. The bounds follow immediately from integrating (both from the left and from the right) the functions specified by the gray lines. The lower half of the left bound and the upper half of the right bound come from cap. The remaining bounds of the p-box come from the bubble.

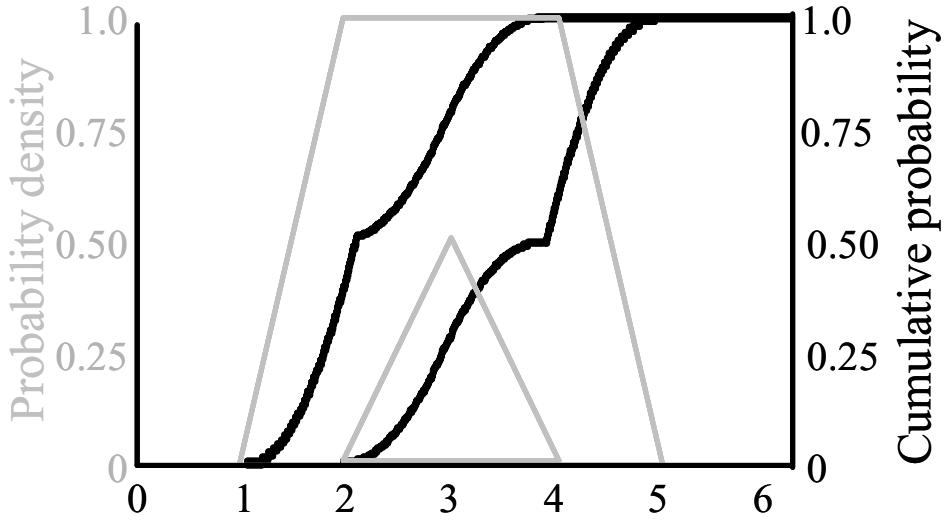


Figure 16: P-box (black) resulting from upper and lower bounds (gray) on a probability density function.

3.4.8 Qualitative shape information

3.4.8.1 Positivity

Knowing that a quantity in question cannot be negative can sometimes tighten the p-box for that quantity. This is perhaps the easiest constraint to account for. It can be done by transforming any p-box $[\overline{F}, \underline{F}]$ to

$$[\min(\overline{F}(x), \mathrm{H}_0(x)),\ \min(\underline{F}(x), \mathrm{H}_0(x))],$$

where $\mathrm{H}_0(x)$ is the unit step function at zero. (This step function is zero for all values of x less than zero and one for all values of x greater to or equal to zero.) The positivity transformation is clearly rigor-preserving, but it may not yield the best possible bounds on the quantity given all the available information. Similar transformations can be developed to reflect a constraint that a quantity must be in specific range, such as 0 to 1 or 1 to 100.

3.4.8.2 Convexity (increasing density)

A function $f(x)$ is convex on an interval $[a, b]$ if, for any two points x_1 and x_2 in $[a, b]$, $f(\frac{1}{2}(x_1+x_2)) \leq \frac{1}{2}(f(x_1)+f(x_2))$. Narumi (1923) found the inequality

$$F(x) \geq 1 - (\mu_r / x^r - (1 - \mu_r / x^r) / r),$$

where μ_r is the r^{th} moment, $r > 0$, for the case when a distribution function F is convex in the open interval $(0, x)$. Even when only the mean $\mu = \mu_1$ is known, the resulting inequality $F(x) \geq 2 - 2\mu_r / x$ is remarkably potent, sometimes beating the Chebyshev

inequality which is based on the first two moments. If the distribution is convex over its support on the positive reals, this inequality limits the range to a finite interval. Further improvement is sometimes possible with the Narumi inequality if the variance is also known, μ_2 being computed as the sum of the variance and the square of the mean. It is hard to imagine that further moments will be available about any variable arising in practical risk assessments.

3.4.8.3 Concavity (decreasing density)

A distribution function F is concave if its complement $1-F$ is convex (see previous section). Concavity means that the density function of the random variable is decreasing over its range, and thus this property can be regarded as a special kind of unimodality. Barlow and Marshall (1964) showed that if a distribution function F is concave over the positive reals, then $F(x) \geq 1 - (\mu_r / x^r) (r / (r + 1))^r$. Like convexity, concavity can imply a strong improvement in bounds. But it is also possible to specify moments, or even a single moment, for which there is no concave function over the positive reals. Analysts must check that the p-box obtained from this inequality is feasible.

3.4.8.4 Monotone hazard function

There is a significant literature in statistics and reliability theory on ordering and bounding distributions based on assumptions about the hazard rate, which is also called the failure rate. The literature is synoptically reviewed in Johnson et al. (1995). Let the random variable be the length of life for some component or system. Hazard rate is thought of as the probability of failure in the next instant divided by the probability of having survived until now. The divisor is called the survival probability and is itself just one minus the cumulative probability distribution. The formal definition of the hazard rate (failure rate) is

$$r(x) = F'(x) / (1 - F(x))$$

where F is the distribution function for the length of life and the prime denotes differentiation. It is usually assumed that $F(0) = 0$, although this restriction can be relaxed. $F(x)$ has an "increasing hazard rate" if $r(x)$ is an increasing function of x. Such functions characterize components or systems that exhibit senescence or aging so that the older they get the more likely they are to fail. $F(x)$ has an "decreasing hazard rate" if $r(x)$ is a decreasing function of x. These functions characterize systems that "burn in" or get more reliable as they age. The exponential distribution (see Section 3.1.1.6) has a constant hazard rate and therefore represents the boundary between these two classes of functions.

Engineers can sometimes confidently assert that the lifetime distribution of a particular kind of component or system is in one category or the other. When they can do this, there will be an exponential distribution that bounds their possible distributions. When the analyst can identify some other aspect of their distribution that specifies this exponential, then it is possible to tighten the probability box used to estimate the length of life random variable. This will be possible in practice if, for instance, the mean (or bounds on it) of the distribution is known (or assumed). It will also be possible if the analyst can empirically specify what the initial hazard rate is (or an upper or a lower

bound on it). Barlow and Marshall (1964) describe several way to use available information to find the limiting exponential distribution.

If the p-box $[F_1, F_2]$ is given as a structure that is certain, by other considerations, to enclose the distribution, then we can account for the additional constraint that the distribution is sure to have an increasing hazard function by replacing $[\overline{F}, \underline{F}]$ with $[\overline{F}^*, \underline{F}^*]$, where

$$\overline{F}^*(x) = \min(\overline{F}(x), H_0(x)),$$
$$\underline{F}^*(x) = \max(\underline{F}(x), E(x)),$$

in which $H_z(x)$ is the unit step function at z and $E(x)$ is the limiting exponential distribution function. If the distribution is known to have a decreasing hazard rate, use the functions

$$\overline{F}^*(x) = \min(\overline{F}(x), E(x)),$$
$$\underline{F}^*(x) = \max(\underline{F}(x), H_\infty(x)),$$

instead.

Even if the hazard rate is known to be montone, this approach cannot be used when there is no way to identify the bounding exponential distribution, such as with a specified mean or initial hazard rate. And, of course, many distributions have hazard rates that are increasing and decreasing over different ranges of x, including, for instance, bathtub-shaped hazard rate functions commonly encountered in reliability analyses. For such functions, the simple strategy outlined above to tighten an estimating p-box won't be useful, although a compound one treating different portions of the curve separately might be useful if knowledge is available about how hazard rate varies with time.

Barlow and Marshall (1964) explained how to obtain even tighter bounds by combining knowledge about monotonicity of the hazard rate with other information, including knowledge about the moments of the distribution, percentiles, and bounds on the hazard rate itself. If a distribution has an increasing hazard rate, it must be the case that the variance divided by the square of the mean must be between 0 and 1 inclusive. If it has a decreasing hazard rate, the quotient must be larger than or equal to 1.

3.4.8.5 Numerical examples

The graph on the left below depicts the p-box that results from applying a positivity transformation to a p-box that was specified by some previous argument (and the methods of Section 3.1.1.5) to be a uniform distribution whose mean is in the interval [2, 5] and whose standard deviation is in the interval [1, 2]. The original p-box had the shape of a rhombus. The resulting p-box differs from it only in having the left tail below zero truncated. The graph on the right below depicts a p-box in black that results from applying a concavity transformation to a p-box that was specified by some previous argument (and the methods of Section 3.4.1.2) to a distribution whose possible range is the interval [0, 15] and whose mean is 1. Before the concavity transformation, the right bound of the p-box was at the location of the gray line. Thus, the difference between the

gray line and the right bound of the black p-box is the improvement that comes from the assumption that the distribution is concave or has decreasing density.

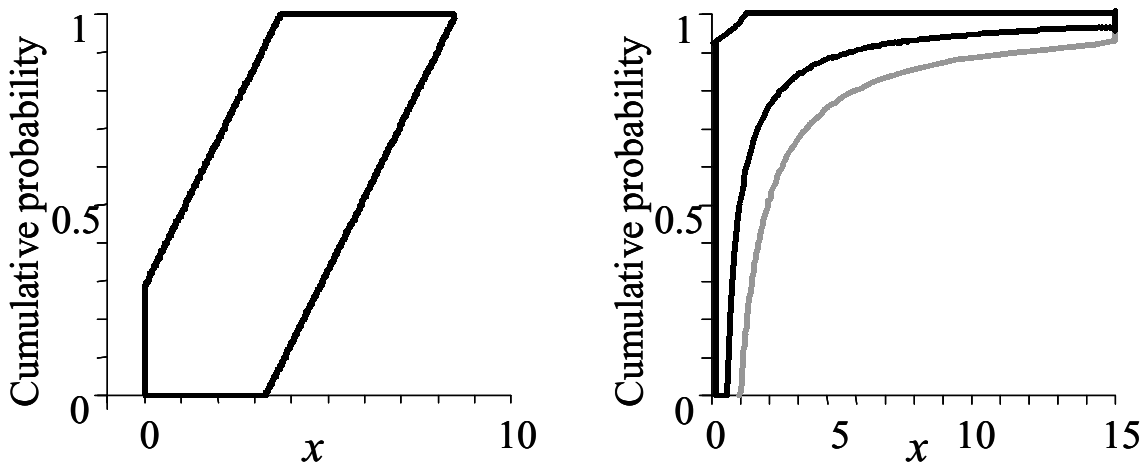


Figure 17: Examples of applying positivity (left) and concavity (right) constraints to p-boxes (see text).

3.5 Experimental measurements

This section addresses how *empirical data* can be used to construct p-boxes and Dempster-Shafer structures. Toward this end, we generalize the notion of an empirical distribution function. The two primary concerns in this generalization are measurement uncertainty and sampling uncertainty.

Measurement uncertainty* (Taylor and Kuyatt 1994) is the incertitude or lack of knowledge about the precise value of a measured quantity. It may involve both random and systematic (bias) components. When measurement uncertainty is negligible, it may be reasonable to consider the results of measurements to be point values. This has been the routine practice in traditional risk analyses. When measurement uncertainty is large, however, or specifically when it is large relative to the variation among individuals in the population, this may not at all be a reasonable strategy. P-boxes and Dempster-Shafer structures allow us to account for this measurement uncertainty and carry it along in calculations in a consistent way.

When measurements are made exhaustively in a population, there is no sampling uncertainty. When there are a great many samples made of a varying population, it may be the case that sampling uncertainty is negligible and analysts may feel confident that they have assembled an accurate picture of the variation in that population. But, in almost all risk analyses outside the insurance industry, sample sizes are typically very

*In the past, measurement uncertainty was commonly called measurement error (Rabinovich 1993) in the scientific and engineering literature. This expression is a misnomer because it is not necessarily "error" of any kind.

small and sampling uncertainty is consequently large. Accounting for sampling uncertainty has been the subject of a great proportion of statistical science over the last century. This section proposes a way to propagate through calculations the uncertainty that comes from both inexhaustive sampling and measurement imprecision.

3.5.1 Intervals are natural models of measurement uncertainty

Measurement uncertainty is associated with almost all measurements (Rabinovich 1993; Taylor and Kuyatt 1994). An interval is a natural model of measurement uncertainty in many if not most situations. Of course, an interval is not the only way to represent the measurement uncertainty of an observation, but it is certainly the most common way in science and engineering today. In many cases, the measurement intervals are given with the data values, either as an explicit interval (e.g., [31.77, 38.83]) or in terms of an appended plus-or-minus range (e.g., 35.3 ± 0.4, or $35.3 \pm 10\%$). Some measurements may be reported without explicit mention of any associated measurement uncertainty. Of course, this is usually a result of carelessness rather than perfect precision of the measurement. Long-standing convention among empiricists allows values to be reported using significant digits to make an implicit statement about error. For instance, the value "12.21" has four significant digits and it has an implied measurement uncertainty of ± 0.005, leading to an interval of [12.205, 12.215]. Likewise, the value "4800" has two significant digits and its implied measurement uncertainty interval is [4750, 4850]. This convention allows us to infer measurement intervals directly from data reports. The measurement uncertainty for integral count data may under some circumstances be given as ± 0.5 units. It is more common among physicists, however, to use plus or minus the square root of the count (this idea appears to be based on a Poisson model of counts).

3.5.2 Basic idea about measurement uncertainty

Measurements are often reported as an interval range described in the previous section together with a nominal point value sometimes called the "best estimate". This triple of numbers can be represented as a triangle whose base represents the interval range and whose peak marks the best estimate. Such a triangle does not represent a triangular probability distribution or anything other than the three values that characterize the best estimate and range of possible values of a single measurement. Suppose that for a particular variable we have some data values represented as triangles distributed along an x-axis shown below. (In the illustration, we've shown the peaks to be centered over their bases and the bases to all have the same length, but neither condition is necessary.) The widths of the triangles represent measurement uncertainty in the data set, which is a kind of *incertitude*. The scatter of the collection of triangles reflects *variability* in the data set. Below the triangles, there is a graph on the same horizontal scale of the p-box implied by these measurements. The x-axis of this graph is the same as for the triangles; the ordinate is cumulative probability. The p-box is formed as two cumulative distribution functions, one based on the left endpoints of the triangle bases, and one based on the right endpoints. These functions correspond to the cumulative plausibility and belief functions for the empirical Dempster-Shafer structure formed by using the intervals of the triangle bases as focal elements and assigning equal probability mass to each. Yager (1986)

recognized these bounds as a generalization of the distribution function for Dempster-Shafer structures.

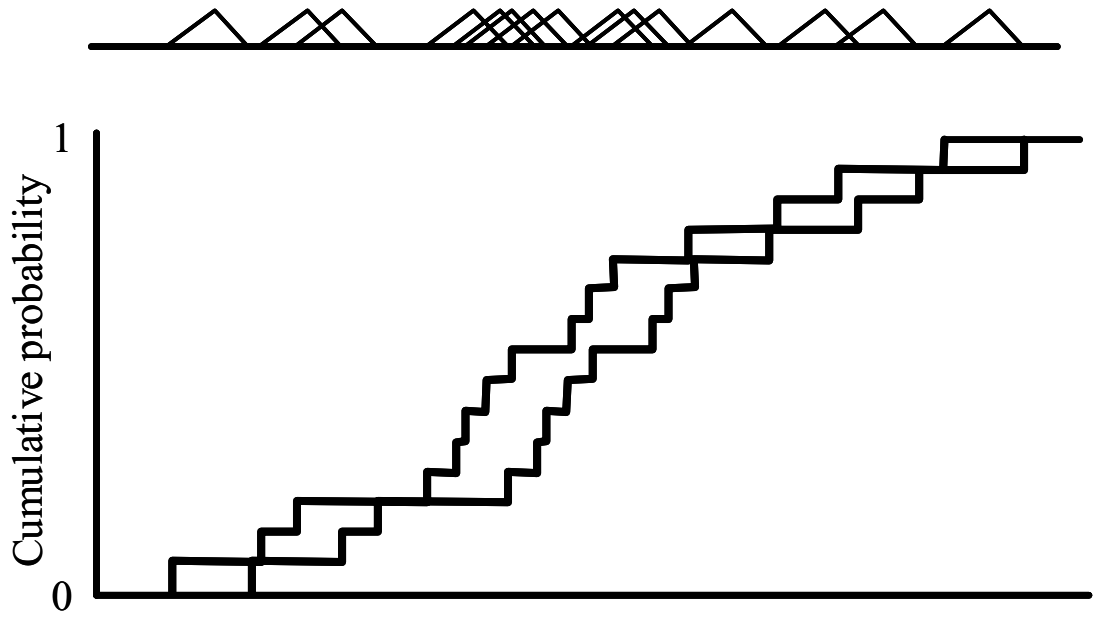


Figure 18: P-box (below) corresponding to a data set (above) consisting of intervals (triangle bases) about best estimates (triangle peaks).

The cumulative empirical distribution function (EDF) associated with these data values, which would traditionally be formed by cumulating the best estimates for each value, would be inside this p-box. If the measurement uncertainties associated with the data values are negligible, then the p-box will approach this EDF. If measurement uncertainties are large, the p-box will be wide. Measurement uncertainty, whether small or large, is commonly ignored by analysts when they construct EDFs. Notice that the p-box, on the other hand, comprehensively expresses the measurement uncertainty exhibited in the measurement values. The resulting p-box and its associated Dempster-Shafer structure, which generalize the empirical distribution function, can be used in subsequent calculations.

3.5.3 Basic idea about censoring

Because they can represent arbitrary bounded measurement uncertainty in a natural and comprehensive way, p-boxes and Dempster-Shafer structures provide an excellent way to account for the uncertainty arising from data censoring. Censoring is simply a kind of measurement uncertainty. It gets a special name because, unlike run-of-the-mill measurement uncertainties, it can often be very large and usually afflicts data within certain value ranges. For instance, censoring in the form of "non-detects" or values "below detection limit" is common in chemical concentration data. Statistical assessments are notoriously sensitive to data censoring. The most common strategy used in traditional statistical analyses to handling censoring uses simple substitution methods

(e.g., replace each censored value by ½ the detection limit). More advanced strategies employ rather elaborate distributional models that attempt to reconstruct the dubious values based on the patterns shown by the remaining values. Helsel (1990) reviews these strategies and notes that the current statistical methods

- Break down when censoring prevalent,
- Become cumbersome or unworkable with multiple detection limits,
- Need assumptions about distribution shapes, and
- Yield approximations only.

Helsel concludes that each of the traditional methods has limitations and none is reliable for general use.

P-boxes and Dempster-Shafer structures, on the other hand, can readily express the uncertainty that arises from many kinds of censoring. As an example, suppose the empiricists who produced the hypothetical data described in the previous section reported that some of the measurements were below their analytical detection limit. This would mean that, because of the resolution of the measurement devices they employed, they cannot be sure the true values are not zeros. Suppose that the thick gray triangles in the graph below represent the censored data values. Censoring means that no matter what lower bounds might have been reported for these data by the measuring device, their left endpoints should really just be set to zero. The right endpoints are the respective detection limits (which may be variable during the experiment). Because the nominal estimate that might have been reported is not really a best estimate in any sense, one might want to redrawn the gray triangles as rectangles ranging from zero to the respective detection limits. The left and right endpoints are then separately cumulated just as in the previous section. This resulting p-box, shown below the data, is trivial to compute, yet it obviously captures in a comprehensive way what this kind of censoring does to the available information.

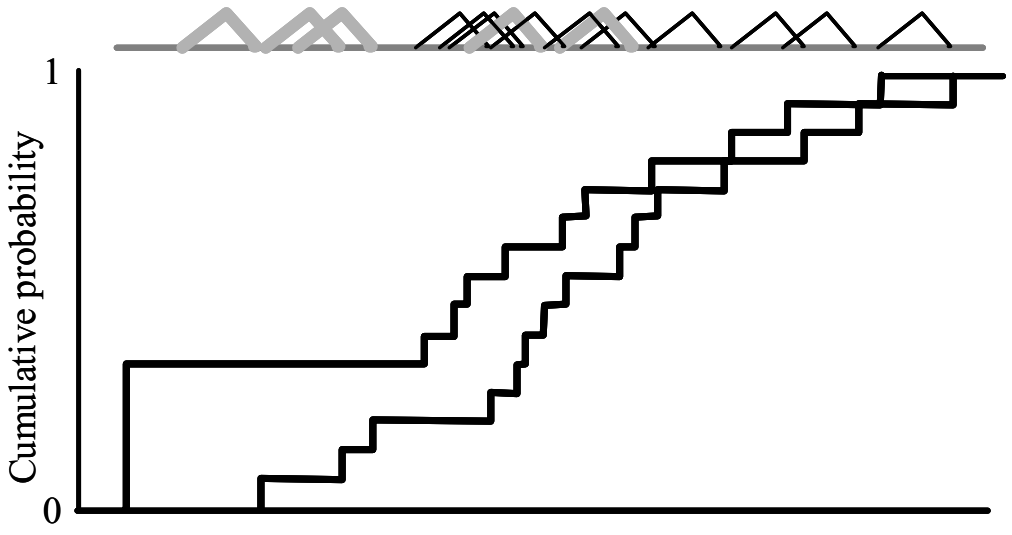


Figure 19: P-box (below) from a data set (triangle, above) in which there is left-censoring (thick gray triangles).

In contrast to the limitations of traditional approaches to censoring, an approach based on p-boxes or Dempster-Shafer structures

- Works regardless of amount of censoring,
- Handles multiple detection limits with no problem,
- Makes no distribution assumptions,
- Uses all available information, and
- Yields rigorous answers.

Moreover, this strategy can be used for interval-censoring or right-censoring as well as left-censoring, or indeed for almost any kind of fundamental or happenstance limitation on mensuration.

3.5.4 Basic idea about sampling uncertainty

The bounding empirical cumulative histograms described above form a complete description of the uncertainty from measurements, which is composed of both variability and incertitude, so long as all the members in a population of interest were measured. The more typical situation, however, is that the available data are just a small sample from a much larger population. If we collected another sample of measurements, the picture of variation and incertitude would probably be somewhat different. In recognition of this fact, one might define sampling error, analogous to measurement error, as the difference between an observed empirical distribution function and the true distribution function for the entire population. We use the expression "sampling uncertainty" to refer to the incertitude about the distribution function that arises because only a portion of the individuals in a population have actually been measured. For the unmeasured individuals in the population, one might hold that our measurement uncertainty about them is infinite, but statisticians have devised arguments based on random sampling that allow us to make some assertions about the population as a whole even though it has not been measured exhaustively.

How should we account for sampling uncertainty that arises from measuring only a portion of the population? It would seem reasonable to inflate the uncertainty about the empirical histograms in some way. The sampling theory for probability bounds analysis and Dempster-Shafer theory needs more development, but one strategy suggests itself. Kolmogorov-Smirnov (KS) confidence limits (Kolmogorov 1941; Smirnov 1939; Feller 1948; Miller 1956) are distribution-free bounds about an empirical cumulative distribution function. Analogous to simple confidence intervals around a single number, these are bounds on a statistical distribution as a whole. As the number of samples becomes very large, these confidence limits would converge to the empirical distribution function (although the convergence is rather slow).

Theoretically, the left tail of the KS upper limit extends to negative infinity. But, of course, the smallest possible value might be limited by other considerations. For instance, there might be a theoretical lower limit at zero. If so, we could use this fact to truncate the upper (left) bound at zero. The right tail of the lower limit likewise extends to positive infinity. Sometimes there are constraints on the largest value of a quantity too. It may be reasonable to select some value at which to truncate the KS limit.

This formula can be extended to the p-boxes described in the previous sections that were formed by integrating left or right endpoints of plus-minus measurement intervals. For instance, consider again the interval sampling data represented as black triangles in Section 3.5.2. The p-box used to characterize the data set is reproduced in the graph below with dotted lines. The 95% KS confidence limits applied to the same data form another p-box shown below with solid black lines. With only a few data points, we'd expect fairly low confidence in the empirical distribution function, but as the number of samples becomes large, the confidence limits get closer together. Note, however, that even for very large samples, the bounds cannot get closer than the incertitude from measurement uncertainty prescribes. Do we need to make a new derivation to justify the use of the KS limits for this application in which we account for measurement uncertainty? This is not necessary, because the extension relies on an elementary bounding argument: We have a bound on the distribution function (that we get when we account for measurement uncertainty), therefore what we compute when we apply the KS method are just bounds on the KS limits.

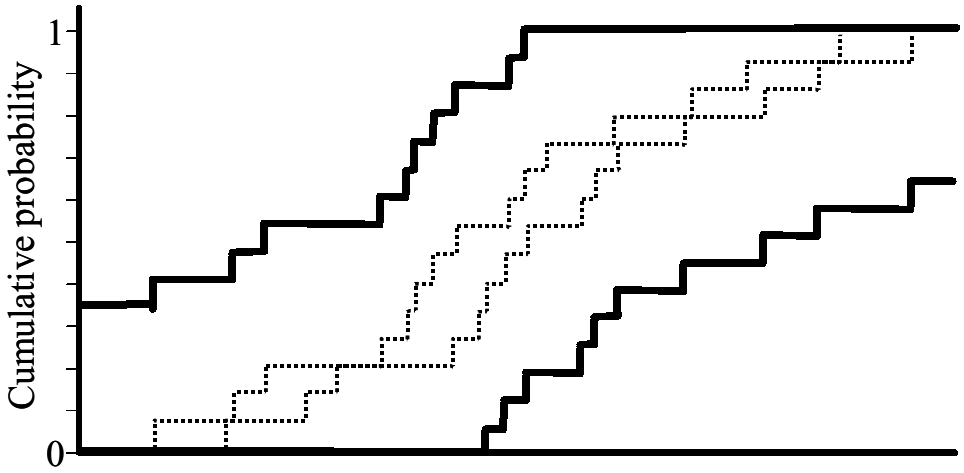


Figure 20: Kolmogorov-Smirnov confidence limits (black) assuming a sample size of 15 associated with an empirical p-box (dotted).

Although we assume that each sample is independent of other samples, this does not imply (nor does this use of the KS limits require) that the locations of the true sample values within their respective measurement uncertainty intervals are independent. Indeed, because we recognize that empirical measurement uncertainty may include systematic errors, we expect that they will generally not be independent of each other. The KS limits make no distributional assumptions, but they do require that the samples are independent and identically distributed. In practice, an independence assumption is sometimes hard to justify.

KS limits are widely used in probability theory and risk analyses, for instance as a way to express the reliability of the results of a simulation. However, it has not heretofore been possible to use KS limits to characterize the statistical reliability of the inputs, just because there has been no way to propagate KS limits through calculations. Probability bounds analysis and Dempster-Shafer theory allows us to do this for the first time in a convenient way.

It is widely suggested that statistical confidence procedures can provide the inputs for rigorous bounding analyses (e.g., Grosof 1986), and this does not seem unreasonable. However, we note that a p-box defined by KS confidence limits is fundamentally different from the sure bounds formed by knowledge of moments or shape information that we discussed above. The KS bounds are not certain bounds, but statistical ones. The associated statistical statement is that 95% (or whatever) of the time the true distribution will lie inside the bounds. It is not completely clear what the consequences of changing the nature of the bounds will be for subsequent calculations based on them. Sampling theory for probability bounds is a current area of research.

3.5.4.1 An alternative treatment of sample uncertainty

This section gives another perspective about representing sample uncertainty that may be useful when sample sizes are extremely small. It also allows us to look carefully at how p-boxes are constructed with point values, non-overlapping intervals and overlapping intervals. In all the graphs of this section, the horizontal axis is the quantity of concern and the vertical axis is cumulative probability. In the three graphs below, there are only two members of the population, and they have both been sampled. Thus, the p-boxes represent population measures rather than mere sample measures. In the top graph, the measured values were 1 and 3, and there was no measurement uncertainty. In the middle graph, the measurements were the intervals [1,2] and [3,4]. In the bottom graph, the measurements were the intervals [1,3] and [2,4]. Notice that the left side of the p-box starts at zero and jumps at each interval's left side by 1/*n*, where *n* is the sample size (in this case *n*=2). The right side starts at zero and jumps by 1/*n* at each interval's right side.

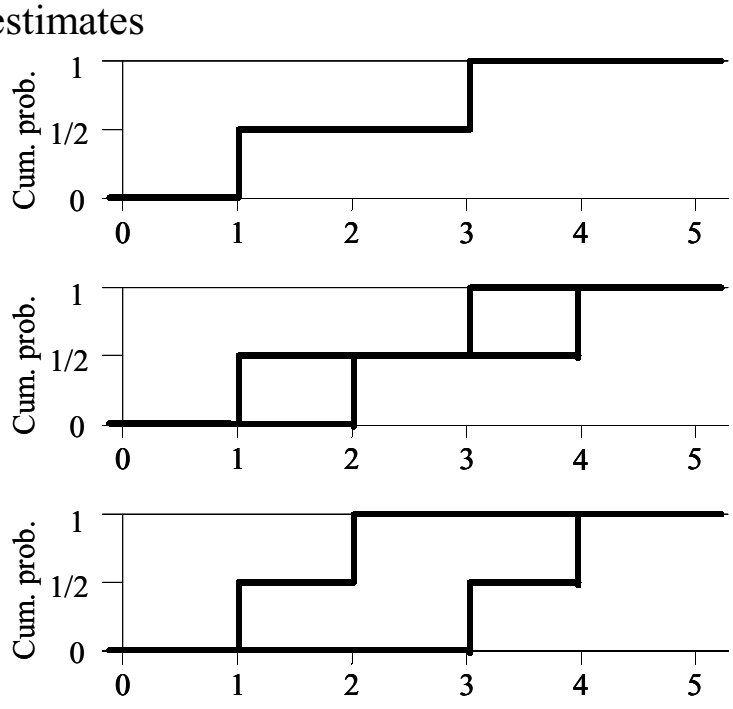


Figure 21: Empirical p-boxes for three data sets (see text), each having sample size equal to 2.

Notice that the assembly of points corresponds to simple *mixing* (see Sections 3.2.1.5 and 4.7) of the indicator functions for the data values. The upper graph is just the (vertical) average of the indicator functions of the two scalars which are the unit step functions at 1 and 3. In terms of the Dempster-Shafer structures, this corresponds to combining {(1, 1)} with {(3, 1)} to yield {(1, 0.5), (3, 0.5)}. Likewise, in the middle graph, we are mixing (vertically averaging) two intervals; the structures {([1,2], 1)} and {([3,4], 1)} are mixed to yield {([1,2], 0.5), [3,4], 0.5)}. Forming a distribution of quantities is, essentially, forming a equally weighted mixture of the elements.

In the next set of three graphs, the measured data are exactly the same: 1 and 3 for the top graph, [1,2] and [3,4] for the middle graph and [1,3], [2,4] for the bottom graph. But in this case, there are many more than two members of the sampled population, none of which have been measured. One perspective, due to Laplace, is that random sampling of n values divides the range into n +1 domains, each of which is equally likely to be the subrange from whence the next sample point will be taken (Solana and Lind 1990). Under this perspective, the left side of each p-box starts at the value $1/(n+1)$ and jumps by $1/(n+1)$ at each interval's left side. Thus, it reaches unity at the largest left side. The right side of the p-box starts at zero and jumps by $1/(n+1)$ at each interval's right side. It therefore only reaches $n/(n+1)$. The tails of the p-boxes in these three cases therefore extend to infinity in both directions.

Sample estimates

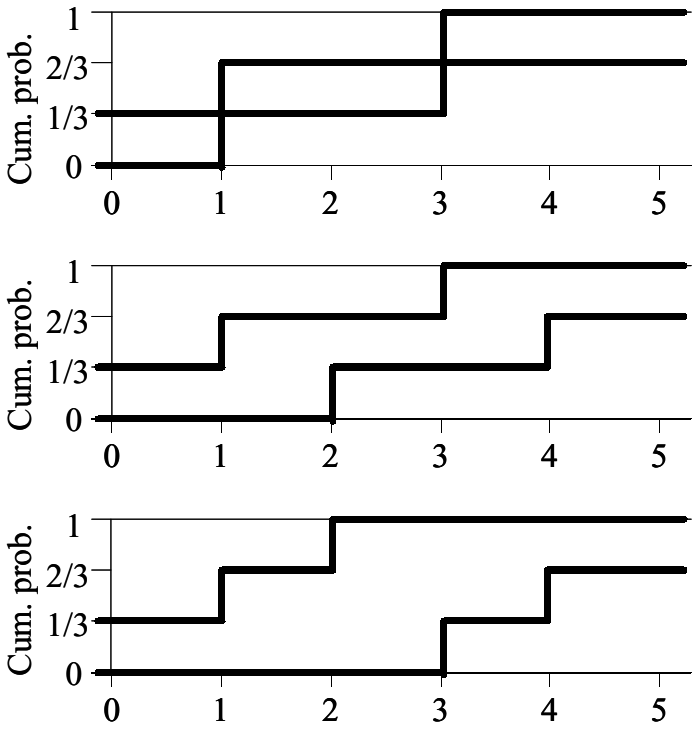


Figure 22: Alternative formulations of p-boxes for data sets of size 2.

This operation can also be interpreted as a mixture, just as the previous, population-estimate constructions were. The difference is that, here, the elements are mixed together with a vacuous element. This element is the infinite interval $[-\infty, +\infty]$. It is also the Dempster-Shafer structure $\{([-\infty, +\infty], 1\}$ and the p-box $[H(-\infty), H(+\infty)]$, where $H(x)$ is the unit step function at x. The weight given each element in the mixture is the same, but, with the addition of the vacuous element, each weight is $1/(n+1)$ rather than $1/n$.

It is clear that the results computed under this perspective differ considerably from those obtained from the Kolmogorov-Smirnov approach. This perspective could be useful if the risk assessment is not about the entire population, but only about the *next sample* to be drawn from the population. Even this application is dubious, however, because Laplace's idea, although attractive, clearly embodies an equiprobability model of what might better be considered incertitude.

3.5.5 Numerical examples

In this section, we will consider several numerical examples, including variation without incertitude of any kind, with sampling uncertainty, with measurement uncertainty, and with both sampling and measurement uncertainty.

3.5.5.1 Variation alone (without measurement or sampling uncertainty)

Suppose we have measured the following 15 data values: 17, 11, 14, 38, 15, 3, 15, 16, 20, 25, 21, 28, 8, 32, 19. These values are plotted below as spikes (note that the two values at 15 overlap), and below that as a cumulative distribution function.

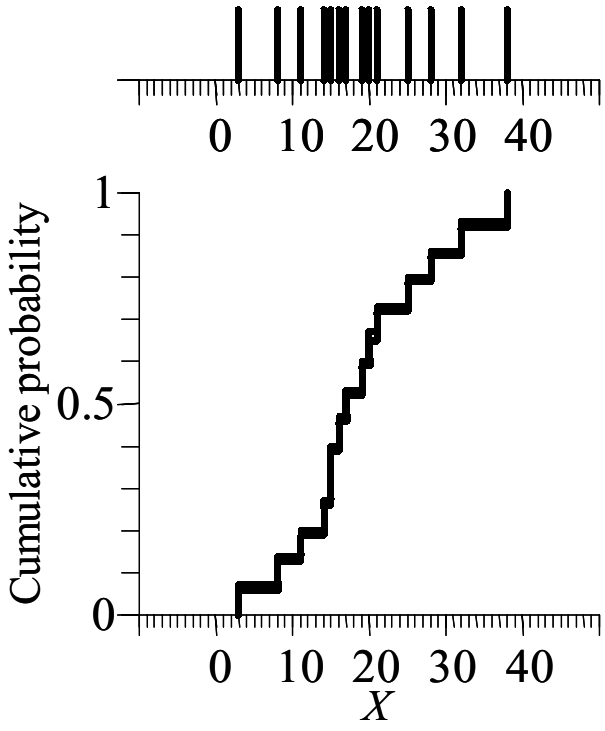


Figure 23: Empirical distribution function (below) corresponding to a data set (spikes, above).

In both plots, the horizontal axis is, say, some physical parameter X such as a transport coefficient or a geometric dimension. In the lower plot, the vertical axis is cumulative probability from zero to one. The step function is the empirical distribution that we get by cumulating the point estimates from the empirical values. This distribution is a complete and comprehensive characterization of the variation among these data values. If these data are an exhaustive sampling of the entire population in question, then the distribution is a complete and comprehensive characterization of the variation in the population. The distribution is, of course, a degenerate case of a p-box for which the distribution serves as both the left and right bound. It is also the Dempster-Shafer structure {(3, 1/15), (8, 1/15), (11, 1/15), (14, 1/15), (15, 2/15), (16, 1/15), (17, 1/15), (19, 1/15), (20, 1/15), (21, 1/15), (25, 1/15), (28, 1/15), (32, 1/15), (38, 1/15)}, which is a degenerate case because the intervals are simple points.

3.5.5.2 Variation with sampling uncertainty

If the data values mentioned in the previous section are not a comprehensive collection from an entire population, but only a sample of the values from that population, then there is incertitude about the precise distribution function. We call this incertitude about the distribution function "sampling uncertainty" (although the expression "sampling error" is sometimes used for the same thing in the statistical literature). Based on only 15 data points, we'd expect relatively low confidence in this empirical distribution function as a characterization of the distribution for the entire population. The 95% Kolmogorov-Smirnov confidence limits are shown below as solid lines around the grayed empirical distribution function. The associated statistical statement is this: 95% of the time such bounds are constructed from random samples, they will totally enclose the true distribution function.

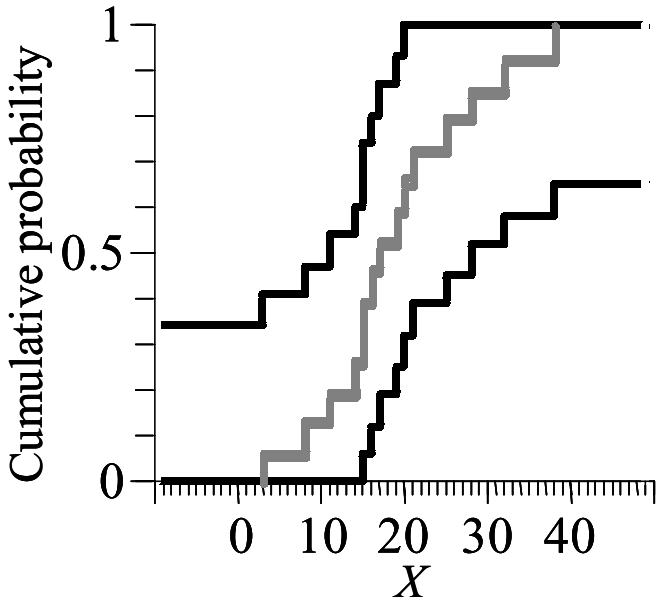


Figure 24: Kolmogorov-Smirnov confidence limits (black) about an empirical distribution function (gray).

The bounds are computed with the expression $\min(1, \max(0, \mathrm{DF}(x) \pm D(\alpha, n)))$, where DF denotes the best estimate of the distribution function, and $D(\alpha, n)$ is the one-sample Kolmogorov-Smirnov critical statistic for intrinsic hypotheses for confidence level $100(1-\alpha)\%$ and sample size n. The values for $D(\alpha, n)$ were tabled by Miller (1956). At the 95% confidence level, for a sample size of 15, the value of $D(\alpha, n)$ is 0.33760. The proof that there is such a number D that can be used to create confidence limits for entire distributions was given originally by Kolmogorov in 1933 (his 1941 paper is the first in English). While solving the related problem of inferring whether the difference between two empirical distribution functions is significant, Smirnov (1939) gave a first-order formula to compute D. Feller (1948) unified and simplified the derivations of both Kolmogorov and Smirnov. Miller (1956) synthesized subsequent work that improved the formulation for D and gave extensive tables that are still in use today.

KS confidence intervals are distribution-free constructions, which means that they do not require any knowledge about the shape of the underlying distribution. They do, however, assume that samples are identically distributed and independent of one another. If these assumptions are justified, then one could consider these bounds as a p-box that characterizes the population. The associated Dempster-Shafer structure can be obtained by discretization in the canonical way.

3.5.5.3 Variation with measurement uncertainty

What if data values are reported with a plus-or-minus range representing the empirical measurement uncertainty? Suppose these are the data: 3±3, 8±5, 11±5, 14±4, 15±4, 15±4, 16±4, 17±3, 19±3, 20±3, 21±2, 25±2, 28±1, 32±1, 38±1. These data are displayed below as little triangles where the locations of the peaks mark the best estimates and the bases mark their associated plus-minus intervals. Below the triangles is a plot of the p-box from these data. It characterizes the uncertainty about the distribution function that results when we account for the measurement uncertainty represented by the plus-minus intervals. The upper bound is found by cumulating the left endpoints of the intervals; the lower bound by cumulating the right endpoints.

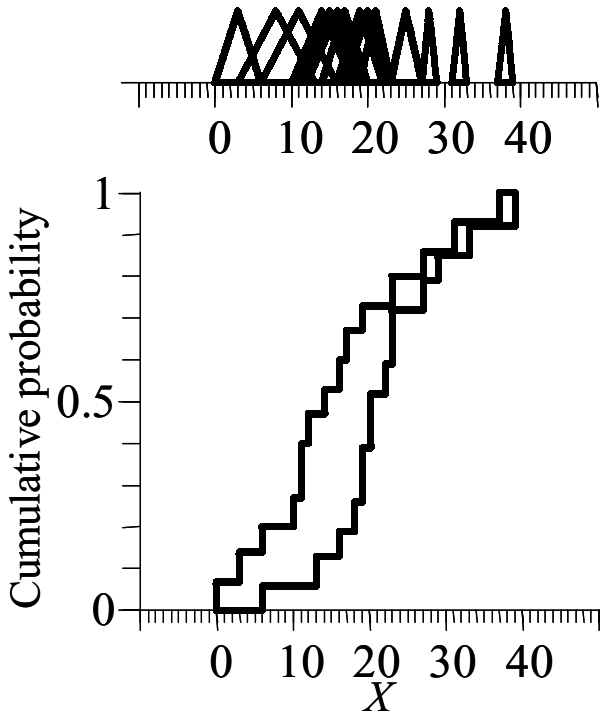


Figure 25: Empirical p-boxes (below) corresponding to data set (triangles, above) containing measurement error.

This pair of bounds, taken together, is analogous to the empirical distribution function. The associated Dempster-Shafer structure is {([0, 6], 1/15), ([3, 13], 1/15), ([6, 16], 1/15), ([10, 18], 1/15), ([11, 19], 2/15), ([12, 20], 1/15), ([14, 20], 1/15), ([16, 22], 1/15), ([17, 23], 1/15), ([19, 23], 1/15), ([23, 27], 1/15), ([27, 29], 1/15), ([31, 33], 1/15), ([37, 39], 1/15)}.

3.5.5.4 Variation with both measurement and sampling uncertainty

Applying the approach described in Section 3.5.4, we can apply the Kolmogorov-Smirnov limits to the measurement uncertainty bounds too. The result is shown below. The associated Dempster-Shafer structure can be obtained by canonical discretization. As can be seen from the graph below, unless there is some ancillary argument that truncates the range of the variable *X*, fully two third of the slivers from the canonical discretization will have infinite endpoints. The discretization is {([–∞, 19], 1/15), ([–∞, 20], 2/15), ([–∞, 22], 1/15), ([–∞, 23], 1/15), ([0, 23], 1/15), ([0, 23], 1/15), ([3, 27], 1/15), ([6, 29], 1/15), ([10, 33], 1/15), ([11, 39], 1/15), ([11, ∞], 1/15), ([12, ∞], 1/15), ([14, ∞], 1/15), ([16, ∞], 1/15), ([17, ∞], 1/15)}.

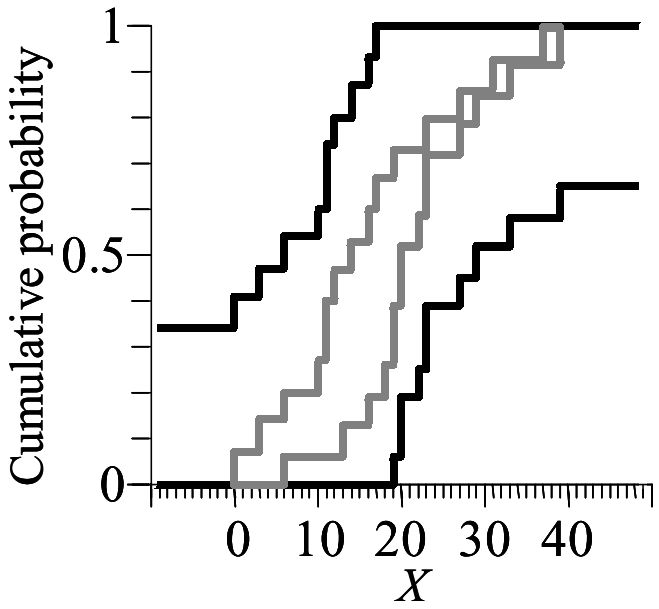


Figure 26: Kolmogorov-Smirnov confidence limits (black) about p-box (gray) accounting for both measurement and sampling uncertainty.

3.5.6 Caveats

The attraction is great for "just using the data" to prescribe the inputs for any analysis. But there are dangers too. The principle danger for risk analyses is that "the data" are usually but a very narrow window on a vast realm of possibility. Have we observed the most extreme values possible for a quantity? Given the smallness of the data sets typically available to analysts, the chances are slim that we can claim that they are representative of the distribution tails. The alternative to constructing the inputs directly from the data would be to model the data with parametric distributions or p-boxes. When we do this, we filter the available data through our engineering judgment to specify the inputs. Although the dangers of modeling are substantial too, it can often help analysts see beyond the limitations and strictures of a data set.

3.5.6.1 Practical limitations of Kolmogorov-Smirnov

There is a practical limitation of the Kolmogorov-Smirnov approach. It is that the bounds outside the data range depend on the choice of the smallest and largest possible x-values, which are used to truncate the KS limits. How does this choice about the tails affect the assessment? As we discussed in Section 3.1.3.3, there is a danger of the tails wagging the distribution in the sense that this choice substantially influences the resulting assessment. In some cases, there are natural or theoretical limits that are easily identified by the engineer or subject matter expert. In other cases, the issue should be studied as in sensitivity analysis. For instance, suppose that one of the results of interest in the example of Section 3.5.5.4 is the mean of the quantity. Given the measurement and sampling uncertainty, this mean would be characterized by an interval whose endpoints correspond to distributions at the left and right sides of the p-box. Clearly, the bounds of this interval would depend on where the tails are truncated. If we suppose that the lowest

possible value of the quantity is zero, we can study the effect of varying the upper limit on the interval estimate of the mean. The graph below illustrates how the choice of the maximum affects the bounds on the mean. This dependence of the upper bound is linear, although the slope would be shallower if the sample size were larger than 15.

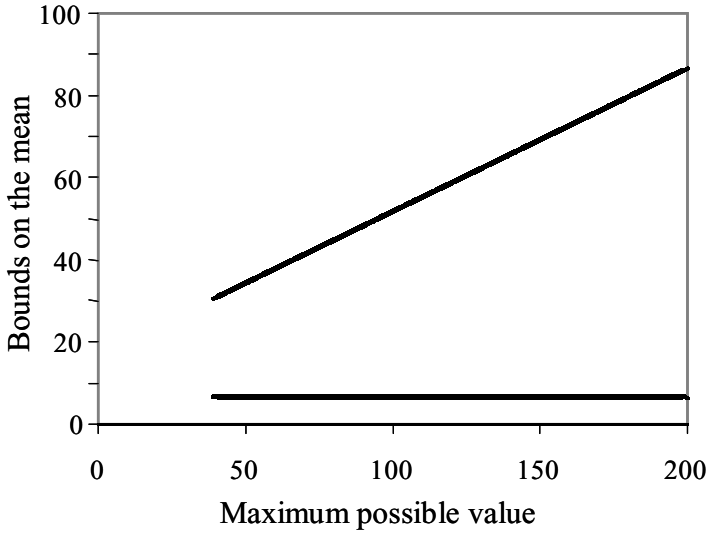


Figure 27: Upper and lower bounds on the mean estimated from Kolmogorov-Smirnov limits as the maximum value (at which the limits are truncated) varies.

If we assume that there could be no larger value in the population than has already been observed in the sample, then the upper bound on the mean would be 30.3, which is the smallest upper bound value plotted in the graph. The lower bound on the mean is not affected by the choice of the maximum possible value at which the KS limits are truncated.

A more fundamental but theoretical limitation of using KS limits to construct p-boxes is discussed in the next section.

3.5.6.2 Sampling theory for p-boxes and Dempster-Shafer structures

Convolutions with p-boxes assume that the true underlying distributions are surely contained within the input p-boxes. If this is the case, then the algorithms are guaranteed to enclose the convolution distributions (Williamson and Downs 1990; Berleant 1993). If, in fact, the p-boxes are not sure bounds, but represent mere statistical claims of the form "95% of the time such bounds are constructed, they will contain the true value (or distribution)", this guarantee will be void. The equivalent convolution operations for Dempster-Shafer structures were likewise defined by Yager (1986) without reference to any sort of sampling theory that would justify the use of confidence intervals in their construction. Consequently, when confidence interval procedures are used to produce intervals, p-boxes, and Dempster-Shafer structures, their use in subsequent calculations is only *contingent* on the implicit hypothesis that the structures appropriately enclose the respective true distributions.

A comprehensive theory that incorporates and generalizes the sampling theory of traditional probabilists would be needed to fully justify the use of confidence procedures

to create p-boxes. Such a theory would allow information based on sample data to be used to make projections about entire populations. The development of such a theory requires further research.

3.5.6.3 Plotting position

All discussions of empirical distribution functions must consider the minor controversy about "plotting position" which addresses exactly how much and where the function is incremented in relation to the observed data. One method commonly employed uses the function

$$F_x(x_i) = i / (n + 1)$$

to set the value of the empirical distribution function. In this formula, n is the sample size and i stands for the rank order (1, 2, …, n) of the measurement. Also commonly employed is the function

$$F_x(x_i) = (i - 0.5) / n$$

which is known as Hazen plotting. Indeed, there are many possible formulas to choose from. Several of these have desirable properties, but none has yet emerged as the standard way to construct the empirical distribution function. When data are very sparse, the choice of plotting position can make a substantial difference for the resulting p-box and therefore on the Dempster-Shafer structure discretized from it. We also have the additional problem of deciding what value should be used for F_x for values of x between measured values. Williamson and Downs (1990) suggested using an outward-directed scheme to conservatively and rigorously capture measurement uncertainty illustrated by the figure below.

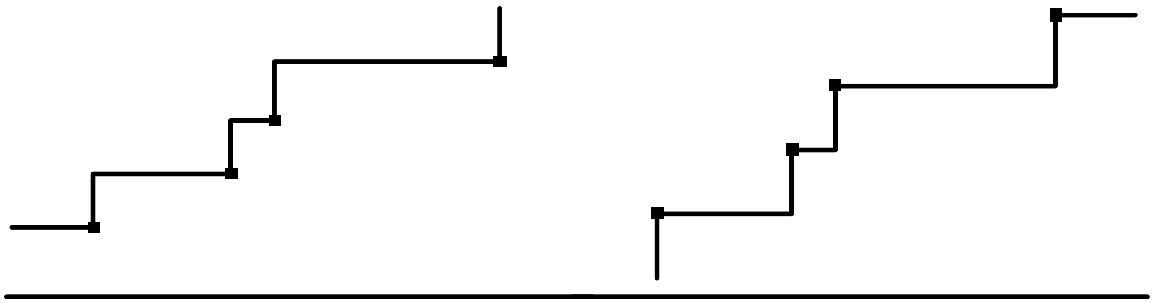


Figure 28: Conservative plotting scheme (step functions) to connect values (squares) of a function. Ordinate is cumulative probability.

In this display, there are two sets of four points, each representing decided values of the empirical distribution function. The four points on the left and the line segments connecting them would be used to characterize the left bound of a p-box (that is, the upper bound on probabilities or lower bound on quantiles). The four points on the right and their line segments would be used to characterize the right bound of a p-box (lower bound on probabilities or upper bound on quantiles). This scheme traffics on the

monotonicity of distribution functions. Given that the points represent reliable samples of the boundaries of uncertainty about a distribution function, these outward-directed line segments surely contain all distribution functions so circumscribed.

3.5.6.4 Confusion about measurement uncertainty

It may not always be obvious what measurement uncertainty is associated with a particular measurement. This is especially likely for historical data that were collected before the widespread appreciation of the importance of such considerations. In tabular summaries of data, the essential information about measurement uncertainty may have been lost. This is particularly true for computerized versions of such tabular summaries where the number of digits originally recorded in a measurement may be obscured or misleading. In such cases, it may be desirable to infer what the likely measurement uncertainty would have been, rather than to neglect it entirely by assuming there was no measurement uncertainty.

3.5.6.5 Non-interval representations of measurement uncertainty

Some theoretical treatments of measurement uncertainty, most notably the ISO standard (Dieck 1997; see also Taylor and Kuyatt 1994), hold that measurement uncertainties should be modeled as a *normal distribution* centered at the best estimate, rather than an interval. This model represents a theory about measurement uncertainty that is entirely different from—and largely incompatible with—that considered here. The normal distribution model is justified theoretically and most reasonable in practice when errors are small and independent and the only interest is in obtaining the best estimate possible. In a context where the concern focuses on tail risks and measurement uncertainties are typically large and often biased and correlated, this approach does not seem to offer the flexibility needed for use in risk analyses. The normal distribution model has been most widely used to characterize the measurement of physical constants such as Plank's constant, etc. It would be possible to replace each given normal distribution with an interval of appropriate width, but the details of how this might best be done would require further investigation.

3.5.6.6 Unbounded censoring

Not all types of censoring are immediately amenable to characterization with p-boxes and Dempster-Shafer structures. Only those forms for which finite bounds on the censored values can be specified are easy to handle. Non-detects which are left-censored values that must be positive and various kinds of window-censored values present no problem. But if the censored data are unbounded and there is no way to bound the values post hoc, then the uncertainty cannot be represented in a finite p-box or Dempster-Shafer structure. Examples of unbounded censored data would include, for instance, unobserved times to failure or arrival times. If such values could be arbitrarily large, then a p-box or Dempster-Shafer structure that is faithful to the data and its uncertainty would necessarily have infinite support.

3.6 Which method should be used?

Most of the problems faced by risk analysts are simultaneously characterized by urgency and lack of relevant data. In this context, the most practical advice about which method should be used to obtain the inputs is to *use anything and everything that works*. This short section presents a synopsis of the properties of the characterization methods considered in this report and a checklist of questions to guide the analyst to methods that may be appropriate for constructing p-boxes and Dempster-Shafer structures.

3.6.1 Summary of properties

In the outline below, we summarize the properties of the various methods for characterizing uncertain numbers with p-boxes and Dempster-Shafer structures. The notation RP (Rigor-preserving) means that the resultant Dempster-Shafer structure or p-box is mathematically rigorous so long as its specifications are. Mathematical rigor implies that, subject to its assumptions, the structure is guaranteed to enclose the underlying distribution or quantity. The notation BP (Best Possible) means that the resultant Dempster-Shafer structure or p-box could not be any tighter without more information. The notation SU (Sample Uncertainty) means that the resultant Dempster-Shafer structure or p-box represents a statistical confidence claim rather than a rigorous statement about the underlying distribution or quantity.

Assumption
- Rigorous bounds on parameters (RP, BP)
- Confidence intervals on parameters (SU)

Modeling
- Envelope (RP, BP)
- Imposition (RP, BP)
- Mixture
  - Weights precise (RP, BP)
  - Weights uncertain
    - Rigorous bounds on weights (RP, BP)
    - Confidence intervals on weights (SU)
- Convolution (RP, BP if no repeated parameters)
- Composition (RP, BP)
- Deconvolution (generally not BP)

Robust Bayes (SU)

Constraint specification
- Markov inequality (RP, BP)
- Rowe's range-mean inequality (RP, BP)
- Chebyshev inequality (RP, BP)
- Inequality of Saw et al. (SU)
- Cantelli inequalities (RP, BP)
- Karlin and Studden's Chebyshev systems
- Positivity constraint (RP)
- Symmetry constraint
- Unimodality constraint (mode known) (RP, BP)

Increasing or decreasing hazard function (RP, BP)

Experimental measurements

Interval measurement uncertainty (RP, BP)

Censoring (RP, BP)

Sampling uncertainty with Kolmogorov-Smirnov confidence limits (SU)

3.6.2 Checklist of questions

The list of questions below can guide the analyst to a useful strategy for transforming the available empirical and theoretical knowledge about a quantity into an estimate of it in terms of a p-box or a Dempster-Shafer structure. If the answer to a question is "yes", follow any questions that may be indented underneath. The number in bold is the number of the section in this report that addresses the method that could be useful to you. This is not a flowchart. There is not necessarily only one path through this thicket of questions. If the answers to more than one series of questions are all yes, it will probably be possible to produce more than one estimate. Because each such estimate is reliable, they can then be combined with the intersection aggregation operator (Section 4.3) to yield an improved estimate.

Do you have measurements of the quantity? (**3.5**)

Of the entire population? (**3.5.2**, **3.5.3**)

Of a sample from a population? (**3.5.4**)

Can you limit the range? (**3.5.6.1**)

Can you limit the possible range of the quantity? (**3.4.1**)

Do you know (bounds on) the mean? (**3.4.1.2**)

Do you know (bounds on) the mode? (**3.4.1.4**)

Do you know (bounds on) the median or other quantiles? (**3.4.1.3**)

Can you put upper limits on the probability density? (**3.4.5**)

Can you put lower limits on the probability density? (**3.4.5**)

Is the quantity necessarily positive? (**3.4.8.1**)

Is the distribution function convex? (**3.4.8.2)**

Is the distribution function concave? (**3.4.8.3**)

Is the hazard function increasing or decreasing? (**3.4.8.4**)

Can you express the quantity in a model in terms of other better known quantities? (**3.2**)

Do you specify a set of prior expectations for the distribution of the quantity?

Can you construct a set of likelihood functions? (**3.3**)

Can you mechanistically justify a particular distribution shape?

Can you bound the parameters? (**3.1**)

Do you have multiple estimates for the quantity? (**4**)

Are all the estimates reliable? (**4.3**)

4 Aggregation: how are different sources combined?

The methods reviewed in the previous section of this report have implicitly assumed that the information used to fashion the Dempster-Shafer structure or p-box was internally consistent. This might be a reasonable assumption when all the information comes from a single source, such as single expert advisor or a single measurement protocol. Of course, risk analysts are not insulated from the proverbial expert who "disagrees with himself" or from apparently inconsistent readings from a measurement protocol device. And the possibility of collating contradictory information becomes all the more likely as multiple experts are consulted and as different or complementary measurement schemes are employed.

The previous sections assumed that the analyst wanted to obtain a Dempster-Shafer structure or p-box. In Section 4, we assume the analyst already has more than one Dempster-Shafer structure or p-box for a single quantity and needs to somehow combine these estimates into a single coherent expression about what is known about the quantity. We discuss several strategies for combining different estimates, including

1. null aggregation,
2. intersection,
3. envelope,
4. Dempster's rule and its modifications
5. Bayes' rule,
6. mixture,
7. logarithmic pool, and
8. averaging.

After a preliminary discussion of the desirable mathematical properties that an aggregation method would have in Section 4.1, a separate section is devoted below to the consideration of each of these strategies. Intersection and enveloping are addressed in Sections 4.3 and 4.4 respectively. Dempster's rule is considered in Sections 4.5, and Bayes' rule is reviewed in Section 4.6. Mixing, logarithmic pooling and averaging are treated in Sections 4.7, 4.8, and 4.9 respectively. The attention given each strategy reflects in part how useful it might be for the kinds of problems we expect to encounter in real-world applications of risk analysis. Section 2 of Oberkampf et al. (2001) gives a synopsis of those problems.

Averaging, mixing and some of the variants of Dempster's rule can be generalized by applying weights to the various estimates being aggregated. In Section 4.10, we consider several different weighting schemes. In Section 4.11, we discuss strategies to handle the case when estimates are only a small sample of a larger population of interest and ways to handle model uncertainty. Section 4.13 offers advice on selecting which aggregation method to use in a particular situation.

In this report, we don't discuss aggregation strategies that lead to fuzzy numbers (Kaufmann and Gupta 1985), hybrid numbers (Ferson and Ginzburg 1995) or fuzzified Dempster-Shafer structures (Yager 1986). For the sake of limiting the report to a

manageable size, neither do we review the larger philosophical questions of data fusion or the more specialized area of statistical meta-analysis.

Many of the estimates to be aggregated will contain some degree of subjectivity. This is obviously true for expert opinions, but it is also true of many empirical estimates. This report is not primarily concerned with aggregating subjective belief or subjectively quantified statements (although the methods described herein may be useful for them). Insofar as they are of interest here, the estimates to be aggregated fulfill roles as quantitative scientific or engineering judgments or claims that we take to be reliable for the most part. We recognize that errors and mistakes are possible in these inputs, and several of the methods we consider will indeed be useful in uncovering such inconsistencies. Nevertheless, we seek analyses appropriate for the case where evidence is objective. We will assume that the evidence we use as input is largely free of personal belief and idiosyncratic opinion. In quantitative risk analyses, we do not care about what one *believes* about a system so much as we are concerned with what one *knows* about it. Clearly, this can be a fine distinction, but we think it is an important one.

A rather large literature has developed over the last thirty years on the aggregation of evidence represented in Dempster-Shafer structures. The focus of this literature has been, of course, Dempster's rule and its various arguments and revisions. Sentz and Ferson (2002) reviewed many of the most important of these rules. Section 4 of the present report, having been thereby freed of the scholarly obligation to review this literature, will focus instead on several aggregation approaches we expect to be useful in synthesizing expert opinion and empirical evidence about real-valued but uncertain quantities for use in risk analysis. For the most part, many of the approaches we review have not previously (or at least widely) been considered for use with Dempster-Shafer structures. This apparent novelty is certainly not the result of any complexity or subtlety or nonobviousness of the method. Instead, it may simply be that this literature has rarely considered the use of Dempster-Shafer structures in the context that motivates the present report, that of risk analyses typically involving convolutions of uncertain but real-valued quantities and a special concern with extreme or tail events.

4.1 Desirable properties of aggregation methods

From a purely mathematical viewpoint, we could consider arbitrary operations for combining estimates involving uncertainty. However, we want to combine different estimates in a sensible and meaningful way. There are some requirements that the aggregation operation should satisfy. These requirements are not absolutely necessary because, in addition to arguments in favor of these requirements, there are usually some counterarguments. However, it is desirable to consider to what extent these aggregation operations satisfy these commonsense requirements. We describe these natural requirements in detail below. We will find that what seems like commonsense in one situation may not be entirely reasonable in another context or when the type of information is different, and therefore the descriptions include any counterarguments or caveats about a requirement that should be kept in mind.

Depending on what is more convenient in a given context, we will use both the notation "$X*Y$" and the notation "$A(X,Y)$" to denote aggregation operations, where X and Y are the individual estimates of some quantity. Given two probability boxes B_1 =

$[\overline{F}_1, \underline{F}_1]$ and $B_2 = [\overline{F}_2, \underline{F}_2]$, we will say that B_1 is tighter than B_2, whenever both $\overline{F}_1(x) \leq \overline{F}_2(x)$ and $\underline{F}_2(x) \leq \underline{F}_1(x)$ for all x. We sometimes express this by saying that B_2 encloses B_1. We will symbolize this fact as $B_1 \subseteq B_2$. (Obviously, this relation is a partial ordering because, given any two p-boxes, it may be the case that neither encloses the other.) If, instead, one or both of B_1 and B_2 are Dempster-Shafer structures, the statement refers to their respective cumulative plausibility and belief functions through the canonical discretization (see Section 2.3). This allows us to say that a Dempster-Shafer structure is tighter than another, or that one encloses another, without ambiguity. Note that the statement has nothing to do with the number of elements in the Dempster-Shafer structures. For instance, the structure {([1,3], 0.25), ([2,4], 0.5), ([3,5], 0.25)} is tighter than the structure {([1,5], 1)} because the pair of plausibility and belief functions of the former fit inside those of the latter.

4.1.1 Generality

To be most useful, an aggregation method should work with real numbers, intervals, probability distributions, p-boxes and Dempster-Shafer structures, and arbitrary finite combinations of these various kinds of objects. Moreover, the aggregation method should generalize operations on these various structures. For instance, if the method gives one answer when applied to real numbers, it should give an equivalent answer when applied to distributions that are delta functions. Likewise, the results from applying it to intervals should be consistent with results from applying it to degenerate p-boxes or Dempster-Shafer structures that are information-theoretically equivalent to the intervals. (See also continuity in Section 4.1.5.)

Counterargument. It might be argued that it is sometimes natural to treat different kinds of objects differently. For instance, it may not be disturbing to an analyst that an aggregation method treats real numbers and probability distributions differently. They are, after all, different things.

4.1.2 Closure

It would be convenient if an uncertain number (Section 2.4) were always the result of applying an aggregation operation to a collection of uncertain numbers. If this is the case, the operation is said to be closed in the space of uncertain numbers.

Counterargument. It may not be entirely reasonable to expect that a single uncertain number can always fully or appropriately express the complexity of an arbitrary collection of uncertain numbers.

4.1.3 Idempotence

What if there were two identical estimates of X? What if two experts or empiricists come up with exactly the same X; how can we combine their knowledge? A natural idea is that if two experts came up with the same uncertainty, this means that this is the right representation of this uncertainty, so both experts are right. In other words, if we combine uncertainty X with itself, we should end up with exactly the same uncertainty X, i.e., we should have $X * X = X$. This "agreement preserving" property of an aggregation operation * is called idempotence. It seems reasonable that an aggregation operation be idempotent.

Clemen and Winkler (1999) mention a related property called unanimity, which can be thought of as a pointwise version of idempotence. If all the estimates agreeing about the probability (or bounds on the probability) for a particular value of x implies that the aggregation will have the same probability (or bounds) for that value, then the aggregation is said to have the unanimity property.

Counterarguments. Suppose that two experts come up with exactly the same description of their uncertainty: that the (unknown) value x of the desired physical quantity belongs to the interval [0, 1] with probability more than 90%. What would the result of aggregating these uncertainties be? If the two experts were using exactly the same sources of information and used the same arguments to process these sources, then the fact that these two experts came up with exactly the same conclusion simply confirms that they both did the correct computations. So, when we aggregate these two uncertainties, we should get the exact same uncertainty. In this case, idempotence is justified. But what if the two experts used independent sources of information and end up with exactly the same conclusion—that $x \in [0,1]$ with probability 90%? In this case the fact that two experts, based on independent sources of information, came up with the same conclusion, increases the reliability of this conclusion. In this case, the result $X*X$ of combining the two identical uncertainties X is that x belongs to the interval [0,1] with some probability $p>90\%$. In other words, in this case, $X*X$ is different from X—so there is no idempotence. Other examples where idempotence should not be demanded are stories told to a police officer or evidence about a historical event. If several independent witnesses tell exactly the same story, its reliability increases.

4.1.4 Commutativity

A seemingly natural requirement is that if we have two sources of information about an uncertain quantity collected from two experts or two empirical devices then the result of aggregating their information X and Y should not depend on the order in which these two different pieces of information are presented. In other words, we should have $X*Y=Y*X$. This property is called commutativity.

Counterarguments. Commutativity makes sense if there is no reason to prefer one of the two sources of information. Commonly, however, one source of information is more reliable than the other. For example, when we combine information coming from two experts, it is normal to give more weight to the opinion of a more respected expert who has a history of better estimates and better predictions. If we weight one input more than another, then clearly switching the inputs (but not switching the weights accordingly) would yield a different numerical result. Consequently, the weighted aggregation operation will not be commutative. Another situation in which one might not expect an aggregation to be commutative is when earlier estimates tend to influence later estimates. The asymmetry of this influence could make commutativity an unreasonable restriction.

4.1.5 Continuity

What if we have two different estimates X and Y, and a third estimate X' which is very close to X, which we symbolize as $X' \approx X$. Because $X' \approx X$, it is reasonable to require that $X*Y$ is close to $X'*Y$. Symbolically, $X*Y \approx X'*Y$. In other words, it is reasonable to

require that a small change in one of the uncertainties X to be aggregated lead to only a small change in the result of the aggregation. This property is called continuity.

Counterarguments. At first glance, continuity seems natural, but there are examples when it is counterintuitive. One such example is the case when each estimate is an interval of all possible values of the desired quantity. In this case, if one piece of information is that the quantity should be in the interval $\boldsymbol{x} = [x^-, x^+]$, and the other piece of information is that this same quantity should be in the interval $\boldsymbol{y} = [y^-, y^+]$, this means that the actual value x should belong to both intervals. The set of all the values which belongs to both intervals $\boldsymbol{x}$ and $\boldsymbol{y}$ is the intersection $\boldsymbol{x} \cap \boldsymbol{y} = [\max(x^-, y^-), \min(x^+, y^+)]$ of these intervals. So, in this case, the aggregation operation is simply an intersection. One might expect this operation to be continuous. After all, both the lower endpoint $\max(x^-, y^-)$ and the upper endpoint $\min(x^+, y^+)$ of the intersection interval are continuous functions of the parameters x^-, x^+, y^-, and y^+ that characterize the intervals to be aggregated. So, a small change in one of these four parameters leads to small changes in the endpoints of $\boldsymbol{x} \cap \boldsymbol{y}$. But intersection is not continuous. Continuity of a function requires that a function both be defined and have a value that is the same as its left and right limits. What happens if we slowly move the interval $\boldsymbol{y}$ so that its intersection with $\boldsymbol{x}$ becomes smaller and smaller and finally disappears? When there is no intersection, the aggregation operation is undefined and thus it fails to be continuous. Any modification of the aggregation operation to make it continuous would, in this case, fail the spirit of intersection and thus be counterintuitive and perhaps unwelcome.

4.1.6 Associativity

It is natural to require that if we have three different sources of information X, Y, and Z, then the result of aggregating the corresponding pieces of information should not depend on the order in which we aggregate these three pieces. We can first combine X and Y into a combined knowledge $X*Y$, and then combine the resulting aggregation with Z to obtain in $(X*Y)*Z$. Alternatively, we can combine Y and Z into the aggregation $Y*Z$ first, and then combine the result with X, yielding $X*(Y*Z)$. It is reasonable to expect that both of these ways will lead to identical results, $(X*Y)*Z = X*(Y*Z)$. In mathematical terms, this requirement is called associativity.

Counterarguments. There are several aggregation operations that are not associative but which are nevertheless considered eminently reasonable ways to combine estimates. The best example is perhaps numerical averaging. For example, if $X*Y$ is taken to mean the average of X and Y, then $(0*1)*2 = 0.5*2 = 1.25$, although $0*(1*2) = 0*1.5 = 0.75$, which is not the same as 1.25. Because associativity is not a feature of numerical averaging, the property must not be essential for a useful aggregation operation.

4.1.7 Symmetry

When an aggregation operation can take several operands at once, it may be too cumbersome to talk about it only in terms of how it behaves for only two at a time. The property that generalizes those of associativity and commutativity for the case with many arguments is called symmetry. Let A denote an aggregation operator that can take a variable number of arguments, which are denoted X_i. If $A(X_1, X_2, \ldots, X_n) = A(X_{\sigma(1)}, X_{\sigma(2)},$

…, $X_{\sigma(n)}$) for all permutations σ of the arguments, the operation is called symmetric (in its arguments).

Caveat. Symmetry makes sense if the estimates should all be weighted equally, or if any needed weighting can be taken care of in a step prior to the aggregation itself.

4.1.8 Quasi-associativity

The fact that simple averaging is not associative belies the larger truth that it is usually used in an associative way. Few analysts would actually average each new datum with the running average with equal weights. Instead, they would typically weight the running average and the new datum in a way that reflects their respective reliabilities. For instance, if M is the current value of the running average and it is based on a total of n separate estimates $X_1, X_2, \ldots, X_n$, then it might combined with a new datum X_{n+1} with the formula $(nM + X_{n+1})/(n+1)$. Of course, this is equivalent to pooling all the data together and forming an average once as in the expression $(X_1 + X_2 + \ldots + X_n, + X_{n+1})/(n+1)$. These operations *are* associative and they suggest a way to disassemble the parts of a non-associative operation and reconstruct them in an associative way. Yager (1987b) used the notion of quasi-associativity to describe operations for which this could be done. If there exists an multiargument operation $A^n(X_1, X_2, \ldots, X_n)$ that is symmetric, i.e., $A^n(X_1, X_2, \ldots, X_n) = A^n(X_{\sigma(1)}, X_{\sigma(2)}, \ldots, X_{\sigma(n)})$ for all permutations σ of the arguments, and $A^{n=2}(X, Y) = X*Y$, then the operation * is said to be quasi-associative.

Counterargument. The only counterargument for quasi-associativity is that, like associativity (Section 4.1.6), it is evidently not an essential property of a useful aggregation operation.

4.1.9 Intersection-preserving property

One might expect that any agreement that may exist among the estimates to be aggregated would persist in the result of the aggregation. One aspect of agreement is covered by the idempotence property (Section 4.1.3), but another aspect of agreement concerns the mutual uncertainty among the estimates. In particular, consider the intersection of the three interval estimates [1, 8], [3, 5], and [4, 12], which is the interval [4,5]. Each of the three estimates considers the values in this interval as possible. Of course, if they didn't all overlap, the intersection would not exist. But, given that it does, perhaps the intersection should be a part of the result of an aggregation operation applied to the three intervals.

Intersections of intervals, p-boxes and Dempster-Shafer structures can be computed with the intersection operation (Section 3.2.1.4). Because of their inherent precision, a collection of real numbers wouldn't have an intersection unless the real numbers all happened to be identical. Likewise, a collection of probability distributions would not have an intersection unless all of the probability distributions are identical. It is clearly possible, then, for estimates to broadly overlap without having an intersection. The intersection identifies the portion of epistemic uncertainty common among the estimates.

It may be reasonable to expect an aggregation operation to preserve an intersection, if one exists, in its result. Given that none of the estimates makes a specific claim about the quantity within the region of intersection, one might be surprised to find that, together, somehow they do. An operation that respects the intersection of uncertain numbers would tend to preserve any agreement about the epistemic form of uncertainty

that exists among the estimates. In the special case that the X_i are intervals, if $X_1 \cap X_2 \cap \ldots \cap X_n \subseteq A(X_1, X_2, \ldots, X_n)$ whenever $X_1 \cap X_2 \cap \ldots \cap X_n \neq \varnothing$, the operation A is said to be intersection-preserving. More generally, for uncertain numbers X_i, the aggregation operation is intersection-preserving if the aggregation of X_1, X_2, …, and X_n encloses their intersection whenever one exists (intersection is taken in the sense of Section 3.2.1.4 and enclosure is taken in the sense of Section 4.1).

Counterargument. Bayes' rule, Dempster's rule, logarithmic pooling and averaging under independence are not generally intersection-preserving. If one nevertheless considers these to be useful aggregation operators, then this property is evidently not an essential one for aggregations. This property will usually be irrelevant when the estimates are all real numbers or all probability distributions because they will generally not have intersection in the first place.

4.1.10 Enclosure-preserving property

It might also be reasonable to expect that tightening the uncertainty in any input estimate could only tighten the aggregation result that used it. In other words, it would be surprising if the result of an aggregation could get wider (more uncertain) even though the input estimates were narrower. If $X_i' \subseteq X_i$ necessarily implies that $A(X_1, X_2, \ldots, X_i', \ldots, X_n) \subseteq A(X_1, X_2, \ldots, X_i, \ldots, X_n)$, the operation A is said to enclosure-preserving.

Counterargument. Like the notion of the preserving intersection, the idea of preserving enclosure is only useful for intervals, p-boxes and Dempster-Shafer structures. Precise real numbers wouldn't enclose each other unless they happened to be identical, and the same is true for precise probability distributions.

4.1.11 Insensitivity to vacuousness

What should be done if one of the experts consulted has no opinion at all on the question asked? Or, what if the expert's opinion is that *no one knows* the answer to the question. If one of the estimates to be aggregated is the vacuous interval $[-\infty, +\infty]$, or its equivalent p-box $[H_{-\infty}(x), H_{+\infty}(x)]$ or Dempster-Shafer structure $\{([-\infty, +\infty], 1)\}$, one might expect that the aggregation operation should ignore this estimate. How else would an abstention be treated? If $A(X_1, X_2, \ldots, X_n, [-\infty, +\infty]) = A(X_1, X_2, \ldots, X_n)$, the operation is said to be insensitive to vacuousness.

Counterargument. Although it might be appropriate to ignore an abstention, it might not be so reasonable to ignore a positive claim of general ignorance. For instance, if one of the experts asserts that none of the other experts' estimates are tenable and that the scientific discipline is too immature to support any claim on the subject, it would be hard to dismiss this assertion simply because it is inconvenient. Many aggregation operators are sensitive to vacuousness. Indeed, no operation could be otherwise if it reflects the number of estimates being aggregated.

4.1.12 Narrowness

One might think it desirable that an aggregation operation not offer results that go beyond the span of the original estimates. To do so would be to engage in a kind of aggregation activism that could seem anti-empirical in spirit. For instance, if one estimate is [1, 5] and the other estimate is [2, 6], one might look askance at an answer that suggested the

value might be 10. An operation A is called narrow if $A(X_1, X_2, \ldots, X_n) \subseteq \text{envelope}(X_1, X_2, \ldots, X_n)$, where envelope is the operation defined in Section 3.2.1.3. Some property like narrowness would seem to be a basic feature of any well behaved aggregation method.

Counterargument. Some aggregation operations such as Dempster's rule or certain modifications of intersection reflect inconsistency among the estimates by partially enlarging uncertainty beyond the span of their inputs. Likewise, all strategies to account for statistical sampling uncertainty among the estimates inflate the uncertainty beyond that seen in the original estimates. Insisting that an aggregation operation be narrow would incapacitate any such strategy.

4.2 Null aggregation

An alternative idea to aggregation is to keep all the estimates separate and perform multiple, contingent analyses for each separately. Although this is hardly an approach to aggregation, it deserves mention only to point out that it is an option available to the analyst who cannot decide on an aggregation. The drawback of course is that this approach yields no overall synthesis about the result of calculations, but it can provide direct answers to some fundamental questions that beset analysts.

Usually, analyses would include a computationally intensive sensitivity or 'what-if' study. The computational costs increase as a combinatorial function of the number of variables for which no aggregation is made, and, of course, these costs can become quite significant very quickly. For instance, if there were three variables, each of which had only five possible alternative values, then null aggregation would require $5^3 = 125$ separate analyses. It is sometimes possible to obtain perhaps most of the utility of a what-if study that makes no decisions about aggregation at a fraction of what would be its computational cost by using an enveloping strategy (Section 4.4) instead of null aggregation.

4.3 Intersection

When the estimates to be aggregated represent enclosures of the uncertainty about a quantity, that is, when each comes with a claim that the quantity is sure (in some strong sense) to lie within limits given by the estimate, then intersection is perhaps the most natural kind of aggregation to use. The idea is simply to use the smallest region that all estimates agree is possible with high confidence as the aggregation result. For instance, if we know for sure that a real value a is within the interval $\boldsymbol{x} = [1, 3]$, and we also know, by some separate argument or evidence, that a is also within the interval $\boldsymbol{y} = [2, 4]$, then we may conclude that a is certainly within the interval $\boldsymbol{x} \cap \boldsymbol{y} = [2, 3]$.

For interval inputs $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, the formula for intersection is the familiar form

$$\boldsymbol{x}_1 \cap \boldsymbol{x}_2 \cap \ldots \cap \boldsymbol{x}_n = [\ \max(x_1^-, x_2^-, \ldots, x_n^-), \min(x_1^+, x_2^+, \ldots, x_n^+)\ ]$$

where $x_i^- = \inf(x \mid x \in \boldsymbol{x}_i)$ and $x_i^+ = \sup(x \mid x \in \boldsymbol{x}_i)$ are the respective endpoints of the intervals. (If the intervals are all closed, then inf and sup can be replaced by min and max respectively.)

The most general definition of intersection can be specified in terms of probability boxes. If there are n p-boxes $F_1 = [\overline{F_1}, \underline{F_1}]$, $F_2 = [\overline{F_2}, \underline{F_2}]$, …, $F_n = [\overline{F_n}, \underline{F_n}]$, then their intersection is defined to be

$$F_1 * F_2 * \ldots * F_n = [\,\overline{F}\,*(x), \underline{F}*(x)]$$

where

$$\overline{F}\,*(x) = \min(\,\overline{F}_1(x),\ \overline{F}_2(x),\ \ldots,\ \overline{F}_n(x)\,),$$
$$\underline{F}\,*(x) = \max(\,\underline{F}_1(x), \underline{F}_2(x), \ldots, \underline{F}_n(x)\,).$$

The operation is defined whenever $\underline{F}\,*(x) \leq \overline{F}\,*(x)$ for all x. This operation is used when the analyst knows (or is highly confident) that *each* of multiple p-boxes encloses the distribution of the quantity in question. The argument that leads to intersection for p-boxes is exactly the same bounding argument that is used for intervals (Rowe 1988).

This formulation extends to Dempster-Shafer structures* immediately. The cumulative plausibility and belief functions of such structures form p-boxes. The result of aggregating these p-boxes can then be translated back into a Dempster-Shafer structure by canonical discretization (see Section 2.3).

As mentioned in Section 3.2.1.4, the intersection operation on uncertain numbers is rigor-preserving in that if the claims represented by the separate estimates are true, then the result of the intersection is also sure to enclose the quantity (Rowe 1988). This method is also best possible in the sense that it could not be any tighter given the stated information. Intersection also enjoys most of the properties outlined in Section 4.1. It is idempotent, commutative, and symmetric in its arguments. It is intersection-preserving, enclosure-preserving, insensitive to vacuousness, and narrow.

When the intersection operation is applied to collections of estimates within the various subcategories of uncertain numbers, the result, if it exists, will be another uncertain number from the same class. For instance, the intersection of a collection of intervals, if it exists, is another interval. Likewise, the intersection of p-boxes will be a p-box. Intersection is a general operation (sensu Section 4.1.1) in that it can be applied to, and it yields coherent results for, all the kinds of estimates we consider in this report. However, it will usually not be useful for real numbers or precise probability distributions. The reason, of course, is that these objects express only variability and no incertitude. In this regard, they are making very specific statements about what a quantity is or how it varies. Unless the inputs happen to be identical, the intersection among real numbers, or among precise probability distributions, usually does not exist. Where it is defined, intersection is continuous, but it is not a continuous operation generally. The reason is that as the overlap becomes smaller and eventually disappears, the operation suddenly fails to yield a result.

Despite its several desirable properties, some analysts feel that intersection has only limited utility for aggregation because it requires the very strong assumption that the

---

*The idea of intersection is also integral in the definition of Dempster's rule and most of its variants (see Section 4.5). However, in that context, it is used at an entirely different level.

individual estimates are each absolutely correct. It might be ill advised, for instance, to apply this operation to expert opinions if any of the experts might be wrong. In real problems accounting for uncertainty, wrong opinions may of course often be more abundant than correct ones.

4.3.1 Numerical example

Suppose we want to aggregate the following three estimates, each of which we take as a completely reliable estimate of a quantity:

$$X = [0, 50],$$

$$Y \sim \text{normal}([5, 10], [2, 3]),$$

$$Z = \{ ([-\infty, 122], 0.1), ([-1, 100], 0.4), ([8, 13], 0.25), ([12, 20], 0.15), ([0, 40], 0.1) \}.$$

The first estimate X is just an interval. The second Y is a p-box (see Section 3.1) specified by the class of normal distributions whose means are between 5 and 10 and whose standard deviations are between 2 and 3. The third estimate Z is a Dempster-Shafer structure that may have come from expert opinion or empirical observations. The sequence of three graphs below depict the p-boxes corresponding to these three inputs. The tails of the graph for Y extend to infinity in both directions. The left tail of the graph for Z extends to negative infinity. These three graphs are depicting the cumulative plausibility and cumulative belief functions of the corresponding Dempster-Shafer structures. The Dempster-Shafer structure for X is $\{([0, 50], 1)\}$. For computational purposes, this is the same as $\{([0,50], 0.01), \ldots, ([0,50], 0.01)\}$ in which there are 100 redundant intervals, each with mass 0.01. The canonical Dempster-Shafer structure for Y is $\{([-\infty, 20.3], 0.01), ([-6.6, 20.9], 0.01), ([-5.3, 21.2], 0.01), ([-4.4, 21.5], 0.01), \ldots, ([8.5, 34.4], 0.01), ([8.8, 35.3], 0.01), ([9.1, 36.6], 0.01), ([9.7, \infty], 0.01)\}$. Z is already expressed as a Dempster-Shafer structure. For computational purposes, it can be canonically discretized so that it has 100 (partially redundant) focal elements, each with mass 0.01. This would mean, for instance, that there would be 10 redundant copies of the first focal element $[-\infty, 122]$, each with a mass of 0.01, and 40 copies of the second focal element and so on.

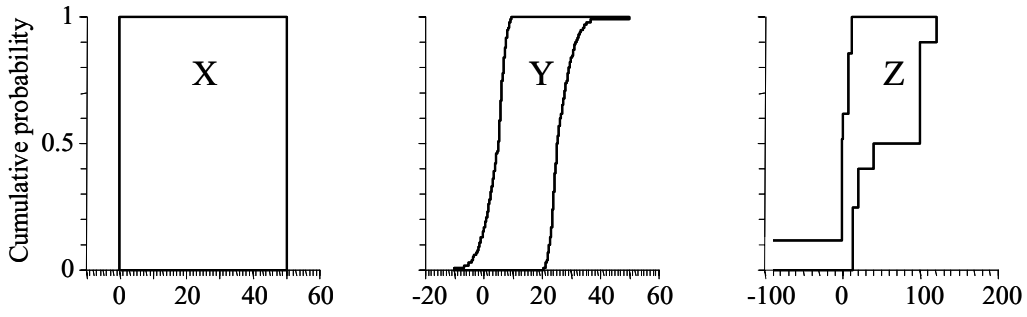


Figure 29: Three uncertain numbers *X*, *Y*, *Z* (see text).

The graph below depicts the intersection $X * Y * Z$ of the three estimates in terms of the p-box. The left and right sides of the p-box are also the cumulative plausibility and belief functions, respectively, of the associated Dempster-Shafer structure. The structure itself is { ([0, 13], 0.16), ([0.0277, 13], 0.01), ([0.229, 13], 0.01), ([0.423, 13], 0.01), ([0.611, 13], 0.01), ..., ([12, 35.27], 0.01), ([12, 36.63], 0.01), ([12, 50], 0.01) }. This structure was obtained simply by intersecting the respective focal elements from the canonical discretizations of the three inputs, and then accumulating the masses of any redundant elements.

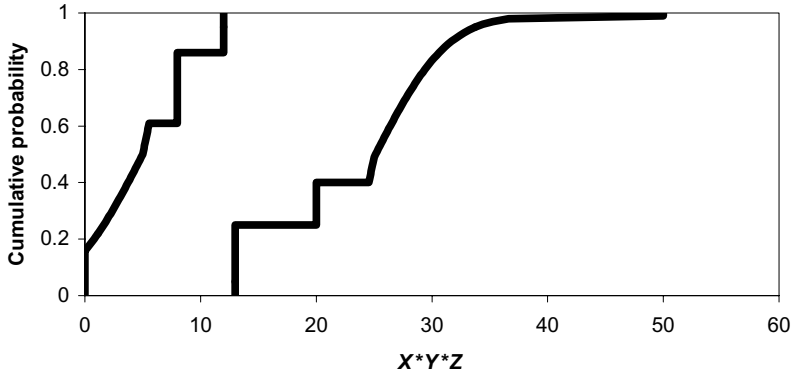


Figure 30: Intersection of *X*, *Y*, and *Z*.

The curved portions of this result come from the input p-box *Y* of course. The portions that are step functions come from the Dempster-Shafer structure *Z*. The interval estimate *X* is also important because it defines the range and constrained both the left and right tails more strongly than either the p-box *Y* (which is theoretically unbounded) or the Dempster-Shafer structure *Z*.

4.3.2 Strategies when the intersection is empty

Intersection satisfies so many of the desirable properties for aggregations that it seems reasonable to try to fix what is essentially the single problem with it: that it doesn't give any answer at all when the estimates don't overlap. Let's consider the question in the simplest situation of aggregating two intervals $\boldsymbol{x}$ and $\boldsymbol{y}$. What happens if we slowly move the interval $\boldsymbol{y}$ so that its intersection with $\boldsymbol{x}$ becomes smaller and smaller and finally disappears? So long as it is not empty, because we assume both uncertainty intervals to be 100% reliable, we conclude that the result of the aggregation is the intersection. However, when the intersection becomes empty, it clearly means that the two intervals cannot both reliably contain the (unknown) value of the estimated quantity. Consequently, at least one of these two intervals is erroneous. If we do not know which of the two intervals is erroneous, then we might conclude about the actual value of the underlying quantity is that either it belongs to the first interval $\boldsymbol{x}$ (and the second interval is erroneous), or it belongs to the second interval $\boldsymbol{y}$ (and the first interval is erroneous). But, if this is all that can be said, then the best we can say about the quantity is that it lies somewhere in the union $\boldsymbol{x} \cup \boldsymbol{y}$ of the two intervals. This union is not an interval, so, if we want an interval that is guaranteed to contain the quantity, then we might take the

smallest interval that contains this union. This is the envelope (convex hull) described in the next section. Whether we take the union itself or the envelope, the aggregation result is not at all close to the intersection and so the continuity property does not hold. It does, at least, always yield *some* answer and that may be the most important thing in practice.

Another argument holds that the lack of an overlap for estimates that are both supposed to be rigorous enclosures suggests a fundamental error somewhere that deserves the analyst's special attention. Under this argument, perhaps the result that should be given when there is no overlap is the vacuous interval $[-\infty, +\infty]$. This doesn't repair the continuity of the operation either, but it may be more generally appropriate than the union or envelope as the default. It is clear, in any case, that an empty intersection will require some sort of reconsideration by the analyst.

4.4 Envelope

In the previous section on intersection, it was presumed that all of the estimates to be aggregated were completely reliable. When the analyst instead is confident only that *at least one* of the input estimates encloses the distribution of the quantity, but doesn't know which estimate it is that does, enveloping should be used to aggregate the estimates into one reliable characterization. Such knowledge could arise from a mechanistic understanding of a system. For instance, if there is water in a hermetically sealed compartment, an engineer might be able to conclude that it must have gotten there either from condensation of water vapor or a leak in the containment. Knowledge that at least one of several scenarios must be correct could be obtained from a process of elimination. Enveloping is a strategy that allows a risk analysis to proceed even though the eliminations could not be taken to completion to identify a single scenario.

In general, when the estimates to be aggregated represent claims about the true value of a quantity that have uncertain reliability individually, enveloping is often a prudent aggregation method to use. The idea is to identify the region that *any* estimate suggested might be possible as the aggregation result. For instance, if one opinion or measurement says the value is 1 and another says it's 2, we might elect to use the interval [1,2] as our aggregated estimate.

If there are n p-boxes $F_1 = [\overline{F_1}, \underline{F_1}]$, $F_2 = [\overline{F_2}, \underline{F_2}]$, …, $F_n = [\overline{F_n}, \underline{F_n}]$, then their envelope is defined to be

$$F_1 * F_2 * \ldots * F_n = [\overline{F}*(x), \underline{F}*(x)]$$

where

$$\overline{F}*(x) = \max(\overline{F}_1(x), \overline{F}_2(x), \ldots, \overline{F}_n(x)),$$
$$\underline{F}*(x) = \min(\underline{F}_1(x), \underline{F}_2(x), \ldots, \underline{F}_n(x)).$$

This operation is always defined. It is used when the analyst knows that at least one of multiple p-boxes describes the distribution of the quantity in question.

This formulation extends to Dempster-Shafer structures immediately. The cumulative plausibility and belief functions of such structures form p-boxes. The result of aggregating these p-boxes can then be translated back into a Dempster-Shafer structure by canonical discretization (see Section 2.3).

For interval inputs $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, the general formulation for enveloping reduces to the convex hull

$$\text{envelope}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n) = [\ \min(x_1^-, x_2^-, \ldots, x_n^-), \max(x_1^+, x_2^+, \ldots, x_n^+)\ ]$$

where $x_i^- = \inf(x \mid x \in \boldsymbol{x}_i)$ and $x_i^+ = \sup(x \mid x \in \boldsymbol{x}_i)$ are the respective endpoints of the intervals.

Enveloping is not the only way that the estimates could be aggregated under the assumption that at least one of the inputs encloses the distribution of the quantity. One could, instead, simply take the union of the estimates. Using union for this purpose is roughly equivalent to null aggregation (Section 4.2) in that it can quickly become computationally unwieldy. The advantage of the envelope is that it is closed in the set of uncertain numbers. The envelope of reals or intervals is an interval. The envelope of probability distributions, p-boxes or Dempster-Shafer structures is a p-box.

As mentioned in Section 3.2.1.3, the envelope operation is rigor-preserving and best possible. Like intersection, enveloping enjoys most of the properties outlined in Section 4.1. It is general, idempotent, commutative, and symmetric in its arguments. It is intersection-preserving, enclosure-preserving, and narrow. Unlike intersection, it is continuous and always produces an answer whenever the inputs are uncertain numbers.

However, enveloping is not insensitive to vacuousness. In other words, it is sensitive to claims of general ignorance and even abstentions that make no claim about the quantity. This means that if only one informant expert or empiricist offers an inconclusive opinion or the vacuous interval as the measurement result, it will determine the result of the aggregation. The overall result of enveloping will be as broad as the broadest input. The simplistic strategy of just omitting any vacuous estimates before computing the envelope would not be sufficient in practice. This is because any estimate that is not vacuous but just very wide could still swamp all other estimates.

4.4.1 Numerical example

We aggregate the following five estimates with the envelope operation:

$$V = 5.3,$$

$$W = [8, 10],$$

$$X = \text{weibull}(\ 15, 3),$$

$$Y = \text{normal}(\ [3, 4], [2, 3]),$$

$$Z = \{\ ([-1, 5], 0.25), ([8, 13], 0.5), ([12, 20], 0.15), ([0, 40], 0.1)\ \}.$$

The first two estimates are a real number and an interval respectively. The third is a precise probability distribution from the Weibull family with a scale (characteristic life) parameter of 15 and a shape parameter of 3. The fourth estimate is a p-box specified by the class of normal distributions whose means are between 3 and 4 and whose standard deviations are between 2 and 3 (see Section 3.1). The fifth estimate is a Dempster-Shafer structure that may have come from expert opinion or empirical observations. The display below depicts the envelope $V * W * X * Y * Z$ of these five estimates in terms of the resulting p-box. The left and right sides of this p-box are also the cumulative plausibility and belief functions, respectively, of the associated Dempster-Shafer structure. The

structure itself is { ([−4.73, 10], 0.01), ([−3.98, 10], 0.01), ([−3.16, 10], 0.01), ..., ([5.07, 20], 0.01), ([5.16, 20], 0.01), ([5.25, 20], 0.01), ([5.3, 40], 0.12) }. This result was obtained in a manner very similar to the computation described for intersection in Section 4.3.1. Each of the five inputs were canonically discretized into data structures composed 100 intervals (each with mass 0.01), and convex hulls were computed from the five intervals at each of the 100 discretization levels. The result was condensed into a Dempster-Shafer structure by accumulating the mass from any redundant intervals.

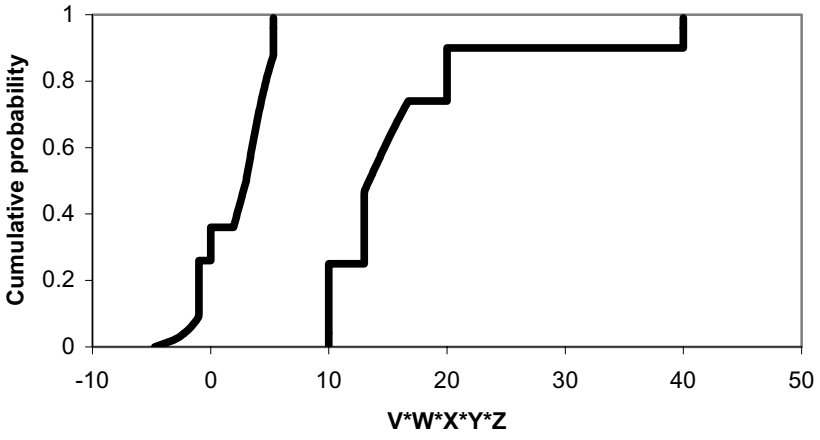


Figure 31: Envelope of five uncertain numbers.

4.4.2 Using envelope when estimate reliabilities are uncertain

Enveloping is commonly employed in a variety of decision-making contexts where the reliability of individual estimates is uncertain. For example, local police responding to a reported skyrage incident at the Albuquerque Sunport initially arrested 14 people. All but one were released the same day. The reasoning of the arresting officers must have been similar to that offered by the envelope aggregation. In the section on intersection, it was presumed that each of the estimates to be aggregated was completely reliable. The reasoning behind the intersection operation is like that of a logician who knows two things with certainty and can conclude from them something surer. But when a police officer arrives at the scene of a disturbance and hears several conflicting stories, he or she may not be able to judge which witnesses are credible. The cop doesn't have the surety of the logician about any of these stories. Thinking that one of the stories is probably true, or at least feeling that none of the stories can be assumed to be false, an officer might choose enveloping (that is, arresting everyone implicated by anyone) as a prudent strategy.

An important limitation of using enveloping for the case when the reliability of the estimates is unknown is that it relies on the hypothesis that some estimate is true.

Enveloping is sure to totally enclose the distribution of a quantity only if at least one* of the original estimates does so. Yet, in the case where the reliability of each of the estimates is unclear, it may not be certain that any of the estimates actually does this. All of the witnesses could be lying or be simply mistaken. Therefore, when the reliability of the individual estimates to be aggregated is uncertain, enveloping will often be a useful strategy, but it will not be an infallible one.

4.4.3 Caveat

Given that enveloping will be applied in situations such as that described in the previous section where one is not certain that *any* of the estimates will enclose the distribution of the quantity, how can an analyst ensure that the envelope will be broad enough to actually do so? The analyst must somehow ensure that the envelope is broad enough to enclose the true variation and uncertainty about the quantity. Unfortunately, this responsibility is not easily satisfied with any simple recipe.

What if the estimates to be aggregated are only samples collected from some larger population and it is really this population as a whole that is our interest? The question of accounting for sampling uncertainty as a part of the aggregation process is addressed in Section 4.11.

4.5 Dempster's rule and its modifications

The central method in the Dempster-Shafer theory of belief functions is Dempster's rule for combining evidence (Shafer 1976; Dempster 1967). Because the rule has some counterintuitive properties, various alternative versions of the rule have been offered by various authors (e.g., Yager 1983;1985; Zadeh 1986; Yager 1987; Halpern and Fagin 1992; Baldwin 1994a; Chateauneuf 1994; Dubois and Prade 1994; Kreinovich et al., 1994; Kruse and Klawonn 1994; Saffiotti 1994; Spies 1994; Yager et al., 1994; Zhang 1994; Mahler 1995; Srivastava and Shenoy 1995; Yager 2001). In this section, we briefly describe only three of the most important of the versions of the combination rule for Dempster-Shafer structures. A considerably more extensive review of this literature is available in Sentz and Ferson (2002), which also contains several numerical examples of these and other combination rules.

4.5.1 Dempster's rule

The combination via Dempster's rule of two independent Dempster-Shafer structures having basic probability assignments m_1 and m_2 is another Dempster-Shafer structure whose basic probability assignment is (Shafer 1986)

$$m(D) = \frac{\sum_{B \cap C = D} m_1(B) m_2(C)}{1 - K},$$

*Because enveloping is often broader than union, it might also enclose the true distribution even if *none* of the original estimates does. However, such an outcome cannot be assured.

where D, B and C are, in this report, closed intervals of the real line, and

$$K = \sum_{B \cap C = \varnothing} m_1(B) m_2(C)$$

is the mass associated with the conflict present in the combined evidence. Because all of the focal elements of one Dempster-Shafer structure are convolved with those of the other (that is, all possible pairs are considered), Dempster's rule can be characterized as a convolutive intersection (cf. convolutive averages in Section 4.9.1). For this reason, the question of the dependence between the Dempster-Shafer structures to be aggregated arises. Shafer (1984) emphasizes that Dempster's rule applies when the arguments or bodies of evidence to be aggregated are independent. Voorbraak (1991) pointed out that Shafer's notion of independence is not the same as that of stochastic independence in probability theory. As Couso et al. (2000) explore, there are many variations of the notion of independence once the strictures of probability theory are relaxed. In principle, it would be possible to fashion a version of Dempster's rule that makes a different assumption about the dependence between the estimates being aggregated, or even one that does not make any assumption about the dependence between the estimates.

Dempster's rule can be applied to p-boxes by first converting them to Dempster-Shafer structures by canonical discretization. An interesting fact is that the vertical distance between the left and right bounds of the resulting p-box at any point x will be proportional to the product of the comparable vertical distances at the same point of all the p-boxes being aggregated. When the rule is applied to precise probability distributions, it is equivalent in the limit* to logarithmic pooling (Section 4.8). In this case, the density function of the aggregation result is proportional to the product of the density functions of the distributions being aggregated. When the rule is applied to intervals, it reduces to simple intersection.

Like intersection, Dempster's rule technically satisfies the generality criterion discussed in Section 4.1.1, but it will not be useful for aggregating real numbers. The rule is commutative and associative, but not idempotent or continuous. Dempster's rule is narrow in the sense that its results will always be within the envelope of the input estimates, but, because it can give results that are tighter than the intersection of these estimates, it is not intersection-preserving. A counterexample can be constructed with the Dempster-Shafer structures A = {([4,14], 1/3), ([8,18], 1/3), ([12,22], 1/3)} and B = {([8,18], 1/3), ([12,22], 1/3), ([16,26], 1/3)}. The intersection (sensu Section 4.3) of A and B is just C = {([8,14], 1/3), ([12,18], 1/3), ([16,22], 1/3)}, as is clear from inspection. The result of applying Dempster's rule to A and B yields D = {([8,14], (1/9)(9/8)), ([8,18], (1/9)(9/8)), ([12,14], (1/9)(9/8)), ([12,18], (2/9)(9/8)), ([12,22], (1/9)(9/8)), ([16,18], (1/9)(9/8)), ([16,22], (1/9)(9/8))}, where the multiplier (9/8) accounts for the fact that one of the pairs of intervals does not have an intersection. The result of the aggregation does not enclose the intersection, thus the operation is not intersection-preserving.

---

*Actually, there are different ways to go to the limit (Halpern and Fagin 1992; cf. Shafer 1986). One of these ways makes Dempster's rule and Bayes' rule equivalent. Another way suggests that Dempster's rule applied to precise probability distributions should yield a density function that is the *maximum* of the density functions of the estimates being aggregated.

Dempster's rule is enclosure-preserving and insensitive to vacuousness. However, the rule will sometimes produce counterintuitive answers when there is substantial conflict among the aggregated estimates (Zadeh 1986). The following example illustrates the problem. Suppose we are to aggregate two Dempster-Shafer structures that have the forms $D_1 = \{(A, 0.999), (B, 0.001), (C, 0.0)\}$ and $D_2 = \{(A, 0.0), (B, 0.001), (C, 0.999)\}$, where $A = [1,2]$, $B = [3,4]$, and $C = [5,6]$. Both D_1 and D_2 agree that B is very unlikely, yet the result of the aggregation under Dempster's rule is $\{(B, 1.0)\}$, just because B is the only area of agreement between the two input structures. Sentz and Ferson (2002) review this issue. Voorbraak (1991) reviews several other weaknesses of Dempster's rule.

4.5.2 Yager's rule

Dempster's rule forgets all the mass that was lost to inconsistency between the bodies of evidence. This essentially ignores the conflict that may be present. Yager's (1987) modification to the rule assigns the mass that would otherwise be lost to the universal set instead. For the purposes of this report, the universal set is always the real line $\Re$. The basic probability assignment of the result is

$$m(D) = \begin{cases} q(D), & \text{if } D \neq \Re \\ q(D) + \mathrm{K}, & \text{if } D = \Re, \end{cases}$$

where D is a closed interval of the real line,

$$q(D) = \sum_{B \cap C = D} m_1(B) m_2(C),$$

and again

$$K = \sum_{B \cap C = \varnothing} m_1(B) m_2(C).$$

Like Dempster's rule, this rule is commutative, but neither idempotent nor continuous. It fails to be associative, but Yager shows the obvious generalization that makes the rule quasi-associative and symmetric in its arguments. The lack of idempotence means that Yager's rule is not intersection-preserving. It is easy to find counterexamples that show that it is not enclosure-preserving either. But it is narrow and insensitive to vacuousness. Yager's rule is an important contribution to the literature of Dempster-Shafer structures because it offers a reasoned way to handle large disagreement between bodies of evidence.

4.5.3 Disjunctive consensus

In another attempt to correct the shortcomings of the original version of Dempster's rule, Dubois and Prade (1986; 1992) define an alternative that is based on unions rather than intersections

$$m(D) = \sum_{B \cup C = D} m_1(B) m_2(C)$$

where D, B and C are, in this report, closed intervals of the reals. In the same way that Dempster's rule is a convolutive intersection, this rule is a convolutive union. Because this rule produces Dempster-Shafer structures whose focal elements are not closed intervals whenever B and C happen not to overlap, this rule is not closed in the space of uncertain numbers defined in Section 2.4. If we replace the union condition for the summation with an enveloping condition, then the disjunctive consensus rule becomes

$$m(D) = \sum_{\text{env}(B,C)=D} m_1(B)m_2(C)$$

where env(B, C) denotes the convex hull of the intervals B and C. This "envelope variant" of Dubois and Prade's disjunctive consensus is closed in the space of uncertain numbers. This means that every time it is applied to real numbers, intervals, probability distributions, p-boxes or finite Dempster-Shafer structures with interval focal elements, it would produce a result that is also from this same assemblage.

We mention that the connection between Dubois and Prade's rule and this envelope variant is very close. If all the focal elements of all the Dempster-Shafer structures to be aggregated overlap with each other, then the two rules would yield the same result. Even if they don't, the cumulative plausibility and belief functions of the Dempster-Shafer structure arising from disjunctive consensus would be exactly the same as that arising from the envelope variant. This means that the associated p-boxes resulting from the two rules would be the same.

Both disjunctive consensus and the envelope variant of disjunctive consensus are defined for any Dempster-Shafer structure and therefore are general in the sense of Section 4.1.1. Interestingly, the results of simple enveloping (Section 4.4) are *tighter* than the results of disjunctive consensus because the latter involves a convolution, the Cartesian product of all unions, rather than only the level-wise unions. Disjunctive consensus is commutative and associative, but it is not idempotent. Because simple enveloping produces aggregations that can be tighter than the results of disjunctive consensus, this rule is not narrow. It is also sensitive to vacuousness, but it is intersection-preserving and enclosure-preserving.

This aggregation method yields results that can be as counterintuitive as those of Dempster's rule. For instance, Jon Helton (pers. comm.) notes that aggregating two Dempster-Shafer structures, each of which has the form {([0,1], 0.5), ([−∞, ∞], 0.5)}, yields the answer {([0,1], 0.25), ([−∞, ∞], 0.75)}. Thus, although both inputs agree that the quantity is within the unit interval with mass 0.5, their aggregation is somehow much less sure, offering only 0.25 mass for being in the same interval. This behavior is perhaps even worse than the counterintuitive behavior that Zadeh (1986) criticized in the original version of Dempster's rule. Nevertheless, the aggregation rule based on disjunctive consensus remains an important version of Dempster's rule that is widely used in the literature of Dempster-Shafer structures.

4.6 Bayes' rule

The Bayesian aggregation methods described in the literature are usually applied to probability distributions reflecting expert opinion. In this section, we imagine that

Bayes' rule is used to aggregate uncertain numbers that might represent expert opinion or empirical evidence, or both. Bayesians generally hold that how one should collect and analyze data depends on what use will be made of those data. This perspective might therefore look disapprovingly on the intention of this report to catalog methods by their general utility rather than for a specific purpose.

As mentioned in the earlier discussion in Section 3.3.3, the computational burdens associated with applying Bayes' rule can be substantial. In particular, there is usually no closed-form solution available for computing the integral in the denominator of Bayes' rule, unless the prior and likelihood happen to constitute a "conjugate pair" for which the analytical details work out nicely. For instance, under particular assumptions, the following pairs of likelihood (from which observations are drawn) and prior yield the given posterior distribution:

Likelihood	*Prior*	*Posterior*
Bernoulli	Beta	Beta
Binomial	Beta	Beta
Poisson	Gamma	Gamma
Negative binomial	Beta	Beta
Normal	Normal	Normal
Normal	Gamma	Gamma
Exponential	Inverse-gamma	Inverse-gamma

For these pairs, updating rules permit the immediate specification of the posterior's parameters from those of the prior and statistics from the data. For the assumptions underlying the use of these conjugate pairs and details on exactly how the calculations are to be made, consult standard references on Bayesian methods (e.g., Lee 1997; Sander and Badoux 1991; Berger 1985; DeGroot 1970; Gelman et al. 1995). Naturally, the existence of these conjugate pairs greatly simplifies the demands of applying the rule and are widely used for the sake of convenience, but of course they are very restricted in scope and obviously require distributional assumptions.

Bayesian methods are regarded by many analysts as the best (or even the only coherent) way to represent and manipulate uncertainty in the context of decision making (French 1985; Genest and Zidek 1986; cf. Cooke 1991; Clemen and Winkler 1999). Unfortunately, it seems clear that the touted advantages of Bayesian methods (collectively called "rationality") do not generally extend to the problem of aggregating uncertain numbers. For instance, Mongin (1995) showed that, under certain conditions, Bayesians cannot aggregate the preferences of multiple rational decision makers in a way that simultaneously satisfies the Bayesian axioms and the Pareto property (that is, whenever all decision makers agree about a preference, the aggregation result also has that preference). French (1985) pointed out the impossibility of any aggregation method simultaneously satisfying all the propertiesone would hope for. Jaffray (1992) considered the even nastier problems of generalizing Bayesian updating to belief functions and interval probabilities.

There are actually several ways that Bayes' rule could be used to combine uncertain numbers estimating a single quantity. Winkler (1968) provided a Bayesian framework for addressing the aggregation probability distributions and formulating a weighting

scheme. Lindley et al. (1979) described two ways to do aggregation in a Bayesian context. The first way, which they call the "internal" approach, presumes that there is a true probability and the various inconsistent estimates of it constitute the available evidence about this true probability. The second approach, which they call "external", does not explicitly concern itself with the true probabilities, but instead addresses only the problem of deriving coherent probabilities based on the original set of incoherent assessments. Both approaches require the participation of an analyst—different from any of the experts providing estimates—who performs the aggregation. It is the coherent probabilities of this analyst, as informed by the available evidence from the various sources or by opinions of the various experts, that is to be the result of the calculation. In this sense, these approaches are not so much methods of aggregation per se, but rather merely another avenue for elicitation.

Clemen and Winkler (1999) note that the Bayesian paradigm is very difficult in practice to apply to the problem of aggregation. The biggest problem is creating the conditional likelihood function, which must itself be a comprehensive model of the interrelationships between θ and the various estimates about it. In particular, it must characterize the precision and bias of the individual estimates and model the dependence among them. This dependence involves the degree and manner in which the estimates are associated with or related to each other. Clemen and Winkler (1999) review several different Bayesian models for aggregation that have been suggested by various authors. Typically, for instance, we might expect different experts to have correlated estimates.

Space limitations preclude a comprehensive review of the possible Bayesian approaches to the problem of combining potentially conflicting estimates. In any case, if they require the existence of an analyst whose beliefs must be elicited separately, we cannot create numerical examples that fairly compare the Bayesian approaches with the other methods of aggregation described in this report.

The mathematical properties of a Bayesian method of aggregation will depend on which of the many possible models is actually used in calculation. Nevertheless, there are some behaviors that would usually be associated with any Bayesian approach. For instance, because there are no convenient general algorithms applying Bayes' rule, practical problems are often computationally challenging. The Bayesian aggregation methods that have been described are defined for precise probability distributions, although they could perhaps be extended to probability boxes via robust Bayes methods (see Section 3.3). They could be applied to Dempster-Shafer structures by first converting them to p-boxes. When applied to intervals, Bayesian aggregation typically reduces to intersection (Section 4.3). It is not really general in the sense of Section 4.1.1, however, because it gives no answers except in trivial cases when it's applied to real numbers. Because Bayes' rule strongly emphasizes agreements, aggregations would not be idempotent. It is not continuous; it fails to produce an answer if the prior and likelihood do not overlap. It is likely to sensitive to vacuousness and not be intersection-preserving, but is probably would be enclosure preserving and narrow.

See Section 3.3.2 for a numerical example of Bayes' rule applied to p-boxes.

4.7 Mixing

The idea of a stochastic mixture is that there are multiple values of a quantity that are expressed at different times, or in different places or under different situations. In the proverb of the blind men who encountered an elephant, very different stories were recounted. One, feeling the trunk, said the elephant was like a snake. One, feeling the elephant's leg, insisted the animal was like a tree. A third, feeling the animal's side, asserted that an elephant was like a wall. The point of the proverb is that all of these things are true at the same time. Stochastic mixture offers a perspective that can see how a quantity, like an elephant, can manifest different or conflicting values. Using mixtures to aggregate estimates treats any disagreement among these estimates as though it is variability. Unlike averaging (Section 4.9) mixing does not erase the disagreement, but rather condenses it into a single distribution (or p-box or Dempster-Shafer structure) that fully expresses it. In the context of a quantitative risk analysis that intends to carefully distinguish epistemic and aleatory uncertainty, it would be inappropriate to apply this approach to a case where the disagreement among the sources represents amorphous incertitude. That would be the same mistake as modeling incertitude about a quantity with a random variable. Our position is that mixtures are only appropriate when the disagreement among the various estimates represents actual variability.

4.7.1 Mixture

The simplest mixture is an unweighted mixture using arithmetic averaging of the distribution functions. The mixture of n p-boxes $[\overline{F_1}, \underline{F_1}]$, $[\overline{F_2}, \underline{F_2}]$, …, $[\overline{F_n}, \underline{F_n}]$ is

$$[\overline{F_1}, \underline{F_1}] * [\overline{F_2}, \underline{F_2}] * \ldots * [\overline{F_n}, \underline{F_n}] = [\overline{F}*, \underline{F}*],$$

where

$$\overline{F}*(x) = (\overline{F}_1(x) + \overline{F}_2(x) + \ldots + \overline{F}_n(x)) / n$$

and

$$\underline{F}*(x) = (\underline{F}_1(x) + \underline{F}_2(x) + \ldots + \underline{F}_n(x)) / n .$$

The mixture of n finite Dempster-Shafer structures with basic probability assignments m_1, m_2, …, m_n has the basic probability assignment

$$m*(A) = \frac{1}{n} \sum_i m_i(A) .$$

When the focal elements are closed intervals, a mixture of two finite Dempster-Shafer structures is essentially a pooling of the focal elements with the masses halved. For instance, the even mixture of the Dempster-Shafer structures

$$\{([a_1, b_1], p_1), \ldots, ([a_n, b_n], p_n)\}$$

and

$$\{([c_1, d_1], q_1), \ldots, ([c_r, d_r], q_r)\}$$

is the structure

$$\{([a_1, b_1], p_1/2), \ldots, ([a_n, b_n], p_n/2), ([c_1, d_1], q_1/2), \ldots, ([c_r, d_r], q_r/2)\}.$$

This formula is correct so long as the two structures have no focal elements in common. If they do have common elements, a condensation step would sum the masses for identical focal elements.

Unlike most of the aggregation methods considered so far in this report, there is no difficulty whatever for the mixture algorithm if the input estimates have inconsistencies. Even total disagreement can be captured. If the mixing aggregation is applied to real numbers, the result is a discrete probability distribution whose masses are at those same real values. If intervals are used as inputs, one gets a p-box. A mixture of probability distributions is another probability distribution. And mixing is a closed operation in the space of p-boxes or of Dempster-Shafer structure whose focal elements are closed intervals. Thus, mixing is general in the sense of Section 4.1.1. It is clearly also idempotent, commutative and continuous. It is not associative, but it is quasi-associative, and the multiargument version is symmetric in its arguments. Mixtures are intersection-preserving, enclosure-preserving and narrow, but they are sensitive to vacuous inputs.

4.7.2 Weighted mixture

The previous section discussed even mixtures, that is, mixtures whose components had equal frequencies. It is also possible to aggregate estimates with a weighted mixture. Because we use mixtures as representations of actual variability, it would not be appropriate to base weights on the mere credibility of the estimate. Instead, the weights should correspond to the frequencies with which each estimate occurs in the overall population modeled with the mixture. For instance, if we have an estimate of the performance of a material during summer months and an estimate of its performance during winter months, the weights with which the two estimates might be aggregated would reflect the relative frequencies of summer and winter months.

The result of mixing n p-boxes $[\overline{F_1}, \underline{F_1}]$, $[\overline{F_2}, \underline{F_2}]$, …, $[\overline{F_n}, \underline{F_n}]$, with respective weights $w_1, w_2, \ldots, w_n$, is $[\overline{F}^*, \underline{F}^*]$, where

$$\overline{F}^*(x) = (w_1\ \overline{F}_1(x) + w_2 \overline{F}_2(x) + \ldots + w_n\ \overline{F}_n(x)) / \Sigma\ w_i$$

and

$$\underline{F}^*(x) = (w_1\ \underline{F}_1(x) + w_2 \underline{F}_2(x) + \ldots + w_n\ \underline{F}_n(x)) / \Sigma\ w_i.$$

The weights must be positive. See Section 4.10 for a discussion on how weights can be selected to reflect the relevance of different estimates.

The weighted mixture of n finite Dempster-Shafer structures with basic probability assignments $m_1, m_2, \ldots, m_n$ has the basic probability assignment

$$m^*(A) = \frac{1}{\sum_i w_i} \sum_i w_i m_i(A)$$

where the weights must be positive. The aggregation for Dempster-Shafer structures can be accomplished, again, by a straightforward pooling algorithm that weights the masses appropriately.

In case the uncertain numbers are represented by precise probability distributions, this weighted mixture is also known as the "linear opinion pool" (Stone 1961)

$$p(x) = \sum_{i=1}^{n} w_i p_i(x)$$

where p denotes a probability density and the positive w_i sum to one. (It turns out that mixtures are computed in the same way whether in the density or the cumulative realm.) In principle, the weights could be different for different values of x, but this generality is probably not needed for most risk analysis problems (cf. Section 4.10). The linear opinion pool is perhaps the most commonly used aggregation method in probability theory.

Weighted mixing is closed in the space of uncertain numbers (Section 2.4) and general in the sense of Section 4.1.1. It is idempotent and continuous, but it is neither commutative nor associative. It is neither quasi-associative nor symmetric in its arguments. Weighted mixtures may not be intersection-preserving or enclosure-preserving, but they are narrow. They can be sensitive to vacuous inputs.

When the mixture weights are known precisely, the mixture aggregation is rigor-preserving in the sense that the result will surely contain the distribution of the quantity in question. This presumes, of course, that the input estimates enclose their respective conditional distributions and the weights are an accurate reflection of the relative frequencies of the various conditions. The mixture aggregation is also best possible in the sense that it could not be any tighter given only this information (i.e., the input estimates and their frequencies as weights). It is also possible to compute mixtures when the weights are known only to within intervals.

Weighted mixture models have also been suggested in another aggregation context. Moral and Sagrado (1998) suggested a scheme for aggregating imprecise probabilities* for the case of distributions on a discrete event space. They intended it for aggregating expert opinion, but it could be useful for combining uncertain numbers whatever their origin. Their scheme is essentially a generalization of linear pooling and envelope (Section 4.4) via a one-parameter linear combination. If their parameter c is set to zero, then the result of the aggregation would be the same as a linear pooling with uniform weights. The result is a precise probability distribution whenever the input estimates are precise. The aggregation also reflects any agreement among the estimates by tending toward the favored estimate. If the parameter c is set to unity, the aggregation yields the convex set of the imprecise probabilities. In this case, the number of estimates that may be in agreement is irrelevant.

*Imprecise probabilities are related to, but much more complex than, either probability boxes or Dempster-Shafer structures. They are often expressed in terms of closed convex sets of probability distributions.

4.7.3 Numerical examples

The graphs below depict three different mixtures. The mixture in far left graph is an unweighted arithmetic mixture of the intervals [1,2] and [3,4]. This aggregation would be appropriate if, for instance, the quantity varies with equal frequency between two states. The states are represented by the two intervals, whose widths represent the incertitude about the exact value or values taken by the quantity in each of the two states. This aggregation would be inappropriate if we did not know the relative frequencies of the two states. As the incertitude within each state decreases so that the intervals approach point values, the resulting mixture approaches a two-mass discrete distribution.

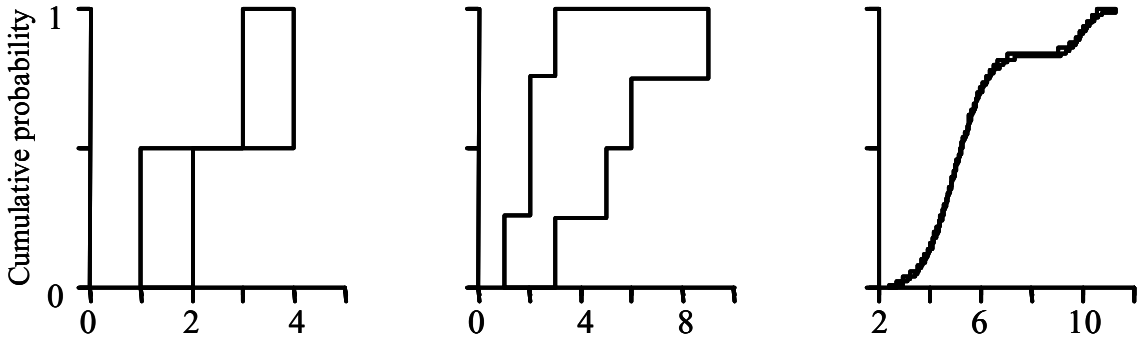


Figure 32: Three different mixture results.

The middle graph above depicts the (unweighted, arithmetic) mixture of two Dempster-Shafer structures: D_1 = { ([1, 5], 0.5), ([3, 6], 0.5) } and D_2 = { ([2, 3], 0.5), ([2, 9], 0.5) }. As in the previous example, this aggregation is appropriate only if the frequencies of the two states are known and equal. In this case, our empirical knowledge about the value(s) taken on by the quantity in each of the two states is more detailed than it was for the example that mixed two intervals. Evenly mixing the two Dempster-Shafer structures, which are themselves even mixtures of two intervals, implies considerable information about the quantity. It says that a quarter of the time the quantity is between 1 and 5, a quarter of the time it is somewhere in the interval [3,6], a quarter of the time it is in [2,3] and the rest of the time it is in [2,9]. But this is all that is being said. The quantity may be varying within these intervals, or it may have some fixed value in an interval that it always takes on. We don't know when the quantity is any particular interval, only that it occupies each interval with an overall frequency 0.25.

The far right graph above is the mixture of two normal distributions. The distribution with mean is 5 and a standard deviation of 1 has a weight of 5/6. The other distribution with mean is 10 and a standard deviation of 0.5 has a weight of 1/6. The tails of the resulting mixture distribution are truncated for the sake of the display. The result of this mixture is essentially perfect statistical knowledge about the behavior of the quantity. The mixture is a precise distribution that characterizes the frequencies of all possible values it takes on.

4.8 Logarithmic pooling

When the inputs estimates are all probability distributions, a commonly used aggregation method is the logarithmic opinion pool

$$p(x) = k \prod_{i=1}^{n} p_i(x)^{w_i}$$

where $p(x)$ denotes the probability associated with the value x, w_i are weights, and k is an normalization factor that ensures that all the probabilities add to one. This formulation requires certain restrictions on the weights to ensure that the result is a probability distribution, but actually permits negative values for some weights. Note this aggregation is defined in the density domain. (Pointwise multiplication of cumulative distributions yields the same result as the maximum convolution.)

The use of pointwise multiplication for all values along the x-axis and the normalization harken to Bayes' rule (Sections 3.3 and 4.6), but the method lacks the interpretation of the operands as prior and likelihood. It does however mimic Bayes' rule in certain respects, such as exhibiting the zero preservation problem. This aggregation method also exhibits what Clemen and Winkler (1999) call "external Bayesianity". Suppose an aggregation has already been computed from several estimates. If new information becomes available about the quantity, one could update each estimate separately and then re-aggregate them, or one could simply update the previously computed aggregation. Logarithmic pooling satisfies external Bayesianity because it doesn't make any difference which of these approaches is used.

In principle, this formula can be applied to to p-boxes by applying it for each member of the class of probability density functions that are compatible with each p-box, cumulating the class of resulting density functions and finding their envelope. One might think that the formula could be applied to Dempster-Shafer structures by computing for each value x the associated interval of possible probability densities. The upper bound of this interval would be the sum of the masses of all focal elements of the Dempster-Shafer structure that intersect with the value x, and the lower bound would be the mass associated with any singleton focal element at x, if there is one, and zero otherwise. (The lower bound is so low because all of the density could flow to other points in a focal element that is not a singleton.) One would expect that interval arithmetic could then be used to do the multiplication, but this approach stumbles on the normalization factor which introduces further complexity to the problem. In any case, the formula can be applied to Dempster-Shafer structures by first converting them to their associated p-boxes by canonical discretization.

There is no convenient formula for finding the normalizing factor k, but the normalization is straightforward for probability distributions in a computer discretization. The problem becomes considerably more challenging computationally for other uncertain numbers however. Logarithmic pooling for aggregation is usually limited to precise probability distributions, but it could be extended to probability boxes and Dempster-Shafer structures via analystical approaches comparable to robust Bayes methods (see Section 3.3). When applied to intervals, logarithmic pooling reduces to intersection (Section 4.3). It is not really general in the sense of Section 4.1.1 because it gives no

answers except in trivial cases when it's applied to real numbers. Because logarithmic pooling strongly emphasizes agreements between its operands, it is not idempotent. It is associative however, so it is also quasi-associative and symmetric in its arguments. But it is not continuous; it fails to produce an answer if the estimates do not overlap. Logarithmic pooling is, however, commutative. It is sensitive to vacuousness and is not intersection-preserving, but is enclosure preserving and narrow.

See Section 4.12 for some numerical examples of logarithmic pooling.

4.9 Averaging

Averaging is very widely used as a way to simplify the clutter of multiple opinions and evidence into a manageably concise representation. In the process, all of the variation present in the data is usually erased. Consequently, we believe that the appropriate uses of averaging in risk analyses are fairly rare. Indeed, without specific justification, averaging is perhaps the *least appropriate* way to aggregate disparate evidence or opinions.

The aggregation methods described in this section generalize what is done with scalars when they are averaged together. Mixing (Section 4.7) of probability distributions is sometimes called "averaging" too because the densities or distribution functions are averaged, but this operation does not generalize what we do when we average scalars. To see why this is so, consider the mixture of two distributions, each of which is a delta distribution (i.e., each is a scalar value without variability or incertitude) such as is shown on the graph below on the left. The distribution functions of the two scalars are depicted as gray lines. Each is a degenerate distribution function; it is a spike rising from zero to one at the value of the scalar. (In principle, the distribution function is a step function having value zero for all values less than the scalar and one for all values greater than or equal to the scalar. To keep the figures simple, only the vertical spikes of these step functions are shown.) The 50:50 mixture of these delta distributions is obtained by vertically averaging the values of the two distribution functions. To the left of gray spike corresponding to the smaller scalar, both distribution functions are zero, so their average is zero too. To the right of the spike for the larger scalar, both functions are one, so their average is too. Between the two spikes, the distribution function for the smaller scalar is one but the distribution function for the larger scalar is zero. The average of zero and one in this region is one half. Thus the distribution function for the mixture, which is shown as a dotted line on the left graph below, is a step function with two steps. The first step is at the smaller scalar from zero to one half, and the second step is at the larger scalar from one half to one. It corresponds to a distribution having equal mass at two points, namely, the two scalars. But, notice that this distribution function is utterly unlike what we would expect from simply computing the numerical average of two scalar values. As depicted in the graph on the right, the scalar average of these two real values yields another real value, that is, another delta distribution, half way between the two scalars. It is shown as a dotted vertical spike. Unlike mixtures, the aggregation operations we call averages (which are defined in the subsections below) do generalize scalar averaging in the sense that, when they are applied to delta distributions, they yield delta distributions that correspond to the scalar averages.

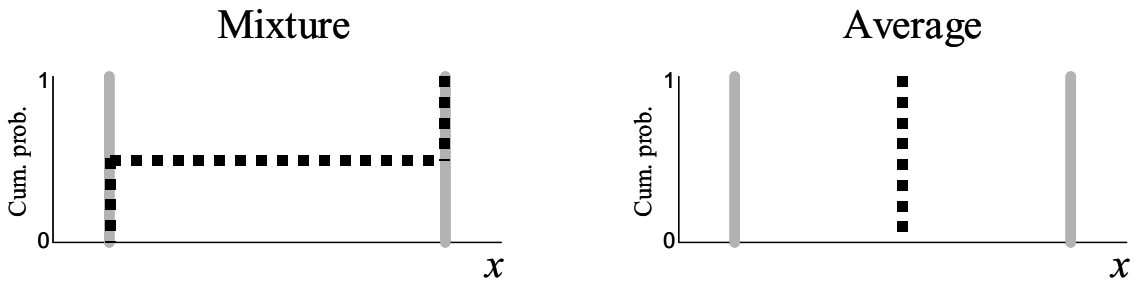


Figure 33: Mixture (dotted, left) and average (dotted, right) of two scalar numbers (gray spikes, left and right).

Nota bene: In this report, the unmodified word "average" does not refer to averaging distribution functions. We are *averaging uncertain numbers*. In this context, we think that our definitions for average and mixture are etymologically and mathematically more reasonable, for the reasons explained above, than any convention that would ambiguously use one word to refer to both averages and mixtures or would confuse language by defining average so that it has different meanings depending on which kind of uncertain numbers are to be combined. To merit the name, an averaging operation should give results equivalent to the results of analogous scalar averaging whenever it is applied to real numbers. Analysts who are familiar with usage in which "average" usually denotes functional averaging are cautioned that our context here is different from conventional probability theory and the difference requires this slight change in terminology. To denote functional averaging of distribution functions, we use the term "mixture" or, sometimes, "vertical averaging".

4.9.1 Convolutive averaging

When real numbers are averaged, they are added together and the sum is divided by the number of addends. Because the generalization of addition for uncertain numbers such as p-boxes and Dempster-Shafer structures is convolution (see Sections 3.2.1.1 and 3.2.3), it makes sense to generalize averaging for uncertain numbers by using this operation. Suppose that we are given n Dempster-Shafer structures, specified in terms of their basic probability assignment functions m_1, m_2, …, m_n. Then their (unweighted) convolutive average under independence has the basic probability assignment defined by

$$m(C) = \sum_{C=(D_1+D_2+\ldots+D_n)/n} \prod_i^n m_i(D_i) \,.$$

This formulation follows directly from the definition of convolution for Dempster-Shafer structures on the real line given by Yager (1986). The convolutive average for p-boxes is obtained immediately by first canonically discretizing them and applying the same formula. The cumulative plausibility and belief functions of the result form the average p-box. If the separate estimates should be weighted, the formula becomes

$$m(C) = \sum_{C = w_1 D_1 + w_2 D_2 + \ldots + w_n D_n} \prod_{i}^{n} m_i(D_i)$$

where w_i is the weight associated with the i^{th} Dempster-Shafer structure such that $\Sigma w_i = 1$ and $0 < w_i$.

The convolutive average can also be defined for cases in which the estimates to be aggregated are not independent of one another. If the dependence is known, then the average will be based on a convolution that expresses the distribution of the sum given that dependence. If the dependence is not known, then bounds on the distribution of the sum (Frank et al. 1987; Williamson and Downs 1990) can be used instead.

Convolutive averages are depicted in the four graphs below. In each of these graphs, the abscissa is the axis for the quantity of concern. In each of the four cases, the input estimates are shown with gray lines and the resulting average is shown with black lines. Each object is shown as a cumulative distribution function or a p-box (i.e., a cumulative plausibility and a cumulative belief function). The upper, left-hand graph depicts the average of two real numbers. The Dempster-Shafer structures for these real numbers are, say, $\{(a, 1)\}$ and $\{(b, 1)\}$, where a and b are the real numbers. The formula for the convolutive average gives the mass for any interval C that is the set-average of focal elements from the two Dempster-Shafer structures. Because there is only one focal element in each structure, there is but one interval C that gets any mass. It is of course the degenerate interval $[(a+b)/2, (a+b)/2]$, and it gets all of the mass. Consequently, the convolutive average is just the degenerate Dempster-Shafer structure $\{((a+b)/2, 1)\}$, which is equivalent to the simple scalar average $(a+b)/2$. The upper, right graph depicts the average of intervals. Assuming the input intervals are $[a_1, a_2]$ and $[b_1, b_2]$, a similar calculation shows that the answer is equivalent to the set average $[(a_1+b_1)/2, (a_2+b_2)/2]$. This answer is just the interval whose endpoints are the averages of the respective endpoints from the two inputs. The two lower graphs depict the average of precise probability distributions and the average of Dempster-Shafer structures.

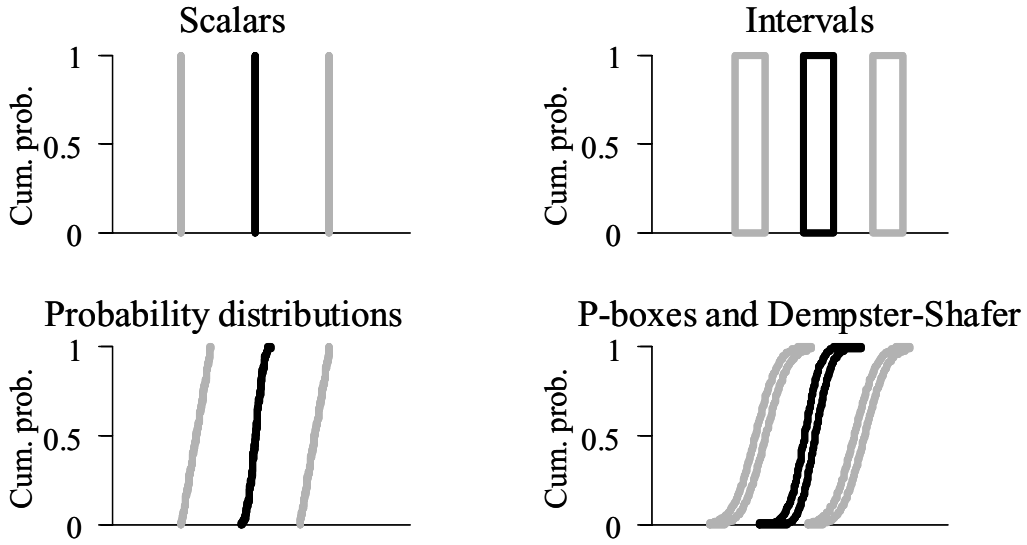


Figure 34: Convolutive averages (black) of different kinds of inputs (gray).

You may be able to detect that the averages for the lower two graphs are slightly steeper than their respective input estimates. This is a result of the independence assumption. The difference in the slantedness of the averages compared to the input estimates becomes more and more exaggerated as the number of estimates increases. If there are many probability distributions of roughly comparable dispersions to be averaged under independence in this way, the convolutive average will approach a scalar (vertical bar) relative to the breadths of the inputs. This is another consequence, of course, of the central limit theorem. As asymptotically many p-boxes of roughly similar dispersion are combined under independence, the convolutive average tends to an interval whose width reflects the overall incertitude (horizontal breadth) of the input estimates.

The convolutive average is general in the sense of Section 4.1.1. Although not simply associative, it is commutative and quasi-associative, and it is symmetric in its arguments when applied to multiple estimates simultaneously. It is easy to show with counterexamples that convolutive averaging is not intersection-preserving. The operation is enclosure-preserving, but it is sensitive to vacuousness.

A single example will suffice to show that the convolutive average is not idempotent, narrow or intersection-preserving. Consider the unit uniform distribution uniform(0,1). Its convolutive average with itself is the triangular distribution ranging on the unit interval with a mode at ½. Because the average of the uniform distribution with itself is not the same distribution, the operation fails to be idempotent. In this example, the envelope of the inputs is just the same uniform distribution. Because this envelope does not enclose the average, the operation fails to be narrow. Likewise, the intersection of the inputs is again just the uniform distribution.

4.9.1.1 Numerical examples

Suppose that the estimates depicted in the three graphs below are to be averaged together. The one of the far left is the cumulative distribution function for a uniform distribution ranging from 4 to 9. The middle graph depicts the cumulative plausibility and belief functions for the Dempster-Shafer structure { ([2,3], 1/3),([5,8], 1/3), ([−1,12], 1/3) }. The graph on the right is a p-box for a unimodal random variable ranging between 0 and 10 with mode 2.

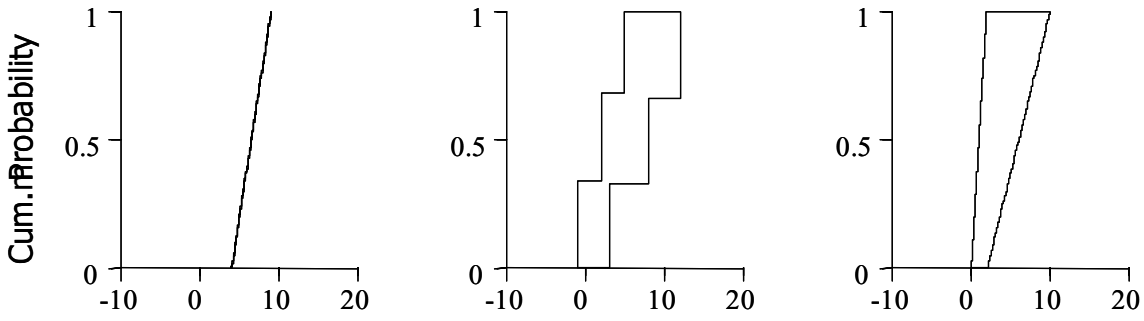


Figure 35: Three uncertain numbers.

The convolutive average assuming independence among all three estimates is shown below on the left in terms of its cumulative plausibility and belief functions. A similar

display on the right depicts the result from the general convolution that does not assume anything about the dependence among the three estimates. The graph on the left was obtained by canonically discretizing the three uncertain numbers displayed above and applying the formula for (unweighted) convolutive average. The graph on the right was obtained using algorithms described by Williamson and Downs (1990) and based on methods developed by Frank et al. (1987). A discussion of these algorithms is outside the scope of this report, but this example was included to demonstrate that such averages can be computed without any independence assumptions.

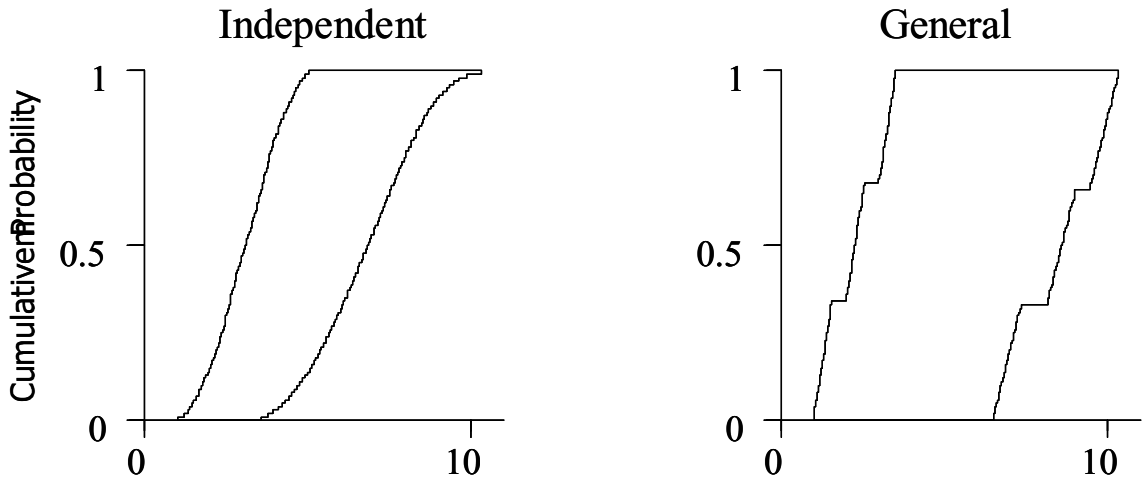


Figure 36: Convolutive averages of three uncertain numbers, assuming independence among the inputs (left) and assuming nothing about their dependence (right).

4.9.2 Horizontal average

How could an average for uncertain numbers be defined so that it would be idempotent? One way is to define it in terms of averaging quantiles, rather than in terms of a convolution. This would amount to averaging cumulative distributions (or p-boxes or Dempster-Shafer structures) *horizontally*. It turns out that horizontal averaging is equivalent to a convolutive average under the assumption that the inputs are perfectly dependent on each other so that they have maximal correlation (Whitt 1976). If there are n p-boxes $[\overline{F}_1, \underline{F}_1]$, $[\overline{F}_2, \underline{F}_2]$, …, $[\overline{F}_n, \underline{F}_n]$, then their (unweighted) horizontal (arithmetic) average is defined to be $[\overline{F}^*, \underline{F}^*]$, where

$$(\overline{F}^*)^{-1}(p) = (1/n)\,((\overline{F}_1)^{-1}(p) + (\overline{F}_2)^{-1}(p) + \ldots + (\overline{F}_n)^{-1}(p))$$

and

$$(\underline{F}^*)^{-1}(p) = (1/n)\,((\underline{F}_1)^{-1}(p) + (\underline{F}_2)^{-1}(p) + \ldots + (\underline{F}_n)^{-1}(p)),$$

and the superscripted "–1" denotes the inverse function. This operation then simply horizontally averages the respective edges of the p-box. A p-box can be obtained from any Dempster-Shafer structure as its cumulative plausibility and belief functions. In the

case of averaging a collection of precise probability distributions, the right and left bounds are the same and the expressions above reduce to a single one that computes the quantile of the average $F^{-1}(p)$ as the simple average of the quantiles $F_i^{-1}(p)$. In the case of interval inputs $[x_i^-, x_i^+]$, this just becomes the weighted average of the endpoints,

$$[x_1^-, x_1^+] * \ldots * [x_n^-, x_n^+] = (1/n)\,[x_1^- + \ldots + x_n^-,\ x_1^+ + \ldots + x_n^+].$$

This operation is clearly idempotent and, like convolutive averaging, generalizes the averaging of scalars. There is a weak precedent for using this operation in risk analyses. Apparently, for a brief period during the 1980s, horizontal averaging was used to aggregate the probability distributions representing expert opinion in the NUREG-1150 studies (J. Helton, pers. comm.; Hora and Iman 1989). In addition to being idempotent, the horizontal average is also general in the sense of Section 4.1.1. Although not simply associative, it is quasi-associative and symmetric in its arguments when applied to multiple estimates simultaneously. Horizontal averaging is commutative. It is clearly intersection-preserving, and it is also narrow and enclosure-preserving, although it is sensitive to vacuousness.

4.9.2.1 Weighted horizontal average

In computing this average, we may elect to weight the various estimates differently. In this case, the quantile intervals are given by

$$(\overline{F}*)^{-1}(p) = w_1\,(\overline{F}_1)^{-1}(p) \ + \ w_2\,(\overline{F}_2)^{-1}\,(p) \ + \ \ldots \ + \ w_n\,(\overline{F}_n)^{-1}(p)$$

and

$$(\underline{F}*)^{-1}(p) = w_1\,(\underline{F}_1)^{-1}(p) \ + \ w_2\,(\underline{F}_2)^{-1}\,(p) \ + \ \ldots \ + \ w_n\,(\underline{F}_n)^{-1}(p)$$

where the weights w_i must be positive and sum to unity. The first expression will always be less than or equal to the second so long as $(\overline{F}_i)^{-1}(p)$ is less than or equal to $(\underline{F}_i)^{-1}(p)$ for all p, which will be true so long as $\overline{F}_i(x) \geq \underline{F}_i(x)$ for all x for each i. In the case of precise probability distributions, the right and left bounds are the same and the expressions above reduce to a single one that computes $F^{-1}(p)$ in terms of $F_i^{-1}(p)$. In the case of interval inputs $[x_i^-, x_i^+]$, the expressions collapse to the elementwise weighted average of the endpoints,

$$[x_1^-, x_1^+] * \ldots * [x_n^-, x_n^+] = [w_1\,x_1^- + \ldots + w_n\,x_n^-\ ,\ w_1\,x_1^+ + \ldots + w_n\,x_n^+].$$

Weighted horizontal averaging is general, idempotent and continuous. It is neither associative nor commutative. Like other weighted schemes, it is neither symmetric in its arguments nor quasi-associative. Like other averages, it is sensitive to vacuousness.

It may make sense to select weights to reflect the reliability of each input estimate. One might argue that intervals, p-boxes and Dempster-Shafer structures wear their reliabilities on their sleeves in that those with greater reliability have tighter bounds. If wider bounds correspond to worse estimates, then it may be reasonable to assign smaller weights to inputs that have larger incertitude so that their effect on the average is less. On the other hand, as was mentioned earlier, one of the most common facts about science

is that empirical uncertainty is typically underestimated. This may mean that the wider an estimate is, the more reliable it is. After all, the broadest error bounds are given by the most sophisticated experts, who have seen how variable the world is and know how limited their own knowledge is. Because of this we might trust estimates with greater uncertainty more than those with less and therefore weight them accordingly.

4.9.3 Numerical examples

The two graphs below depict various kinds of horizontal averages of the same three input estimates used in the numerical example described in Section 4.9.1.1. The graph labeled "Horizontal average" shows the cumulative plausibility and belief functions (the p-box) for the unweighted arithmetic horizontal average of the three uncertain numbers. It should be compared with the convolutive averages depicted in Section 4.9.1.1. The weighted average shown in the graph to the right was based on weights for the three inputs of 0.2 for the uniform distribution, 0.7 for the Dempster-Shafer structure, and 0.1 for the p-box. Unsurprisingly, it fairly closely resembles the Dempster-Shafer structure with the largest weight.

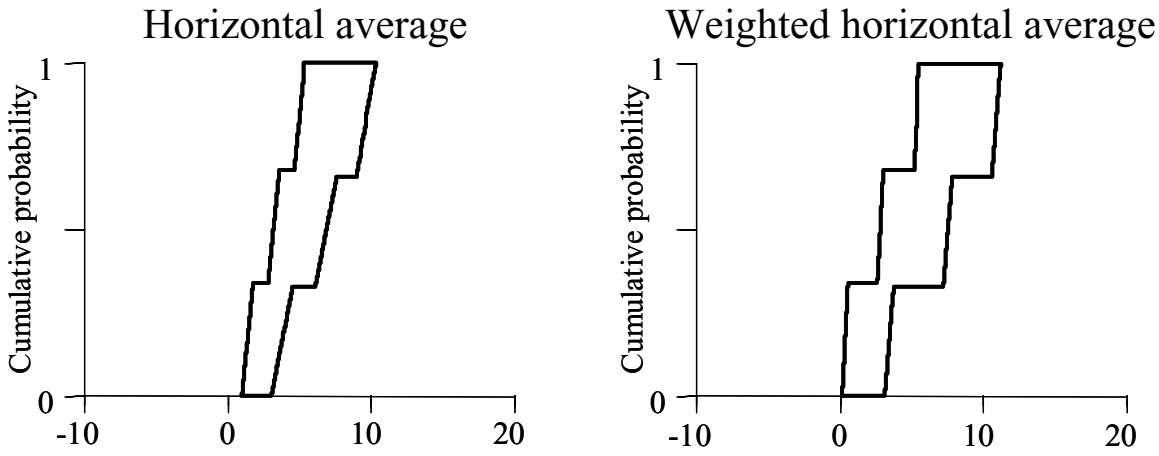


Figure 37: Unweighted (left) and weighted (right) horizontal averages of three uncertain numbers.

4.9.4 Caveats

All averaging operations tend to erase disagreement among the estimates, rather than capture this disagreement so it can be propagated through subsequent calculations in a comprehensive risk analysis. As would be expected, they generally contract the range of possible answers compared to other aggregation techniques. When applied to discrete distributions, they give results about intermediate values that may be impossible. For these reasons, averaging may not be a suitable default technique for aggregation.

4.10 Where do the weights come from?

Several approaches to aggregation, including mixing, averaging and some versions of Dempster's rule, allow estimates to be weighted before they are combined. The ability to use weights when combining estimates in an aggregation greatly increases the flexibility of the methods to account for a variety of ancillary considerations in an analysis. Weighting is usually used to account for differences in the *relevance* of the different inputs as estimates for a particular quantity. There are various considerations an analyst might wish to capture in a weighting scheme. Among the most common include (1) reliability, (2) frequency, (3) temporal sequence, and (4) magnitude order.

Reliability refers to things like sample size used to derive empirical estimates, the trustworthiness of informants, and the stature of experts. Sometimes the weights representing reliability are easy to quantify, such as when they are determined by sample size or closeness to a quantitation limit. In other case, such as when they represent subjective judgments about expert advice or testimony, they may be extremely difficult to quantify. The analyst's temptation in the latter cases is to assume equal weights. But this choice is rarely an appropriate default in the case of uncertainty.

One might think to use the specificity of the estimate itself as a weight representing reliability. For intervals, specificity might be some decreasing function the interval's width; for p-boxes or Dempster-Shafer structures it might be a function of the area between the bounds or the integral of cumulative plausibility minus cumulative belief. After all, these objects ostensibly manifest their own reliability.As discussed in Section 4.9.2.1, however, this idea does not always withstand careful scrutiny. It would seem to require, for instance, that real numbers be given very large weight, even though their apparent precision is probably illusory.

Frequency weighting is often used for forming stochastic mixtures. For instance, if it is known that 75% of the time a quantity is characterized by one set of estimates and 25% of the time it is characterized by a different set of estimates, it makes sense to use these numbers as weights in computing the mixture distribution. Another example of frequency weighting is the use of area weighting to reflect non-uniform spatial sampling. For instance, suppose that a spatially varying quantity has been sampled irregularly across geographic space. The areas of the Thiessen polygons* about each measurement might be used as spatial weights when forming the mixture estimate of the distribution for the spatial variation in that quantity.

Sometimes the temporal sequence in which estimates are collected is very important. For instance, reliability and relevance of historical documents are often strongly weighted by their age relative to some event. In particular, documents dating close to (but not before) the event are weighted heavily and more recent documents are given lesser weights. An opposite temporal weighting is common for military reconnaissance. For instance, a general might reasonably give more credence to fresh reports. In this case, the weighting might be very sharp if current reports are weighted with a value near 1 and old reports are weighted near zero.

---

* A Thiessen polygon about a sample location (among a collection of such sample locations) is the set of points that are closer to that location than to any other location from the collection. Thus 'polygon' is a misnomer because it is the points interior to the polygon that form the set.

Another source of weights is the magnitude of the estimates themselves. The archetypal example of this is Olympic scoring of athletic performance in which the highest and lowest scores are thrown out (i.e., weighted with zero) before an average is computed. Such a weighting scheme might be reasonable in contexts beyond the Olympic games. For instance, it will commonly be useful in other situations involving expert elicitation if the analyst suspects certain kinds of bias or prejudice among the experts. It can even arise in situations involving only objective measurements. For example, sometimes the measurement protocols used in chemical laboratories depend in part on the concentration present. Very low concentrations and very high concentrations are typically hard to measure well. In such cases, there is nothing in particular about a measurement, other than its magnitude, that makes it questionable in any way. Nevertheless, a low or high value might deserve a diminished weight if they are likely to be associated with greater imprecision. Yager (1988; Yager and Kacprzyk 1997; Klir and Yuan 1995) describes a class of aggregation operators based on ordered weighted averages (OWA) of surprising generality.

4.11 Accounting for small sample size

What if the estimates to be aggregated are only a small number of samples taken from a population of values? In that case, even the envelope aggregation, which yields very broad results, may not be broad enough to account for the distribution tails of the underlying population that we cannot usually observe because of our small sample size. The uncertainty of any aggregation result should be inflated somehow to account for small sample size. We have already broached the issue of accounting for sample uncertainty in Sections 3.5.4 and 3.5.6.2, and we acknowledged there that the theoretical foundations for a sampling theory for uncertain numbers have not yet been developed. Still, some analytical strategies that might be useful in practice present themselves.

If the samples are independent of each other and collected at random from the same population that is the focus of interest, then one could employ Kolmogorov-Smirnov confidence intervals (Section 3.5.4) to inflate the uncertainty about an estimate obtained by a mixing aggregation. As already discussed, these samples could be real numbers or intervals. It seems reasonable to think that more general uncertain numbers could also be employed, but no algorithms currently exists to do so. Of course, without independence, Kolmogorov's approach makes no statistical claim. And without random sampling, there is no assurance that the picture emerging from sampling will be representative of the population of interest.

4.12 Summary of properties and operations

The properties of the aggregation methods reviewed in Section 4 are summarized in the table beginning on page 102. Mathematical properties are arranged in columns of the table in the order of their importance (generality is most important). The aggregation operations are arranged in order of their usefulness as generic tools in risk analysis (enveloping is most useful). The letter Y in a cell means that the aggregation operation of that row has the property of that column; the letter N means it does not.

Motivations and counterarguments for the various properties relevant for an aggregation operator are detailed in Section 4.1. They are restated synoptically here. Let A denote an aggregation operator, and let X, Y, and Z denote uncertain numbers (e.g., intervals, probability distributions, p-boxes or Dempster-Shafer structures) which are the arguments of A.

- A is **general** if it can be applied to real numbers, intervals, probability distributions, p-boxes and Dempster-Shafer structures and combinations thereof in a consistent way.
- A is **closed** if $A(X_1, \ldots, X_n)$ is an uncertain number whenever $X_1, \ldots, X_n$ are.
- A is **symmetric** in its arguments if $A(X_1, X_2, \ldots, X_n) = A(X_{\sigma(1)}, X_{\sigma(2)}, \ldots, X_{\sigma(n)})$ for all permutations σ.
- A is **idempotent** if $A(X, X) = X$ for all X.
- A is **intersection-preserving** if intersection$(X_1, \ldots, X_n) \subseteq A(X_1, \ldots, X_n)$ whenever the intersection exists.
- A is **enclosure-preserving** if $X_1' \subseteq X_1$ implies $A(X_1', \ldots, X_n) \subseteq A(X_1, \ldots, X_n)$.
- A is **narrow** if $A(X_1, \ldots, X_n) \subseteq$ envelope$(X_1, \ldots, X_n)$.
- A is **insensitive** to vacuousness if $A(X_1, \ldots, X_n, [-\infty, +\infty]) = A(X_1, \ldots, X_n)$,
- A is **continuous** if $A(X, Y) \to A(Z, Y)$ as $X \to Z$ and $A(X, Y) \to A(X, Z)$ as $X \to Y$ for all X, Y, Z.
- A is **commutative** if $A(X, Y) = A(Y, X)$ for all X and Y.
- A is **associative** if $A(X, A(Y, Z)) = A(A(X, Y), Z)$ for all X, Y, Z.
- A is **quasi-associative** if there exists an operation $A^n(X_1, X_2, \ldots, X_n)$ that is symmetric in its arguments, and $A^{n=2}(X, Y) = A(X, Y)$.

None of these properties is absolutely essential because, as we have shown, there are reasonable examples for which requiring any of these properties would be counterintuitive. See Section 4.1 for a discussion of each of these properties and arguments for and against using them as criteria for aggregation operators.

Table 1 Summary table of mathematical properties of aggregation methods.

Aggregation operator	*General*	*Closed*	*Symmetric*	*Idempotent*	*Intersection-preserving*	*Enclosure-preserving*
Envelope	Y	Y	Y	Y	Y	Y
Intersection	Y	N	Y	Y	Y	Y
Mixture, unweighted	Y	Y	Y	Y	Y	Y
Mixture, weighted	Y	Y	N	Y	N	N
Horizontal average, unweighted	Y	Y	Y	Y	Y	Y
Horizontal average, weighted	Y	Y	N	Y	Y	N
Logarithmic pooling	Y	N	Y	N	N	Y
Convolutive average	Y	Y	Y	N	N	Y
Dempster's rule of combination	Y	N	Y	N	N	Y
Yager's rule	Y	Y	Y	N	N	N
Disjunctive consensus	Y	N	Y	N	Y	Y
Null aggregation	Y	N	Y	N	Y	Y
Aggregation operator	*General*	*Closed*	*Symmetric*	*Idempotent*	*Intersection-preserving*	*Enclosure-preserving*

Aggregation operator	*Narrow*	*Insensitive*	*Continuous*	*Commutative*	*Associative*	*Quasi-assoc.*
Envelope	Y	N	Y	Y	Y	Y
Intersection	Y	Y	N	Y	Y	Y
Mixture, unweighted	Y	N	Y	Y	N	Y
Mixture, weighted	Y	N	Y	N	N	N
Horizontal average, unweighted	Y	N	Y	Y	N	Y
Horizontal average, weighted	Y	N	Y	N	N	N
Logarithmic pooling	Y	N	N	Y	Y	Y
Convolutive average	N	N	Y	Y	N	Y
Dempster's rule of combination	Y	Y	N	Y	Y	Y
Yager's rule	Y	Y	N	Y	N	Y
Disjunctive consensus	Y	N	Y	Y	Y	Y
Null aggregation	Y	N	Y	Y	Y	Y
Aggregation operator	*Narrow*	*Insensitive*	*Continuous*	*Commutative*	*Associative*	*Quasi-assoc.*

Although the various aggregation operations can sometimes agree, they generally yield very different results. Figure 38 through Figure 40 over the following pages depict three sets of nine aggregations of two inputs *A* and *B*. The operations are intersection (Section 4.3), mixture (4.7), envelope (4.4), the logarithmic pool (4.8), convolutive average (4.9.1), horizontal average (4.9.2), Dempster's rule (4.5.1), Yager's rule (4.5.2) and the envelope variant of disjunctive consensus (4.5.3). In the first set of nine aggregations, *A* and *B* are overlapping intervals. In the second set of nine, they are precise probability distributions whose ranges overlap. In the third set of nine aggregations, the two inputs are broadly overlapping p-boxes (or Dempster-Shafer structures). The inputs used for each set of nine aggregations are depicted in the pair of graphs on the left of each page. (For the illustrative purposes of these graphs, the aggregation operations have been applied to two inputs of the same kind, but the methods can generally be applied to many inputs and these inputs can be arbitrary combinations of any kind of uncertain number, including reals, intervals, probability distributions, p-boxes, and Dempster-Shafer structures.) All inputs and all twenty-seven results are displayed as p-boxes or cumulative plausibility and belief functions. Thus, the abscissas are the *x*-variable (that is, whatever *A* and *B* are estimates of), and the ordinate in cumulative probability. Note that there is no solution for intersection when the inputs are probability distributions. The result for Yager's rule for this case is the vacuous result (because the "agreement" between two probability distributions has vanishingly small measure). In the case of p-boxes, Yager's rule has infinite tails. All twenty-seven results were computed from their respective inputs by straightforward* application of the algorithms for each method as described in Section 4.

*The only result that might be surprising is that of logarithmic pool for p-boxes. The following example explains the left bound on the aggregation that results from using robust methods on the p-boxes. Suppose that the distribution for *A* is actually an even stochastic mixture of a uniform distribution over the interval [0.0, 1.01] and a uniform distribution over [3.01, 4]. This distribution falls within the p-box given for *A*. Suppose the distribution for *B* is actually a uniform distribution over [1, 3]. This is the left side of the p-box for *B* and is therefore consistent with its specification. Because these two distributions jointly have nonzero mass only over the interval [1, 1.01], the result of applying the logarithmic pool to them yields the (uniform) distribution over [1, 1.01]. The overlap can be made as small and as close to one as desired, leading to a limiting delta distribution at the *x*-value of one. A similar example can be constructed for a delta distribution at 4. The zero preservation property of the logarithmic pool guarantees that all the distributions that could possibly result from its application are constrained to the intersection of the supports of *A* and *B*. In this case, that intersection is [0, 4], so the bounds of the logarithmic pool aggregation cannot be any larger than those shown.

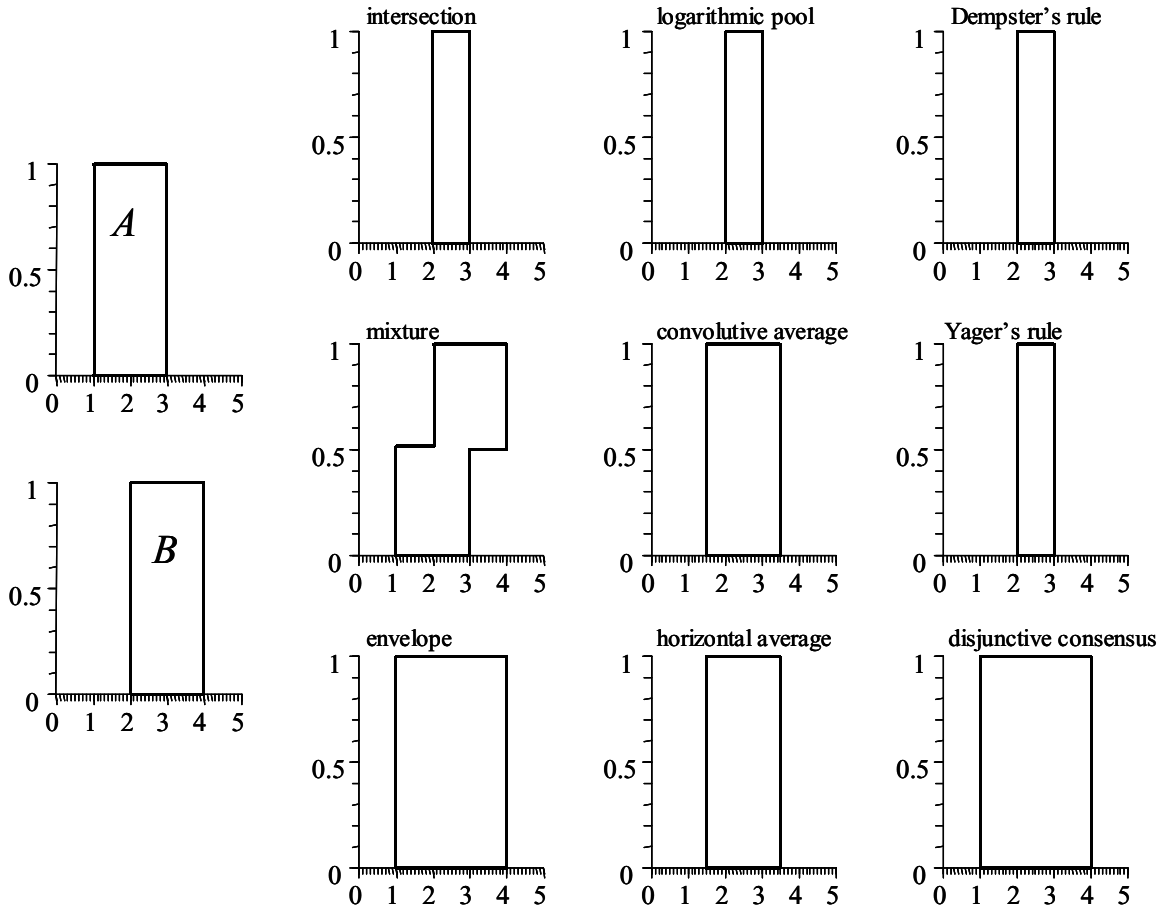


Figure 38: Nine aggregation operations applied to interval inputs *A* and *B* (left).

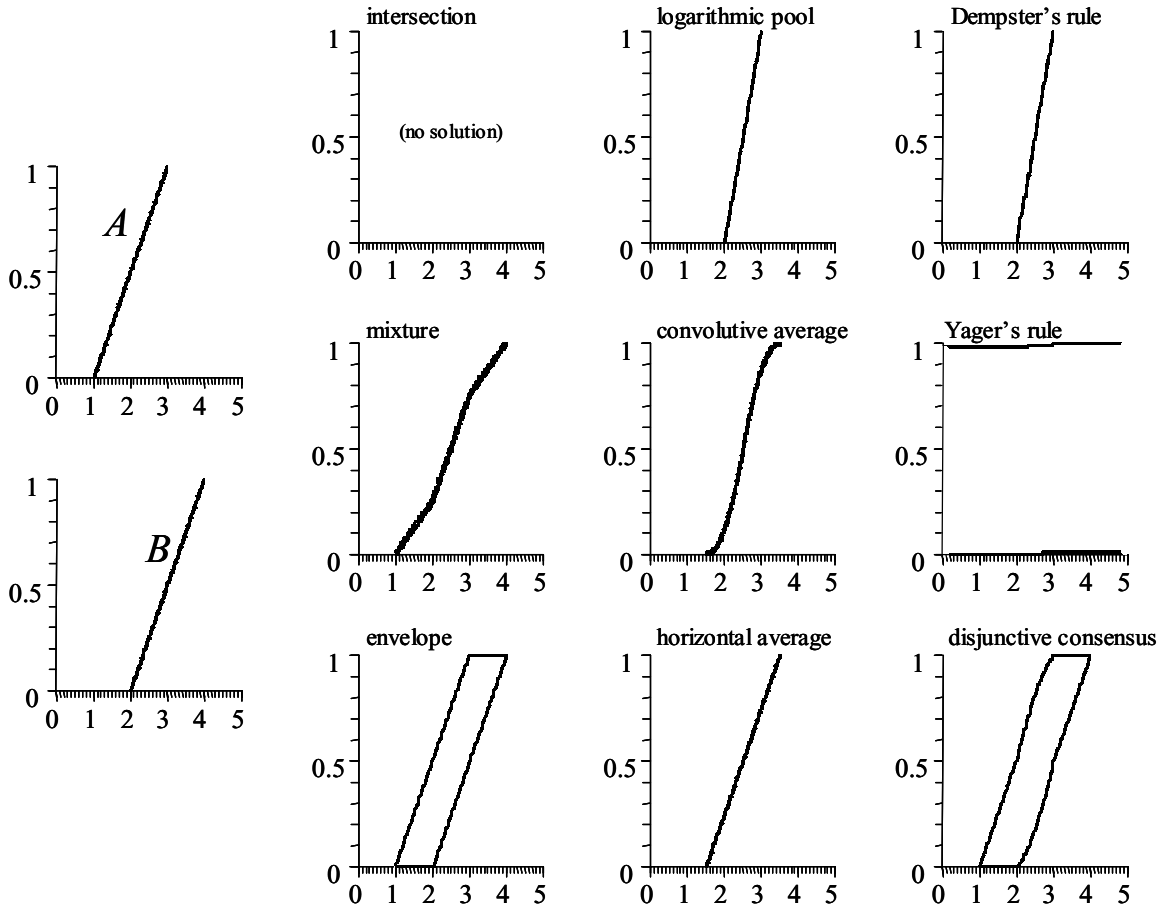


Figure 39: Nine aggregation operations applied to probability distributions *A* and *B* (left).

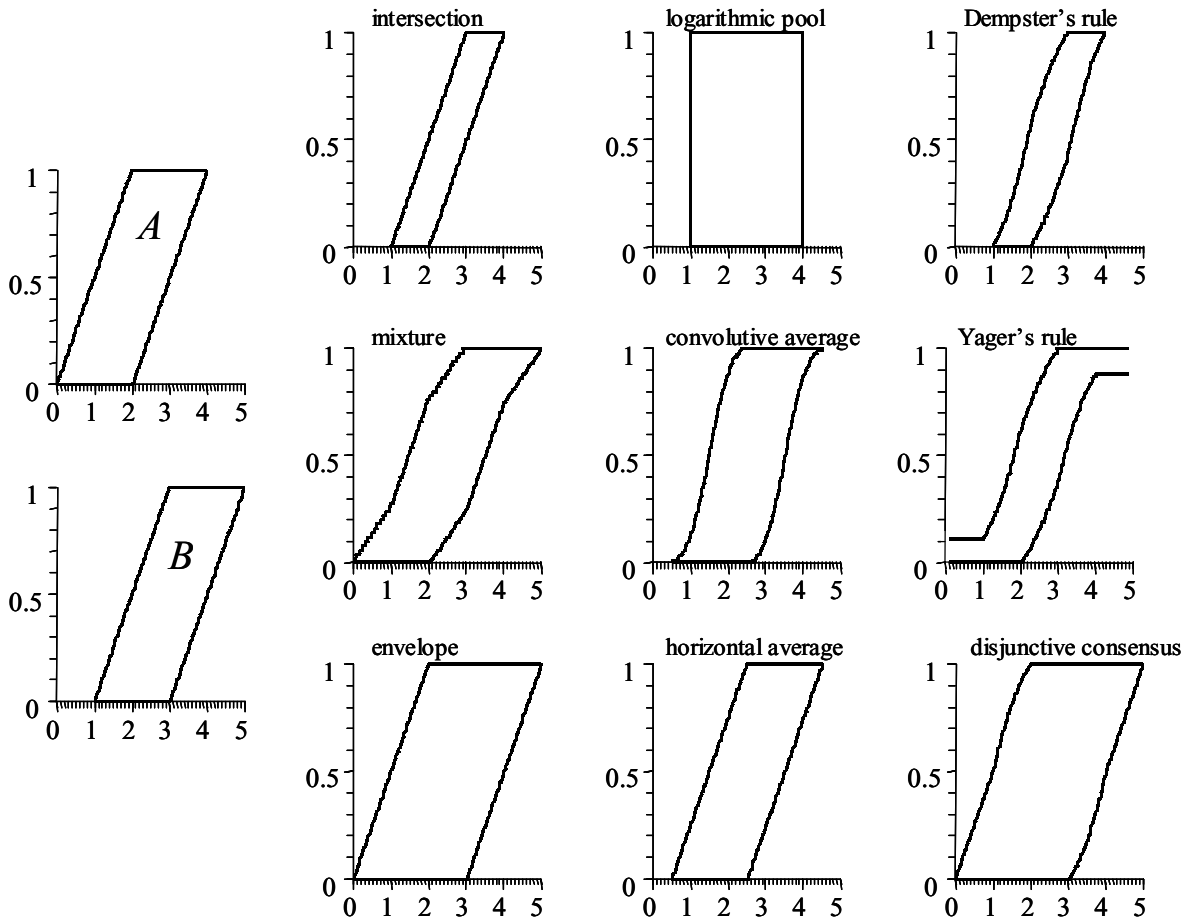


Figure 40: Nine aggregation operations applied to p-boxes *A* and *B* (left).

4.13 Choosing an aggregation operator

The conclusion at the end of Section 3 on characterization was to use any and all of the methods outlined there. The choice of aggregation operators is a very a different story. Here, the selection of the appropriate operator to use is a more delicate matter, whose consequences will typically be important for an assessment. The selection ought to be a considered decision that is based on what is understood about the various information sources and their interrelationships.

The list of questions below can guide you to a useful strategy for selecting the aggregation operator most appropriate for a given circumstance. If the answer to a question is "yes", follow any questions that may be indented underneath. The number in bold is the number of the section in this report that addresses the method that could be useful to you. This is not a flowchart, but there may be a question whose answer determines the best strategy to use. This means that, unlike the suggestions in Section 3.6.2, an analyst is generally *not* free to follow multiple paths and use different methods according to one's taste or the method's performance. The different operators are appropriate for different contexts. In particular, an operation that is appropriate for aggregating variability will probably not be suitable for aggregating incertitude. Likewise, one that is well adapted to handle incertitude may be suboptimal for handling variability.

Do the different estimates represent variability (aleatory uncertainty)? (**4.7**)
    Should a weighting scheme be used? (**4.7.2**)
Are the estimates logical or reliable empirical enclosures? (**4.3**)
    Do the intersections fail to exist? (**4.3.2**)
Is at least one of the estimates an enclosure? (**4.4**)
Are the reliabilities of the estimates uncertain? (**4.4.2**)
Are the estimates samples of a larger population of interest? (**4.11**)
Are the estimates independent of one another? (**4.5**)
Is it important in incorporate the prior beliefs of the analyst? (**4.6**)
Will it be necessary to do sensitivity studies for each estimate anyway? (**4.2**)
Is it appropriate to ignore variability? (**4.9**, **4.9.2**)
    Is a weighting scheme needed? (**4.9.2.1**)
    Are the estimates independent? (**4.9.1**)
    Are the estimates independent of each other? (**4.9.1**)

5 Model uncertainty

In many areas of physics, the appropriate model to use for a particular situation is well established. However, in some areas, there is still controversy among scientists about how to best describe the physical interactions in a system. This is often the case in new areas of science and in scientific fields where empirical investigation is difficult or expensive. It is also especially true in domains of study involving performance of new materials or new system designs, and behavior of systems under abnormal or extreme conditions. These domains are very common in risk assessments.

By "model uncertainty", we refer to that incertitude about the correct form that the model should take. Are the mathematical expressions right? Are the dependencies and interactions among physical components reasonably and accurately represented? Are the submodels appropriate for the situation and do they mesh together coherently? The model in a risk assessment includes all the structural decisions made by the analyst (or modeler) that govern how the parameters interact. Each of these decisions is in principle subject to some measure of doubt. Model uncertainty is about whether or not those parameters are combined together in the right way. In most cases, it is a form of incertitude or epistemic uncertainty because we are unsure whether their constructions are reasonable and complete.

Model uncertainty is distinguished from parametric uncertainty, which is the uncertainty about the value(s) of a particular constant or variable. Risk analysts have many computational tools available to them to assess the consequences of parametric uncertainty. But analyses consist of statements about both parameter values and the relationships that tie the parameters together. These relationships are expressed in models. This means that model uncertainty could be just as important, or even more important, than parametric uncertainty. Despite this, almost all risk analyses and, indeed, statistical analyses in general neglect this source of uncertainty entirely. Ignoring model uncertainty could lead to over-confident inferences and decisions that are more risky than one thinks they are. A risk analyst who constructs a single model for use in an assessment and then uses it to make forecasts is behaving as though the chosen model is actually correct. Draper (1995) argued that model uncertainty should be taken very seriously in computing forecasts and calculating parameter estimates. How should this be done?

A traditional Monte Carlo analysis might handle model uncertainty by creating a new parameter, say m, to represent which model to use (e.g., Apostolakis 1995; Morgan and Henrion 1990; cf. Cullen and Frey 1999). If there are two possible models, this parameter would be represented as a Bernoulli random variable taking on both possible values with probability determined by the relative likelihoods that either model is the right one. If this probability is unknown, the traditional approach is to assume both models are equiprobable. If there are several possible models, then the parameter m would be a more general discrete variable, whose values would again be equiprobable unless the relative probabilities of the different models were known. Finally, if there are infinitely many models possible, but they can be parameterized in a single-dimensional family, then a continuous version of the parameter m can be used. In all cases, values for

this variable are randomly generated in the Monte Carlo simulation. Which model is used is determined by the random value. Typically, the model selection would happen in the outer loop of a two-dimensional simulation (in which the inner loop simulated variability), but this is not essential. The result of the Monte Carlo simulation depends then on a randomly varying model structure. This approach requires that the analyst know, and be able to enumerate or at least continuously parameterize, all the possible models.

The Bayesian approach to handling model uncertainty, which is called Bayesian model averaging (Raftery et al. 1997; Hoeting et al. 1999), has essential similarities to the Monte Carlo approach, and it will typically produce similar if not identical results. Until very recently, analysts chose a single model and then acted as though it had generated the data. Bayesian model averaging recognizes that conditioning on a single selected model ignores model uncertainty, and therefore can lead to underestimation of uncertainty in forecasts. The Bayesian strategy to overcome the problem involves averaging over all possible models when making inferences about quantities of interest. Draper (1995) suggested employing standard techniques of data analysis, but when a good model is found, embedding it in a richer family of models. By assigning prior probabilities for the parameters of this family of model and treating model selection like other Bayesian parameter estimation problems, this approach produces a weighted average of the predictive distributions from each model, where the weights are given by the posterior probabilities for each model. By averaging over many different competing models, this approach incorporates model uncertainty into conclusions about parameters and predictions. In practice, however, this approach is often not computationally feasible because it can be difficult to enumerate all possible models for problems with a large number of variables. However, a variety of methods for implementing the approach for specific kinds of statistical models have been developed. The approach has been applied to many classes of statistical models including several kinds of regression models (Hoeting et al. 1999).

The Monte Carlo strategy to account for model uncertainty and Bayesian model averaging are similar in that they both use what is essentially a mixture of the competing models. Aside from the technical burden of parameterizing the space of possible models and assigning a probability to each, there is a far greater problem with the approach that these strategies use. In representing model uncertainty as a stochastic mixture of the possible models, this approach effectively *averages together incompatible theories* (Finkel 1995). It is equivalent in this respect to the approach to modeling what is fundamentally incertitude as an equiprobable stochastic mixture (the uniform distribution). This approach is due originally to Laplace, but when it is applied in risk analysis to the study of distributions (rather than estimating point values), it can underestimate* the true tail risks in an assessment. The potential results are distributions that no theories for any of the models would consider reasonable.

In light of this problem, it may be a more reasonable strategy to use an envelope of the models rather than an average or mixture of models. Because model uncertainty

*Some probabilists maintain that one can use stochastic mixtures to represent model uncertainty and that this does not average alternative models so long as the results are presented properly. It is hard for us to see how this is a tenable position if we want to be able to interpret the probabilities in the output of a quantitative risk analysis as frequencies.

typically has the form of doubt about which of a series of possible models is actually the right one, such an approach would propagate precisely this doubt through subsequent calculations. An enveloping approach would clearly be more comprehensive than the traditional approach based on model averaging. We note that it would even be able to handle non-stationarity of distributions, which is another important source of uncertainty that is usually ignored in traditional assessments for lack of a reasonable strategy to address it. Unlike the Monte Carlo and Bayesian model averaging strategies, an enveloping approach will work even if the list of possible models cannot be enumerated or parameterized. So long as we can somehow *bound* the regions in any output or intermediate variables that depend on the choice of the model, we can represent and propagate the uncertainty about the model in a comprehensive way.

How can one bound a class of models without enumerating or parameterizing them? There are several examples of how this can be done. The manifestations of model uncertainty are extremely numerous, but there are some particular forms for which useful bounding approaches have been developed. These include uncertainty about distribution family, dependence, choice among specific competing theories, and choice among unknown theories when consequences are bounded. We briefly mention the relevant strategies for handling these situations below in turn.

Model uncertainty about distribution family is the focus motivating the development of both robust Bayes methods (Section 3.3; Berger 1985) and the analytical techniques for probability boxes (Ferson 1996; 2002; Berleant 1996). It is fair to say that an analyst using these techniques could conduct a distribution-free risk analysis that makes no assumptions about the families of statistical distributions from which variables are drawn. Obviously some assumptions or evidence is necessary (such as constraints on the range or moments, or empirical sampling data), but no assumption about the statistical *family* of distributions need be required. These techniques allow an analyst to fully explore the consequences of model uncertainty about distribution shape.

The available techniques are just as comprehensive for model uncertainty about the dependencies among random variables. In general, dependency is captured in a copula which is a real-valued function of two (or more) variables (the inverses of the distribution functions for the two random variables). Solving a problem originally posed by Kolmogorov, Frank et al. (1987) developed the mathematical tools to compute bounds on arithmetic functions of random variables about which only marginal information is available. Williamson and Downs (1990) implemented these tools and extended them to the case when only bounds on the marginals are known. Note that the resulting bounds on the answer are not equivalent to those obtained by merely computing the result under all possible correlation coefficients from −1 to +1; they are slightly wider. Nevertheless, the resulting bounds are known to be best possible (Frank et al. 1987).

An accounting for model uncertainty can be done by enumeration if it can be narrowed to a controversy among a finite list of specific competing theories. In a variety of cases, there are two or a few models that have appeared in the scientific literature as descriptions of an incompletely understood phenomenon. In some cases these models are extreme cases of a continuum that captures the possible ways that the phenomenon could work. In other cases, the models are the idiosyncratic products of individual scientists, and the resulting collection of models cannot be claimed to be comprehensive in any sense. In either case, so long as the model uncertainty in question is about only these

specific models, then surveying each model and enveloping the results they produce will suffice to construct the bounds needed for analysis. If the number of competing models is small, a null aggregation strategy that studies each possible model in turn might even be workable.

Even in situations where we cannot list the possible models because there are infinitely many of them, it may still be possible to bound the consequences of model choice. Trivial cases include models that influence a probability value. We know that the probability is constrained to the interval [0,1], no matter what the model is. Nontrivial cases depend on specific knowledge of the physics or engineering of the system under study. Note that it is often possible to *bound* the results in such situations, even though we could not possibly form a mixture distribution.

The strategies of enveloping (Section 4.4) and null aggregation (Section 4.2) are generally useful for representing model uncertainty. In some cases where different mechanisms appear to be acting in different circumstances, mixing (Section 4.7) may be appropriate. Bayesian model averaging (Hoeting et al. 1999) is similar to mixing in that functional averages of distributions are computed. The other aggregation operations considered in this report (intersection, Dempster's rule and its variants, and averaging) would probably not be reasonable for use with model uncertainty because they really focus on parametric uncertainty per se.

Many analysts make a reflexive criticism that bounding approaches may lead to very wide uncertainties. Certainly it is true that bounding can lead to large uncertainties, but the width of the uncertainty is irrelevant if those bounds can be shown to be best possible or are otherwise small enough to lead to useful risk analyses that support effective decision making. In any case, it is better to have a correct analysis that honestly distinguishes between variability and incertitude than an analysis that depends on unjustified assumptions and wishful thinking. We argue that a correct analysis does not mistake ignorance on the part of the analyst for variability in a natural system. Assumptions that incertitude in one variable will tend to cancel out incertitude in another variable, or that extremes from one model will be tempered by outputs from another model are little more than wishful thinking unless they can be supported by affirmative evidence of the supposed variability. Not knowing the value of a quantity is *not* equivalent to the quantity varying. Likewise, not knowing which model is the right one to use does *not* mean that it is reasonable to suppose that each model governs the system part of the time. Indeed, this seems like the most unlikely situation of all. It has always been easy to get tight bounds on uncertainty if we're not constrained to account for what is known and what is not known. If the price of a correct assessment is broad uncertainty as a recognition or admission of limitations in our scientific knowledge, then we must pay that price.

Although there are several important forms of model uncertainty that are amenable to a complete assessment by enumeration or bounding, there are, of course, other forms of model uncertainties that remain difficult to address in comprehensive way, such as choice of what parameters to use and choice about the level of abstraction and depth of detail to incorporate into the model. For such uncertainties, the family of possible models may be infinite-dimensional and the analyst may lack any natural way to bound the parameters that depend on the model selection.

6 Conclusions

Uncertain numbers are a class of objects that include intervals, probability distributions, probability boxes, and Dempster-Shafer structures on the real line. Real numbers are a degenerate special case of uncertain numbers. By their ability to simultaneously represent incertitude (or epistemic uncertainty) and variability (or aleatory uncertainty) in a single data structure, uncertain numbers are especially suited for use in quantitative risk analyses. Incertitude is expressed in the widths of intervals that are the focal elements composing a Dempster-Shafer structure. Variability is expressed by the scatter of those intervals. Incertitude is also expressed in the breadth between the left and right sides of a probability box, and variability is expressed in the overall range and steepness of the probability box. An interval is a special case because it expresses pure incertitude. Likewise, a probability distribution expresses pure variability. (These ideas are contrary to the common notion championed by Bayesians that probability is fully capable of expressing both incertitude and variability by itself.)

A Dempster-Shafer structure on the real line can always be converted into a probability box, although some information about the internal arrangement of masses may be lost in the conversion. Conversely, a probability box (specified by a pair of distribution functions) can be identified with an equivalence class of Dempster-Shafer structures. The equivalence class consists of all of those Dempster-Shafer structures whose cumulative plausibility function is the same as the left bound of the probability box and whose cumulative belief function is the same as the right bound of the probability box. The interconvertibility of probability boxes and Dempster-Shafer structures is very important because it allows analysts to use tools developed in either theory to solve risk assessment problems.

Section 3 reviewed the variety of ways that data and expert knowledge can be quantified as uncertain numbers. These ways include

direct assumption,
modeling,
robust Bayes methods,
constraint specification, and
synthesis of experimental measurements.

These are the approaches available to a risk analyst for selecting the inputs to be used in the risk assessment.

Direct assumption is the derivation (or sometimes merely the intuitive choice) by the analyst of a probability distribution or other uncertain number based on the analyst's understanding of the engineering principles or underlying physics involved. For instance, the normal distribution is often motivated by an appeal to the central limit theorem. The disadvantage of direct assumption as a way to specify uncertain numbers is that it cannot usually be justified by any empirical evidence or explicit model. Because humans routinely and severely underestimate their own uncertainties, even about processes within their professional domains, inputs selected by direct assumption should probably receive careful review.

The use of modeling to choose an uncertain number for use in a risk analysis divides the problem into subproblems. Instead of directly selecting the representation for the quantity of interest, it is analyzed in terms of other quantities which (it is hoped) are

easier to quantify. The quantity of interest is then computed from these estimates via the model. The advantage of modeling over direct assumption is that the analyst's mental reasoning leading to the choice is made explicit by the model. This explicitness improves the transparency and the reviewability of the selection. There are a variety of models that can be used, including finite combinations of convolutions (additions, subtractions, multiplications, divisions, etc.), transformations (logarithm, square root or absolute value), aggregations (enveloping, intersection, stochastic mixtures, etc.), compositions and deconvolutions. Algorithms to effect all of these operations are available for uncertain numbers.

Section 3.3 considered the use of robust Bayes methods for the purpose of obtaining uncertain numbers for use in risk calculations. Although Bayes rule is often considered a method for updating estimates with new evidence, in this context it can also serve as a means of characterizing an uncertain number from a single source. Ordinary Bayesian updating is accomplished with a single prior distribution and a precise likelihood function. In robust Bayes, the requirement to specify a particular prior and likelihood is relaxed and an entire class of priors can be combined with a class of likelihood functions. Because the resulting class of posterior distributions can be represented by a probability box, robust Bayes constitutes a means of obtaining an uncertain number.

The maximum entropy criterion is used by many analysts to select probability distributions for use in risk assessments from limited quantitative information that may be available about a random variable. For instance, if only the minimum and maximum values of the random variable are known with confidence, an argument dating back to Laplace asserts that a uniform distribution over that range is the best model for the variable. This distribution, out of all of the distributions that have the same range, has the largest statistical entropy. Likewise, if the mean and variance are somehow the only pieces of available information about a random variable, a normal distribution has the largest entropy and would be selected for use under the maximum entropy criterion.

Section 3.4 described a parallel strategy for selecting uncertain numbers from limited quantitative information. Instead of selecting a single distribution from the class of all distributions matching a given set of constraints, this approach simply uses the entire class of those distributions. For instance, if the risk analyst only knows the minimum and maximum possible values for a random variable, then the set of distributions having the specified range will be represented as a probability box. In this case, the probability box is equivalent to an interval with the same range and also to a degenerate Dempster-Shafer structure with one focal element. Best possible probability boxes have been derived for many different cases that might arise in practice. These include, for instance, cases in which the following sets of information are known (or can be assumed): {min, max}, {min, max, mean}, {min, max, median}, {min, max, mode}, and {sample mean, sample variance}. Qualitative constrains on the shape of the distribution may also be accounted for to tighten the uncertain number. For instance, techniques are known that can account for positivity (nonnegativity) of the random variable, convexity or concavity of the distribution function, and monotonicity of the hazard function. It is also possible to derive bounds on the distribution function, and thus specify an uncertain number, by limiting the probability density over various real values. Such information is sometimes available to analysts from their previous scientific study of a system. Sometimes the constraints represent design constraints that are to be built into a system. The extensive

available library of special cases forms a ready alternative to appeals to the maximum entropy criterion.

The use of experimental measurements in developing uncertain numbers was described in Section 3.5. Uncertain numbers can represent the variability in a data set, in essentially the same way as an empirical distribution function does. Unlike empirical distribution functions, however, they can also represent the natural measurement uncertainty that may have accompanied the collection of the original data. This measurement uncertainty could include the ordinary plus-minus intervals commonly reported with most kinds of measurements. It could also include many forms of statistical censoring. The uncertainty could even include the sampling uncertainty that arises from measuring only a sample of a larger population of quantities. All of these kinds of uncertainty can be represented in a comprehensive way with a single uncertain number.

Section 3.6 reviewed strategies for selecting methods for characterizing uncertain numbers and proffered a checklist of questions to guide analysts in the use of these methods.

Section 4 considered the issue of how different estimates of a single quantity or distribution should be combined. It reviewed twelve properties of aggregation methods and considered the arguments for and against requiring each property for an aggregation method. Although none of the properties was clearly essential for an aggregation operator, we think that some of the properties are especially important to consider when choosing a method. In particular, it seems reasonable that the aggregation method should

1) be applicable to all kinds of uncertain numbers (generality),
2) yield an uncertain number as a result whenever its arguments are uncertain numbers (closure),
3) yield the same answer no matter what order the estimates appear in (symmetry in arguments),
4) preserve any agreement that may exist among the estimates (idempotence and the intersection-preserving property), and
5) never widen answers if inputs get tighter (enclosure-preserving property).

The generality and closure properties are obviously matters of convenience. Nevertheless, they are, just as obviously, important considerations. The other properties seem both entirely reasonable and usefully broad in scope.

The symmetry and idempotence properties are controversial among some probabilists. However, the controversy may simply be the result of prejudice in favor of established aggregation methods. For instance, if Bayes' rule and logarithmic pooling were idempotent, it seems doubtful that many would have any serious objection to this property. To those who are not already committed to the notion that Bayes' rule is *the* way that uncertainty should be aggregated, it is hard to imagine that a reasonable method should give different results depending merely on the order in which the evidence is considered. Some have also argued against symmetry. For instance, they suggest that a general's conclusion about the battlefield depends on the order in which he receives reconnaissance. Although it may be that the general's conclusion does depend on the order in which messages are received, it is not clear that this means that it *should* be. In the context of dispassionate risk analyses based on empirically measured quantities about engineered devices, it seems highly implausible to think that which answer is correct

should depend on which of a series of measurements happen to be taken first. Bayesians recognize the intrinsic attraction of this symmetry when they emphasize that Bayes' rule asymptotically converges to the right answer after many data sets have been combined. Their difficulty, of course, is that risk analysts rarely have (asymptotically) many data sets to combine. They typically must make forecasts and draw inferences in the short term.

In practice, however, analysts rarely select an aggregation strategy based solely on its mathematical properties. The different methods that are available are *doing different things* and each may be useful in a particular circumstance. Intersection (Section 4.3) may be clearly most useful if all the estimates are completely reliable. Enveloping (Section 4.4) may be more reasonable if the reliability of the estimates is unknown. Mixing (Section 4.7) may be appropriate if the discrepancies among the estimates represents variability that should be factored into the analysis. The least useful general strategy for aggregating uncertain numbers is perhaps averaging (Section 4.9), because it tends to erase uncertainty rather than capture it for subsequent propagation. But even averaging may be appropriate if the relevant feature of the system is an integral of the variable measured. Bayes' rule (Section 4.6) might be superior when subjective probabilities are involved. Three of the many variants of Dempster's rule (Section 4.5) were discussed, but it is not completely clear under what circumstances any of these variants would be the most appropriate method of aggregation. Nevertheless, some version of Dempster's rule will no doubt continue to be widely used whenever estimates are expressed as Dempster-Shafer structures. Section 4.13 offered a list of questions that may serve as a helpful roadmap to choosing the aggregation method to use in a particular situation.

Which aggregation operation yields the tightest result for a given collection of uncertain numbers depends in part on whether the uncertainty in the inputs is primarily incertitude or variability. Nevertheless, one can recognize the following rough order of the aggregation operations in terms of the breadth of uncertainty in their results.

Tightest: Dempster's rule,
intersection,
convolutive averaging,
horizontal averaging,
mixing,
logarithmic pooling,
enveloping,
Yager's rule,
Broadest: disjunctive consensus.

Note, however, that this ranking is not hard and fast. If the uncertain numbers being aggregated are all intervals, or if they are all precise probability distributions, the order shifts somewhat.

Section 5 considered the treatment of model uncertainty. Its main thesis is that, just like parametric uncertainty, model uncertainty can be either incertitude or variability. When it is the latter, it is best to represent it with a stochastic mixture. When it is the former, a mixture model is inappropriate and it should be represented using a bounding approach such as enveloping.

7 Future research directions

As already mentioned in Sections 3.5.4, 3.5.6.2 and 4.11, research is needed to develop a comprehensive theory that incorporates and generalizes the sampling theory of traditional probabilists and fully justifies the use of confidence procedures to create p-boxes and Dempster-Shafer structures and their subsequent use in risk assessment models. Such a theory would allow information based on sample data to be used to make projections about entire populations. One possible strategy to account for small sample size would be to use the inequality of Saw et al. (1988) described in Section 3.4.3 to obtain non-parametric bounds on the underlying distribution. This approach assumes that the samples are real numbers. It might be possible to generalize it to the case of interval estimates, but clearly fundamental research would probably be necessary to develop the approach to accept more general uncertain numbers as the estimates.

Another conceivable strategy to account for small sample size would be to combine ordinary statistical confidence procedures with a modeling assumption about the distribution shape. As discussed in Section 3.1, the shape of the overall distribution family could be posited by direct assumption, or perhaps justified by mechanistic knowledge. If the sample values are all real numbers, then ordinary statistical procedures can be used to obtain confidence limits on, for instance, their mean and variance. In conjunction with the assumption about shape, these interval parameters would specify p-boxes that account for sampling variation. This approach extends Grosof's (1986) idea of using statistical confidence procedures to come up with intervals to an even more ambitious one that would empower confidence procedures to specify p-boxes as well. If the sampled estimates are intervals rather than point values, then calculating the needed confidence intervals becomes much more difficult. In fact, the computation is NP-hard, although there are algorithms that may produce close-to-best-possible bounds more conveniently (Ferson et al. 2002). The problem for more general uncertain numbers has not yet been considered. Coming up with a mean might not be very difficult, but what is the variance of a collection of p-boxes or Dempster-Shafer structures? Further research is needed in this direction.

8 Glossary

aleatory uncertainty The kind of uncertainty resulting from randomness or unpredictability due to stochasticity. Aleatory uncertainty is also known as variability, stochastic uncertainty, Type I or Type A uncertainty, irreducible uncertainty, objective uncertainty.

best possible A upper bound is best possible if it is the smallest such bound possible. A lower bound is best possible if it is the largest lower bound possible.

bound An upper bound of a set of real numbers is a real number that is greater than or equal to every number in the set. A lower bound is a number less than or equal to every number in the set. In this report, we also consider bounds on functions. These are not bounds on the range of the function, but rather bounds on the function for every function input. For instance, an upper bound on a function $F(x)$ is another function $B(x)$ such that $B(x) \geq F(x)$ for all values of x. $B(x)$ is a lower bound on the function if the inequality is reversed. If an upper bound cannot be any smaller, or a lower bound cannot be any larger, it is called a best possible bound.

composition The formation of one function by sequentially applying two or more functions. For example, the composite function $f(g(x))$ is obtained by applying the function g to the argument x and then applying the function f to this result.

conjugate pair In Bayesian estimation, when the observation of new data changes only the parameters of the prior distribution and not its statistical shape (i.e., whether it is normal, beta, etc.), the prior distribution on the estimated parameter and the distribution of the quantity (from which observations are drawn) are said to form a conjugate pair. In case the likelihood and prior form a conjugate pair, the computational burden of Bayes' rule is greatly reduced.

convolution The mathematical operation which finds the distribution of a sum of random variables from the distributions of its addends. The term can be generalized to refer to differences, products, quotients, etc. It can also be generalized to refer to intervals, p-boxes and Dempster-Shafer structures as well as distributions.

copula The function that joins together marginal distributions to form a joint distribution function. For the bivariate case, a copula is a function $C: [0,1]\times[0,1]\rightarrow[0,1]$ such that $C(a, 0) = C(0, a) = 0$ for all $a\in[0,1]$, $C(a, 1) = C(1, a) = a$ for all $a\in[0,1]$, and $C(a_2, b_2)-C(a_1, b_2)-C(a_2, b_1)+C(a_1, b_1) \geq 0$ for all $a_1,a_2,b_1,b_2\in[0,1]$ whenever $a_1\leq a_2$ and $b_1\leq b_2$. For any copula C, $\max(a+b-1,0) \leq C(a,b) \leq \min(a,b)$.

cumulative distribution function For a random variable X, the probability $F(x)$ that X will take on a value not greater than x. If the random variable takes on only a finite set of values, then $F(x)$ is the sum of the probabilities of the values less than or equal to x. Also known as a distribution function.

delta distribution The cumulative distribution function associated with a precisely known and constant scalar value. This distribution therefore lacks all variability and incertitude. Its shape is that of the step function $H_c(x)$, where c is the value of the scalar.

Dempster-Shafer structure A kind of uncertain number representing indistinguishability within bodies of evidence. In this report, a Dempster-Shafer structure is a finite set of closed intervals of the real line, each of which is associated with a nonnegative value m, such that the sum of all such m's is one.

distribution function For a random variable X, the probability $F(x)$ that X will take on a value not greater than x. If the random variable takes on only a finite set of values, then $F(x)$ is the sum of the probabilities of the values less than or equal to x. Also known as a cumulative distribution function.

epistemic uncertainty The kind of uncertainty arising from imperfect knowledge. Epistemic uncertainty is also known as incertitude, ignorance, subjective uncertainty, Type II or Type B uncertainty, reducible uncertainty, and state-of-knowledge uncertainty.

focal element A set (in this report, a closed interval of the real line) associated with a nonzero mass as a part of a Dempster-Shafer structure.

$\mathbf{H_c(x)}$ The step function that is zero for all values of x less than c and one for all values equal to or greater than c.

imprecise probabilities Any of several theories involving models of uncertainty that do not assume a unique underlying probability distribution, but instead correspond to a set of probability distributions (Couso et al. 2000). An imprecise probability arises when one's lower probability for an event is strictly smaller than one's upper probability for the same event (Walley 1991). Theories of imprecise probabilities are often expressed in terms of a lower probability measure giving the lower probability for every possible event from some universal set, or in terms of closed convex sets of probability distributions (which are generally much more complicated structures than either probability boxes or Dempster-Shafer structures).

incertitude The kind of uncertainty arising from imperfect knowledge. Incertitude is also known as epistemic uncertainty, ignorance, subjective uncertainty, Type II or Type B uncertainty, reducible uncertainty, and state-of-knowledge uncertainty.

infimum The greatest lower bound of a set of values. When the set consists of a finite collection of closed intervals, the infimum value is the same as the minimum value.

interval A kind of uncertain number consisting of the set of all real numbers lying between two fixed numbers called the endpoints of the interval. In this report, intervals are always closed so that the endpoints are always considered part of the set.

inverse function For a function $y = F(x)$, an inverse function F^{-1} takes y-values in the range of the function F to x-values in the domain of F in such a way that $F^{-1}(F(x)) = x$ and $F(F^{-1}(y)) = y$. For instance, if $F(x)$ is the distribution function for a random variable X giving the probability associated with the event $X \leq x$, then the inverse function $F^{-1}(p)$ is the value of x associated with any value p. An inverse function does not necessarily exist for any function, but any one-to-one function will have an inverse.

lower probability The maximum rate for an event A one would be willing to pay for the gamble that pays 1 unit of utility if A occurs and nothing otherwise.

mean The probability-weighted average of a set of values or a probability distribution. The mean is also called the expected value or the expectation of a random variable. It is the first moment of a probability distribution.

measurement error The difference between a measured quantity and its actual or true value is called measurement error. The term is also used to refer to the imprecision or uncertainty about a measurement, although the term measurement uncertainty is now preferable for this meaning.

measurement uncertainty The uncertainty (incertitude) about the accuracy of a measurement.

median A magnitude that is greater than half of the values, and less than half of the values, taken on by a random variable. This value is the 50^{th} percentile and the 0.5 quantile of a probability distribution.

mode A value that occurs most frequently in a set of values or a probability distribution.

Monte Carlo simulation A method of calculating functions (often convolutions) of probability distributions by repeatedly sampling random values from those distributions and forming an empirical distribution function of the results.

p-box A probability box.

precision A measure of the reproducibility of a measured value under a given set of conditions.

probability box A kind of uncertain number representing both incertitude and variability. A probability box can be specified by a pair of functions serving as bounds about an imprecisely known cumulative distribution function. The probability box is identified with the class of distribution functions that would be consistent with (i.e., bounded by) these distributions.

quantile A number that divides the range of a set of data or a distribution such that a specified fraction of the data or distribution lies below this number.

random variable A variable quantity whose values are distributed according to a probability distribution. If the potential values of the random variable are a finite or countable set, the random variable is said to be discrete. For a discrete random variable, each potential value has an associated probability between zero and one, and the sum of all of these probabilities is one. If the random variable can take on any value in some interval of the real line (or any rational value within some interval), it is called a continuous random variable.

real number A real number is an element from the real line consisting of positive and negative integers, rational numbers, irrationals and transcendental numbers. A real number is a rational number or the limit of a sequence of rational numbers. Real numbers are sometimes called scalars.

rigorous Exact or sure, as opposed to merely approximate.

robust Bayes A school of thought among Bayesian analysts in which epistemic uncertainty about prior distributions or likelihood functions is quantified and projected through Bayes rule to obtain a class of posterior distributions.

sampling uncertainty The incertitude about a statistic or a probability distribution arising from incomplete sampling of the population characterized by the statistic or distribution.

scalar A real number.

stationary Characterized by an unchanging distribution function.

support The subset of the domain of a distribution function over which the function is neither perfectly zero nor perfectly one.

supremum The least upper bound of a set of values. When the set consists of a finite collection of closed intervals, the supremum value is the same as the maximum value.

two-dimensional Monte Carlo A kind of nested Monte Carlo simulation in which distributions representing both incertitude and variability are combined together. Typically, the outer loop selects random values for the parameters specifying the distributions used in an inner loop to represent variability.

uncertain number A numerical quantity or distribution about which there is uncertainty. Uncertain numbers include intervals, probability distributions, probability boxes, Dempster-Shafer structures as special cases. Uncertain numbers also include scalars (real numbers) as degenerate special cases.

uncertainty The absence of perfectly detailed knowledge. Uncertainty includes incertitude (the exact value is not known) and variability (the value is changing). Uncertainty may also include other forms such as vagueness, ambiguity and fuzziness (in the sense of border-line cases).

upper probability The minimum rate for an event A one would be willing to pay for the gamble that pays 1 unit of utility if A does not occur and nothing otherwise.

variability The fluctuation or variation due to randomness or stochasticity. Variability is also associated with aleatory uncertainty, stochastic uncertainty, Type I or Type A uncertainty, irreducible uncertainty, objective uncertainty.

9 References

Allen, A.O. (1990) *Probability, Statistics and Queueing Theory with Computer Science Applications,* second edition. Academic Press, Boston.

Apostolakis, G.E. (1995). A commentary on model uncertainty. *Proceedings of Workshop on Model Uncertainty*, A. Mosleh, N. Siu, C. Smidts and C. Lui (eds.), Center for Reliability Engineering, University of Maryland, College Park, Maryland.

Appriou, A. (1998). Uncertain Data Aggregation in Classification and Tracking Processes. *Aggregation and Fusion of Imperfect Information*. B. Bouchon-Meunier, Physica-Verlag Heidelberg.

Ash, C. (1993). *The Probability Tutoring Book*. IEEE Press, New York.

Ayyub, B.M. (2001). *Elicitation of Expert Opinions for Uncertainty and Risks*, CRC Press.

Baldwin, J.F. (1994). Mass assignments and fuzzy sets for fuzzy databases. *Advances in the Dempster-Shafer Theory of Evidence*, R.R. Yager, J. Kacprzyk and M. Fedrizzi (eds.). John Wiley and Sons, Inc.

Baldwin, J.F. (1994a). A calculus for mass assignments in evidential reasoning. *Advances in the Dempster-Shafer Theory of Evidence*, R.R. Yager, J. Kacprzyk and M. Fedrizzi (eds.). John Wiley and Sons, Inc.

Barlow, R.E. and A.W. Marshall (1964). Bounds for distributions with monotone hazard rate, I and II. *Annals of Mathematical Statistics* 35: 1234–1257, 1258–1274.

Barlow, R.E. and A.W. Marshall (1965). Tables of bounds for distributions with monotone hazard rate. *American Statistical Association Journal* 60: 872–890.

Barlow, R.E. and A.W. Marshall (1967). Bounds on interval probabilities for restricted families of distributions. *Fifth Berkeley Symposium on Mathematical Statistics and Probability* 3: 229–257.

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.

Berleant, D. (1993). Automatically verified reasoning with both intervals and probability density functions. *Interval Computations* 48–70.

Berleant, D. (1996). Automatically verified arithmetic on probability distributions and intervals. *Applications of Interval Computations*, B. Kearfott and V. Kreinovich (eds.), Kluwer Academic Publishers, pages 227–244.

Berleant, D. and C. Goodman-Strauss. (1998). Bounding the results of arithmetic operations on random variables of unknown dependency using intervals. *Reliable Computing* 4: 147–165.

Berleant, D. and B.J. Kuipers (1997). Qualitative and quantitative simulation: bridging the gap. *Artificial Intelligence* 95(2): 215–255.

Bezdek, J.C., D. Dubois, et al. (1999). F*uzzy Sets in Approximate Reasoning and Information Systems*, Kluwer Academic Publishers.

Bezdek, J.C., J. Keller, et al. (1999a). *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. Toulouse, France, Kluwer Academic Publishers.

Bloch, I. and H. Maitre (1998). Fusion of Image Information under Imprecision. *Aggregation and Fusion of Imperfect Information*. B. Bouchon-Meunier, Physica-Verlag Heidelberg.

Boole, G. (1854). *An Investigation of the Laws of Thought, On Which Are Founded the Mathematical Theories of Logic and Probability*. Walton and Maberly, London.

Burmaster, D.E., K.J. Lloyd, and K.M. Thompson. (1995). The need for new methods to backcalculate soil cleanup targets in interval and probabilistic cancer risk assessments. *Human and Ecologial Risk Assessment* 1:89–100.

Burmaster, D.E. and K.M. Thompson. (1995). Backcalculating cleanup targets in probabilistic risk assessments when the acceptability of cancer risk is defined under different risk management policies. *Human and Ecological Risk Assessment* 1:101–120.

Caselton, W. F. and W. Luo (1992). Decision making with imprecise probabilities: Dempster-Shafer theory and application. *Water Resources Research* 28(12): 3071–3083.

Castillo, E. (1988). *Extreme Value Theory in Engineering.* Academic Press, San Diego.

Chang, S.L. (1974). On risk and decision making in a fuzzy environment. U.S.-Japan Seminar on Fuzzy Sets and Their Applications, University of California, Berkeley, 1974, Academic Press, Inc.

Chateauneuf, A. (1994). Combination of compatible belief functions and relation of specificity. *Advances in the Dempster-Shafer Theory of Evidence*, R.R. Yager, J. Kacprzyk and M. Fedrizzi (eds.). John Wiley and Sons, Inc.

Chebyshev [Tchebichef], P. (1874). Sur les valeurs limites des integrales. *Journal de Mathematiques Pures Appliques*, Ser 2, 19: 157-160.

Clemen, R.T. (1989). Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting* 5: 559–583.

Clemen, R.T. and R. L. Winkler (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis* 19(2): 187–203.

Cooke, R.M. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science.* Oxford University Press, New York.

Cooper, J.A., S. Ferson, and L. R. Ginzburg (1996). Hybrid processing of stochastic and subjective uncertainty data. *Risk Analysis* 16: 785–791.

Corliss, G.F. (1988). How can you apply interval techniques in an industrial setting? Technical Report No. 301, Department of Mathematics, Statistics and Computer Science, Marquette University, Milwaukee.

Couso, I., S. Moral, and P. Walley (2000). A survey of concepts of independence for imprecise probabilities. *Risk Decision and Policy* 5: 165–181.

Cox, L. A., Jr. (1999). Internal Dose, Uncertainty Analysis, and Complexity of Risk Models. *Environment International* 25(6/7): 841–852.

de Cooman, G., T. L. Fine, et al. (2001). *ISIPTA '01. Second International Symposium on Imprecise Probabilities and Their Applications*, Robert Purcell Community Center of Cornell University, Ithaca, NY, USA, Shaker Publishing BV.

Cullen, A.C. and H.C. Frey. (1999). *Probabilistic Techniques in Exposure Assessment*. Plenum Press.

DeGroot, M.H. (1970). *Optimal Statistical Decisions*, McGraw-Hill.

Dempster, A.P. (1967). Upper and lower probabilities induced by a multi-valued mapping. *Annals of Mathematical Statistics* 38: 325–339.
Dharmadhikari, S. W. and K. Joag-Dev (1986). The Gauss-Tchebyshev inequality for unimodal distributions. *Theory of Probability and its Applications* 30: 867-871.
Dieck, R.H. (1997). *Measurement Uncertainty Methods and Applications.* Instrument Society of America, Research Triangle Park, North Carolina.
Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society Series B* 57: 45–97.
Dubois, D. and H. Prade (1986). A set-theoretic view on belief functions: Logical operations and approximations by fuzzy sets. *International Journal of General Systems* 12: 193–226.
Dubois, D. and H. Prade (1992). On the combination of evidence in various mathematical frameworks. *Reliability Data Collection and Analysis*. J. Flamm and T. Luisi (eds.), Brussels, ECSC, EEC, EAFC**:** 213–241.
Dubois, D. and H. Prade (1994). Focusing versus updating in belief function theory. *Advances in the Dempster-Shafer Theory of Evidence*, R.R. Yager, J. Kacprzyk and M. Fedrizzi (eds.). John Wiley and sons, Inc.
Dubois, D. and H. Prade (2000). *Fundamentals of Fuzzy Sets*, Kluwer Academic Publishers.
Dubois, D. and H. Prade, et al., (eds.) (1997*). Fuzzy Information Engineering: A Guided Tour of Applications*. New York, John Wiley and Sons, Inc.
Feller, W. (1948). On the Kolmogorov-Smirnov limit theorems for empirical distributions. *Annals of Mathematical Statistics* 19: 177–189.
Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*. Volume 2, Second edition. John Wiley & Sons, New York.
Ferson, S. (1996). What Monte Carlo methods cannot do. *Human and Ecological Risk Assessment* 2: 990–1007.
Ferson, S. (2002). *RAMAS Risk Calc 4.0 Software: Risk Assessment with Uncertain Numbers*. Lewis Publishers, Boca Raton, Florida.
Ferson, S. and Ginzburg, L. (1995). Hybrid arithmetic, *Proceedings of the 1995 Joint ISUMA/NAFIPS Symposium on Uncertainty Modeling and Analysis*, IEEE Computer Society Press, Los Alamitos, California, 619–623.
Ferson, S. and L.R. Ginzburg (1996). Different methods are needed to propagate ignorance and variability. *Reliability Engineering and Systems Safety* 54:133–144.
Ferson, S., L. Ginzburg, V. Kreinovich, L. Longpre and M. Aviles (2002). Computing variance for interval data is NP-hard. [submitted]
Ferson, S. and T.F. Long (1995). Conservative uncertainty propagation in environmental risk assessments. *Environmental Toxicology and Risk Assessment, Third Volume, ASTM STP 1218*, J.S. Hughes, G.R. Biddinger and E. Mones (eds.), American Society for Testing and Materials, Philadelphia, pp. 97–110.
Ferson, S., L.R. Ginzburg and H.R. Akçakaya. Whereof one cannot speak: when input distributions are unknown. *Risk Analysis* [to appear].
Finkel, A.M. (1995). A second opinion on an environmental misdiagnosis: the risky prescriptions of *Breaking the Vicious Circle*. *New York University Environmental Law Journal* 3: 295–381.

Fodor, J. and R. R. Yager (2000). Fuzzy Set-Theoretic Operators and Quantifiers. *Fundamentals of Fuzzy Sets*. D. Dubois and H. Prade, Kluwer Academic Publishers.

Fox, D., M. Schmidt, M. Koshelev, V. Kreinovich, L. Longpré, and J. Kuhn (1998). We must choose the simplest physical theory: Levin-Li-Vitányi theorem and its potential physical applications. In: G. J. Erickson, J. T. Rychert, and C. R. Smith (eds.), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht, 238–251.

Frank M.J., R.B Nelsen, B. Schweizer (1987). Best-possible bounds for the distribution of a sum—a problem of Kolmogorov. *Probability Theory and Related Fields* 74:199–211.

Fréchet, M., (1935). Généralisations du théorème des probabilités totales. *Fundamenta Mathematica* 25: 379–387.

French, S. (1985). Group consensus probability distributions: a critical survey. In: J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith (eds.), *Bayesian Statistics 2*, North_Holland, Amsterdam, 183-197.

Frey, H.C. and D.S. Rhodes (1998). Characterization and simulation of uncertain frequency distributions: effects of distribution choice, variability, uncertainty, and parameter dependence. *Human and Ecological Risk Assessment* 4: 423–468.

Galambos, J. (1978). *The asymptotic theory of extreme order statistics*. Wiley, N.Y.

Gebhardt, J. and R. Kruse (1998). Parallel combination of information sources. *Handbook of Defeasible Reasoning and Uncertainty Management Systems*. D. M. Gabbay and P. Smets. Dordrecht, The Netherlands, Kluwe Academic Publishers. 3: 393–439.

Gelman, A., J.B. Carlin, H.S. Stern, D.B. Rubin (1995). *Bayesian Data Analysis*. CRC Press, Boca Raton.

Genest, C. and J.V. Zidek (1986). Combining probability distributions: a critique and annotated bibliography. *Statistical Science* 1: 114–148.

Gigerenzer, G. (2002). *Calculated Risks: How to Know When Numbers Deceive You.* Simon and Schuster, New York.

Godwin, H.J. (1955). On generalizations of Tchebychef's inequality. *Journal of the American Statistical Association* 50: 923-945.

Godwin, H.J. (1964). *Inequalities on Distribution Functions*. Hafner, New York.

Goodman, I. R., R. P. S. Mahler, et al. (1997). *Mathematics of Data Fusion*. Dordrecht, The Netherlands, Kluwer Academic Publishers.

Grabisch, M., S. A. Orlovski, et al. (1998). Fuzzy aggregation of numerical preferences. *Fuzzy Sets in Decision Analysis, Operations Research and Statistics*. D. Dubois and H. Prade (eds.), Kluwer Academic Publishers.

Grandy, W.T., Jr. and L.H. Schick (1991). *Maximum Entropy and Bayesian Methods*. Kluwer Academic Publishers, Dordrecht.

Grosof, B.N. (1986). An inequality paradigm for probabilistic knowledge: the logic of conditional probability intervals. *Uncertainty in Artificial Intelligence*. L.N. Kanal and J.F. Lemmer (eds.), Elsevier Science Publishers, Amsterdam.

Gumbel, E.J. (1958). *Statistics of Extremes.* Columbia University Press, New York.

Hailperin, T. (1986). *Boole's Logic and Probability*. North-Holland, Amsterdam.

Halpern, J. Y. and R. Fagin (1992). Two views of belief: belief as generalized probability and belief as evidence. *Artificial Intelligence* 54: 275–317.

Hammer, R., M. Hocks, U. Kulisch, D. Ratz (1993). Numerical toolbox for verified computing. I. Basic numerical problems. Springer Verlag, Heidelberg, N.Y.

Hammitt, J.K. (1995). Can more information increase uncertainty? *Chance* 15–17,36.

Hansen, E. R. (1992). *Global optimization using interval analysis*. Marcel Dekker, N.Y.

Helsel, D.R. (1990). Less than obvious: statistical treatment of data below the detection limit. *Environmental Science and Technology* 24: 1766–1774.

Henrion, M. and B. Fischhoff (1986). Assessing uncertainty in physical constants. *American Journal of Physics* 54: 791–798.

Hoeting, J., D. Madigan, A. Raftery and C. Volinsky (1999). Bayesian model averaging. *Statistical Science* [to appear].

Hora, S. C. and R. L. Iman (1989). Expert opinion in risk analysis: The NUREG-1150 methodology. *Nuclear Science and Engineering* 102: 323–331.

Hyman, J.M. (1982). *FORSIG: An Extension of FORTRAN with Significance Arithmetic*. LA-9448-MS, Los Alamos National Laboratory, Los Alamos, New Mexico. See also the website http://math.lanl.gov/ams/report2000/significancearithmetic.html

Insua, D.R. and F. Ruggeri (eds.) (2000). *Robust Bayesian Analysis*. Lecture Notes in Statistics, Volume 152. Springer-Verlag, New York.

Jaffray, J.-Y. (1992). Bayesian updating and belief functions. *IEEE Transactions on Systems, Man, and Cybernetics* 22: 1144–1152.

Jaffray, J.-Y. (1994). Dynamic decision making with belief functions. *Advances in the Dempster-Shafer Theory of Evidence*, R.R. Yager, J. Kacprzyk and M. Fedrizzi (eds.). John Wiley and Sons, Inc.

Jaynes, E.T. (1957). Information theory and statistical mechanics. *Physical Review* 106: 620–630.

Johnson, N.L., S. Kotz and N. Balakrishnan (1995). *Continuous Univariate Distributions, Volume 2.* Second edition, Wiley, New York.

Jouini, M. N. and R. T. Clemen (1996). Copula models for aggregating expert opinions. *Operations Research*, 44(3): 444–457.

Jousselme, A.-L., D. Grenier, et al. (2001). A new distance between two bodies of evidence. *Information Fusion* 2: 91–101.

Kacprzyk, J. and M. Fedrizzi (1990). *Multiperson Decision Making Models Using Fuzzy Sets and Possibility Theory*, Kluwer Academic Publishers.

Kahneman, D. P. Slovic and A. Tversky (1982) *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge.

Kaplan, S. (1981). On the method of discrete probability distributions in risk and reliability calculations—applications to seismic risk assessment. *Risk Analysis* 1: 189-196.

Karlin, S. and W.J. Studden. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics*. Interscience Publishers, John Wiley and Sons, New York.

Kaufmann, A. and M.M. Gupta (1985). *Introduction to Fuzzy Arithmetic: Theory and Applications.* Van Nostrand Reinhold, New York.

Kearfott, R. B. (1996). *Rigorous global search: continuous problems*. Kluwer, Dordrecht.

Kearfott, R. B. and V. Kreinovich (1996). *Applications of Interval Computations*. Dordrecht, The Netherlands, Kluwer Academic Publishers.

Kelman, A. and R. R. Yager (1998). Using priorities in aggregation onnectives. *Aggregation and Fusion of Imperfect Information*. B. Bouchon-Meunier, Physica-Verlag Heidelberg.
Kirchner, T.B. (1992). *QS-CALC: An Interpreter for Uncertainty Propagation*. Quaternary Software, Fort Collins, Colorado.
Klir, G. and B. Yuan (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications.* Prentice Hall, Upper Saddle River, NJ.
Kolmogorov [Kolmogoroff], A. (1941). Confidence limits for an unknown distribution function. *Annals of Mathematical Statistics* 12: 461–463.
Kosko, B. (1986). Fuzzy Knowledge Combination. *International Journal of Intelligent Systems* 1: 293–320.
Kreinovich, V. (1997). Random sets unify, explain, and aid known uncertainty methods in expert systems. *Random Sets, Theory and Applications*. A. Friedman and R. Gulliver. New York, Springer. 97.
Kreinovich, V., A. Bernat, W. Borrett, Y. Mariscal and E. Villa (1994). Monte-Carlo methods make Dempster-Shafer formalism feasible. *Advances in the Dempster-Shafer Theory of Evidence*, R.R. Yager, J. Kacprzyk and M. Fedrizzi (eds.). John Wiley and Sons, Inc.
Kreinovich, V., A. Lakeyev, J. Rohn, and P. Kahl (1997). Computational complexity and feasibility of data processing and interval computations. Kluwer, Dordrecht, 1997.
Kruse, R. and F. Klawonn (1994). Mass distributions on L-fuzzy sets and families of frames of discernment. *Advances in the Dempster-Shafer Theory of Evidence*, R.R. Yager, J. Kacprzyk and M. Fedrizzi (eds.). John Wiley and Sons, Inc.
Kuznetsov, V. P. (1991). Interval statistical models. Moscow, Radio i Svyaz Publ. (in Russian).
Lee, P.M. (1997). *Bayesian Statistics: An Introduction*. Edward Arnold (Hodder Arnold) London.
Lee, R.C. and W.E. Wright (1994). Development of human exposure-factor distributions using maximum-entropy inference. *Journal of Exposure Analysis and Environmental Epidemiology* 4: 329–341.
Levi, I. (2000). Imprecise and indeterminate probabilities. *Risk Decision and Policy* 5: 111–122.
Levine, R.D. and M. Tribus (1978). *The Maximum Entropy Formalism*. MIT Press, Cambridge.
Li, M. and P. Vitányi (1997). *An Introduction to Kolmogorov Complexity and its Applications* Springer-Verlag, NY.
Lindley, D.V. A. Tversky and R.V. Brown. (1979). On the reconciliation of probability assessments. *Journal of the Royal Statistical Society* A 142 (Part 2): 146–180.
Luo, W. B. and B. Caselton (1997). Using Dempster-Shafer theory to represent climate change uncertainties. *Journal of Environmental Management* 49: 73–93.
Mahler, R. P. S. (1995). Combining ambiguous evidence with respect to ambiguous a priori knowledge. Part II: Fuzzy Logic. *Fuzzy Sets and Systems* 75: 319–354.
Mallows, C.L. (1956). Generalizations of Tchebycheff's inequalities. *Journal of the Royal Statistical Society*, Series B, 18: 139–176.
Manes, E.G. (1982). A class of fuzzy theories. *Journal of Mathematical Analysis and Its Applications* 85.

Markov [Markoff], A. (1886). Sur une question de maximum et de minimum proposée par M. Tchebycheff. *Acta Mathematica* 9: 57−70.
Markov, A.A. (1898). On the limiting values of integrals in connection with interpolation. *Zap. Imp. Akad. Nauk. Fiz.—Mat. Otd.* (8) 6: 146–230 [Russian].
Miller, L.H. (1956). Table of percentage points of Kolmogorov statistics. Journal of the American Statistical Association 51: 111–121.
Milner, P. M. (1970). *Physiological psychology*. Holt, New York.
Mongin, P. (1995). Consistent Bayesian aggregation. *Journal of EconomicTheory* 6: 313−351.
Mood, A.M., F.A. Graybill and D.C. Boes. (1974). *Introduction to the Theory of Statistics*. McGraw-Hill, New York.
Moore R.E. (1966). *Interval Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey.
Moore, R. (1979). *Methods and Applications of Interval Analysis*. SIAM, Philadelphia.
Morgan, M.G. and M. Henrion. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* Cambridge University Press, Cambridge.
Mosleh, A. and V. Bier (1992). On decomposition and aggregation error in estimation: some basic principles and examples. *Risk Analysis* 12: 203−214.
Narumi, S. (1923). On further inequalities with possible application to problems in the theory of probability. *Biometrika* 15: 245−253.
Nelsen, R. B. (1999). *Introduction to Copulas*. Springer Verlag, New York.
Neumaier A., (1990). *Interval Methods for Systems of Equations.* Cambridge University Press, Cambridge.
Nguyen, H. T. and V. Kreinovich (1996). Nested Intervals and Sets: Concepts, Relations to Fuzzy Sets, and Applications. In: R. B. Kearfott and V. Kreinovich (eds.), *Applications of Interval Computations*, Kluwer, Dordrecht, 245–290.
Nguyen, H. T. and V. Kreinovich (1997). *Applications of continuous mathematics to computer science*. Kluwer, Dordrecht.
Nguyen, H. T. and V. Kreinovich (1997a). Kolmogorov's Theorem and its impact on soft computing, In: Yager, R. R. and J. Kacprzyk. *The Ordered Weighted Averaging Operators: Theory and Applications*, Kluwer, Boston, MA, 1–17.
Nguyen, H. T., V. Kreinovich, and Q. Zuo (1997). Interval-valued degrees of belief: applications of interval computations to expert systems and intelligent control". *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems* (IJUFKS) 5(3): 317–358.
Nguyen, H. T. and M. Sugeno (1998). *Fuzzy Systems Modeling and Control*, Kluwer Academic Publishers.
Nguyen, H. T. and E. A. Walker (1999). *First Course in Fuzzy Logic*. CRC Press, Boca Raton, Florida.
Oberkampf, W.L., J.C. Helton and K. Sentz (2001). Mathematical representation of uncertainty. American Institute of Aeronautics and Astronautics Non-Deterministic Approaches Forum, Seattle, WA, Paper No. 2001-1645, April, 2001.
Parysow, P. and D.J. and Tazik. Assessing the effect of estimation error in population viability analysis: an example using the black-capped vireo. *Ecological Modelling* [in review].

Rabinovich, S. (1993). *Measurement Errors: Theory and Practice*. American Institute of Physics, New York.
Raftery, A.E., D. Madigan, and J.A. Hoeting (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* 92: 179–191.
Regan, H.M., S. Ferson, and D. Berleant. (2002). Equivalence of five methods for bounding uncertainty. *International Journal of ApproximateReasoning* [to appear].
Rohlf, F.J. and R.R. Sokal. (1981). *Statistical Tables*. Freeman, New York.
Ross, T. J., B. Wu, and V. Kreinovich (2000). Optimal elimination of inconsistency in expert knowledge: Formulation of the problem, fast algorithms. *Proceedings of an International Conference on Intelligent Technologies*, Bangkok, Thailand, December 13–15, 2000, 450–458.
Rowe, N.C. (1988). Absolute bounds on the mean and standard deviation of transformed data for constant-sign-derivative transformations. *SIAM Journal of Scientific Statistical Computing* 9: 1098–1113.
Royden, H.L. (1953). Bounds on a distribution function when its first *n* moments are given. *Annals of Mathematical Statistics* 24: 361-376.
Saffiotti, A. (1994). Issues of knowledge representation in Dempster-Shafer's theory. *Advances in the Dempster-Shafer Theory of Evidence*, R.R. Yager, J. Kacprzyk and M. Fedrizzi (eds.). John Wiley and Sons, Inc.
Sander, P. and R. Badoux (eds.) (1991). *Bayesian Methods in Reliability*. Kluwer, Dordrecht.
Savage, I.R. (1961). Probability inequalities of the Tchebycheff type. *Journal of Research of the National Bureau of Standards–B. Mathematics and Mathematical Physics* 65B: 211-222.
Saw, J.G., M.C.K. Yang and T.C. Mo (1984). Chebyshev inequality with estimated mean and variance. *The American Statistician* 38: 130–132.
Saw, J.G., M.C.K. Yang and T.C. Mo (1988). Corrections. *The American Statistician* 42: 166.
Schubert, J. (1994). *Cluster-based specification techniques in Dempster-Shafer theory for an evidential intelligence analysis of multiple target tracks*. Ph.D. Dissertation, Royal Institute of Technology, Department of Numerical Analysis and Computer Science, Stockholm, Sweden.
Sentz, K. and S. Ferson (2002). *Combination of Evidence in Dempster-Shafer Theory.* SAND2002-0835 Technical Report. Sandia National Laboratories, Albuquerque, New Mexico.
Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ.
Shafer, G. (1984). A mathematical theory of evidence. *The AI Magazine*: 81–83.
Shafer, G. (1986). The combination of evidence. *International Journal of Intelligent Systems* 1: 155–179.
Slowinski, R. (1998). *Fuzzy Sets in Decision Analysis, Operations Research and Statistics*, Kluwer Academic Publishers.
Smirnov [Smirnoff], N. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bulletin de l'Université de Moscou, Série internationale (Mathématiques)* 2: (fasc. 2).

Smith, C.D. (1930). On generalized Tchebycheff inequalities in mathematical statistics. *The American Journal of Mathematics* 52: 109-126.
Smith, J.E. (1990). *Moment Methods for Decision Analysis*. Department of Engineering-Economic Systems (dissertation), Stanford University.
Smith, J.E. (1995). Generalized Chebychev inequalities: theory and applications in decision analysis. *Operations Research* 43: 807-825.
Sokal, R.R. and F.J. Rohlf. (1981). *Biometry.* Freeman, New York.
Solana, V. and N.C. Lind (1990). Two principles for data based on probabilistic system analysis. *Proceedings of ICOSSAR '89, 5th International Conferences on Structural Safety and Reliability*. American Society of Civil Engineers, New York.
Spies, M. (1994). Evidential reasoning with conditional events. *Advances in the Dempster-Shafer Theory of Evidence*, R.R. Yager, J. Kacprzyk and M. Fedrizzi (eds.). John Wiley and Sons, Inc.
Srivastava, R., P. (1997). Audit decisions using belief functions: a review. *Control and Cybernetics* 26(2): 135–160.
Srivastava, R. P., P. P. Shenoy, et al. (1995). Propagating Belief Functions in AND-Trees. *International Journal of Intelligent Systems* 10: 647–664.
Stone, M. (1961). The opinion pool. *Annals of Mathematical Statistics* 32: 1339–1342.
Strat, T. M. (1984). Continuous Belief Functions for Evidential Reasoning. Proceedings of the National Conference on Artificial Intelligence, University of Texas at Austin.
Stuart, A. and J.K. Ord (1994). *Kendall's Advanced Theory of Statistics*. Edward Arnold, London.
Suppes, P., D. M. Krantz, R. D. Luce, and A. Tversky (1989). *Foundations of measurement. Vol. I–III*, Academic Press, San Diego, CA.
Taylor, B.N. and C.E. Kuyatt (1994). *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*. NIST Technical Note 1297, National Institute of Standards and Technology, Gaithersburg, Maryland. Available on line at http://physics.nist.gov/Document/tn1297.pdf.
Tchebichef [Chebyshev], P. (1874). Sur les valeurs limites des integrales. *Journal de Mathematiques Pures Appliques.* Ser 2, 19: 157-160.
Trejo, R. and V. Kreinovich (2001). Error estimations for indirect measurements: randomized vs. deterministic algorithms for 'black-box' programs. In: S. Rajasekaran, P. Pardalos, J. Reif, and J. Rolim (eds.), *Handbook on Randomized Computing*, Kluwer, 673–729. Available on line at http://www.cs.utep.edu/vladik/2000/tr00-17.pdf.
Tsoukalas, L. H. and R. E. Uhrig (1997). *Fuzzy and Neural Approaches in Engineering*, John Wiley and Sons, Inc.
Uhrig, R. E. and L. H. Tsoukalas (1999). Soft computing technologies in nuclear engineering applications. *Progress in Nuclear Energy* 34(1): 13–75.
Voorbraak, F. (1991). On the justification of Dempster's rule of combination. *Artificial Intelligence* 48: 171–197.
Wadsworth, H. M., ed. (1990). *Handbook of Statistical Methods for Engineers and Scientists*, McGraw-Hill Publishing Co., N.Y.
Wald, A. (1938). Generalization of the inequality of Markoff. *Annals of Mathematical Statistics* 9: 244–255.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.
Walley, P. and G. de Cooman (2001). A behavioral model for linguistic uncertainty. *Information Sciences* 134: 1–37.
Walley, P.and T.L. Fine (1982). Towards a frequentist theory of upper and lower probability. *Annals of Statistics* 10: 741–761.
Whalen, T. (1994). Interval probabilities induced by decision problems. *Advances in the Dempster-Shafer Theory of Evidence*, R.R. Yager, J. Kacprzyk and M. Fedrizzi (eds.). John Wiley and Sons, Inc.
Whitt, W. (1976). Bivariate distributions with given marinals. *The Annals of Statistics* 4: 1280–1289.
Wilks, S.S. (1962). *Mathematical Statistics*. John Wiley & Sons, New York.
Williamson, R.C. (1989). *Probabilistic Arithmetic*. Ph.D. dissertation, University of Queensland, Austrialia. http://theorem.anu.edu.au/~williams/papers/thesis300dpi.ps
Williamson R.C. and T. Downs. (1990). Probabilistic arithmetic I: numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning* 4:89–158.
Winkler, R.L. (1968). The consensus of subjective probability distributions. *Management Science* 15: B61–B75.
Woo, G. (1991). A quitting rule for Monte-Carlo simulation of extreme risks. *Reliability Engineering and System Safety* 31: 179–189.
Yager, R.R. (1983). Hedging in the combination of evidence. *Journal of Information and Optimization Sciences* 4(1): 73–81.
Yager, R.R. (1985). On the relationship of methods of aggregating evidence in expert systems. *Cybernetics and Systems: An International Journal* 16: 1–21.
Yager, R.R. (1986). Arithmetic and other operations on Dempster-Shafer structures. *International Journal of Man-Machine Studies* 25: 357–366.
Yager, R.R. (1987). On the Dempster-Shafer framework and new combination rules. *Information Sciences* 41: 93–137.
Yager, R.R. (1987b). Quasi-associative operations in the combination of evidence. *Kybernetes* 16: 37-41.
Yager, R.R. (1988). On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics* 18(1): 183–190.
Yager, R.R. (2001). Dempster-Shafer belief structures with interval valued focal weights. *International Journal of Intelligent Systems* 16: 497–512.
Yager, R.R., J. Kacprzyk and M. Fedrizzi (eds.) (1994). *Advances in the Dempster-Shafer Theory of Evidence*, John Wiley and Sons, Inc.
Yager, R. R. and J. Kacprzyk (1997). *The Ordered Weighted Averaging Operators: Theory and Applications*, Kluwer, Boston, MA.
Yager, R. and A. Kelman (1996). Fusion of fuzzy information. *International Journal of Approximate Reasoning* 15: 93–122.
Young, D.M., J.W. Seaman, D.W. Turner et al. (1988). Probability-inequalities for continuous unimodal random-variables with finite support. *Communications in Statistics: Theory and Methods* 17: 3505–3519.
Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control* 8: 338–353.

Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics SMC*-3(1): 28–44.
Zadeh, L. A. (1986). A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination. *The AI Magazine*: 85–90.
Zhang, L. (1994). Representation, independence, and combination of evidence in the Dempster-Shafer theory. *Advances in the Dempster-Shafer Theory of Evidence*. R.R. Yager, J. Kacprzyk and M. Fedrizzi (eds.). John Wiley and Sons, Inc.
Zimmerman, H.-J. (1999). *Practical Applications of Fuzzy Technologies*, Kluwer Academic Publishers.