Verified computation with probabilities

Scott Ferson and Jack Siegrist

Applied Biomathematics

Abstract. Because machine calculations are prone to errors that can sometimes accumulate disastrously, computer scientists use special strategies called verified computation to ensure output is reliable. Such strategies are needed for computing with probability distributions. In probabilistic calculations, analysts have routinely assumed (i) probabilities and probability distributions are precisely specified, (ii) most or all variables are independent or otherwise have well-known dependence, and (iii) model structure is known perfectly. These assumptions are usually made for mathematical convenience, rather than with empirical justification, even in sophisticated applications. Probability bounds analysis computes bounds guaranteed to enclose probabilities and probability distributions even when these assumptions are relaxed or removed. In many cases, results are best-possible bounds, i.e., tightening them requires additional empirical information. This paper presents an overview of probability bounds analysis as a computationally practical implementation of the theory of imprecise probabilities that represents verified computation of probabilities and distributions.

Keywords: probability bounds analysis, probability box, p-box, verified computation, imprecise probabilities, interval analysis, probabilistic arithmetic

1 Introduction

Many high-profile disasters are attributable to numerical errors in computer calculations. The self-destruction of the Ariane 5 rocket on its maiden test flight was caused by integer overflow (ESA 1996). The crash of the Mars Climate Orbiter during orbital insertion resulted from a units incompatibility (Isbell et al. 1999). The Sleipner A offshore drilling platform sank because of an inaccurate finite element approximation (Selby et al. 1997). The Aegis cruiser USS *Yorktown* was dead in the water for several hours because of a propagated divide-by-zero error (Slabodkin 1998). The Flash Crash in which the Dow Jones Industrial Average plunged 9% almost instantaneously was due to runaway computerized trading mediated by the interacting algorithms used by high-frequency traders (CFTC/SEC 2010). These errors can be worse than costly or embarrassing. The failure of a Patriot missile to intercept the SCUD missile that killed 28 people and injured 100 more was supposedly due to accumulated round-off error (GAO 1992). Miscalculations arising from a race-condition error in the medical

software controlling the Therac-25 used for radiation therapy caused multiple fatal radiation overdoses to patients (Baase 2008, 425).

The corny adage "To err is human, but to really foul things up requires a computer" is not merely bitterness of the underemployed. Because computers are so fast, errors can propagate and accumulate very quickly, and because they often lack a machine analog of human contextual common sense, dramatic errors can go unnoticed until damage is unavoidable. Subtle, even minor features can, in unlucky situations, interact to create disastrously bad numerical results. Ironically, the appearance of precision in computer results can often induce a human error in which users place undue trust in the computer's output.

To overcome these problems, computer scientists have developed methods for 'verified computing' by which users will always get reliably accurate results, or at least will be made aware of the problem when their results are not reliable. One basic task in verified computing is to find an enclosure that surely contains the exact result of a calculation. This problem is often addressed using the methods of interval analysis, which is a mathematically rigorous form of arithmetic that can be implemented in software even though computers can represent numbers with only finite precision (Kulisch et al. 1993; Hammer et al. 1997; Popova 2009; Tucker 2011). In fact, these interval calculations can have rigor corresponding to that of a mathematical proof, in spite of the fact that they are done automatically by machine. The approach guarantees that rounding error is limited, integer overflow is prevented, and division by zero as well as similar impossible operations are handled appropriately to ensure the integrity of the affected calculation. Of course, this means that real-valued answers cannot generally be represented precisely in finite machine number schemes. Instead, the answers are represented by enclosures consisting of two bounding machine-representable values. If this enclosure interval is narrow, we know the answer reliably and accurately. If the interval is wide, we have a transparent warning that the associated uncertainty is large, which implies that a more careful reanalysis may be useful.

Interval analysis is often offered as the primary—and one might think the only—method for verified computation, but verified computing requires a panoply of methods. Consistent application of mathematical rigor in the design of the algorithm, in the arithmetic operations it uses, and in the execution of the program allow an analyst to guarantee that a problem has a solution somewhere in the computed enclosing interval (or that no solution exists). To enable such consistency, methods must be developed for the wide variety of numerical and other operations that computers do for us. For instance, basic mathematical operations on floating-point numbers are replaced by interval analysis on intervals guaranteed to enclose scalar real values. Likewise, methods for vector and matrix operations have been developed that extend and generalize interval analysis with multidimensional arrays of interval ranges.

Similar methods of verified computing are needed for representing and calculating with probabilities and probability distributions on finite-precision machines. Unfortunately, the properties of probability distributions and the features of the laws of probability complicate the effort considerably. For instance, even

representing a univariate continuous distribution is an infinite-dimensional problem, because it is a continuous function whose values at an infinity of points must be captured in the finite storage accessible by the computer. The critical role of assumptions about the stochastic dependence among variables is particularly complicating, and this is an issue even for total probabilities that can be represented by single scalar values. For example, if two events have probabilities 0.2 and 0.3 respectively, the AND operator commonly used in fault trees would only yield the product 0.06 when the probabilities are independent. Without specifying the dependence between the events, the result of the operator is undefined (although it can be bounded). The role of dependence assumptions is much more complicated still for distributions of random variables (Ferson et al. 2004; Nelsen 1999).

Nevertheless, we must undertake the effort, whatever its complexity. There is a pronounced need for verified computing methods to use with probabilities and probability distributions because they are becoming more and more pervasively used in engineering calculations including uncertainty analyses, risk assessments, sensitivity studies, and modeling of quantities with intrinsic aleatory uncertainties. They are being used across a host of fields as diverse as financial planning (Hertz 1964; Boyle 1977), human health risk analyses (McKone and Ryan 1989), ecological risk assessments (Suter 1993), materials and weapons safety calculations (Elliott 2005; Cooper 1994), extinction risk analysis for endangered species (Burgman et al. 1993; Ferson and Burgman 2000), and probabilistic risk assessment for nuclear power (Hickman et al. 1983) and other engineered systems (Vick 2002).

Engineers routinely face three crucial issues when they develop probabilistic models. The first is that their model uncertainty, i.e., their doubt about the proper mathematical form the model should have, is almost never articulated, much less accounted for in any comprehensive way. Modelers may recognize and acknowledge the limitations induced by this problem, yet they rarely conduct the sensitivity studies needed to fully assess the consequences of the uncertainty on model results. The second crucial problem is that there is often little or no quantitative information about possible correlations among the input variables, and in many cases the nature of the intervariable dependencies may not have been empirically studied at all. The typical response of analysts, even if they are aware of their uncertainty, is to nevertheless assume independence among variables, even though this assumption may be neither realistic nor conservative. In fact, using an incorrect assumption about dependence can strongly distort the output distributions, especially in their tails (Ferson et al. 2004; contra Smith et al. 1992).

The third crucial problem faced by engineers developing probabilistic models is that it is often impossible to fully justify a precise probability distribution to be used as input in the model, and sometimes the family of the distribution is only a guess. There is a huge literature on the subject of estimating probability distributions from empirical data, and there are several methods available for use including the method of matching moments, maximum likelihood es-

timation, the maximum entropy criterion and Bayesian methods to compute posterior predictive distributions. But these standard approaches are of limited practical reliability when few relevant data exist. Even when confidence limits are computed, little use can be made of them without an elaborate sensitivity study that is cumbersome to organize, computationally intense, and difficult to interpret. With limited empirical information, all of these methods for selecting input distributions require assumptions that cannot be justified by appeal to evidence and therefore may be false. These unsubstantiated assumptions can make a difference in the results. As Bukowski et al. (1995) showed, the choice about distribution shape can have a sizeable effect on the output distributions, especially at the tails.

It is generally assumed that the only solution to incomplete information is additional empirical effort to measure correlations, develop input distributions, and validate the model. As a practical matter, since such empirical information is typically incomplete—and indeed often quite sparse—analysts are forced to make assumptions without empirical justifications, leading to diminished credibility for the assessment and any subsequent decisions. There are, however, computational methods that allow analysts to sidestep a lack of information about the correlation and dependency structure among variables to obtain partial or complete solutions in many practical cases without having to make unjustified and possibly false assumptions. Likewise, when empirical information about the input distributions is limited, far more appropriate representations of uncertainty can be developed than are currently obtainable using techniques such as the maximum entropy criterion. These new methods allow us to compute bounds on estimates of probabilities and probability distributions that are guaranteed to be correct even when one or more of the assumptions is relaxed or removed. In many cases, the results obtained are the best possible bounds, which means that tightening them would require additional empirical information. This paper reviews probability bounds analysis (PBA, Ferson et al. 2003), as a computationally practical calculus of the theory of imprecise probabilities (IP, Walley 1991), that combines ideas from both interval analysis and probability theory to sidestep the limitations of each. Probability bounds analysis is logically and morally equivalent to a sensitivity analysis. Objecting to PBA implies an objection to sensitivity analysis. PBA uses exactly the same mathematical approach used in sensitivity analysis, but its computational methods are applicable to broader questions and are vastly more efficient.

2   Kinds of uncertainty

In the past, uncertainty analysis considered the source of uncertainty to be its salient aspect, so modelers talked, for example, about their parametric uncertainty or their model-form uncertainty. A more modern view is that the nature of the uncertainty, rather than its source, is a more important characteristic. We can distinguish between two main forms of uncertainty: variability and incertitude. *Variability* refers to the stochastic fluctuations in a quantity through time,

variation across space, manufacturing differences among components, genetic or phenotypic differences among individuals, or similar heterogeneity within some ensemble or population. Engineers often refer to variability as aleatory uncertainty, harkening to *alea*, the Latin word for dice. This is considered to be a form of uncertainty because the value of the quantity can change each time one looks, and one cannot predict precisely what the next value will be (although the distribution of values may be known). *Incertitude*, on the other hand, refers to the lack of full knowledge about a quantity that arises from imperfect measurement, limited sampling effort, or incomplete scientific understanding about the underlying processes that govern a quantity. Many engineers refer to incertitude as epistemic uncertainty. We might simply and non-euphemistically call it 'ignorance', except for the embarrassment or confusion that word might evoke should professionals need to mention it in front of their bosses or the laity.

These two forms of uncertainty have important differences. Incertitude can in principle be reduced by empirical effort; investing more in measurement should yield better precision. Variability, in contrast, can perhaps be better characterized, but cannot generally be reduced by empirical effort. Incertitude depends on the observer and the observations made. Variability does not depend on an observer at all. It exists whether or not anyone witnesses it, like the sound waves emanating from the proverbial tree falling unseen in the forest. Although variability and incertitude can sometimes be like ice and snow in that their distinction can be difficult to discern through complicating details, and sometimes one can change into the other depending on the scale and perspective of the analyst, the macroscopic differences between these two forms of uncertainty are usually obvious and often significant in practical settings.

There is a crucial difference between a quantity actually varying and our simply not being sure about its magnitude, and this difference affects how we should do calculations. Consider, for example, the following elementary question: Suppose we are told that a quantity A is some value or values between 2 and 4, and that B is a quantity inside the range between 3 and 5. What can be said about their sum $A + B$? When this exemplar question was posed on the Riskanal electronic mailing list, half the respondents suggested the proper answer can be computed by modeling A as a uniform distribution between 2 and 4, and modeling B with another uniform distribution between 3 and 5, and convolving these two uniforms together with Monte Carlo simulation to obtain the triangular distribution ranging between 5 and 9 with a mode at 7 shown as a probability density function in Fig. 1. This is the traditional answer from probabilists for such a question. Indeed, it is the answer that Laplace (1820) himself would have suggested. This answer says that the value 7 is the most likely magnitude of the sum, and also that the extreme values of 5 and 9 have vanishing probabilities. There is more than two-thirds probability that the sum falls in the middle interval [6.1, 7.9].

But what exactly justifies this concentration of probability mass in the central range? There is nothing in the statement of the elementary question that suggests that 2 is not a perfectly possible value of A, and likewise nothing to suggest that
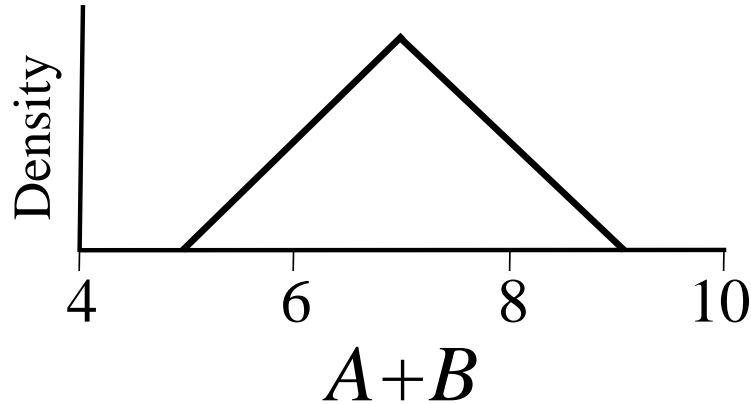
Fig. 1. Triangular distribution which is the traditional probabilist answer to the elementary question "What is the sum $A+B$ where A and B are in the respective intervals [2,4] and [3,5]?"

B might not simply just be 3. If so, then the sum is the scalar value 5, and the triangular distribution seems hard to explain. Given what is expressly known about the inputs, there is no reason to deprecate any of the possible values of the sum, or to distinguish one value as more probable than any other. But that may be a far cry from saying that all the values are equally probable. The other half of the respondents to the Riskanal poll said that the proper answer to the elementary question can be computed simply by adding together the intervals [2,4] + [3,5] using interval arithmetic (Moore 1966) to yield the interval [5,9]. Notice that this answer offers no concentration of mass in the central range, and suggests that the sum might simply be 5, and likewise might simply be 9, and there is nothing to suggest that these values, although extreme, are in any way unlikely.

The interval answer is a much looser statement than is any probability distribution. For instance, modeling the sum with a uniform probability distribution would say that all possible values within the range [5,9] are equally probable. Taking such a model seriously would suggest that one could profitably make bets about future values of the sum based on the probability. For instance, a probabilist would presumably be disposed to bet favorably, and big, on a gamble that the sum is larger than 5.01. A more sanguine view is that one had better not place any such bets, other than those that can be actually justified by the given knowledge. All that can be justified is that the probability distribution of the sum has its support within the range [5,9], but this admits a whole host of possible distributions. The interval answer can be identified with the entire class of such distributions.

Our view is that only one of these two answers to the elementary question is correct. We think the right answer is clearly the interval and not the triangular distribution, at least in practical contexts such as risk analysis and most uncertainty assessments. The triangular distribution traditionally given by probabilists is wrong because it implies or appears to imply more is known than is actually justifiable. It is the incertitude in the elementary problem that must be propagated by interval analysis or other bounding methods. Although these assertions are commonly met with nodding agreement from engineers and biologists, they sometimes evoke agitated criticism from probabilists. So let us hasten to point out some important tempering caveats. We are surely not saying we should only use intervals in risk or uncertainty analysis. We are not even saying that all uncertainty is incertitude. In fact, we would not be surprised that most of the uncertainty in some setting is not incertitude, and we agree that sometimes incertitude is entirely negligible, in which case probability theory is perfectly sufficient for modeling uncertainties and risks.

What we are saying, however, is that some analysts face non-negligible incertitude and handling this incertitude with standard probability theory requires assumptions that may not be tenable, including unbiasedness, uniformity or equiprobability, and independence. Because it will often be useful in practical situations to know what difference incertitude might make, it is important to have methods that can make probabilistic calculations without requiring the traditional assumptions. It turns out that this is possible with the theory of imprecise probabilities (Walley 1999) and a practical calculus for making computations with imprecisely specified probability distributions such as probability bounds analysis (Ferson 2002) which combines probability theory with interval analysis.

3   P-boxes and probability bounds analysis

A probability box, or p-box, is a characterization of an uncertain number which may have variability (aleatory uncertainty) or incertitude (epistemic uncertainty), or both. A p-box is specified by left and right bounds on the cumulative probability distribution function of a quantity and, optionally, additional information about the quantity's mean, variance and distributional shape (family, unimodality, symmetry, etc.). A p-box represents a class of probability distributions consistent with these constraints. Fig. 2 depicts an example for an uncertain number X consisting of a left (upper) bound and a right (lower) bound on the probability distribution for X. The bounds are coincident for values of X below 2 and above 29. The bounds may have almost any shapes, including step functions, so long as they are monotonically increasing and do not cross each other. A p-box simultaneously expresses incertitude (epistemic uncertainty), which is represented by the breadth between the left and right edges of the p-box, and variability (aleatory uncertainty), which is characterized by the overall slant of the p-box. This p-box suggests that the probability that X is below 10 is less than 25%. It might be as low as zero. We cannot say more than this because of
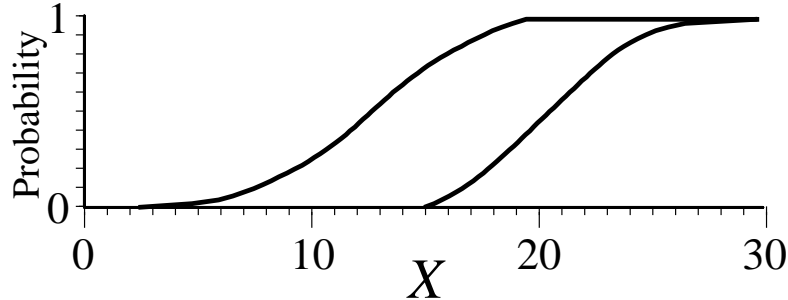
Fig. 2. A p-box specified by left and right bounding cumulative distribution functions and representing a class of probability distributions whose cumulative distribution functions can be drawn within the bounds.

the epistemic uncertainty about X's distribution function. The 95th percentile is somewhere between 18.5 and 26. We don't know where in that range it is because of the associated incertitude.

There are many ways that p-boxes can be constructed from the available information about uncertain numbers (Ferson et al. 2003). Fig. 3 illustrates six of these ways. The top, left graph depicts a distributional p-box for which the shape or family of the distribution is known (e.g., normal, uniform, beta, Weibull, etc.) but the parameters are known only to within intervals. For example, an analyst may know from mechanistic or physical considerations that the distribution is normal, but not be able to precisely identify the two parameters needed to specify it exactly. If the parameters can be bounded, then a distributional p-box can easily be constructed from enveloping all the possible distributions.

The top, right graph in Fig. 3 depicts what might be considered the opposite situation where the analyst is confident about some parameters describing the uncertain number, but is unsure about what shape or family of distributions it might be from. Such a situation arises frequently when distributions are developed from information obtained from scientific publications, where summary statistics are often reported without further details or the original data. Even though the available information might seem meager, what is known often suffices to define a nontrivial p-box that can be used in calculation. In some cases, classical results such as the Markov or Chebyshev inequalities can be used to derive formulas for p-boxes from a few parameters. In different situations, different sets of parameters may be known. Ferson et al. (2003) gave formulas for p-boxes for the following common situations:
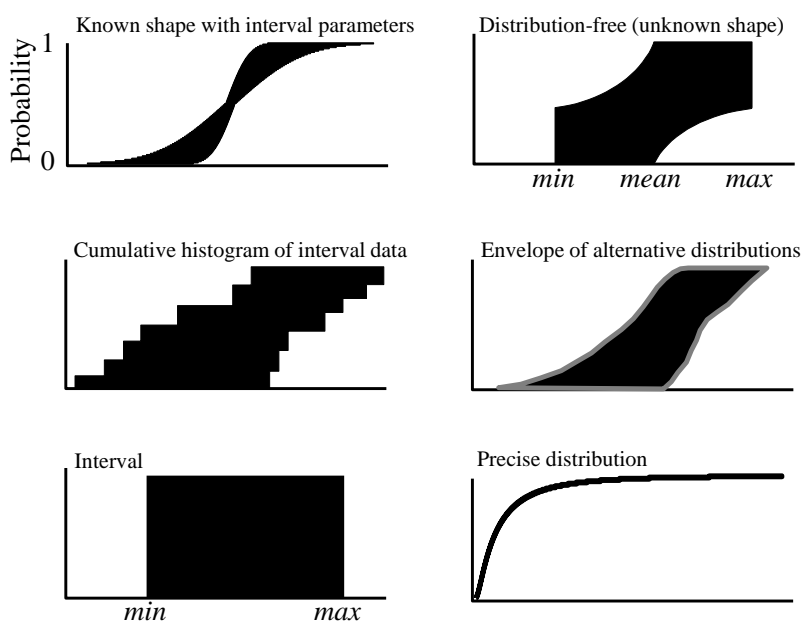
Fig. 3. A few ways p-boxes arise.

{min, mean}	{min, max, mean}
{min, max, median}	{min, max, mean=median}
{min, max, mode}	{min, max, median=mode}
{mean, variance}	{min, mean, variance}
{min, max, mean, variance}	{min, max, mean, variance, mode}

These define what might be called distribution-free p-boxes because they make no assumption whatever about the family or shape of the uncertain distribution and yet enclose all distributions which match the given parameters. The p-boxes are somewhat wider when the parameters are only known to within intervals. When qualitative information is available, such as that the distribution is symmetric or unimodal, the p-boxes can often be tightened substantially.

The middle, right graph in Fig. 3 depicts a situation in which one of two distributions is the correct one, but the analyst cannot discern which. By enveloping them into a single p-box, the analyst can represent the uncertainty in a single structure that does not require a cumbersome sensitivity analysis to propagate. This facility can become very important when there are multiple possible distributions and several variables have such uncertainty because exploring them in a sensitivity study requires a combinatorially complex effort. Collapsing the uncertainty into a single p-box per variable can simplify the problem considerably.

The middle, left graph in Fig. 3 shows a p-box from a situation in which there is no sampling uncertainty because the entire population has been measured, but there is substantial mensurational uncertainty that comes from our inability to measure individual values precisely. A similar p-box with both sampling and mensurational uncertainty can be formed by enclosing the empirical histogram of interval data with Kolmogorov–Smirnov confidence bands. These bands are distribution-free and merely assume independence of the sample data. Alternatively, there are potentially tighter confidence bands that can be used which do make assumptions about the shape of the distribution.

P-boxes include the special cases of intervals and precise probability distributions too. For example, in some situations an analyst may have no information about a distribution except its potential range, that is, knowledge that its values must be larger than min and smaller than max. In this case, Laplace (1820) used the Principle of Insufficient Reason (sometimes called the Principle of Indifference) to select a uniform distribution for the variable but, as argued above, an interval is a fuller characterization of this uncertainty than any particular probability distribution could be. An interval, illustrated in the bottom, left graph of Fig. 3, is a special case of a p-box whose left and right bounds are step functions at min and max respectively. Finally, it is also possible that the distribution for some variable actually is well specified. The bottom, right graph illustrates this case where the left and right bounds of the p-box are coincident.

This idea of bounding probability has a very long tradition throughout the history of probability theory. Indeed, George Boole (1854; Hailperin 1986) used the notion of interval bounds on probability. The classical inequality attributed to Chebyshev (1874) described bounds on a distribution when only the mean and variance of the variable are known, and the related inequality attributed to Markov (1886) found bounds on a positive variable when only the mean is known. Keynes (1921) argued that probabilities of some propositions cannot be ordered because they overlap due to uncertainty. Fréchet (1935) discovered how to bound calculations with total probabilities without assuming independence or making other dependence assumptions. Bounding probabilities has continued to the present day (e.g., Berger 1985; Walley 1991). Kyburg (1999) reviewed the history of interval probabilities and traced the development of the critical ideas over the last century.

Several authors have described strategies for computing with bounds on distribution functions (e.g., Makarov 1981; Yager 1986; Frank et al. 1987; Williamson and Downs 1990; Berleant 1993; 1996; 1998; Ferson 2002; Ferson et al. 2003; inter alia). Williamson and Downs (1990) described explicit algorithms to compute sums, products, differences and quotients. Since their effort, algorithms for essentially all the standard mathematical operations have been derived and implemented (Ferson 2002). These methods, collectively called probability bounds analysis (PBA), have been used to propagate p-boxes through mathematical expressions of widely varying complexity, ranging from simple arithmetic formulas common in risk analyses and logical expressions summarizing fault or event trees to finite-element computations (Zhang et al. 2010; 2012) and evaluations of non-

linear ordinary differential equations (Enszer et al. 2011). The calculations made with these methods can be shown to be *rigorous*, i.e., they are guaranteed to enclose the true outcome distribution whenever the input p-boxes enclose their respective distributions. In many cases, the calculations can also be shown to be pointwise *best-possible* in the sense that they could not be any narrower without excluding distributions that might arise as results given the inputs.

These calculations account for, and preserve the integrity of, both the incertitude and variability expressed by a p-box. Combining a p-box with a scalar, interval, probability distribution or other p-box in any arithmetic or logical calculation generally yields another p-box. Combining an interval with a probability distribution also generally yields a p-box, as the incertitude of the interval combines with the variability of the probability distribution. P-boxes also arise when two precise probability distributions are combined whenever their intervariable dependency is unknown or only partially known. Such combinations produce precise probability distributions only when the dependence function or copula (Nelsen 1999) is completely specified.

The approach of Williamson and Downs (1990) includes a way to rigorously represent continuous probability distributions by using outward-directed rounding on finitely many interval discretizations of the bounds on cumulative distribution functions. Bounding in the cumulative domain, rather than the density domain, allows representation error to be completely contained and propagated using the same algorithms that handle mathematical combinations. When operations on the interval discretizations are handled with interval analysis, the method constitutes verified computation for probability distributions.

The methods of probability bounds analysis are available in several software implementations, including multiple free demonstration programs (e.g., Berleant and Zhang 2004), a full-featured stand-alone commercial program (Ferson 2002), an advanced add-in for Microsoft Excel developed for NASA (Ferson et al. 2011), and a package in development for the statistical computing language R (R Development Core Team 2010).

4   Correlations and dependencies

Independence can be a dangerous assumption for analysts to make. Stochastic dependence is far more pervasive—and important—than many analysts seem to recognize. For instance, placing backup generators side by side makes their failure probabilities dependent and reduces the redundancy they were intended to provide because they become susceptible to the common-cause failure from flooding, as was realized too late in New Orleans and Fukushima. Even in relatively sophisticated analyses of uncertainty, the most common assumption about the dependence among variables is independence, although there may be no actual empirical evidence or serious theoretical justification to support this assumption. In truth, despite warnings about falsely assuming statistical independence, some analysts routinely ignore correlations for the sake of computational convenience. And conscientious analysts who would like to include them in analyses are often

stymied by the difficulty of measuring correlations when data are sparse. As a consequence, correlations are commonly omitted from analyses and the default assumption of independence is used even when there is no evidence whatsoever in support of this assumption.

Although central tendencies may be generally insensitive to correlations of small to moderate strength (Smith et al. 1992), the *tails* of distributions can be extremely sensitive to even small or moderate correlations. Of course, decision makers are often especially concerned with these tails. They represent the risks of extreme events, which might be the probability of some mechanical stress exceeding the engineered strength intended to resist it, or the probability of exposing people to large doses of a carcinogen, or perhaps the risk of extinction for an endangered species. It is these extreme adverse events in the distribution tails that are often the whole focus of the analysis, so it may be very important that the tail probabilities not be underestimated. Unfortunately, the common practice of assuming independence among all input variables can lead directly to such underestimations.

Moreover, many analysts seem to be unaware that consideration of correlation is only the tip of the iceberg. The issue of dependence is much broader than correlation because it includes all nonlinear relationships. This is the reason, of course, that lack of correlation does not guarantee independence. What we might call linear dependence, which can be fully characterized by a single correlation coefficient, is only a small subspace of the forms of statistical dependence. Consequently, it is impossible to use a sensitivity study to characterize the effect of uncertainty about dependence on an uncertainty projection. Varying correlations, even all the way from +1 to –1, over a single family of dependence functions does not come close to capturing the diversity of possible interactions the variables may have. This is similar to supposing that one has characterized the variability of *all possible functions* through a point simply by representing *all linear functions* through a point. (And there is no analog of Taylor's theorem for dependence, so the mistake is severe on all scales.) We note that no popular software packages for probabilistic calculations support more than a single family of dependence functions, if they support intervariable dependence at all.

Makarov (1981) and Frank et al. (1987) showed, however, that it is possible to compute bounds on results of probabilistic calculations no matter what correlations or statistical dependencies may exist among the variables. The algorithms of Williamson and Downs (1990) include this no-assumptions case. The top, left-hand graph of Fig. 4 shows an example calculation. In this example, X and Y are random variables each drawn from uniform distributions between 1 and 25. Any distribution of the sums $X + Y$ that could result from adding these uniformly distributed random values together must lie entirely inside the p-box shaped like a parallelogram ranging between 2 and 50. This does not mean of course that any distribution within the bounds could be the sum of these two distributions, but the bounds are pointwise best-possible, which means the black region could not be any smaller without excluding some distributions that could arise as sums of these two uniform variables. This elementary calculation shows

that the probability that $X + Y$ is smaller than 10 could be as high as one third, or as low as zero. This is a very different characterization of the distribution tail than what comes from a conventional Monte Carlo analysis that falsely assumes independence which, in this case, would suggest the chance the sum is less than 10 is about 5%. The PBA result makes no false assumptions about independence because it makes no assumptions at all about dependence, which potentially makes it very useful in applications such as risk analysis where it is critical not to underestimate tail risks.
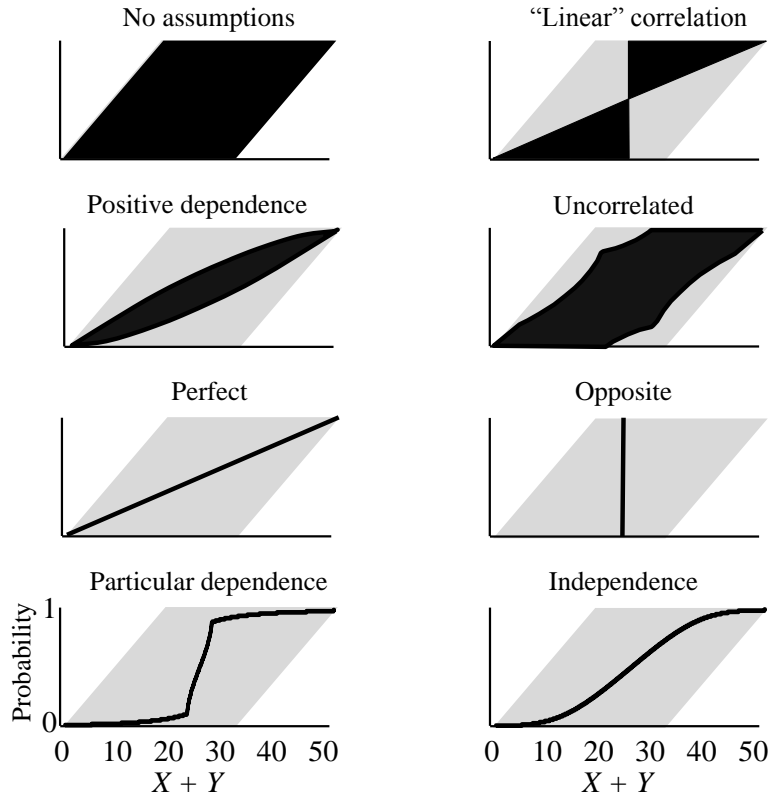
Fig. 4. The effect of various assumptions about the dependence between X and Y on the convolution $X + Y$, where X and Y are both uniformly distributed over the range $[1, 25]$. For comparison, the no-assumptions p-box is shown in gray.

The simple parallelogram shape of the example result is a consequence of the uniformity of the marginal distributions and the simplicity of the addition function that combines them. The algorithms can be applied equally well to virtually any finite distributions, including theoretically infinite distributions such as the normal that are truncated to some practical range, and extend beyond ad-

dition to all the basic mathematical operations. Surprisingly, the algorithms for the no-assumptions cases are computationally less expensive than Monte Carlo simulation methods that assume independence. More importantly, Monte Carlo methods cannot generally compute these no-assumptions bounds, no matter how many replications are used. Strategies that vary the correlation coefficient in a Monte Carlo sensitivity study will be limited to outputs like the cone shown in the top, right graph of Fig. 4, which grossly understates the uncertainty including possible tail risks.

Partial knowledge about the dependence can tighten the output p-box substantially (Ferson et al. 2004). For example, the left-hand graph in the second row of Fig. 4 shows the bounds on the distribution than can be inferred from the qualitative knowledge that the dependence function has positive sign. It is also possible to compute bounds on the result implied by a given correlation coefficient, which might have been reported among summary statistics published without original data. Berleant and Goodman-Strauss (1998) showed how the bounds in this case can be computed using mathematical programming. The right-hand graph of the second row of Fig. 4 shows an example with correlation zero. Remarkably, the resulting p-box almost fills up the no-assumptions parallelogram, which confirms the intuition that knowing only that variables are uncorrelated actually tells us very little about their possible dependence. In fact, the bounds are widest when the correlation is close to zero. When the correlation is extremal, corresponding to the third row of graphs in the figure, and the dependence is either perfect (comonotonic) or opposite (countermonotonic), then the resulting p-box becomes very tight. In the case of the example with precise addends, the result is also a precise distribution. This is true whenever the dependence function (copula) is precisely specified, as when it is given by some particular function such as that yielding the result shown in the bottom, left graph, or when it is taken to be independence as in the bottom, right graph. If the addends had been p-boxes rather than precise uniform distributions, the lower four graphs would also depict p-boxes rather than distributions.

As required in verified computation, all of these outputs are rigorous so they are sure to enclose the true output distribution even when the dependence is not precisely known, and their representations are pointwise best-possible so they could not be any tighter without excluding some possible distributions. It is possible to mix and match dependence assumptions within an analysis, assuming independence where it really is appropriate and making weaker assumptions where it is not.

5   A calculus for uncertainty beyond probability

It is possible to conceive of probability bounds analysis as an entirely traditional application of sensitivity analysis for conventional probability theory. Indeed, the same can be said of the whole of the theory of imprecise probability. Although this conception should therefore be acceptable even to strict Bayesians, PBA and IP are sometimes met with distrust by probabilists. The reason is that they

recognize in it an essential heterodoxy: that there exists a kind of uncertainty that should not be characterized by a unique probability measure, which is a notion they perceive as wrong, or subversive.

There are many misconceptions about what justifies what we might call the precisionist school of probability. Some people have argued that probability theory is the only logically possible calculus of uncertainty (e.g., Lindley 1982). Lindley (2006, 71) asserted, "Whatever way uncertainty is approached, probability is the *only* sound way to think about it." Neapolitan (1992) pointed out, however, that the arguments about the inevitability of probability theory say nothing to deny the utility of interval bounding. Some in the precisionist school mention a 'proof' due to Cox (1946), but Cox's theorem has been proven to be incorrect (Halpern 1999), and perhaps more importantly it has been shown to be irrelevant to the question of whether probability is the only possible calculus for uncertainty because it uses as assumptions the very questions that are at stake in the debate (Colyvan 2004).

Some critics of p-boxes and imprecise probabilities in general argue that precisely specified probability distributions are in any case *sufficient* to characterize uncertainty of all kinds. These critics argue that it is therefore meaningless to talk about 'uncertainty about probability' and that traditional probability is a complete theory. Under this criticism, users of p-boxes have simply not made the requisite effort to identify the appropriate precisely specified distribution functions. This argument may have been reasonable when probability distributions were only used to characterize uncertain scalar values, that is, real numbers. In modern uncertainty analyses, however, the objects of study are often themselves probability distributions deployed in risk assessments already involving aleatory uncertainty. In such cases, a richer characterization of uncertainty seems useful.

Other people think that requiring inferences to be consistent and coherent is logically equivalent to requiring the use of precise probabilities. This is not the case, as was suggested in the expansive review of the axiomatics of decision theory under subjective probability by Fishburn (1986). Walley (1991) showed how the theory of imprecise probabilities enjoys the same properties of consistency and coherence championed by probabilists as the definition of rationality, and how proper inferences in the IP context also avoid sure losses from Dutch books just as coherent Bayesian decisions do.

Luce (1992) noted that a major finding in the theory of subjective expected utilities is that humans do not make decisions according to the theory of subjective expected utilities. One of the reasons for this is that a traditional probabilist must always be able to discern which of two events is more probable, unless they are equally probable, and which of two options is preferable, unless she is indifferent to the choice between them. The axiom that this is always possible is called "completeness" by Fishburn (1986), or the "ordering postulate" by Seidenfeld (1988). A relaxed, more general theory of uncertainty recognizes that some events or options may be *incomparable* and does not demand an agent be able to distinguish between any two, even beyond indifference. The notion that one might not be able to compare every two probabilities dates back at least to

Keynes (1921), and it is another of his great ideas that has remarkable salience today.

Getting rid of the completeness axiom—which some might argue is far from self-evident anyway—induces the theory of imprecise probabilities. In this case, the definition of the probability of an event can be operationalized as the interval between the highest buying price and the lowest selling price for a gamble that pays one dollar if the event occurs (but nothing otherwise). This generalizes de Finetti's notion of a fair price to an interval whenever the probability is uncertain. Traditional Bayesians in principle *must agree to either buy or sell any gamble at the same fair price.* Under a relaxed theory, an agent may elect to neither buy nor sell a gamble if the price is not sufficiently favorable. Of course, if one knows all the probabilities (and utilities) perfectly, then IP reduces to Bayes.

Seidenfeld (1988) holds that a relaxed theory based on imprecise probabilities can provide a unified treatment of group decisions, where Bayesians admit their theory does not apply. This is important because most engineering decisions cannot be described as personal decisions, but rather must be in line with collective verdicts by teams of collaborators. Under IP, their decisions can be rational and coherent so long as indecision is admitted occasionally.

A parallel to the status of imprecise probability today can be found in the non-Euclidean revolution in geometry nearly two hundred years ago (Bardi 2009). The basic question at the time was this: Given a line in a plane, how many parallel lines in the plane can be drawn through a point in the plane not on the line? This was the subject of Euclid's fifth axiom, which had prescribed the answer 'one'. For over twenty centuries in mathematics since Euclid, this was the only answer to the question. Although the answer 'one' was itself never in doubt, the fifth axiom had long been controversial because it did not seem to be self-evident in the same way the other axioms were.

Many geometers attempted to prove the fifth axiom as a consequence of the other Euclidean axioms, and, in the early 1800s, tried a proof by *reductio ad absurdum* in which they denied the fifth axiom and looked for logical inconsistencies. But they found no logical inconsistencies from denying the fifth axiom, and, almost accidentally, they developed non-Euclidean geometries in which the answer to the question about the number of parallels could also be either 'zero' or 'many' rather than 'one'.

The advent of non-Euclidean geometry created a tumult in mathematics. Because it rests on what seemed to many to be an obvious fallacy, it was controversial and maddening to some mathematicians. Some proponents of the new geometry were ignored or faced condescension, and others, including even the eminent Gauss, hid what they had realized for fear of ridicule (Bardi 2009). Nevertheless, the new non-Euclidean geometries were in time accepted as legitimate and eventually heralded as greatly enriching the discipline of geometry by vastly expanding its scope and broadening its applications. Famously, Einstein used non-Euclidean geometry in his general theory of relativity.

We believe that probability theory may be in the early throes of a revolution similar in some ways to the non-Euclidean revolution in geometry. The debate today between the traditional precisionists and the imprecise probability community hinges on the adoption or rejection of a single axiom. Omitting that axiom relaxes the earlier, unnecessarily strict theory and opens it up to a richer perspective on a wider array of applications. The new theory, although an obvious generalization of the older theory, has met some strong and sometimes dismissive criticism. The advantages of the newer theory arise because of its greater flexibility, and its usefulness will likely be cemented by applications to significant problems beyond the reach of the older theory, even though the older theory persists as an important and still extremely widely used special case.

6   Conclusions

Given the increasing and critical use of Monte Carlo simulation and other probabilistic calculations, it is important to be able to assess the reliability of their numerical outputs as a function of the error in finite computer representation of probabilities and distributions, as well as the express uncertainty about the input distributions, their interdependencies, and model assumptions. Until recently, the only way to assess this reliability has been to develop an elaborate sensitivity study, which, because of its combinatorial complexity, is rarely even attempted. Consequently, in practice residual uncertainties about the selection of input probability distributions and the nature of interdependencies among the variables in an analysis are usually neglected, and the sufficiency of the numerical methods used in the calculations is not even addressed.

Practical computational methods in probability bounds analysis now exist to assess the potential impact on probabilistic calculations from (i) computer representation error of distributions, (ii) incomplete knowledge about the parameters and shapes of input distributions, (iii) imperfect understanding of the correlation and dependency structure among variables, and (iv) several kinds of model-form uncertainty. Probability bounds analysis is essentially an efficient and comprehensive form of sensitivity analysis that is compatible with both frequentist and Bayesian perspectives. Its methods allow many analysts to make routine and relatively inexpensive bounding estimates for calculations involving probabilities or probability distributions. In many cases, the bounds will be exact in the sense that they are the pointwise best-possible bounds. In all cases, the bounds will enclose the probability distributions and therefore provide a conservative expression of the reliability of the results.

The Panglossian view that every probabilistic calculation can be solved with precisely specified inputs determined from available information does not seem plausible in many cases, especially in new environments or novel situations with highly constrained scientific knowledge and limited engineering experience. When there is insufficient empirical information available to justify precise probabilistic calculations, the methods of probability bounds analysis naturally yield bounds on output distributions. Different input variables can be specified either

as particular distributions or as wide or narrow bounds on distributions as appropriate to represent the state of empirical information, whatever it may be, that is available about each variable. Likewise, dependence functions can be precisely modeled with a specific copula, qualitatively modeled with available information, or discharged entirely by making no assumptions about the dependence. The degree of specificity about the marginal distributions and dependencies does not affect how they are combined in subsequent arithmetic operations, and the result appropriately represents its own level of uncertainty, rather than a false precision gained by unjustified assumptions.

Sound scientific and engineering analyses ought to be based on objective, documented, and verifiable information. Whenever assumptions are used to make up the difference between what is empirically known and what is needed to obtain a working answer, those assumptions should be subjected to quantitative assessment with methods such as probability bounds analysis. Without such assessment, the calculations are unsound, and the resulting answers are merely wishful thinking.

References

1. Baase, S.: *A Gift of Fire.* Pearson Prentice Hall (2008)
2. Bardi, J.S.: *The Fifth Postulate: How Unraveling a Two-thousand-year-old Mystery Unraveled the Universe.* John Wiley & Sons, Hoboken, New Jersey (2009)
3. Beer, M.: Fuzzy probability theory. In: *Encyclopedia of Complexity and System Science*, volume 6, Meyers, R., (ed.), pp. 4047–4059. Springer, New York (2009)
4. Ben-Haim, Y.: *Information Gap Decision Theory: Decisions under Severe Uncertainty.* Academic Press, San Diego (2001)
5. Berger, J.O.: *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York (1985)
6. Berleant, D.: Automatically verified reasoning with both intervals and probability density functions. *Interval Computations* 1993(2): 48–70 (1993)
7. Berleant, D.: Automatically verified arithmetic on probability distributions and intervals. In: *Applications of Interval Computations*, Kearfott, B., Kreinovich, V. (eds.), pp. 227–244. Kluwer Academic Publishers, Dordrecht, The Netherlands (1996)

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

9. Berleant, D., Zhang, J.: Representation and problem solving with distribution envelope determination (DEnv). *Reliability Engineering and System Safety* 85: 153–168 (2004)

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

11. Boyle, P.P.: Options: a Monte Carlo approach. *Journal of Financial Economics* 4: 323–338 (1977)

12. Bukowski, J., Korn, L., Wartenberg, D.: Correlated inputs in quantitative risk assessment: the effects of distributional shape. *Risk Analysis* 15: 215–219 (1995)

13. Burgman, M., Ferson, S., Akçakaya, H.R.: *Risk Assessment in Conservation Biology.* Chapman & Hall, London (1993)

14. Burmaster, D.E., Harris, R.H.: The magnitude of compounding conservatisms in superfund risk assessments. *Risk Analysis* 13: 131–134 (1993)

15. CFTC/SEC [U.S. Commodity Futures Trading Commission / U.S. Securities and Exchange Commission]: *Findings Regarding the Market Events of May 6, 2010: Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues*, `http://www.sec.gov/news/studies/2010/marketevents-report.pdf` (2010)

16. Chebyshev [Tchebichef], P.: Sur les valeurs limites des intégrales. *Journal de Mathématiques Pures et Appliquées.* Ser 2, 19: 157–160 (1874)

17. Colyvan, M.: The philosophical significance of Cox's theorem. *International Journal of Approximate Reasoning* 37: 71–85 (2004)

18. Cooper J.A.: Fuzzy algebra uncertainty analysis for abnormal environment safety assessment. *Journal of Intelligent and Fuzzy Systems* 2: 337–345 (1994)

19. Cox, R.T.: Probability, frequency and reasonable expectation. *American Journal of Physics* 14: 1–13 (1946)

20. de Finetti, B.: *Theory of Probability.* Wiley, New York (1974)

21. Elliott, G.: US nuclear weapon safety and control. `http://web.mit.edu/gelliott/Public/sts.072/paper.pdf` MIT Program in Science, Technology, and Society (2005)

22. Enszer, J.A., Lin, Y., Ferson, S., Corliss, G.F., Stadtherr, M.A.: Probability bounds analysis for nonlinear dynamic process models. *AIChE Journal* 57: 404–422 (2011)

23. EPA [U.S. Environmental Protections Agency (Region I)]: *Human Health Risk Assessment*, `http://www.epa.gov/region1/ge/thesite/restofriver-reports.html#HHRA` (2005a)

24. EPA [U.S. Environmental Protections Agency (Region I)]: *Ecological Risk Assessment*, `http://www.epa.gov/region1/ge/thesite/restofriver-reports.html#ERA` (2005b)

25. EPA [U.S. Environmental Protections Agency (Region 6 Superfund Program)]: *Calcasieu Estuary Remedial Investigation*, `http://www.epa.gov/region6/6sf/louisiana/calcasieu/la_calcasieu_calcri.html` (2006)

26. ESA [European Space Agency]: Ariane 5 Flight 501 Failure. Report by the Inquiry Board, Paris, `http://esamultimedia.esa.int/docs/esa-x-1819eng.pdf` (1996)

27. Ferson, S.: *RAMAS Risk Calc 4.0 Software: Risk Assessment with Uncertain Numbers.* Lewis Publishers, Boca Raton, Florida (2002)

28. Ferson, S., Burgman, M. (eds.): *Quantitative Methods for Conservation Biology.* Springer-Verlag, New York (2000)

29. Ferson, S., Tucker, W.T.: Probability boxes as info-gap models. In: *Proceedings of the North American Fuzzy Information Processing Society*, IEEE, New York City (2008)

30. Ferson, S., Kreinovich, V., Ginzburg, L., Sentz, K., Myers, D.S.: *Constructing probability boxes and Dempster–Shafer structures*. Sandia National Laboratories, SAND2002-4015, Albuquerque, New Mexico, `http://www.ramas.com/unabridged.zip` (2003)

31. Ferson, S., Nelsen, R., Hajagos, J., Berleant, D., Zhang, J., Tucker, W.T., Ginzburg, L., Oberkampf, W.L.:. *Dependence in Probabilistic Modeling, Dempster–Shafer Theory, and Probability Bounds Analysis*. Sandia National Laboratories, SAND2004-3072, Albuquerque, New Mexico, `www.ramas.com/depend.pdf` (2004)

32. Ferson, S., Mickley, J., McGill, W.: Uncertainty arithmetic on Excel spreadsheets: add-in for intervals, probability distributions and probability boxes. *Vulnerability, Uncertainty, and Risk: Analysis, Modeling, and Management, Proceedings of the ICVRAM 2011 and ISUMA 2011 Conferences*, ASCE Reston, Virginia (2011)

33. Fishburn, P.C.: The axioms of subjective probability. *Statistical Science* 1: 335–358 (1986)

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

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

36. Fréchet, M.: Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université de Lyon. Section A: Sciences mathématiques et astronomie* 9: 53–77 (1951)

37. GAO [General Accounting Office]: *Patriot Missile Defense: Software Problem Led to System Failure at Dhahran, Saudi Arabia*. GAO/IMTEC-92-26, `http://www.fas.org/spp/starwars/gao/im92026.htm` (1992)

38. Grove, A.J., Halpern, J.Y.: Updating sets of probabilities. In: *Proceedings of the Fourteenth Conference on Uncertainty in AI*, pp. 173–182, `http://arxiv.org/abs/0906.4332` (1998)

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

40. Halpern, J.Y.: Cox's theorem revisited. *Journal of AI Research* 11: 429–435 (1999)

41. Hammer, R., Hocks, M., Kulisch, U., Ratz, D.: *C++ Toolbox for Verified Computing I: Basic Numerical Problems*. Springer, Heidelberg (1997)

42. Hertz, D.B.: Risk analysis in capital investment. *Harvard Business Review* 42: 95–106 (1964)

43. Hickman, J.W. et al.: *PRA Procedures Guide: A Guide to the Performance of Probabilistic Risk Assessments for Nuclear Power Plants*, in two volumes. NUREG/CR-2300-V1 and -V2. National Technical Information Service, Washington (1983)

44. Isbell, D., Hardin, M., Underwood, J.: Mars Climate Orbiter team finds likely cause of loss. `http://mars.jpl.nasa.gov/msp98/news/mco990930.html` (1999)

45. Jeffreys, H.: *Scientific Inference*. Cambridge University Press, Cambridge (1931)

46. Kaufmann, A., Gupta, M.M.: *Introduction to Fuzzy Arithmetic: Theory and Applications*. Van Nostrand Reinhold Company, New York (1991)

47. Keynes, J.M.: *A Treatise on Probability*. Macmillan, New York, `http://www.archive.org/details/treatiseonprobab007528mbp` (1921)

48. Kolmogorov, A.N.: *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Julius Springer, Berlin (1933). Translated into English as Kolmogorov, A.N.: *Foundations of the Theory of Probability*. Chelsea Publishing, New York (1950)

49. Kriegler, E., Held, H.: Utilizing belief functions for the estimation of future climate change. *International Journal of Approximate Reasoning* 39: 185–209 (2005)
50. Kulisch, U., Hammer, R., Ratz, D., Hocks, M.: *Numerical Toolbox for Verified Computing I: Basic Numerical Problems: Theory, Algorithms, and Pascal-XSC Programs.* Computational Mathematics, Volume 21, Springer, Heidelberg (1993)
51. Kyburg, H.E., Jr.: Interval valued probabilities. *SIPTA Documention on Imprecise Probability*, http://www.sipta.org/documentation/interval_prob/kyburg.pdf (1999)
52. Laplace, Marquis de, P.S.: *Théorie analytique de probabilités* (edition troisième). Courcier, Paris (1820). The introduction (*Essai philosophique sur les probabilités*) is available in an English translation in *A Philosophical Essay on Probabilities*, Dover Publications, New York (1951)
53. Lavine, M.: Sensitivity in Bayesian statistics: the prior and the likelihood. *Journal of the American Statistical Association* 86: 396–399 (1991)
54. Lindley, D.V.: Scoring rules and the inevitability of probability. *International Statistical Review* 50: 1–26 (1982)
55. Lindley, D.V.: *Understanding Uncertainty.* John Wiley & Sons, Hoboken, New Jersey (2006)
56. Luce, R.D.: Where does subjective expected utility fail descriptively? *Journal of Risk and Uncertainty* 5: 5–27 (1992)
57. Makarov, G.: Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory of Probability and its Applications* 26: 803–806 (1981)
58. Markov [Markoff], A.: Sur une question de maximum et de minimum proposée par M. Tchebycheff. *Acta Mathematica* 9: 57–70 (1886)
59. McKone, T.E., Ryan, P.B.: Human exposures to chemicals through food chains: an uncertainty analysis. *Environmental Science and Technology* 23: 1154–1163 (1989)
60. Möller, B., Beer, M.: *Fuzzy Randomness—Uncertainty in Civil Engineering and Computational Mechanics.* Springer, Berlin (2004)
61. Moore, R.E.: *Interval analysis.* Prentice Hall, Englewood Cliffs, New Jersey (1966)
62. Neapolitan, R.E.: A survey of uncertain and approximate inference. In: *Fuzzy Logic for the Management of Uncertainty*, Zadeh, L., Kacprzyk, J., (eds.), pp. 55–82. John Wiley & Sons, New York (1992)
63. Nelsen, R.B.: *An Introduction to Copulas.* Lecture Notes in Statistics 139, Springer-Verlag, New York (1999)
64. Nong, A., Krishnan, K.: Estimation of interindividual pharmacokinetic variability factor for inhaled volatile organic chemicals using a probability-bounds approach. *Regulatory Toxicology and Pharmacology* 48: 93–101 (2007)
65. Popova, E.: Mathematica connectivity to interval libraries filib++ and C-XSC. In: Cuyt, A., et al. (eds.), *Numerical Validation.* LNCS, vol. 5492, pp. 117–132. Springer, Heidelberg (2009)
66. R Development Core Team: *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org (2010)
67. Seidenfeld, T.: Decision theory without "independence" or without "ordering": what is the difference? *Economics and Philosophy* 4: 267–290 (1988)
68. Seidenfeld, T., Wasserman, L.: Dilation for sets of probabilities. *The Annals of Statistics* 21: 1139–1154 (1993)
69. Selby, R.G., Vecchio, F.J., Collins, M.P.: The failure of an offshore platform. *Concrete International* 19: 28–35 (1997)

70. Slabodkin, G.: Smart ship inquiry a go. *Government Computer News* (August 31). `http://gcn.com/articles/1998/08/31/smart-ship-inquiry-a-go.aspx` (1998)
71. Smith, A.E., Ryan, P.B., Evans, J.S.: The effect of neglecting correlations when propagating uncertainty and estimating the population of risk. *Risk Analysis* 12: 467–474 (1992)
72. Suter, G.W., II: *Ecological Risk Assessment.* Lewis Publishers, Boca Raton (1993)
73. Tucker, W.: *Validated Numerics: A Short Introduction to Rigorous Computations.* Princeton University Press, Princeton, New Jersey (2011)
74. Tuyl, F., Gerlachy, R., Mengersen, K.: Posterior predictive arguments in favor of the Bayes-Laplace prior as the consensus prior for binomial and multinomial parameters. *Bayesian Analysis* 4: 151–158 (2009)
75. Vick, S.G.: *Degrees of Belief: Subjective Probability and Engineering Judgment.* American Society of Civil Engineers, Reston, Virginia (2002)
76. Walley, P.: *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London (1991)
77. Walley, P.: Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B* 58: 3–57 (1996)
78. Walley, P., Gurrin, L., Barton, P.: Analysis of clinical data using imprecise prior probabilities. *The Statistician* 45: 457–485 (1996)
79. Williamson, R.C., Downs, T.: Probabilistic arithmetic I: numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning* 4: 89–158 (1990)
80. Yager, R.R.: Arithmetic and other operations on Dempster–Shafer structures. *International Journal of Man-Machine Studies* 25: 357–366 (1986)
81. Zhang, H., Mullen, R.L., Muhanna, R.L.: Finite element structural analysis using imprecise probabilities based on p-box representation. *Proceedings of the 4th International Workshop on Reliable Engineering Computing: Robust Design—Coping with Hazards, Risk and Uncertainty.* 3–5 March 2010, Singapore, Beer, M., Muhanna, R.L. Mullen, R.L., (eds.). Research Publishing, `http://rpsonline.com.sg/rpsweb/`. `http://www.eng.nus.edu.sg/civil/REC2010/documents/papers/013.pdf` (2010)
82. Zhang, H., Mullen, R., Muhanna, R.: Safety structural analysis with probability-boxes. *International Journal of Reliability and Safety* 6: 110–129 (2012)

DISCUSSION

Speaker: Scott Ferson

Jon Helton : What is the current status of computational capability to propagate p-boxes through complex models, e.g., long running and/or containing repeated variables?

Scott Ferson : I would say it is significantly better than the computational capability currently available for propagating Dempster–Shafer (DS) structures, and in some instances, better than that for Monte Carlo simulations. There are several reasons for this. Firstly, as a bounding approach it can effectively propagate some kinds of uncertainties that cannot be comprehensively addressed by sampling approaches even with infinitely many samples. For instance, if an analyst does not know the distribution family for some input, she can use a distribution-free p-box that bounds all possible distribution families consistent with the other information available about that variable. Zhang et al. (2012) noted that the computational burden for p-box propagation when no assumptions are made about intervariable dependencies is actually smaller than even that for simple Monte Carlo simulation, because it does not require a full convolution of all possible combinations of values for the various input variables.

There are also a variety of tricks and shortcuts available for p-box calculations, including and extending the conjugacy rules familiar to Bayesians. If a problem is computationally challenging when p-boxes are used to characterize the inputs, one or more of the p-boxes can be coarsened in a way that preserves conservativism yet radically lessens the computational burden. This coarsening is also possible with DS structures, but it will alter the internal features of the output uncertainty structure, the elucidation of which is the whole point of using DS structures in the first place.

The computational capability to propagate p-boxes can be judged by its practical applications. Probability bounds analysis has been used in cases with many dozens of inputs, although I have seen no practical applications yet with many hundreds of inputs. It has been used in uncertainty analyses of substantial complexity in a wide variety of contexts ranging from Superfund human health and ecological risk assessments (EPA 2005a; 2005b; 2006) to finite element models in engineering (Oberguggenberger et al. 2007; Zhang et al. 2010; 2012), and on scales from lab bench chemistry and pharmacokinetics (Enszer et al. 2011; Nong and Krishnan 2007) to the planet's climate in global circulation models (Kriegler and Held 2005).

Certainly there are challenges in computing with p-boxes, especially concerning the appearance of repeated uncertain quantities, and developing strategies to meet these challenges is a current area of research. There are no particular computational challenges associated with "long-running" calculations per se, and p-box operations can be applied iteratively in a straightforward way. However, in cases where sequential iterations involve repeated uncertain quantities such as in calculating solutions to differential equations, difficulties can arise. Enszer

et al. (2011) described special methods to solve nonlinear ordinary differential equations with parameters or initial conditions expressed as p-boxes. This is another area where what we can do computationally with p-boxes exceeds what can be done with Monte Carlo simulation which quickly runs into instability in such problems.

William Kahan : Your acceptance of diverse bounds upon probabilities (wherever those bounds may come from) reminds me of the "degrees of belonging" of fuzzy sets and Lotfi Zadeh's fuzzy logic; but you do not mention fuzzy sets at all. Why not?

Scott Ferson : The theory I'm talking about here is purely probabilistic and conforms with the Kolmogorov (1933) axioms. Of course the bounds can come from many places just because there are many sources of information and data and many disparate reasons for uncertainty, but the quantities we work with are bounds on probabilities which are interpreted in only one way, as Kolmogorov probabilities.

I have considered fuzzy sets and possibility theory in the past, and many colleagues still use these ideas extensively (Möller and Beer 2004; Beer 2009). I don't use them now myself mostly because they evoke such a very negative reaction among probabilists. I do not think the visceral reaction from probabilists is legitimate at all, but this is not my battle. My objection to fuzzy numbers and their arithmetic as proposed by Kaufmann and Gupta (1991) is that there is still no way to ensure that the result will be meaningful from level-wise combining fuzzy numbers that came from different formulations with distinct possibility scales. This is the same objection I have for possibility theory and, by the way, for info-gap decision theory (Ben-Haim 2001).

Pasky Pascual : How does one use p-boxes to formulate priors within a Bayesian framework?

Scott Ferson : The most common way is to use p-boxes to characterize an analyst's uncertainty about the appropriate prior to employ. Consider, for example, the problem of estimating a binomial probability which is perhaps the most elementary and fundamental problem in all of risk or uncertainty analysis. Amazingly, Bayesians cannot agree on the prior to use for this problem, even in the basic case when they all agree no relevant prior information is available (Tuyl et al. 2009; Berger 1985, page 89). The search for a so-called "uninformative prior" has produced several candidates, including Haldane's improper prior beta(0,0), Jeffreys' reference prior beta($\frac{1}{2}$,$\frac{1}{2}$), the uniform distribution favored by Laplace which can be modeled as beta(1,1), and other distributions such as Zellner's binomial prior (which is not from the beta family). Unless the sample size is pretty large, which is rare in many practical situations, these different priors yield noticeably different results.

Peter Walley (1991; Walley et al. 1996) has suggested using an imprecise beta model (IBM) which is effectively a p-box of all beta distributions that could be good priors. In the degenerate case, when the sample size is zero, the IBM yields

a vacuous posterior that says the probability could be anywhere in the interval [0,1]. Isn't that a reasonable result for an analysis that *uses no data at all*? When the sample size is very large, the posterior is a tight p-box that tends to the observed frequency, as most all Bayesian analyses do. In the practical intermediate cases of small sample sizes, the posterior from the IBM is a p-box containing a range of beta distributions whose breadth reflects the uncertainty about the prior that a traditional Bayesian analysis ignores. Contrary to what Tony O'Hagan suggests in his comment below, this breadth is not too wide to be useful, but yields answers whose imprecision is roughly what one might expect to see across a community of competent Bayesians.

The imprecise beta model generalizes in the multivariate case to an imprecise Dirichlet model (IDM, Walley 1996). The IBM and IDM are examples of Bayesian sensitivity analysis (Lavine 1991) or robust Bayes analysis (Berger 1985), the idea of which originated with Jeffreys (1931) and de Finetti (1974). Walley (1991) has demonstrated that robust Bayes analysis is part of a more general theory based on imprecise probabilities of very broad scope and flexibility, for which there is a firm theoretical foundation based on respecting consistency and coherence requirements but which avoids making unwarranted assumptions to obtain quantitative answers. Probability bounds analysis is a computationally convenient method within this general theory

Michael Goldstein : While I agree that notions of imprecision have a valuable role in considering uncertainty, I am not convinced that the approach advocated by the speaker can be viewed as a complete theory. In particular, it is quite possible for the analyst to face a situation where there is available data which, for every possible outcome, will increase uncertainty about some key quantity in such a way that the information has negative value to the analyst. Effectively, this turns the analyst into a money-pump, as the analyst appears to need to keep paying to avoid receiving the data. This is counter-intuitive to me, and makes me uncomfortable about the notion that the theory is sufficient to deal with all uncertainty models.

Scott Ferson : Michael is talking about a phenomenon described by Seidenfeld and Wasserman (1993) known as dilation. It occurs when new evidence leads different Bayesian investigators into greater disagreement than they had prior to their getting the new evidence. Such evidence is not merely surprising in the sense that it contradicts one's prior conceptions; it expands everyone's uncertainty. It is counterintuitive because it does not depend on what the new information is actually saying. Michael's criticism is perhaps the pot calling the kettle black, because dilation occurs among Bayesians too. They simply don't recognize it because Bayesians don't have to agree with each other (or with the world for that matter).

It's hard to explain dilation with a simple example, but let me try. Suppose Lucius Malfoy tosses a fair coin twice, but the second 'toss' depends on the outcome of the first toss. It could be that Malfoy just lets the coin ride, and the second outcome is exactly the same as the first outcome. Or he could just

flip the coin over so that the second outcome is the opposite of the first. You don't know what he will do. The outcome of the first toss is either heads H1 or tails T1. Because the first toss is fair (and no spells are cast midair), you judge the probability P(H1) = 0.5. Whether Malfoy lets the coin ride or flips it, you judge the probability the second 'toss' ends up heads to be the same, P(H2) = 0.5. What happens when you see the outcome of Malfoy's first toss? Suppose it was a head. What is your probability now that the second 'toss' will also be a head? It turns out that once you condition on the first observation, the probability of the second toss being a head dilates. It is now either zero or one, but you don't know which. It doesn't depend on chance now; it depends on Malfoy's choice, about which you have no knowledge (unless perhaps you too dabble in the dark arts). Dilation occurs because the observation H1 has caused the earlier precise unconditional probability P(H2) = 0.5 to devolve into the vacuous interval P(H2 | H1) = [0,1].

This issue may be a pretty esoteric theoretical concern. I have yet to see examples of dilation in practice that would create any problems for analysts or decision makers. Although dilation seems highly counterintuitive to some people, others consider it a natural consequence of the interactions of partial knowledge (Walley 1991, 298f). My attitude is far from Michael's worry that the theory of imprecise probabilities might somehow be incomplete because it recognizes this phenomenon. Instead, I think it is rather evidence of its being a *richer* theory.

One way to avoid dilation is not to use conditionalization as the updating rule for new information. Interestingly, it is possible to do this with imprecise probabilities. Grove and Halpern (1998) point out that the standard justifications for conditionalization may no longer apply when we consider sets of probabilities. And it may turn out that conditionalization may not be the most natural way to update sets of probabilities in the first place. Instead, a constraint-based updating rule may sometimes be more sensible. We note that dilation does not occur in interval analysis (Seidenfeld and Wasserman 1993), which is a kind of constraint analysis.

Anthony O'Hagan : First I would like to thank Scott for an entertaining presentation. Unfortunately, much of what he says is in my opinion wrong– entertainingly wrong, but nonetheless *wrong*. In these comments I would just like to pick out what seem to me to be the two most important errors.

First he repeatedly confuses epistemic uncertainty with what he calls incertitude. Epistemic uncertainty relates to a quantity that has a unique (albeit unknown) value but which is not random in the usual sense of that word. In particular, we cannot observe a series of repetitions or 'trials' and so its uncertainty cannot be described by the conventional relative frequency form of probability. Bayesian statistics quantifies epistemic uncertainty with subjective probabilities. Frequentist statistics cannot do this because it only acknowledges relative frequency probability, so its "quantifications" of epistemic uncertainty are oblique, using such convoluted devices as confidence intervals.

What Scott calls incertitude is not completely clear, but I think I can define it as those things that he would represent using intervals of probabilities

or p-boxes. His primary idea of a p-box is to express incertitude about a probability distribution. In his examples, those distributions are often conventional frequency probability distributions (for quantities with aleatory uncertainty or randomness), but he also discuss putting p-boxes on Bayesian prior and posterior distributions, which relate to epistemic uncertainty.

To my mind, his incertitude is an attempt to do something that is perfectly sensible, namely to quantify in some way the fact that when we specify probability distributions (whether they be aleatory sampling distributions or epistemic prior distributions, for instance) we can never do so precisely. All judgments are imprecise, and probability distributions are nearly always specified partly by convenience. So what he calls incertitude seems to me to be addressing imprecision in judgments. And that's an important issue.

My second point, however, is that I have reservations about whether p-boxes and interval arithmetic are the way to handle incertitude. His approach is a kind of half-way house between formal treatment of imprecision with a second-order (or hierarchical) probability quantification on the one hand, and on the other a purely informal 'sensitivity analysis' in which we simply explore a few possible alternative distributions. As such, it is certainly more comprehensive than sensitivity analysis and may indeed have a role to play. However, it requires one to specify bounds, and these bounds are almost always arbitrary. If set very wide so as to be quite sure of encompassing whatever the 'true' distributions might be, then the resulting derived bounds on quantities of interest and decisions are likely to be hopelessly wide. If set narrower so as to encompass just the more likely 'true' distributions, then he can no longer claim that the p-boxes are exhaustive.

Despite these reservations, I welcome the basic idea of p-boxes and interval arithmetic. I just wish that Scott would not oversell it. These are *not* tools for quantifying epistemic uncertainty (although they may have a role in addressing the imprecision in subjective distributions that *do* quantify epistemic uncertainty). I might also add that they have nothing to do with utility theory or prospect theory.

Scott Ferson : Tony seems to want to dismiss the presentation as wrong, wrong, *wrong*, yet he agrees that my concern with incertitude is "perfectly sensible" and that it is an "important issue". Moreover he concedes that p-boxes "may indeed have a role to play" and "welcome[s] the idea of p-boxes". So let us try to clarify the sources and details of our differences.

The first of two disagreements that Tony highlights is my use of the phrase 'epistemic uncertainty'. I'm not at all sure why Tony insists that a quantity about which we are epistemically uncertain must be a unique, fixed quantity. There is nothing in the definition of the phrase that *requires* the quantity to be fixed underneath all our uncertainty about it. A quantity might be a fixed value, but it also might not be, and indeed our epistemic uncertainty about a quantity might often include whether it is in fact fixed or varying. Whether we know it is fixed or not, we could still have epistemic uncertainty about it.

It is true that Bayesians use probability distributions to model epistemic uncertainty, but it is simply not true that these objects represent the only possible way to model epistemic uncertainty. Clearly, intervals and p-boxes are general and flexible tools to quantify and propagate epistemic uncertainty, the latter specifically designed to do so, even when the nature of the underlying quantity is itself unknown. PBA can also handle Tony's case where the unknown quantity does have some unique fixed (but unknown) value. A p-box conveniently represents this case as extra information that the quantity's variance is zero. This is a special case compared to the general situation in which the possible range of the variance includes zero. Knowing a p-box's variance is exactly zero may have implications for the left and right bounds of the p-box, and it will usually have implications for mathematical results that depend on the p-box. You might also know the quantity must vary, in which case the p-box's variance might exclude zero as a possible value, even though you may not know its distribution precisely. Knowing a minimum range for the quantity or knowing a lower bound on its variance can improve the p-box and calculations that depend on it.

I do not know the origin or purpose of Tony's restrictive definition of 'epistemic uncertainty'. Our view is more expansive, and perhaps more useful. In addition to our not requiring the underlying value to be fixed, our usage of the phrase need not refer to a *quantity* at all, but includes uncertainty about the mathematical form of the model, which can also be captured in probability bounds analysis.

The second of Tony's two complaints is about whether intervals and p-boxes are a practical way to handle incertitude (epistemic uncertainty). He asserts that the bounds of p-boxes are "almost always arbitrary" and suggests that setting them very wide to be sure to encompass the true distribution will make the results vacuous, and narrowing them will lose the claim that p-boxes are comprehensive. In fact, however, there are many ways to construct p-boxes, and many of these ways constitute constraint analyses that are best-possible and in no way arbitrary. They don't even depend on parameters that might be varied arbitrarily. The subsequent calculations are also essentially constraint analyses that include no arbitrariness.

There are other ways to construct p-boxes that do involve decisions by the analyst that might have to be made arbitrarily. For instance, picking the confidence level in a p-box defined as a confidence band. Most analysts consider these decisions to be part of the modeling task and therefore to be the responsibility of the analyst, as they are in many exercises involving modeling or analysis. There are various strategies to avoid arbitrariness including appealing to conventions such as Fisher's 0.05 level, or considering would-be arbitrary parameters to be part of the analysis by nesting p-boxes at different levels (Ferson and Tucker 2008) or enveloping all tenable levels. Whether the uncertainty overwhelms the analyst's ability to make decisions depends on the details of the application and the available empirical information. In practical cases, analysts have generally found that useful inferences and decisions can be obtained.

One further fundamental point should be emphasized in reaction to Tony's complaint. Uncertainty analysis shouldn't be a game. Analysts are invited to be honest in expressing what they know and what they don't know. If it turns out that so little is known about a system that the p-boxes characterizing it are wide to the point that the results are vacuous, then it seems to me that a proper uncertainty analysis should reveal this fact. It is, after all, the very point of an uncertainty analysis. The alternative—which is to squeeze *unwarranted* conclusions or decisions out of a tenuous model that is not actually supported by evidence—is of course possible, but does not seem desirable, especially in an engineering context.