SAND2015-4167J

Originally Published by
Sandia National Lagoratories as
SAND2004-3072, October 2004

Dependence in probabilistic modeling, Dempster-Shafer theory, and probability bounds analysis*

Scott Ferson[1], Roger B. Nelsen[2], Janos Hajagos[1], Daniel J. Berleant[3], Jianzhong Zhang[3], W. Troy Tucker[1], Lev R. Ginzburg[1] and William L. Oberkampf[4]

[1]Applied Biomathematics, 100 North Country Road, Setauket, New York 11733
[2]Mathematical Science, Lewis & Clark College, Portland Oregon 97219-7899
[3]Iowa State University, Electrical and Computer Engineering, Ames, Iowa 50011-3060
[4]Validation and Uncerstinay Estimation Department, MS 0828, Sandia National Laboratories, Albuquerque, New Mexico 87175-0828

Abstract

This report summarizes methods to incorporate information (or lack of information) about inter-variable dependence into risk assessments that use Dempster-Shafer theory or probability bounds analysis to address epistemic and aleatory uncertainty. The report reviews techniques for simulating correlated variates for a given correlation measure and dependence model, computation of bounds on distribution functions under a specified dependence model, formulation of parametric and empirical dependence models, and bounding approaches that can be used when information about the inter-variable dependence is incomplete. The report also reviews several of the most pervasive and dangerous myths among risk analysts about dependence in probabilistic models.

---

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

Acknowledgments

Discussions with Peter Walley were helpful in elucidating the problem of representing dependence in the context of imprecise probabilities and the role of probability boxes and Dempster-Shafer structures as models of imprecise probabilities. Discussions with Thomas Fetz of Universität Innsbruck and Gert de Cooman of Universiteit Gent were helpful in elucidating the details about when computational methods assuming different kinds of independence would be conservative. Other discussions with David S. Myers of Stony Brook University, Robert Clemen of Duke University, Vladik Kreinovich of the University of Texas at El Paso, and Roger Cooke of the Technische Universiteit Delft on various other related subjects were also very helpful. We thank Jon Helton, Floyd Spencer and Laura Swiler of Sandia National Laboratories and Jim Hall of the University of Bristol for reviewing the report and providing a number of helpful suggestions. The opinions expressed herein are solely those of the authors. Updates and corrections to this document after its printing may be available on-line at http://www.ramas.com/depend.zip.

Symbols

$\sim$	is distributed as
$\in$	is an element of
$\subseteq$	is a subset of
$\pm$	plus or minus
$+$, $-$, etc.	addition, subtraction, etc. under no assumption about the dependence between the operands
$\|+\|$, $\|-\|$, etc.	addition, subtraction, etc. assuming independence
$/+/$, $/-/$, etc.	addition, subtraction, etc. assuming perfect dependence
$\backslash+\backslash$, $\backslash-\backslash$, etc.	addition, subtraction, etc. assuming opposite dependence
$\varnothing$	the empty set, i.e., the set having no members
$[\underline{F}, \bar{F}]$	probability-box specified by a left side $\underline{F}(x)$ and a right side $\bar{F}(x)$ where $\underline{F}(x) \leq \bar{F}(x)$ for all $x \in \Re$, consisting of all non-decreasing functions F from the reals into [0,1] such that $\underline{F}(x) \leq F(x) \leq \bar{F}(x)$.
$\{(s_1,m_1),\ldots,(s_n,m_n)\}$	an enumeration of the elements of a Dempster-Shafer structure in terms of its focal elements s_i and their nonzero masses m_i
beta(v, w)	a beta distribution with shape parameters v and w
convolve(X,Y,r)	convolution (usually addition) assuming that X and Y have correlation r
convolve(X,Y,C)	convolution (usually addition) assuming the copula C describes the dependence between X and Y
cov(X, Y)	covariance between random variables X and Y
E(X)	expectation (mean) of random variable X
$f{:}A \rightarrow B$	a function f whose domain is the set A and whose range 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)
$M(u, v)$	the copula min(u,v) corresponding to perfect dependence
normal(μ, σ)	a normal distribution with mean μ and standard deviation σ
$\Pi(u, v)$	the copula defined by the product uv corresponding to independence
P(A)	probability of event A
P(A|B)	probability of event A given that event B has occurred
P(A&B)	probability of conjunction (and) of events A and B
P($A \vee B$)	probability of disjunction (or) of events A and B
$\rho_{C,L}(F, G)$	infimal convolution for binary operation L between distribution functions F and G related to each other by the copula C
$\Re$	the set of all real numbers
$\Re^+$	the set of all non-negative real numbers
sgn	sign or signum function

sup	supremum (for a finite set, simply the maximum)
$\tau_{C,L}(F, G)$	supremal convolution for binary operation L between distribution functions F and G related to each other by the copula C
uniform(a, b)	a uniform distribution ranging between a and b, where $a \leq b$
V(X)	variance of random variable X
$W(u, v)$	the copula $\max(u+v-1,0)$ corresponding to opposite dependence
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 analysis and other applications of probabilistic modeling generally require specification of the joint multivariate distribution of the random variables involved in the problem. In practice, the input for the mathematical model used in a probabilistic risk assessment is usually constructed in two steps: (1) the marginal distributions for the input variables are specified and (2) the dependence between the variables is specified. The second step is perhaps just as important as the first, but it has received considerably less attention by theorists and practitioners in risk analysis. Moreover, there are several pernicious myths about dependence that confuse analysts, not the least of which is the idea that it is okay to ignore correlations and dependencies altogether. Even analysts who recognize the importance of dependence sometimes ignore the issue because of a lack of relevant empirical data on which to base a reasoned model.

Risk assessments based on probability models can be expressed in terms of *total probabilities* of events (e.g., the chance that a pump fails to operate) or in terms of *probability distributions* of random variables (e.g., the failure temperature of a population of pumps). Section 2 of this report reviews the modeling of dependence among events in fault and event trees based on point or interval-valued probabilities. It reviews how empirical information or theoretical specification about how events are related to one another can be incorporated into a risk assessment, and it offers strategies to account for a lack of knowledge about dependence between events that can replace inappropriate or unjustified use of independence assumptions in such assessments. This section introduces special, extreme forms of dependence that may be useful as alternatives to a default or reflexive assumption of independence. These extreme kinds of dependence are generalized in the following sections. It also introduces the strategy of *bounding* results when dependence cannot be specified precisely, which also forms an important theme throughout the entire document.

Section 3 addresses the problem of dependence among random variables in more elaborate models based on mathematical functions of probability distributions, Dempster-Shafer structures or probability boxes. It reviews the methodological dangers of assuming all variables in an assessment are independent of one another and shows how different dependencies can lead to quantitatively different results. It includes a discussion of how the very concept of independence disintegrates into distinct notions in the context of imprecise probabilities. It describes several strategies that have been or could be employed to represent knowledge about how the random variables are interrelated.

Section 4 considers how risk assessment models can account for a lack of relevant information about the dependence among random variables, and how partial or qualitative information about dependencies might best be incorporated into the analysis. Extending the idea of bounding results when dependence cannot be specified precisely, it considers approaches based on sensitivity studies and direct analytical methods that

bound results based on theoretical limits governing dependence. This section points out two surprising results that have implications for risk assessments. The first is that the question of dependence is usually moot for a binary operation if one of the inputs is characterized only by an interval. The second is that knowing the correlation coefficient between two random variables generally tells rather little about how they are related without additional knowledge about the functional nature of their dependence.

Section 5 reviews a large variety of myths about correlations and dependencies that are common in risk assessments and probabilistic modeling generally. For the most part, this section merely highlights ideas explored in the earlier sections, emphasizing their practical consequences when misapplied. This section can be viewed as a summary of the essential practical ideas of the report.

The methods considered in this report are to be applied to models involving

1. Real-valued probabilities,
2. Interval probabilities,
3. Probability distributions,
4. Dempster-Shafer structures of the real line, and
5. Probability boxes (p-boxes).

For readers unfamiliar with recent developments in uncertainty analysis, the following three subsections introduce the notions of interval probabilities, Dempster-Shafer structures and probability boxes, respectively, and provide some pointers to their relevant literatures. Section 9 is a glossary of terms that may be new to many readers.

1.1 Interval probabilities

Some of the probabilities needed as inputs for a fault tree may be difficult to specify precisely. In such cases, it may be desirable to express the analyst's uncertainty by using *interval probabilities* to characterize the events. The idea is that one may not be able to give the exact probability of some event but could still given an upper and lower bound on that probability. This notion of using intervals to describe probabilities is a very old idea in the history of probability. George Boole (1854; Hailperin, 1986) considered interval bounds on certain probabilities. Fréchet (1935) discovered the limits of probabilities of conjunctions and disjunctions of events under all possible dependencies. Kyburg (1998) reviewed the history of interval probabilities and traced the development of the critical ideas through the twentieth century. Bounding probabilities has continued to the present (e.g., Walley and Fine, 1982; Loui, 1986; Hailperin 1986; Madan and Owings, 1988; Williamson 1989; Walley, 1991; Tessem, 1992). Bounding probability is different from the approach of second-order or two-dimensional probability (e.g., Hoffman and Hammonds 1994; Cullen and Frey 1999) in which uncertainty about probabilities is itself modeled with probability.

Although ordinary interval arithmetic (Moore 1966; Alefeld and Herzberger 1983; Neumaier 1990) can be used for some calculations with such intervals, the fact that the inputs represent probabilities which are constrained to the unit interval makes these calculations somewhat more subtle. For instance, if A and B are exhaustive events (i.e., at least one must be true), then we know that the sum of their probabilities

P(A)+P(B) must equal one, even if we cannot know the precise value of either individual probability. Hailperin (1986) reviews the kinds of calculations involving interval probabilities that might be used in a risk or reliability assessment.

1.2 Dempster-Shafer structures

Dempster-Shafer theory (Shafer 1976; Klir and Yuan 1995), which is sometimes called evidence theory, can be considered as a variant of probability theory in which the elements of the sample space (to which nonzero probability mass is attributed) are not single points but rather sets which represent the indistinguishability of alternatives within bodies of evidence. For instance, instead of a discrete probability mass function over the real line giving the probability for each of some number of precise points on the line, a comparable Dempster-Shafer structure might give a probability mass that corresponds to an *interval* rather than a point value. The sum of these masses is one, just as for any probability mass function. The sets that get nonzero mass are called focal elements. The breadth of these focal elements represents the uncertainty that might arise from measurement uncertainty or other form of ambiguity. Klir and Yuan (1994) reviewed the mathematical underpinnings of Dempster-Shafer theory for uncertainty assessment. Oberkampf and Helton (2002; Oberkampf et al. 2001; Helton et al. 2004) reviewed and illustrated the use of Dempster-Shafer theory for engineering applications. Sentz and Ferson (2002) reviewed methods for aggregating multiple Dempster-Shafer structures from different information sources. Yager (1986; Ferson et al. 2003) defined the basic algorithms to compute arithmetic operations of Dempster-Shafer structures under independence in risk assessments.

Dempster-Shafer theory is intimately related, if not equivalent, to the theory of random sets (Matheron 1975; Robbins 1944; 1945). Tonon et al. (1999; 2000a; 2000b) have applied random set theory to various problems in reliability and design problems in engineering.

Although focal elements can generally be any subset of some universal set, in this report, we consider only Dempster-Shafer structures for which the universal set is the real line $\Re$ and whose focal elements are closed intervals of the real line.

1.3 Probability boxes

Risk assessments commonly involve calculations with random variables characterized by probability distributions. Like probabilities that describe events, these distributions may sometimes be difficult to specify precisely. A probability box (p-box) is a class of distribution functions delimited by an upper and a lower bound which collectively represent the epistemic uncertainty about the distribution function of a random variable. A p-box is the class of distribution functions $F(x)$ specified by a bounding pair of cumulative distribution functions $\underline{F}(x)$ and $\overline{F}(x)$ such that $\underline{F}(x) \le F(x) \le \overline{F}(x)$ for all x values. P-boxes thus express interval-like uncertainty about a distribution function. Probability bounds analysis is the collection of methods and algorithms that are used to do calculations with, and make inferences from, p-boxes. These methods are essentially

a combination of the methods of standard interval analysis (Moore, 1966; Neumaier, 1990) and classical probability theory (see, inter alia, Feller, 1968; 1971; Mood et al. 1974).

Like interval probabilities, the idea of bounding probability distribution functions has a very long tradition throughout the history of probability theory. Indeed, Chebyshev (1874; Smith 1995) described bounds on a distribution when only the mean and variance of the variable are known, and Markov (1886) found bounds on a positive variable when only the mean is known. Kolmogorov posed the question of what the bounds would be on a sum of random variables if only their marginal distributions and not the dependence between were known (Makarov 1981). Ferson et al. (2003) reviewed the characterization of p-boxes from empirical information, the aggregation of p-boxes from multiple or competing information sources, and the basic algorithms to compute arithmetic operations under independence between p-boxes in risk assessments. P-boxes are a somewhat coarser way to describe uncertainty than are Dempster-Shafer structures on the real line. Every Dempster-Shafer structure specifies a unique p-box and every p-box specifies an equivalence class of Dempster-Shafer structures (Regan et al. 2004; Joslyn and Ferson 2004). Ferson et al. (2003) described the relationship between these two generalizations of probability distributions. P-boxes are also coarse special cases of imprecise probabilities (Walley 1991), which are arbitrary sets of probability distributions. As an interval is a special kind of set of real numbers, a p-boxe is a special kind of imprecise probabilities.

2 Dependence between events

Fault or event trees are commonly used in risk assessments to estimate the reliability or risk of some "top event" such as the failure of a system in terms of the probabilities of failures of the components of the system (Vesely et al. 1981). Compound events in a fault tree are defined as conjunctions, disjunctions or negations of more elementary events. The conjunction of events A with B is denoted by A & B and is sometimes called the "and" event. It is the event in which both A and B occur. For instance, in a safety assessment, the event A might represent an inadvertent fuel spill that produces a combustible vapor in a closed space. The event B might represent the presence of an ignition source such as a spark somewhere in the space caused by the closing of an electrical circuit. The conjunction of these two events would represent the conditions necessary for an explosion. The disjunction of two events is the event in which one or the other of A or B occurs, or they both occur. The disjunction is sometimes called the "or" event and it is denoted with the expression $A \vee B$. For instance, if the events represent the as-designed functioning of redundant safety systems, then only one of the events must occur for some adverse consequence to be avoided.

The purpose of a fault tree is to recursively express the top event as a function involving such conjunctions and disjunctions of more elementary events. The endpoints of these recursions, which are not further decomposed into subevents, are called "basic events". Because they are not defined in terms of other events, basic events need to be characterized by inputs to the analyses that represent empirical observations or theoretical argument. These basic events are often characterized by real-valued probabilities, sometimes called "total probabilities" to distinguish them from probability distributions. This section reviews the representation of dependencies between basic events and their propagation through the logical model such as a fault tree.

In case two events are independent, the probability of their conjunction can be computed as the product of the probabilities of the two events, thus

$$\mathrm{P}(A \,\&\, B) = \mathrm{and}_{\mathrm{independent}}(a, b) = ab,$$

where $a = \mathrm{P}(A)$ and $b = \mathrm{P}(B)$. The probability of the disjunction can also be computed in terms of the probabilities of the separate events with the formula

$$\mathrm{P}(A \vee B) = \mathrm{or}_{\mathrm{independent}}(a, b) = 1 - (1 - a)(1 - b).$$

Although it might be convenient, it is not always possible to assume that the events in a fault tree are independent of one another (Smith and Watson 1980; Hickman et al. 1983). For example, Vesely et al. (1981) described many situations involving common-cause or common-mode failures in which events will not generally be independent. For instance, there can sometimes be a single cause that can precipitate failure of several

components at the same time. Categories of common causes include many things such as impact, manufacture or use history, location, vibration, contamination, humidity, flooding, temperature, fire, etc. If all of the components in a fault tree's minimal cut set are affected by the same cause, that cause can trigger the top event. In this way, the risk of the top event can degenerate to the risk of the occurrence of one of these common causes. The assessment of a system's susceptibility to common-cause failures has become increasingly more important in risk analysis. In engineering practice, common-cause failures can often dominate random hardware failures (Vesely et al. 1981, page XII-12). It will be important to the correct assessment of system performance that involve such phenomena to be able to evaluate risks and reliabilities without necessarily relying on independence assumptions. Moreover, in the context of abnormal operating environments such as fires, the independent functioning of components in a system intended by the designer of the system may actually devolve to dependent behaviors. It becomes a serious question then to be able to estimate the consequences of a lack of independence on the risks and reliabilities being estimated in the assessment.

If the probabilities characterizing events are depicted in a Venn diagram, the dependence between events is completely determined by the area of the overlap between the sets. Consider the five Venn diagrams shown in Figure 1. Each depicts the probabilities of two events, the first represented by a stippled circle and the second represented by a gray circle. The fact that the shapes are circles is irrelevant; only their areas matter. Likewise, the complexity of the shape of the overlapping region does not matter because only its area is significant. These diagrams are all drawn to the same scale, so that the area of the enclosing square for each of the five Venn diagrams is one, and the area of the larger stippled circle is 0.29 while the area of the smaller gray circle is 0.22. Case A depicts the gray event totally inside the stippled event. This represents the strongest possible dependence between these two events, given their marginal probabilities. In this perfect dependence, if the event represented by the gray circle occurs, it is guaranteed that the other event represented by the stippled circle also occurs. Case C shows the independent case. In this case, the area of overlap is given by the product 0.29×0.22, which is 0.0638. In general, the probability of the joint event (in which both events occur) is given by the product of the probabilities of the two events. The events are not independent unless this quantitative relation holds. Case E shows the events as mutually exclusive. Their area of overlap is zero. It represents the other extreme possible case of dependence because it says that the occurrence of one event precludes the occurrence of the other. Cases B and D depict dependencies that are intermediate between the extreme cases and independence.

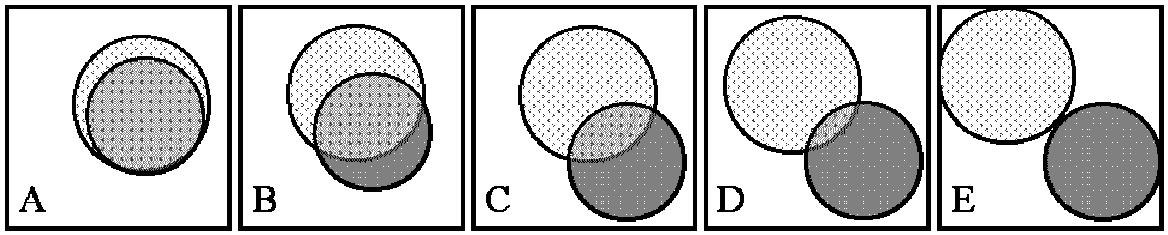

Figure 1: Dependence between two events (gray and stippled) depicted in Venn diagrams where area denotes probability

2.1 Extreme dependence: perfect and opposite

To represent different dependencies between the underlying events, the gray and stippled areas shown in the diagrams of Figure 1 can be moved around within the square and their shapes arbitrarily deformed so long as the areas are not changed. When areas are constrained to given values, certain relationships must hold between the areas of the two sets and their dependence as represented by the degree to which they overlap. For instance, the dependence that leads to the largest area of overlap is that depicted in diagram A. Such an eclipse between the two areas cannot be any greater if the areas are fixed, no matter how the areas are positioned. We can therefore call this dependence "perfect". Because one of the areas is totally inside the other, the area of overlap is the minimum of the two areas. The probabilities of the conjunction and disjunction between the events are very simple to compute for this kind of dependence:

(Perfect)

$$\mathrm{P}(A \,\&\, B) = \mathrm{and}_{\mathrm{perfect}}(a, b) = \min(a, b),$$
$$\mathrm{P}(A \vee B) = \mathrm{or}_{\mathrm{perfect}}(a, b) = \max(a, b),$$

where a and b are the two areas. The probability of the conjunction is measured by the area of the intersection of the two areas. Because they overlap totally, the area of the intersection must be the smaller of the two areas. Because the probability of the disjunction is the area of the union of the two areas, this probability must be the larger of the areas.

The pattern of dependence in Figure 1 that leads to the smallest area of overlap is that shown in diagram E. The area of the overlap is zero because the sets are disjoint. We can call the dependence associated with minimal overlap "opposite" dependence. Note that having opposite dependence does *not* necessarily mean that the events are mutually exclusive. For instance, it may be that both events have probabilities greater than 50%. In such a case, the areas of the events, however they are drawn, must overlap by some amount. Saying that events are mutually exclusive is therefore not just making an assertion about their dependence. It also says something about the probabilities of the events. Saying events have opposite dependence is a claim only about dependence.

The formulas for probabilities of conjunction and disjunction under opposite dependence are

(Opposite)

$$P(A \,\&\, B) = \text{and}_{\text{opposite}}(a, b) = \max(a + b - 1, 0),$$
$$P(A \vee B) = \text{or}_{\text{opposite}}(a, b) = \min(1, a + b).$$

These formulas account for the constraint that probabilities must be no larger than one.

Examples: Suppose that $P(A) = a = 0.29$ and $P(B) = b = 0.22$ and the events A and B are perfectly dependent. This situation is depicted in diagram A of Figure 1 which shows the areas maximally overlapping. The probability of the conjunction $A \,\&\, B$ is $\min(0.29, 0.22) = 0.22$. Despite the use of the minimum operator, this is the *largest* possible value of the probability given these marginals. The probability of the disjunction $A \vee B$ is $\max(0.29, 0.22) = 0.29$. Despite the maximum, this is the *smallest* possible value of the probability for any possible dependency. Now suppose that the dependence is like that depicted in diagram E where the events are oppositely dependent. Now the areas are minimally overlapping. In this case, the probability of the conjunction is $\max(0.29 + 0.22 - 1, 0) = 0$, and the probability of the disjunction is $\min(1, 0.29+0.22) = 0.51$.

These extreme cases are useful mostly as bounds when the analyst has no empirical knowledge or theoretical argument about the dependency (Section 2.3), but it is conceivable that they could be used in an actual assessment in their own right. For instance, in many engineered systems, several of the basic components are often supplied by a single vendor, or have experienced the same inspection, service and repair history. Moreover, components may experience a similar, abnormal condition such as a fire, or the same temporal sequence of environmental conditions. Such commonalities may tend to suggest that probabilities of failure associated with these components may be closer to perfectly dependent than to independent. Opposite dependence, on the other hand, may be suggested by tradeoffs in utilization. For instance, suppose that one of a pair of redundant safety systems always activates first in response to a particular kind of stimulus, and that the kinds of stimuli are not randomly experienced by the systems. If operation of a safety system leads to wear-and-tear aging on that system, then the joint failure of both systems might be better modeled by events that are oppositely dependent than an assumption that they are independent. In such situations, the extreme-dependency formulas might be used in place of the independence formulas because extreme dependence is a somewhat more reasonable assumption than independence.

2.2 Correlation between events

Although the idea of dependence applies to both random variables and simple events, the word *correlation* is often reserved for use only with random variables. Measuring the overall degree of association with some scalar quantity ranging between −1 and +1

can be very useful in risk assessments in general. For this reason, it may be desirable to extend the notion for use with events each of which is characterized by a total probability (that is, a dimensionless real number between zero and one). Because of the simplicity of dependence in the context of events, defining such a scalar measure of the correlation of events is entirely reasonable. Indeed, this use seems far more reasonable than the traditional use of scalar measures of overall dependence between random numbers. The dependence between random numbers is infinite-dimensional (see Section 3), so the single dimension of a correlation coefficient cannot capture the potential complexity of the dependence function. The dependence between events, however, can be characterized completely and without loss of information by a scalar measure.

Lucas (1995; see also Cheng 2003; cf. Cui and Blockley 1990; Davis and Hall 2003) suggests defining correlation between events as the correlation of their indicator functions. To illustrate this idea, imagine throwing darts randomly at a Venn diagram such as those shown in Figure 1 and scoring each throw using two binary values, one for each event. The value of the indicator function would be zero if the dart misses the event's area and one if it hits it. Lucas' measure of dependence between the two events would be the correlation coefficient for a long sequence of such scores for randomly thrown darts. The measure would be the same if, instead of randomly thrown darts, a grid of many pins were uniformly and systematically getting similar pairs of scores. The formula for a (Pearson product-moment) correlation between random variables X and Y is

$$r = \frac{\mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)}{\sqrt{\mathrm{V}(X)}\sqrt{\mathrm{V}(Y)}}$$

where E denotes the expectation and V denotes the variance. The expectation of an indicator function for an event A is the probability of the event P(A). The variance for the indicator function is P(A)(1−P(A)). Consequently, when the formula for the correlation is applied to the indicator functions for events A and B, it becomes

$$r = \frac{\mathrm{P}(A \,\&\, B) - \mathrm{P}(A)\,\mathrm{P}(B)}{\sqrt{\mathrm{P}(A)(1 - \mathrm{P}(A))}\sqrt{\mathrm{P}(B)(1 - \mathrm{P}(B))}}$$

because the expected value of the product of the indicator functions is the probability of the conjunction given the dependence between the two events. With a little rearranging, this leads to the formulation

$$\mathrm{P}(A \,\&\, B) = \mathrm{and}_{\mathrm{Lucas}}(a, b, r) = ab + r\sqrt{a(1-a)b(1-b)}$$

where a = P(A) and b = P(B). We call this formula the Lucas model for correlated events. For instance, suppose that a = 0.29 and b = 0.22 and the correlation r = 0.2.

The Lucas model suggests the probability of the conjunction A & B would be 0.101. If r is set to zero, the Lucas model yields the value 0.0638, which is the probability of the conjunction assuming the events A and B are independent. As r varies, the probability of the conjunction according to the Lucas model changes linearly. This model might be considered desirable except that it can produce values for the conjunction that are outside the Fréchet bounds that limit such probabilities (see Section 4.2). In fact, it can even produce probabilities that are smaller than zero or larger than one. For instance, if we set $r = -1$, the result of the Lucas model is –0.124. No one likes a negative probability.

The problem is not with the formula itself, but with the mistaken idea that correlation can take on any value between positive and negative one. It is well known that Pearson correlation coefficients cannot always range over this entire range (Feller 1968; Nelsen 1999). In this case, the smallest possible correlation coefficient for the indicator functions is not –1, but only –0.339. And the largest value is not +1 but only 0.831. This can be demonstrated by considering two columns of numbers. One column, representing the indicator function for event A has 290 ones and 710 zeros. The second column of numbers for event B has 220 ones and 780 zeros. If the values in the columns are sorted so that all the zeros are at the top of the columns, then the correlation between the two columns will be 0.831. If we then sort one of the columns in the reverse order so that the zeros are all at the bottom, then the correlation will be –0.339. No matter how the zeros and ones are shuffled within the columns, the correlation coefficient between these two columns of numbers cannot be any larger or smaller than these extremes. If the input values for the correlation r are constrained to the interval [–0.339, 0.831], then the probabilities computed by the Lucas model are correctly limited to the possible range that ensures no negative probabilities or probabilities larger than one. The only way to make using the Lucas model reasonable is to limit the input correlations to be no smaller than $\underline{r}$ and no larger than $\overline{r}$, where

$$\underline{r} = \frac{\max(a+b-1, 0)}{\sqrt{a(1-a)b(1-b)}},$$
$$\overline{r} = \frac{\min(a,b) - ab}{\sqrt{a(1-a)b(1-b)}}.$$

Another formulation for correlation of events can be derived from the Frank family of copulas (Section 3.7), which were first introduced by Frank (1979). In the Frank model of correlation between events, the probability of a conjunction of events A and B is given by the formula

$$\mathrm{P}(A \,\&\, B) = \mathrm{and}_{\mathrm{Frank}}(a,b,r) = \begin{cases} \min(a,b) & \text{if } r = +1 \\ ab & \text{if } r = 0 \\ \max(a+b-1,0) & \text{if } r = -1 \\ \log_s\left[1 + (s^a - 1)(s^b - 1)/(s-1)\right] & \text{otherwise} \end{cases}$$

where $s = \tan(\pi(1-r)/4)$, $a = \mathrm{P}(A)$ and $b = \mathrm{P}(B)$. This function is continuous; the special cases when r is +1, 0 or −1 are the limiting values of the bottom expression on the right-hand side of this formula when r tends to these values respectively. Disjunction of correlated events can likewise be defined by Frank's co-copula, so that

$$\mathrm{P}(A \vee B) = \mathrm{or}_{\mathrm{Frank}}(a,b,r) = \begin{cases} \max(a,b) & \text{if } r = +1 \\ 1-(1-a)(1-b) & \text{if } r = 0 \\ \min(a+b,1) & \text{if } r = -1 \\ 1 - \log_s\left[1 + (s^{1-a} - 1)(s^{1-b} - 1)/(s-1)\right] & \text{otherwise.} \end{cases}$$

Example: If we again suppose that $a = 0.29$ and $b = 0.22$ and the correlation $r = 0.2$, the Frank model of dependence suggests that the probability of A & B is 0.0695.

In the Frank model of dependence, no value between −1 and +1 is an impossible correlation. Figure 2 depicts the probabilities of the disjunction (shown as the gray line) and the conjunction (shown as the black line) for various values of the correlation according to the Frank model of correlation among the two events.

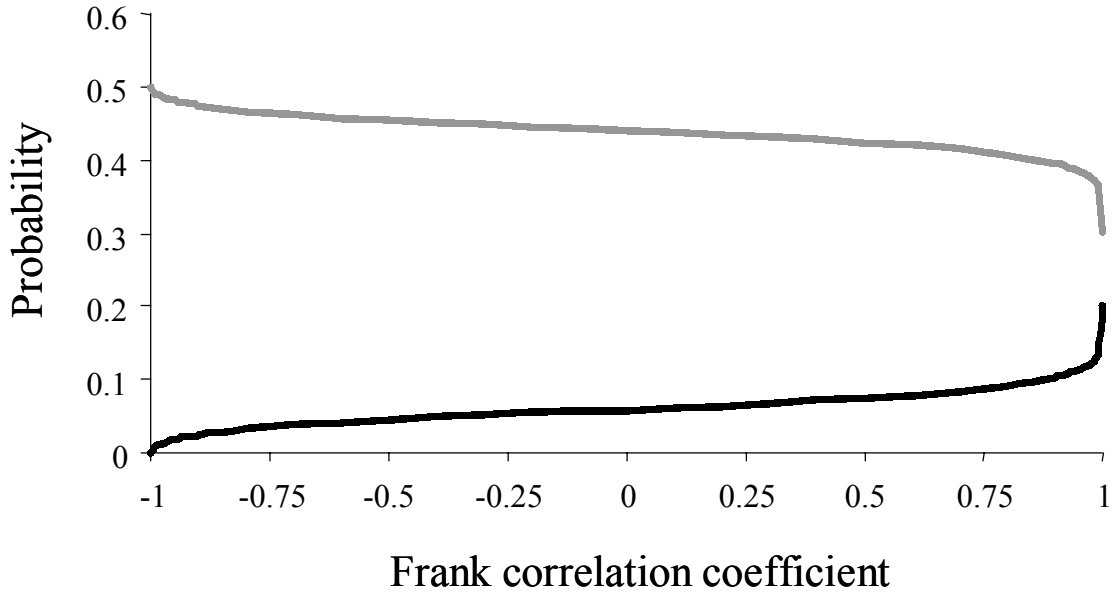

Figure 2: Probability of the disjunction (gray) and conjunction (black) of two correlated events of probability 0.29 and 0.22 as a function of the Frank correlation.

The Frank model is only one of many possible ways to parameterize a notion of correlation for events. Several researchers have suggested using t-norms and t-conorms as models for the "and" and "or" operations (e.g., Joslyn 1995; Klir and Yuan 1994; Schweizer and Sklar 1983). T-norms are also called generalized intersection operators, and t-norms include as special cases the functions $\text{and}_{\text{independent}}$, $\text{and}_{\text{perfect}}$, $\text{and}_{\text{opposite}}$, as well as $\text{and}_{\text{Frank}}$ (for a particular value of the correlation) for evaluating conjunctions under various models of dependence between the events. Likewise, t-conorms, which are called generalized union operators include the functions $\text{or}_{\text{independent}}$, $\text{or}_{\text{perfect}}$, $\text{or}_{\text{opposite}}$, and or_{Frank} (given r). However, it does not seem reasonable to use t-norms and t-conorms to define operations to estimate probabilities of conjunctions and disjunctions. The reason is that some of these functions are incompatible with probability values. For instance, the aptly named "drastic intersection" (Klir and Yuan 1994) yields results that are not possible with probabilities because they are outside the Fréchet limits (Section 2.3). For this reason, it seems more prudent to look to copulas* (Section 3.7) as a model characterizing correlations among dependence. The Frank model is a copula (Frank 1979; Nelsen 1999). Nelsen (1999) describes many other families of copulas and co-copulas that could also be used for this purpose.

*A t-norm T is a copula if and only if it is 2-increasing; that is, if $T(a_2, b_2) - T(a_1, b_2) - T(a_2, b_1) + T(a_1, b_1) \geq 0$ whenever $a_1 \leq a_2$ and $b_1 \leq b_2$ for $a_1, a_2, b_1, b_2 \in [0,1]$. A copula is a t-norm if and only if it is associative.

2.3 Accounting for unknown dependence

It is possible to estimate bounds on conjunctions and disjunctions and other joint events using only the marginal information about the probabilities of the events and no information at all about their dependence. Such calculations make use of the classical Fréchet inequalities

$$\mathrm{P}(A \,\&\, B) = \mathrm{and}_{\text{Fréchet}}(a, b) = [\ \max(0,\ a + b - 1),\ \ \min(a, b)\ ],$$
$$\mathrm{P}(A \vee B) = \mathrm{or}_{\text{Fréchet}}(a, b) = [\ \max(a, b),\ \ \min(1,\ a + b)\ ],$$

where $a = \mathrm{P}(A)$, $b = \mathrm{P}(B)$, and the square brackets indicate that the resulting probability estimates are intervals rather than precise real values even if a and b are precise. This says that, whatever the actual probability is, it must lie within the interval. Williamson (1989, page 131) points out that at least the first of these was known to Boole (1854, page 299). Fréchet (1935) proved not only that they are the extreme cases but also that they are the bounds on all possible cases of dependence and, moreover, that they are the best possible such bounds in the absence of information about the dependence.

The proofs of the Fréchet inequalities are elementary. We consider the proof for the conjunction. The definition $\mathrm{P}(A \vee B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \,\&\, B)$ implies that $\mathrm{P}(A \,\&\, B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \vee B)$. Because $\mathrm{P}(A \vee B) \le 1$, as all probabilities are no bigger than one, it must be the case that $\mathrm{P}(A) + \mathrm{P}(B) - 1 \le \mathrm{P}(A \,\&\, B)$. It's also true that $0 \le \mathrm{P}(A \,\&\, B)$, just because all probabilities are no smaller than zero, so it must also be true that $\max(0, \mathrm{P}(A) + \mathrm{P}(B) - 1) \le \mathrm{P}(A \,\&\, B)$. This establishes the lower bound on the conjunction. To get the upper bound, recall that $\mathrm{P}(A \,\&\, B) = \mathrm{P}(A|B)\ \mathrm{P}(B) = \mathrm{P}(B|A)\ \mathrm{P}(A)$. Because $\mathrm{P}(A|B) \le 1$ and $\mathrm{P}(B|A) \le 1$, as all probabilities are, it follows that $\mathrm{P}(A \,\&\, B) \le \mathrm{P}(A)$ and $\mathrm{P}(A \,\&\, B) \le \mathrm{P}(B)$, so $\mathrm{P}(A \,\&\, B) \le \min(\mathrm{P}(A), \mathrm{P}(B))$. The best-possible nature of these bounds follows by observing that they are realized by some dependency relation between the events A and B.

Example: Suppose that $\mathrm{P}(A) = a = 0.001$ and $\mathrm{P}(B) = b = 0.002$, then $\mathrm{P}(A \,\&\, B)$ is sure to lie in the interval $[\max(0, 0.001+0.002-1), \min(0.001, 0.002)] = [0, 0.001]$. $\mathrm{P}(A \vee B)$ is likewise certain to be somewhere in $[\max(0.001, 0.002), \min(1, 0.001+0.002)] = [0.002, 0.003]$. These intervals are rigorous and true no matter what dependency there may be between A and B.

The Fréchet (1935) inequalities generalize by induction to the multivariate case. The resulting formulas are straightforward extensions, so

$$\mathrm{P}(A_1 \,\&\, A_2 \,\&\, \ldots \,\&\, A_n) = [\ \max(0,\ a_1 + a_2 + \ldots + a_n - (n-1)),\ \ \min(a_1, a_2, \ldots, a_n)\ ],$$
$$\mathrm{P}(A_1 \vee A_2 \vee \ldots \vee A_n) = [\ \max(a_1, a_2, \ldots, a_n),\ \ \min(1,\ a_1 + a_2 + \ldots + a_n)\ ],$$

where $a_i = \mathrm{P}(A_i)$.

Example: Suppose that $\mathrm{P}(A_1) = a_1 = 0.001$, $\mathrm{P}(A_2) = a_2 = 0.002$, and $\mathrm{P}(A_3) = a_3 = 0.003$. Then $\mathrm{P}(A_1 \,\&\, A_2 \,\&\, A_3) \in [\max(0, 0.001+0.002+0.003-(3-1)), \min(0.001, 0.002,$

0.003)] = [0, 0.001]. P($A_1 \vee A_2 \vee A_3$) $\in$ [max(0.001, 0.002, 0.003), min(1, 0.001+0.002+0.003)] = [0.003, 0.006].

Section 2.5 will review how the Fréchet inequalities can be incorporated into more elaborate logical expressions needed for practical fault tree analysis which may involve composition of many conjunctions and disjunctions, as well as negations and other kinds of logical gates.

2.3.1 Using knowledge about the sign of the dependence

Williamson (1989; cf. Wise and Henrion 1986) suggested that, even in situations where one may not know much about the possible dependence between events, it might still be possible to confidently assert knowledge about the sign of the dependence. For instance, it might be reasonable to assert that the similar manufacture and repair histories shared by components in some system means that they could only be positively dependent (that is, no less strongly dependent than independent). If events A and B could only be positively dependent, then

(Positive)

$$\mathrm{P}(A \,\&\, B) = \mathrm{and}_{\mathrm{positive}}(a, b) = [\, ab, \;\; \min(a, b)\,],$$
$$\mathrm{P}(A \vee B) = \mathrm{or}_{\mathrm{positive}}(a, b) = [\max(a, b),\, 1 - (1 - a)(1 - b)].$$

If, on the other hand, an analyst believes that which of two outcomes occurs is the result of a tradeoff such that the events could not be positively dependent, it might be likewise reasonable to tighten the Fréchet inequalities. If events A and B could only be negatively dependent, then

(Negative)

$$\mathrm{P}(A \,\&\, B) = \mathrm{and}_{\mathrm{negative}}(a, b) = [\, \max(a + b - 1, 0),\, ab\,],$$
$$\mathrm{P}(A \vee B) = \mathrm{or}_{\mathrm{negative}}(a, b) = [\, 1 - (1 - a)(1 - b),\, \min(1, a + b)\,].$$

Example: Suppose A and B are sure to be positively dependent, and that P(A) = a = 0.003 and P(B) = b = 0.005, then P(A & B) is sure to lie in the interval [0.003×0.005, min(0.001, 0.002)] = [0.000015, 0.003]. If, however, A and B are negative dependent, then P($A \vee B$) is sure to be within [1–(1–0.003)(1–0.005), min(1, 0.003+0.005)] = [0.007985, 0.008].

2.4 Interval probabilities

As mentioned in Section 1.1, some of the probabilities needed as inputs for a fault tree may be difficult to specify precisely and in such cases the analyst's uncertainty may be expressed by using intervals that contain the inexactly known probabilities to characterize some or all of the events. These representations are called *interval probabilities*. Even if all of the inputs to a fault tree are precisely specified, applying the inequalities described in the previous sections in computing bounds on

subexpressions of the fault tree generally yield interval probabilities rather than point estimates of probability whenever they account for incompletely specified correlations. The formulations to evaluate the probabilities in the tree must therefore be capable of accepting interval inputs. In this section, we briefly describe strategies that can be used to compute with intervals in such situations.

2.4.1 Interval arithmetic

Moore (1966) described the use of interval arithmetic to evaluate the ranges of functions taking interval arguments. The approach is to generalize the definitions of the binary operations out of which the function is composed to handle interval inputs. For instance, we can define some basic interval operations for use with interval probabilities (cf. Moore 1966; Neumaier 1990). For all real numbers w, x, y and z such that $0 \le w \le x \le 1$ and $0 \le y \le z \le 1$, let

$$
\begin{gathered}
[w, x] + [y, z] = [\, w + y, x + z \,], \\
[w, x] - [y, z] = [w - z, x - y], \\
[w, x] \times [y, z] = [\, w \times y, x \times z], \\
[w, x] / [y, z] = [\, w / z, x / y], \text{ assuming } 0<y, \\
\min([w, x], [y, z]) = [\, \min(w, y), \min(x, z) \,], \text{ and} \\
\max([w, x], [y, z]) = [\, \max(w, y), \max(x, z) \,].
\end{gathered}
$$

Note that these formulas for multiplication and division are considerably simpler than those of ordinary interval arithmetic. The simplicity is a consequence of the constraint that probabilities must lie in the interval [0, 1].

Examples: Suppose there is epistemic uncertainty about the probabilities of events A and B so that $a = \mathrm{P}(A) = [\, 0.0015, 0.0025]$* and $b = \mathrm{P}(B) = [\, 0.00025, 0.00035]$. Then the following table gives bounds on the probabilities for the conjunction and disjunction of events A and B under different assumptions (or, in the Fréchet case, no assumption) about dependence between the events.

Dependence	**P(*A* & *B*)**	**P(*A* ∨ *B*)**
independent	[0.000000375, 0.000000875]	[0.001749625, 0.002849125]
perfect	[0.00025, 0.00035]	[0.0015, 0.0025]
opposite	0	[0.00175, 0.00285]

*This just means that we believe that the probability for the event A is some value between 0.0015 and 0.0025 (inclusive). Some purists would prefer we write $\mathrm{P}(A) \in [0.0015, 0.0025]$,but we have found that more readers are distressed by the unfamiliar symbol $\in$ than are perturbed by the idea that a probability can be identified with an interval. Mathematically, an interval probability for an event A can be defined as the interval $[\underline{\mathrm{P}}(A), \overline{\mathrm{P}}(A)]$, where $\underline{\mathrm{P}}(A) = \inf \{\mu(A) : \mu \in \mathcal{P}\}$ and $\overline{\mathrm{P}}(A) = \sup \{\mu(A) : \mu \in \mathcal{P}\}$ for a given set $\mathcal{P}$ of probability measures, all defined on an algebra $\mathcal{F}$ (that contains A as an element) over some sample space (universal set). $\underline{\mathrm{P}}(A)$ is called the lower probability of A, and $\overline{\mathrm{P}}(A)$ is called its upper probability.

positive	[0.000000375, 0.00035]	[0.0015, 0.002849125]
negative	[0, 0.000000875]	[0.001749625, 0.00285]
Fréchet	[0, 0.00035]	[0.0015, 0.00285]

Interestingly, the upper bound for the disjunction is the same (within three significant figures) for all of the dependence cases except perfect.

In many problems, interval arithmetic can be used in a straightforward way to obtain results that are both rigorous and best possible. However, when an uncertain number appears more than once in a mathematical expression, the naïve sequential application of the rules of interval arithmetic may yield results that are wider than they should be. The result is still rigorous in the sense that it is sure to enclose the true range, but it may fail to be best-possible if it is wider than it needs to be to do so. The reason for this loss of optimality is basically that the uncertainty in the repeated parameter is entered into the calculation more than once. The appearance of repeated parameters in expressions is a well-known problem with interval arithmetic and, indeed, with all common uncertainty calculi (e.g., Moore 1966; Manes 1982; Hailperin 1986; Ferson 1996a). Many strategies have been developed to address this problem. One of the most direct methods is to algebraically manipulate the expression to reduce the occurrences of the repeated parameters. Consider, for instance, the probabilistic sum which is usually written as

$$a + b - ab.$$

Using interval arithmetic to evaluate this expression could lead to probabilities appearing to be smaller than zero or larger than one because of the uncertainty inflation from the repetition of a and b in the expression. Fortunately, this expression can always be rearranged to

$$1 - (1 - a)(1 - b),$$

which is algebraically equivalent to the previous expression but has only one occurrence of each parameter. This formulation is therefore safe to use with interval probabilities and will not exhibit inappropriate inflation of uncertainty.

2.4.2 Subinterval reconstitution

Moore (1979; Corliss 1988) described a method to obtain bounds on an interval expression that are arbitrarily close to best possible. The method works by partitioning the input interval into a collection of subintervals, projecting the subintervals through the function separately, and reconstituting the answer as the union of their images. If the expression of a function f has interval arguments $x, y, \ldots, z$ which are repeated, and intervals $u, v, \ldots, w$, which are not repeated, then

$$f(x,y,...z,u,v,...,w) \subseteq \bigcup_i \bigcup_j ... \bigcup_k f(x_i, y_j, ..., z_k, u, v, ..., w).$$

where

$$x = \bigcup_i x_i; \quad y = \bigcup_j y_j; \quad ...; \quad z = \bigcup_k z_k.$$

In general, the reconstituted union may also be overly wide, although it will tend to narrow as the number of subintervals increases. In all cases, however, it will surely rigorously contain the true interval. The tightness of the resulting estimate depends on the fineness of the partitioning of the inputs. This method of subinterval reconstitution allows us to obtain answers that, with sufficiently many subintervals, are arbitrarily close to best possible.

Example: Suppose that events *A* and *B* have a Lucas correlation *r* of 0.2 and uncertain marginal probabilities described by the intervals, $a = P(A) = [0.045, 0.055] = [4.5, 5.5]\%$ and $b = P(B) = [0.005, 0.015] = [0.5, 1.5]\%$. Naively applying interval analysis to the Lucas formula for conjunction, $ab + r\sqrt{(a(1-a)\, b(1-b))}$, we calculate bounds on the probability of the conjunction $A \,\&\, B$ to be $[0.312, 0.642]\%$. Using subinterval reconstitution with a 10×10 regular partition of the input space, gives us a tighter estimate of $[0.315, 0.637]\%$. This result can be contrasted with the bounds on the probability of the conjunction assuming the events are independent, which are $[0.0225, 0.0825]\%$, or the Fréchet limits assuming nothing about dependence between the two events, which are $[0, 1.5]\%$.

Subinterval reconstitution is just one of the many tools that have been developed to handle the 'excess width problem' in evaluating expressions with repeated interval parameters (Moore 1966; 1979). It is the easiest technique to apply and in many situations sufficient to reduce the excess width with modest computational effort. In cases where there are many different repeated parameters or the repeated parameter occurs multiple times in the expression, the rate of convergence to the exact bounds will be reduced. In some cases, alternative computational strategies will be required to obtain good results in a reasonable time. The Moore-Skelboe algorithm (Skelboe 1974; Moore 1979) adaptively subdivides the box defined by the vector of interval inputs to find the global maximum and minimum of the function. The algorithm is most efficient when there are few local minima and maxima located in the interior of the box. Subinterval reconstitution converges to the correct bounds in a linear order. Centered forms such as the mean-centered form (Moore 1979) and Taylor forms (Neumaier 2002) can converge at rates higher than linear order. However, using a mean-centered form entails rewriting the original expression in terms of a central point, usually the midpoint of one or more intervals. For intervals that are wide, as in many of the examples given in this report, applying centered forms can perform poorly, sometimes giving bounds that are wider than those obtained by applying naïve interval arithmetic. Centered forms work best when combined with another technique, such as subinterval reconstitution. In many cases, they can significantly reduce the number of function

evaluations the Moore-Skelboe algorithm needs to compute to find the global maximum and minimum.

2.4.3 Range sampling

An alternative approach to evaluate the range of a function with interval inputs is a variant of Monte Carlo simulation. Scalar-valued "samples" of each input quantity are selected randomly from their respective intervals. These real numbers are combined according to the function being evaluated. The observed range of resulting values is taken to be the estimate of the range of the function. This approach is not guaranteed to bound the true range of the function. However, as the number of replicates increases, the interval of the results will tend to widen toward the range of the function. The resulting interval can never be wider than the true range. It can be narrower than the true range if, for instance, the function has an extreme behavior at particular points in the input space which the sampling strategy happens to miss. Interval analysis techniques such as subinterval reconstitution, in contrast, do not have this disadvantage because they will never underestimate the width of the true range.

Example: The problem described above in Section 2.4.3 can also be solved with range sampling. Using 200 paired values for a and b randomly sampled from their respective intervals, and four additional pairs consisting of the corners (a = 0.045, b = 0.005), (a = 0.045, b = 0.015), (a = 0.055, b = 0.005), and (a = 0.055, b = 0.015), one obtains [0.315, 0.637] as the estimated range on the probability of the conjunction of the correlated events. This answer agrees with the answer previously obtained by subinterval reconstitution, at least in three digits shown.

2.4.4 Mathematical programming

It is also possible to express an interval calculation as a mathematical programming problem. In particular, linear programming can be used to find bounds on Boolean functions with interval inputs. Hailperin (1986; Williamson 1989) has explored this approach and made several significant contributions by use of some duality theorems that simplify many results. Hailperin proved that linear programming can be used to determine the best possible bounds on arbitrary Boolean functions having interval inputs.

2.5 Using inequalities in risk assessments

The methods described in the previous sections motivate a calculus for evaluating fault and event trees. The approach is fairly general and can readily be extended to the other logical operations defined for events characterized by interval probabilities besides conjunction and disjunction. For instance, the exclusive-or operation can be defined by

$$\mathrm{P}(A \vee B \;\&\; (\mathrm{not}(A \;\&\; B) = \mathrm{xor}_{\text{Fréchet}}(a, b) = [\, \max(a - b, b - a), \min(a + b, 2 - a - b) \,].$$

Unary operations may also be defined, such as negation P(not A) = 1 – a, although such operations are not affected by any dependency considerations. When there are more than two events, calculations proceed by building up the expression by composition of binary operations. For instance, P(A & B & C) can be estimated by solving first for the A & B event and then for (A & B) & C. It can also be solved by solving for the B & C event first and then for A & (B & C). In many practical problems, these straightforward methods can be used to make calculations that are guaranteed to be rigorous bounds on the risks of interest. Moreover, in many cases these bounds will also be best possible (i.e., as narrow as can be justified by the input assumptions).

The pairwise composition of the formulas will produce the best possible bounds when there are no repeated parameters and dependence can be fully accounted for in the binary operations. When the uncertainty about the probabilities of the basic events is large, or when the events to be combined are numerous, uncertainty can sometimes become large. Williamson (1989, page 135) argued, "Even when there are not repeated variables, the bounds can rapidly become quite loose…. This should not be taken as an argument against [interval probabilities] though. What it does show is the danger, even in simple problems, of assuming independence in order to obtain a unique value at the end." If the intervals are as tight as they can be given the inputs, it is up to the analyst to justify a narrower answer with empirical information about the input probabilities or the dependencies between them rather than engaging in a mathematical exercise that amounts only to wishful thinking.

The challenge for the risk analyst who eschews wishful thinking and adopts an interval probability approach is to produce the best possible intervals that can be justified by the available information embodied in the given inputs and assumptions. The effective use of this calculus for interval probabilities depends on the analyst being able to make use of the information that is known about dependence to obtain the tightest justified intervals. The Fréchet case, as well as the positive or negative cases, tend to widen intervals. Therefore, it is best to use these cases only when one cannot justify (on either empirical or theoretical grounds) assuming independence or some particular dependence which would produce tighter intervals. The optimal strategy then is to arrange the expression being evaluated to one's best advantage. In many analyses, dependence is only an issue for a few of the components. Typically (though certainly not always) components that are likely to be dependent will naturally appear close to one another in the logical expression. The rules of Boolean algebra can permit the analyst to rearrange the terms in an expression to bring dependent events together in one binary operation. When the expressions cannot be arranged to permit the convenient calculation of best possible results, then an analyst must choose between a suboptimal (but still rigorous) result and the extra computational burden of subinterval reconstitution, mathematical programming, or some other approach to obtain the best possible answer.

2.6 Caveat: best-possible calculations are NP-hard

The problem of calculating optimal bounds for arbitrary logical functions with interval parameters and dependencies belongs to a class of problems known to be NP-hard, which is to say that there is no general algorithm that can produce a solution with computational effort that grows as some polynomial function of the number of inputs. Indeed, it is well known that finding the reliability of the top event is already NP-hard when the fault tree is not in fact a graph-theoretic tree because of links induced by common cause events (Rosenthal 1975). This does not necessarily imply practical insolubility, but it surely suggests that big problems such as may arise in complex assessments with hundreds of components and complicated dependencies may be daunting. When repeated parameters are sparse and when complex dependencies are isolated to a few variables or to subportions of the logical expression, fairly large problems can be tackled in moderate time. Moreover, if one can be satisfied with answers that are sure bounds on the risk or reliability, but may not be the tightest or best possible such bounds, then the requisite calculations needed will not be NP-hard. In fact they will require on the order of only $4n$ calculations, where n is the number of interval inputs (Moore 1966; 1979).

2.7 Numerical example: fault tree for a pressure tank

This numerical example illustrates how one could account for partial knowledge about the dependencies between events as well as epistemic uncertainties about the probabilities of these events. The example will involve the calculation of the probabilities of conjunctions and disjunctions of both point and interval probabilities and a variety of dependence assumptions, including independence, perfect dependence, (Frank) correlation and the Fréchet case of assuming nothing about dependence.

The probabilistic fault-tree model we use is based on an example due to Vesely et al. (1981). It concerns the risk of rupture associated with the operation of the pressure tank system depicted in Figure 3. This diagram depicts the control system regulating a pump that compresses air into a pressure tank. The system is depicted in the "off" state with the pressure tank empty. The state diagram in Figure 4 shows the operational modes and the transitions between these modes. The system is turned on by briefly depressing switch S1. This energizes the coil of relay K1, which closes its contacts, so that relay K1 becomes electrically self-latched. When relay K1's contacts are closed, the coil of relay K2 is energized and its contacts close and start the pump motor. This starts a pressurization cycle. When the tank is empty or not fully pressurized, the pressure switch S is closed. When the tank is fully pressurized, pressure switch S opens, which de-energizes relay K2 and opens its contacts. This cuts off power to the pump motor. When the tank's outlet valve is opened, the tank depressurizes, the pressure switch closes, and the pressurization cycle begins again.

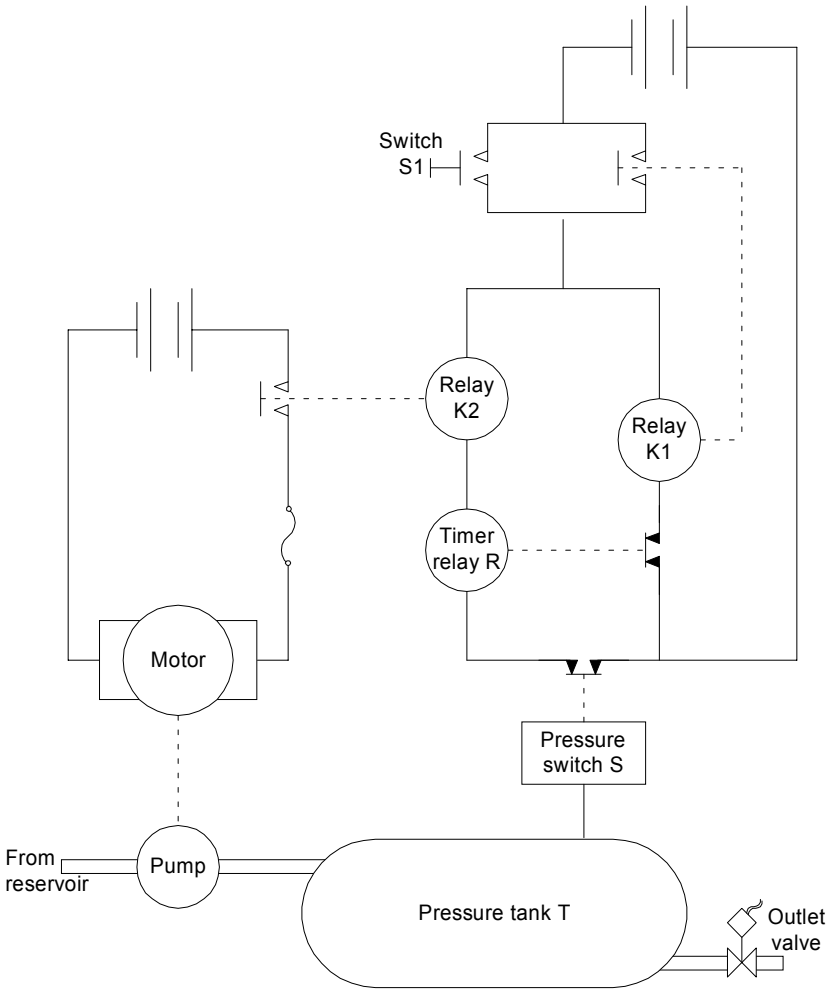

Figure 3: Pressure tank system.

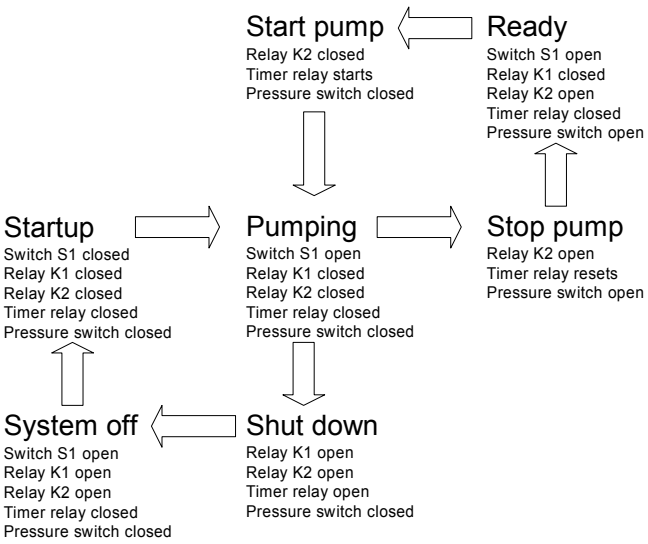

Figure 4: State diagram for the pressure tank system.

The design of the system provides for an emergency shutdown. When relay K1's contacts are first closed, power is also applied to the timer relay, which begins counting. After 60 seconds of continuous power, the timer relay's contacts open and break the current to relay K1 so the system shuts down. However, if the tank fully pressurizes before the timer relay times out, the pressure switch contacts open, which causes the contacts for relay K2 to open which in turn causes the timer relay to reset itself. In this way, the system continues in ready mode.

Vesely et al. (1981) derived a fault tree, shown in Figure 5, for the pressure tank rupturing under pumping. The circles indicate basic events corresponding to primary failures of the components. For instance, the circle enclosing T represents the failure of the tank to withstand pressures according to its specifications. Other events are denoted by symbols beginning with E. These include the top event E1 (which is the tank rupturing under pumping) and the various intermediate events E2, E3, E4, and E5. The conjunctions and disjunctions are indicated with "and" and "or" gates. For instance, event E4 is a disjunction that occurs if either the S1 fails or the fault E5 occurs. This fault tree is completely equivalent to the logical expression

$$E1 = T \vee (K2 \vee (S \,\&\, (S1 \vee (K1 \vee R)))).$$

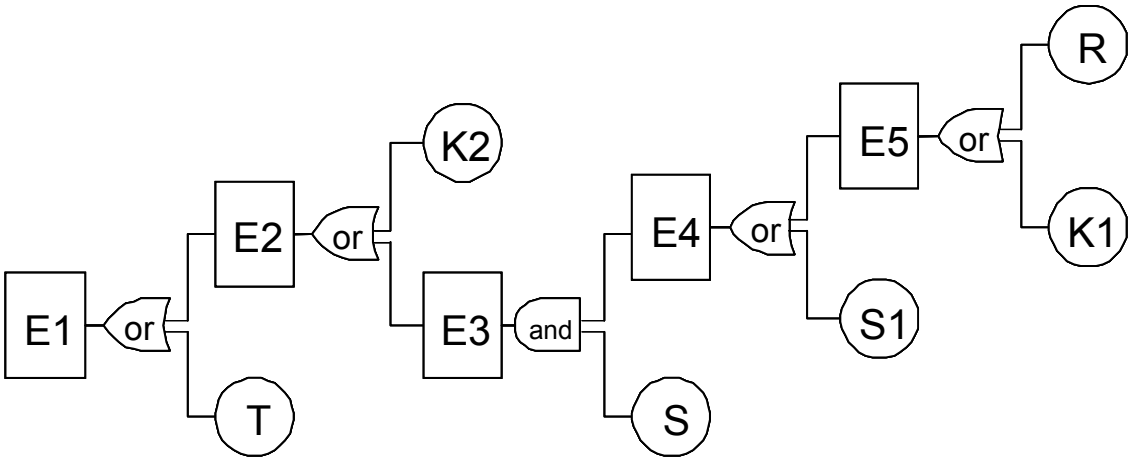

Figure 5: Fault tree for the pressure tank system.

Quantitative evaluation of the fault tree requires estimates of failure rates for the components corresponding to the inputs in the logical expression. Suppose the analyst has the following estimates for the probabilities of primary failure per demand of each of the major components:

Component	*Probability*
Pressure tank T	5×10^{-6}
Relay K2	3×10^{-5}
Pressure switch S	1×10^{-4}
Relay K1	3×10^{-5}
Timer relay R	1×10^{-4}
On-switch S1	3×10^{-5}

Traditionally, a risk analyst would combine these values to estimate the probability of a tank rupture under pumping to be about 3.5×10^{-5} per demand. This calculation assumes that all failures are statistically independent of one another. However, this implies that failure in one component doesn't somehow make failure in another more likely. It also assumes that multiple failures are not the result of a common cause or mechanism.

Often the risk analyst has various degrees of information regarding dependencies between variables and wishes to incorporate this knowledge into the risk calculation. Continuing with the pressure tank example, suppose that the analyst feels confident that the tank's likelihood of failure due to pressure levels that do not exceed specifications (T) is independent of all other events in the system. The analyst is also confident that the failure of the on-switch (S1) is independent of relay K1 or the timer relay (R), and that failure of K1 and R are perfectly correlated. The analyst might have knowledge of some specific correlations as well, for example, pressure switch failure (S) might be known to be correlated under the Frank copula with the conjunction of the on switch or relay K1 or timer relay failure, by a specific coefficient, *r*. Finally, the

analyst might know nothing regarding dependency or lack thereof between the relay K2 and any events further down in the tree, and might want this lack of knowledge to be fully expressed by allowing for any and all possible dependencies (including but not limited to independence). The event tree model can be written to incorporate this collection of information regarding dependencies between events as

$$E1 = T \,|\vee|\, (K2 \vee (S \,\&_r\, (S1 \,|\vee|\, (K1 \,/\vee/\, R)))),$$

where the vertical lines around a logical operator indicate that the operation is to be carried out assuming independence, the forward (back) slashes around a logical operator indicate the operation is to be carried out assuming perfect (opposite) correlation, logical operators with subscripts indicate the operation is to be carried out assuming correlation between operands as specified by the subscript, and logical operators with no additional notational elements indicate operations to be carried out making no assumption whatsoever regarding dependency. Equivalently, one could write

$$E1 = \text{or}_{\text{independent}}(T, \text{or}_{\text{Fréchet}}(K2, \text{and}_{\text{Frank}}(S, \text{or}_{\text{independent}}(S1, \text{or}_{\text{perfect}}(K1, R)), r))).$$

The result of this calculation, with r specified to be equal to 0.15, is the interval $[3.499\times10^{-5}, 3.504\times10^{-5}]$. The result is an interval even though all of the input failure probabilities are point values. This is the result of not knowing the dependence among some of the components. In this case, the interval turns out to be a narrow one. For this particular calculation, then, the independence assumptions have a rather small effect on the answer.

It is certainly not true that dependence assumptions always (or even usually) have minor consequences on a calculation. For instance, suppose the analyst knows nothing about the dependence between the pressure switch S and the components S1, K and R. This can be expressed by replacing the Frank correlation with Fréchet limits:

$$E1 = \text{or}_{\text{independent}}(T, \text{or}_{\text{Fréchet}}(K2, \text{and}_{\text{Fréchet}}(S, \text{or}_{\text{independent}}(S1, \text{or}_{\text{perfect}}(K1, R))))).$$

The result of this calculation is the wider interval $[3.50\times10^{-5}, 1.35\times10^{-4}]$. As the amount of information regarding dependencies between variables decreases, the interval grows wider, reflecting the analyst's increasing uncertainty. In the face of a complete lack of knowledge, the analyst can evaluate the fault tree allowing for any and all dependencies between all events, expressed as

$$E1 = \text{or}_{\text{Fréchet}}(T, \text{or}_{\text{Fréchet}}(K2, \text{and}_{\text{Fréchet}}(S, \text{or}_{\text{Fréchet}}(S1, \text{or}_{\text{Fréchet}}(K1, R))))),$$

which reflects the analyst's level of ignorance more accurately than the more traditional assumption of independence. The result of this calculation is the interval $[3\times10^{-5}, 1.4\times10^{-4}]$. These bounds are both rigorous and best possible. They are *rigorous* because they are sure to contain the true frequency of the top event, so long as the input

intervals contain the true failure rates of the respective components and, of course, that the model by which we combined these inputs is correct. The bounds are *best possible* in the sense that they could not be any narrower and still contain all outcomes that might actually arise given the uncertainties about the dependencies. If, on the other hand, one assumes independence when it's not justified, the error can result in estimates that are too tight and may therefore lead to underestimation of risks. For this reason, the default should be no assumption rather than an independence assumption. Figure 6 summarizes the effect of the different treatments of dependency in the fault tree model discussed thus far.

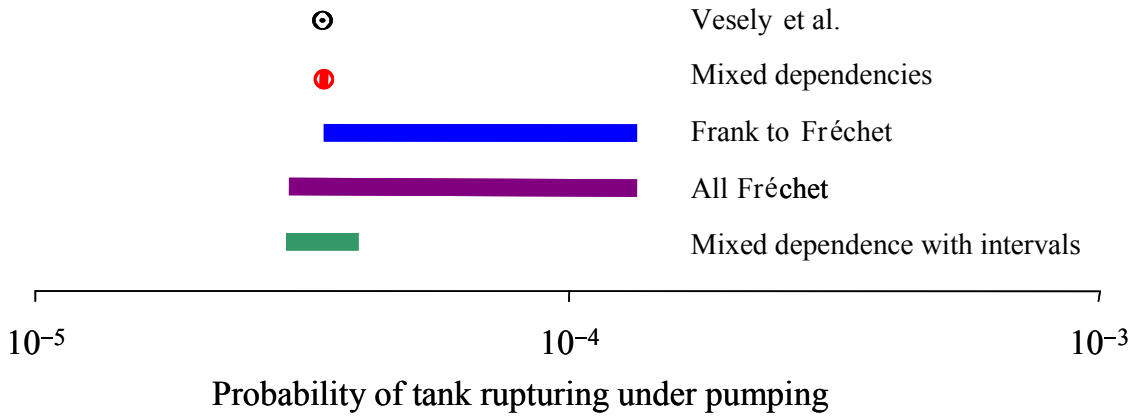

Figure 6: Comparison of the effect of assumptions regarding dependence between variables on the outcome of the probability calculation.

All the calculations thus far in the event tree example assumed that the inputs were precise point estimates. An analyst should of course use whatever information is available to get estimates that are as precise as possible. As Vesely et al. (1981) acknowledge, however, extreme precision is not required for these failure rate and importance calculations, and it would probably not be believed if it were provided. Instead, what are typically sought are order-of-magnitude results. How does uncertainty in the inputs affect the result? To find out, we can replace the precise point estimates with intervals representing that uncertainty. Suppose the analyst now has these inputs to the event tree:

Component	*Probability interval*
Pressure tank T	$[4.5 \times 10^{-6}, 5.5 \times 10^{-6}]$
Relay K2	$[2.5 \times 10^{-5}, 3.5 \times 10^{-5}]$
Pressure switch S	$[0.5 \times 10^{-4}, 1.5 \times 10^{-4}]$
Relay K1	$[2.5 \times 10^{-5}, 3.5 \times 10^{-5}]$
Timer relay R	$[0.5 \times 10^{-4}, 1.5 \times 10^{-4}]$
On-switch S1	$[2.5 \times 10^{-5}, 3.5 \times 10^{-5}]$

How does the uncertainty in the input failure rates affect the assessment of the failure probability of the top event? The event tree equation incorporating the mix of dependencies

$$E1 = or_{independent}(T, or_{Fréchet}(K2, and_{Frank}(S, or_{independent}(S1, or_{perfect}(K1, R)), r))),$$

and using all interval inputs yields the answer $[2.9\times10^{-5}, 4.1\times10^{-5}]$. This can be considered the comprehensive answer because it incorporates what is known and what is not known about both the marginal probabilities and their dependencies. Its range is of course much wider than the probability interval based on the same dependencies but with point estimates for the event probabilities. The answer can also be compared with the traditional answer of 3.5×10^{-5} computed by Vesely et al. (1981) which was based on point estimates and independence among all the inputs. The breadth of the range represents uncertainty of about 30% of the point estimate's magnitude. In this case, the uncertainty is roughly symmetric around the point estimate. Given the stated uncertainty in the inputs, the probability of the tank rupturing under pressurization might be as large as 0.000041, or about one in 25,000. Or it might be as low as 0.000029 or about one in 34,000.

3 Dependence between random variables

This section reviews the most important strategies a risk analyst can use to account for knowledge about correlations and dependencies between the random variables in an assessment. Several recent reviews have considered strategies to model inter-variable dependencies in probabilistic models (Helton and Davis 2003; 2002; Henderson et al. 2000; Clemen and Reilly 1999; Cullen and Frey 1999; Cario and Nelson 1997; Cooke 1997; Hutchinson and Lai 1990; Morgan and Henrion 1990). These reviews have addressed the issue with respect to variables that are characterized by (precise) probability distributions. This section focuses on those techniques that can also be applied to Dempster-Shafer structures and probability boxes. Section 4 addresses methods that can be used when information about dependence is incomplete.

Sections 3.2 through 3.8 consider a variety of possible approaches, including both mechanistic and phenomenological modeling of dependence. In the former, a risk model involving dependent variables is re-expressed in terms of one involving variables that are mutually independent. In the latter, methods to generate correlated random deviates are used in a statistical model that does not seek to represent the underlying physical relationships behind the dependencies. The strategies for handling dependencies that are discussed in this section include counterfactually assuming independence (Section 3.1), functional modeling of dependence (Section 3.2), stratification (3.3), conditioning (3.4), assuming perfect or opposite dependence (3.5), simulating observed correlations (3.6), using parameterized copulas (3.7), and appealing to empirical copulas (3.8) or constructed copulas (3.9).

3.1 Assume independence anyway

The zeroth strategy for dealing with dependencies among variables is to simply ignore them and assume all variables are independent of one another. In fact, it is still common practice among risk analysts in many quarters to assume independence among variables even when this assumption is known to be false. The reasons for this are many, ranging from mathematical convenience and the laziness of the analyst, to the preliminary nature of the analysis, to a lack of ready and workable alternative strategies and misconceptions about how important dependence can be in risk assessments. We mention this zeroth strategy mostly to emphasize its wrongness (Section 3.1.1) and dangers for risk assessments (Section 3.1.2).

3.1.1 Caveat: uncorrelatedness does not imply independence

Setting aside, for the moment, the cases of correlated variables, this section considers the cases in which random variables are *uncorrelated*. One might expect or hope that dependence, at least in these cases, would have a negligible or small effect on convolutions. This turns out to be false. Although most risk assessors recognize that

uncorrelatedness between variables does not formally imply their independence, many are apparently not aware of how much difference dependence can make in their calculations. Many patterns of dependence produce the same correlation, and, in particular, there are a *lot* of ways a joint distribution can yield uncorrelated variables.

Consider a small example in which X and Y have the same marginal distribution, which is a discrete uniform on the integers from 1 to 25. Thus, the chance that X is 1 is 1/25; the chance that it's 2 is 1/25, and so on, and the same for Y. What can be said about the sum $X+Y$ if we suppose that X and Y are uncorrelated? Consider the ten dependence patterns in Figure 7. The abscissa of each plot is the value of X and the ordinate is the value of Y. Because the distributions are discrete, there is a mass (of size 1/25) allotted for each of twenty-five columns in each plot. Likewise, the same amount of mass is allotted for each of twenty-five rows. To make the illustration easy to understand, let's further suppose that all of each row's mass is concentrated into a single slug of density located at some x-value, and all of each column's mass is likewise condensed at one y-value. (This is different from our previous assumption that the mass in the marginal distributions were at discrete points. We're not only saying that the mass has to be at the integer points, but also that there is only one y-value that has mass for each x-value.) By rearranging these masses on a 25×25 grid, we can create different joint distributions between X and Y. We will consider only those distributions that respect the marginal distributions by constraining our arrangements so that each row has only one mass and each column has one mass. Of these, we consider only those patterns that also have a correlation equal to zero (or so close to zero as to be appropriate for our example). Even with all of these constraints, there are still many possible arrangements. Figure 7 depicts only a few of them.

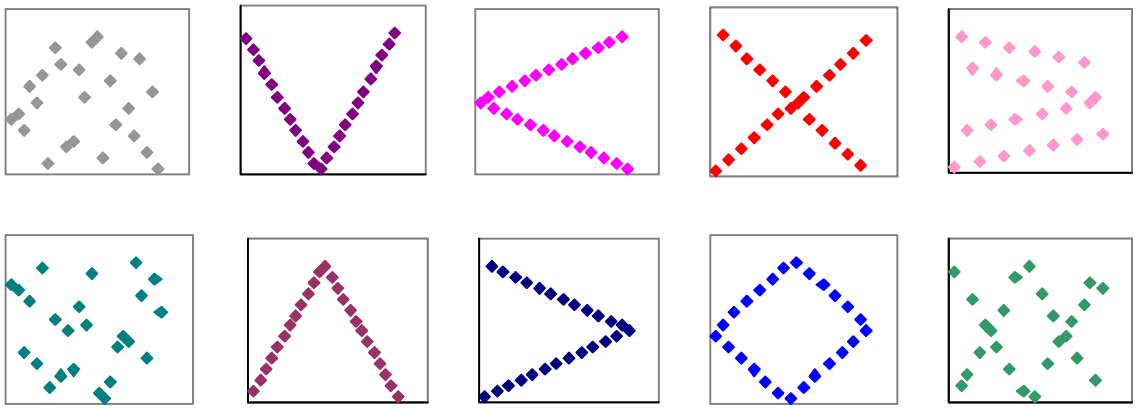

Figure 7: Some possible patterns of dependence between uncorrelated variables.

Now consider how these patterns of dependencies, all of which are uncorrelated, influence the distribution of an arithmetic combination of X and Y. Figure 8 shows the distributions of $X+Y$ associated with each of the ten patterns of dependence shown in Figure 7. Also shown in this figure is the distribution under independence (it's the one

going down the middle with somewhat smoother tails). It should probably not be surprising that the distribution of $X+Y$ depends on the dependence between X and Y, but many analysts are surprised to see the magnitude of its potential influence. Note, for instance, that the smallest possible value of the sum ranges between 2 and 14, depending on which pattern of dependency exists between the addends. This range is a quarter of the entire support of the distribution! Around the value 14, the cumulative probability ranges between zero and almost 30%. In other words, there might be no chance that the sum is smaller than 14, or there might be a 30% chance that it's smaller than 14.

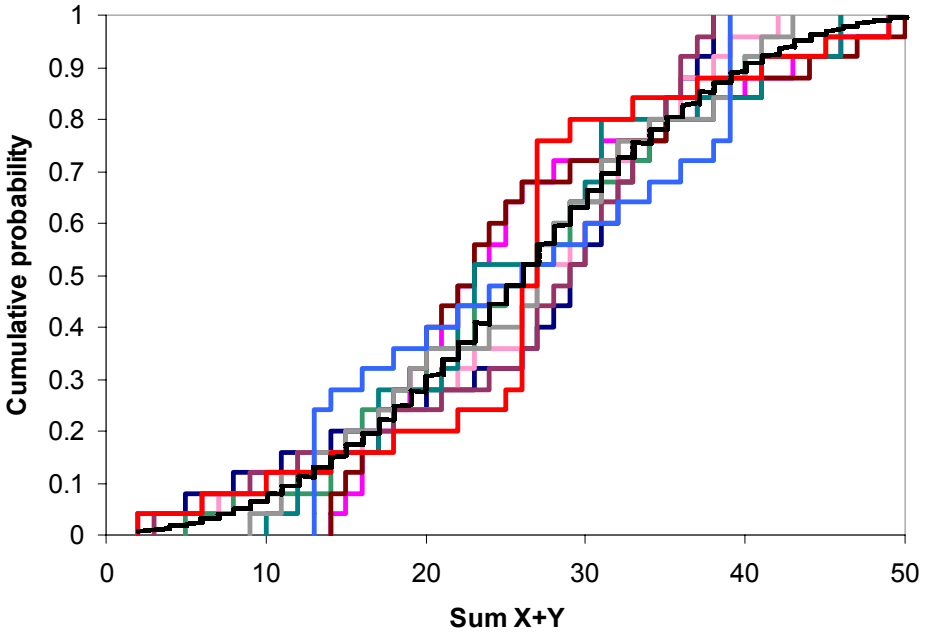

Figure 8: Some possible distributions of a sum of uncorrelated random variables.

In fact, the true uncertainty about the distribution of the sum is even larger than that depicted in Figure 8 which only depicts distributions for the independent case and 10 dependent cases. There are many more patterns of dependency that would lead to uncorrelated variables. For instance, the mass need not be concentrated in unit slugs in the joint distribution. A column's mass could be distributed throughout the column without altering the discreteness of the distributions. The results depicted in Figure 8 are only a few of the infinitely many possible outcomes that are consistent with the uncorrelatedness of X and Y and their given marginal distributions. As we shall see in Section 4.3.3, it can be shown that the region depicted in Figure 9 represents bounds on all distributions of the sum $X+Y$ that could arise when X and Y are uncorrelated and both distributions are uniform on the integers from 1 to 25. We see in Figure 9 that the

minimum value of the sum can be any integer between 2 and 21, and there could be as much as a 40% chance that the sum is less than 14. All of this uncertainty surrounds the sum of X and Y, even though their marginal distributions are precisely specified and *even though the variables are exactly uncorrelated*.

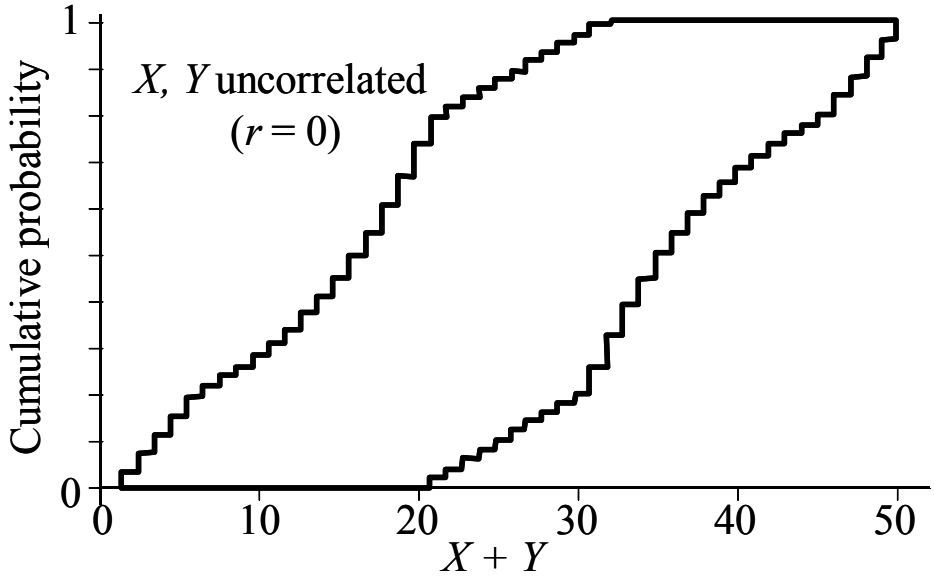

Figure 9: Bounds on the distribution of the sum $X+Y$ given that X and Y are uncorrelated and identically distributed as discrete uniforms on [1,25].

3.1.2 Unjustified independence assumptions harmful

Many risk analysts reflexively assume independence among all events or random variables even when they have no particular justification for doing so other than mathematical convenience. It is improper, however, to assume independence among variables in an analysis unless there is reliable evidence or a compelling argument that this is a reasonable assumption. If a dependency is neglected, the answer obtained by an analysis assuming independence will generally be wrong. Under certain conditions, the central tendency of output distributions could be approximately correct (Smith et al. 1992). However, the estimated dispersion and especially the tail probabilities can be highly inaccurate (Bukowski et al. 1995; Ferson and Burgman 1995; Ferson 1994). In some cases, the dispersion will be larger than it should be. In some cases, it will be smaller. In the latter, the probabilities of extreme events will likely be underestimated. These extreme events are often the primary focus of the risk assessment. They may represent very large stresses or threatening conditions that correspond to system failures and structural collapses that the risk analysis was intended to assess. In such cases, it is therefore crucial that these tail probabilities be accurately characterized and, in no circumstance, underestimated. Assuming independence without proper justification

amounts to wishful thinking and is therefore detrimental to the purposes of risk assessment to be a dispassionate and reasoned accounting of the possible adverse consequences and their probabilities.

3.1.3 Caveat: independence in imprecise probabilities

In probability theory, there are several ways to define the concept of independence between events and between random variables. For events A and B characterized by real-valued probabilities $P(A)$ and $P(B)$, independence between A and B is implied by any of the following conditions:

i) $P(A \,\&\, B) = P(A)\, P(B)$,
ii) $P(A \vee B) = P(A) + P(B) - P(A)\, P(B)$,
iii) $P(A \mid B) = P(A)$ if $0 < P(B)$,
iv) $P(B \mid A) = P(B)$ if $0 < P(A)$,

where P denotes the probability of an event. It is an elementary exercise in mathematical probability to prove that each of these four conditions implies the others (Mood et al. 1974, page 40). The case of random variables similarly has several possible definitions for independence. For random variables X and Y characterized by the joint distribution H with marginals F and G such that $P(X \le x) = F(x)$, $P(Y \le y) = G(y)$ and $P(X \le x, Y \le y) = H(x, y)$, then independence between X and Y implies, and is implied by, each of the following conditions:

i) $H(x,y) = F(x)\, G(y)$, for all values x and y,
ii) $P(X \in I, Y \in J) = P(X \in I)\, P(Y \in J)$, for any subsets I, J of the real line,
iii) $h(x,y) = f(x)\, g(y)$, for all values x and y,
iv) $P(X \le x \mid Y) = P(X \le x)$ and $P(Y \le y \mid X) = P(Y \le y)$,
v) $E(w(X)\, z(Y)) = E(w(X))\, E(z(Y))$, for arbitrary functions w and z, and
vi) $\varphi_{X,Y}(t,s) = \varphi_X(t)\, \varphi_Y(s)$.

where P is probability, f, g and h are density analogs of F, G and H respectively, E is expectation, and φ denotes the Fourier transform (characteristic function). As was true for events, when probabilities are precise these various definitions of independence between random numbers are all *equivalent*. Each definition implies all the others. Therefore, there's a single concept of independence that simultaneously embodies all of these possible definitions.

There is a decidedly different story in the context of imprecise probabilities (Walley 1991). Here, the special case of independence, which is unique in probability theory, disintegrates into several different cases when probabilities are imprecise. Couso et al. (2000) pointed out that, for imprecise probabilities (which includes both Dempster-Shafer structures and probability boxes as special cases), the various possible definitions of independence are no longer equivalent to each other. The different definitions induce distinct concepts of independence for imprecise probabilities. Couso

et al. (2000) recognized several concepts that might be called independence, of which four seem especially germane for risk assessments: repetition independence, strong independence, epistemic independence, and random-set independence. We briefly describe these four ideas in turn:

- *Repetition independence* arises when there is stochastic independence (in the traditional sense) between random variables that are identically distributed, although their distribution may be imprecisely known. Repetition independence is thus the analog in the context of imprecise probabilities of the constraint in probability theory that variables are independent and identically distributed (iid).
- *Strong independence*, on the other hand, is the complete absence of any relationship between random variables. Variables X and Y are strongly independent if the set of possible joint distributions is the largest set such that each joint distribution $H(x, y) = F(x)\ G(y)$, where F is one of the possible distribution functions characterizing X and G is one of the possible distribution functions characterizing Y. Variables X and Y should be characterized as strongly independent if (*i*) X and Y result from random experiments, each governed by a unique albeit possibly unknown probability distribution, (*ii*) the random experiments are stochastically independent (in the traditional sense), and (*iii*) there is no known relationship between the variables that would preclude some possible combinations of the possible marginal distributions.
- *Epistemic independence* arises when an analyst's uncertainty about either of two outcomes of a random experiment does not change when some information about the outcome of one of them becomes known. Random variables X and Y are epistemically independent if the conditional probability of each given the other is equal to its unconditional probability, so that $P(X|Y) = P(X)$ and $P(Y|X) = P(Y)$. In the context of imprecise probabilities, epistemic independence is defined in terms of lower bounds on expectations such that $\underline{E}(f(X)|Y) = \underline{E}(f(X))$ and $\underline{E}(f(Y)|X) = \underline{E}(f(Y))$ for all functions f where $\underline{E}(Z)$ denotes the infimum of all expectations of Z over all possible probability distributions that could characterize Z.
- *Random-set independence* is the kind of independence embodied in the Cartesian products between Dempster-Shafer structures originally proposed by Yager (1986) and used in Berleant (1993; 1996), Williamson and Downs (1990) and Ferson et al. (2003). Dempster-Shafer structures X and Y with mass functions m_X and m_Y respectively are random-set independent if the Dempster-Shafer structure for their joint distribution has mass function $m(A_1 \times A_2) = m_X(A_1)\ m_Y(A_2)$ when A_1 is a focal element of X and A_2 is a focal element of Y, with $m(A) = 0$ for all subsets not of the form $A = A_1 \times A_2$.

Couso et al. (2000) review these definitions and give simple examples of each of these definitions (and more). Couso et al. (1999) gave examples of how the definitions could influence numerical calculations. Fetz (2001) illustrated the consequences of the various independence definitions in a probabilistic assessment for an engineering

system. Cozman and Walley (2001) explored the properties of epistemic irrelevance and epistemic independence.

The various possible definitions of independence correspond to nested sets of probability distributions, as depicted with Venn diagrams in Figure 10. Couso et al. (2000) explained that the set of joint distribution functions that are repetition-independent is a subset of the set of joint distribution functions for variables that are strongly independent, which is a subset of distributions for variables that are epistemically independent which, in turn, is a subset of the distributions under random-set independence. All of these joint distributions are members of the set of uncorrelated distributions, but, as emphasized in Section 3.1.1, uncorrelatedness is not a category of independence and, therefore, it is the outer ring in gray in the figure. Beyond the category of uncorrelated lie joint distributions for correlated variables, distributions for variables with functional dependencies and more complex patterns which we collectively call simply the Fréchet case. One very important advantage of this nesting is that methods and algorithms that can compute bounds assuming random-set independence (such as those described in Ferson et al. 2003) can be used to compute sure bounds on results that are epistemically or strongly independent. In such cases, the bounds may not be best possible of course, but they may still be useful if it is practical or expedient to have conservative results.

It is no doubt possible to divide categories of independence even more finely than this picture suggests, and the discussion in Couso et al. (2000) hints at this. It is also conceivable that researchers will identify other, perhaps non-nested categories. If such definitions turn out to be important in engineering contexts, the analyses they will require will almost certainly demand a much fuller treatment than can be currently mustered in either Dempster-Shafer theory or probability bounds analysis.

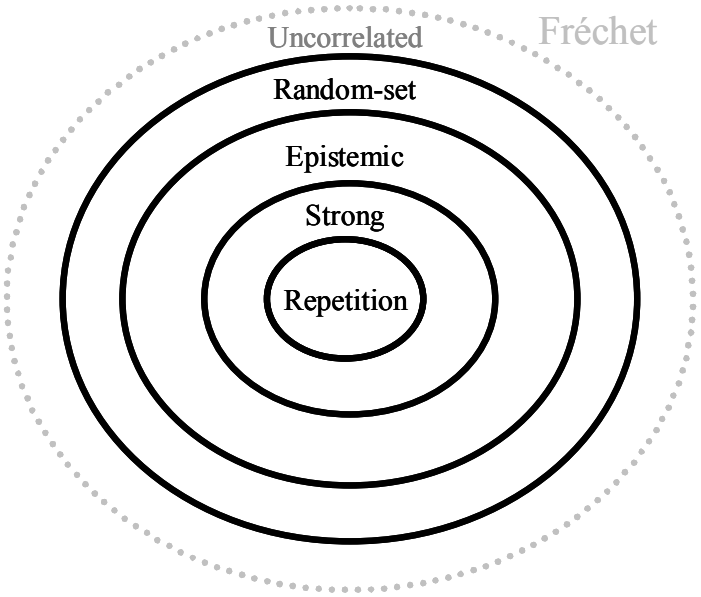

Figure 10: Categories of 'independence' for imprecise probabilities.

3.2 Functional modeling (reducing to independence)

Some methodologicians (e.g., Morgan and Henrion 1990) argue that it would be best for an analyst to reduce any risk assessment problem involving dependent variables into one involving only independent variables. This strategy, which can be viewed as an extreme form of conditioning (Section 3.4), does not try to characterize the dependencies statistically, but rather tries to sidestep the problem altogether. In the case of a risk expression involving correlated variables X and Y, this strategy would replace the Y with some function of X based on the physical relationship that produced the dependency between the variables in the first place. If this relationship is completely specified, the value of Y can be precisely determined solely by the value of X. Of course, cases of such complete predictability are very rare in science and engineering, and generally the function will involve a random error term that represents the residual uncertainty about Y after accounting for X. By construction, however, this error term can be made independent of X, and therefore the problem with two correlated variables has been changed into a different problem with two, or possibly more, independent variables.

Although this approach can require considerably more scientific understanding about the modeled system than is commonly available in risk assessments, some analysts feel this strategy is the best way to treat dependencies. For instance, the developers of the probabilistic modeling software package Analytica* suggest that any dependencies present should be accounted for and modeled explicitly (Morgan and Henrion 1998). In fact, their package does not even support user-defined correlations, so it forces users to untangle any dependencies before they can begin an analysis.

This purist approach does not always provide a workable strategy however. For example, suppose an analyst has been charged with conducting a risk assessment for vegetation wildfire in the Everglades that might be sparked by a malfunction and explosion of solid-propellant boosters used at Cape Canaveral. Such an assessment would likely be very complex and might involve considerations about current weather patterns such as a wind rose, humidity distributions, recent weather's impact on the vegetation's fire risk, and a host of sundry design and mission parameters. The model of the explosion's effects on the ground vegetation might require probability distributions for the mass and surface area for fragments of the propellant and the housing vehicle. These variables are clearly unlikely to be stochastically independent of one another. A functional modeling approach to accounting for their dependence would be to develop a submodel about how the fragments were produced in the explosion process itself. Obviously, this could significantly enlarge the modeling effort.

Even if the analyst were game to undertake the challenge of modeling the generation of explosion fragments, there could be other pesky correlations and

*Analytica is the successor to the Demos software (Morgan and Henrion 1990).

dependencies among the weather parameters. For instance, wind speed and humidity may not be independent meteorological variables in south Florida. Vegetation fire risk tends to vary over the course of a year. Therefore the timing of launches may tend to covary with fire risk on the ground. To explicate all of these dependencies by functional modeling, the analyst would need to employ (or become) a meteorologist. At some point, the analytical demands of the functional modeling approach will likely become prohibitive.

Besides the obvious disadvantage owing to the extra modeling effort that may be required by the use of functional modeling to account for dependencies in a risk assessment, there is one further caveat: it is not generally sufficient to transform the model into uncorrelated variables (Section 3.1.1); they must be statistically *independent* variables.

3.3 Stratification

Some risk analysts find it useful to stratify the assessment by creating relatively homogeneous subgroups that have similar characteristics to reduce dependencies among variables (Frey 1992; Cullen and Frey 1999). For these cases, one isolates the covariance into the difference between the groups. Within groups, the assumption of independence is more reasonable and workable. For instance, if some components were manufactured in Oak Ridge, Tennessee, and some were manufactured in Paducah, Kentucky, it may be reasonable to treat these two subgroups in completely separate analyses rather than trying to pool them together into a heterogeneous population of components manufactured at two facilities. Such stratification by age group or gender is often employed in human health assessments in part to avoid having to specify and model correlations. The separate treatment of different receptor wildlife species can also be viewed as an example of this strategy. The cost of this strategy is that the analysis becomes more complex and cumbersome because it must be repeated for each new group in the stratification.

3.4 Conditioning

A standard approach in probability theory for modeling a joint distribution has been to specify the joint distribution as a product of marginals and conditional distributions (Clemen and Reilly 1999). In this way, arbitrary intervariable dependencies can be expressed in terms of conditioning, at least in principle. For instance, it may be convenient to use distributions that are conditional on the values sampled from other distributions. This approach has been useful in hierarchical simulations (e.g., Voit et al. 1995). This strategy extends to making the parameters or even the shape of a distribution depend on the value of other random variable(s). This use of conditioning to account for dependence is essentially similar to functional modeling described above in Section 3.2 and shares its main disadvantage. The task of specifying all the necessary conditional distributions grows combinatorially with the number of variables, and Clemen and Reilly (1999) suggest that this may make the approach unwieldy for

large assessment problems. Unless most of the underlying variables are independent or conditionally independent, this strategy is information-intensive strategy and may not often be practical for risk assessments where empirical knowledge is limiting.

3.5 Perfect and opposite dependence

Random variables have perfect dependence if each is almost surely a non-decreasing function* of the other. Some authors call this relationship comonotonicity (e.g., Müller 1997; Vyncke et al. 2000; Goovaerts et al. 2000; Kaas et al. 2000). This is the distributional analog of the concept of perfect dependence for events introduced in Section 2.1. In many cases, assuming variables are perfectly dependent may be a better default strategy than one assuming they are independent. For instance, suppose the variables are component mass and surface area and the population under study includes a wide variety of components. Presuming the variables covary perfectly is probably considerably better than assuming they are statistically independent which is manifestly false. Bernat et al. (2004) describe applications in timing analysis for computer codes where perfect dependence is a good assumption because execution times of different program blocks can depend on common parameter settings.

In other circumstances, the dependence may be similar to perfect but opposite in sign. In this case, we say that the variables have opposite dependence. This often happens, for example, with reciprocal losses and gains, as well as with quantities that are constrained to add to a fixed sum. When variables are oppositely dependent, knowing that one variable is at the upper end of its range tells us that the other variable is surely at the lower end of its range. Opposite dependence between random variables is also called countermonotonicity because each variable is almost surely a non-increasing function of the other (Kaas et al. 2000).

It is generally easy to simulate perfect dependence between probability distributions (Bratley et al. 1983; Whitt 1976; Fréchet 1951; Hoeffding 1940). Saying that variables perfectly covary in this way means that knowing that one variable is large with respect to its statistical distribution implies the other variable is surely large to the same degree with respect to its own statistical distribution. This suggests the relationship $Y = G^{-1}(F(X))$ where F and G are the cumulative distribution functions for the random variables X and Y respectively. In principle, one could use this relationship to simulate variates from a specified marginal distribution that are perfectly dependent with sampled X-values. Alternatively, one could simulate values for both variables from a single uniform deviate with the assignments $X = F^{-1}(u)$ and $Y = G^{-1}(u)$ where $u \sim$ uniform(0,1). If the dependence is opposite, on the other hand, the relationship between the variates is $Y = G^{-1}(1 - F(X))$ instead. In simulations this relationship can be expressed by generating $X = F^{-1}(u)$ and $Y = G^{-1}(1-u)$ where $u \sim$ uniform(0,1). Convolutions under perfect or opposite dependence can then be estimated by operating on these simulated variates.

---

*Assuming variables are perfectly dependent is different from assuming that either is completely dependent on the other, which is a more general situation.

Simulating perfect or opposite dependence is quite different from simulating a functional dependence because the relationship is expressed through the distribution functions. And, in fact, perfect dependence between variables is not the same as a functional relationship between them. Y can be completely dependent on X and yet the dependence between X and Y is neither perfect nor opposite.

Both perfect and opposite dependence imply the quantities have extremal correlations. Perfect dependence is associated with maximal correlation for the given marginal distribution shapes. Opposite dependence induces the minimal correlation possible given the margins. In both situations, the dependence is as strong as it can be given the marginal distributions of the variables.

Example: Suppose that the random variable X is normally distributed with mean zero and unit variance, and that random variable Y is uniformly distributed between zero and one. Further suppose that X and Y are oppositely dependent, so that large values of one are invariably associated with small values of the other and vice versa. Figure 11 depicts the cumulative distribution function for the product XY. The expression "$X \backslash\times\backslash Y$" denotes the product of X and Y under opposite dependence. The slashes are mnemonic for the underlying non-increasing relationship between the variables. The distribution of this product has a mean of about −0.28 and a variance of about 0.33. This asymmetric distribution contrasts strongly with the product under independence, which is symmetric about zero.

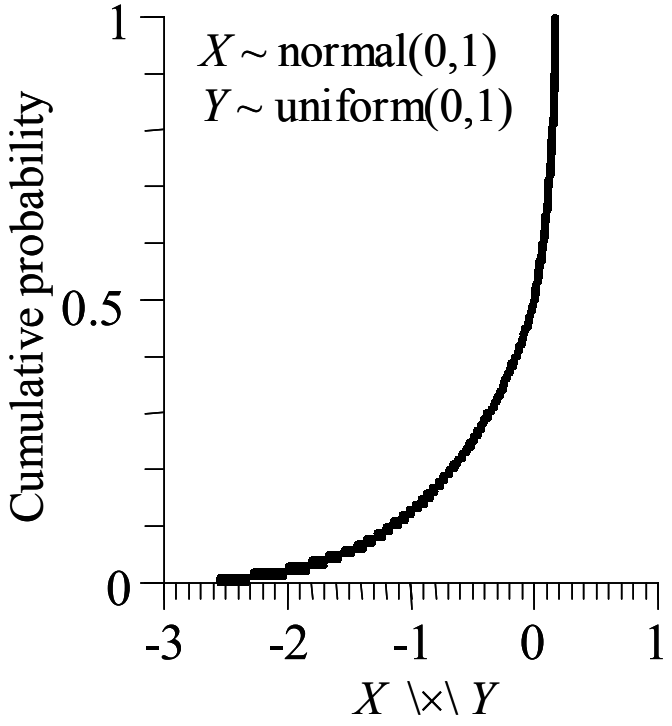

Figure 11: Distribution function of products of random variables with opposite dependence.

3.5.1 Extreme dependence with p-boxes

Perfect and opposite dependence can be assumed for convolutions involving probability boxes. For instance, consider the sum of X and Y under perfect dependence, which we can symbolize as X /+/ Y. The right bound of the p-box for X /+/ Y is defined by the inverse of

$$\left(\underline{F}_Z\right)^{-1}(p) = \left(\underline{F}_X\right)^{-1}(p) + \left(\underline{F}_Y\right)^{-1}(p)$$

where $0 \le p \le 1$, and $(\underline{F})^{-1}$ denotes the inverse or quantile functions of the bounding function. Likewise, the left bound of the p-box for X /+/ Y is defined by the inverse of

$$\left(\overline{F}_Z\right)^{-1}(p) = \left(\overline{F}_X\right)^{-1}(p) + \left(\overline{F}_Y\right)^{-1}(p).$$

Note that the functions that are added together are quantile functions of the bounds, not bounds on the quantile function (Williamson and Downs 1990). This distinction can make a difference when discretization is coarse. Figure 12 shows how this addition is done for hypothetical p-boxes A and B. For every probability level p, the corresponding values of the right bound for A and the right bound for B are added together to get the right bound at the same probability level for the sum A /+/ B. For instance, we see in the figure that, at the level p=0.21, 4 + 6.4 = 10.4. As p is varied, the right bound is traced out. Similar calculations are used to obtain the left bound.

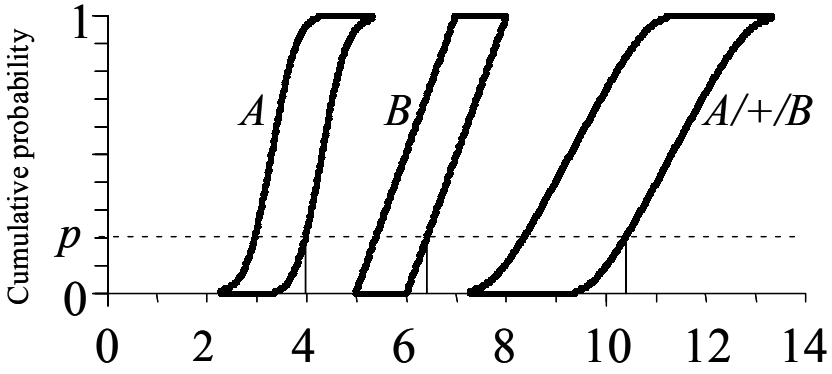

Figure 12: Addition of probability boxes under perfect dependence.

The p-box for the sum of X and Y under *opposite* dependence, which we symbolize X \+\ Y, is defined on the other hand by the inverses of the pair of functions

$$\left(\underline{F}_Z\right)^{-1}(p) = \left(\underline{F}_X\right)^{-1}(p) + \left(\underline{F}_Y\right)^{-1}(1-p)$$
$$\left(\overline{F}_Z\right)^{-1}(p) = \left(\overline{F}_X\right)^{-1}(p) + \left(\overline{F}_Y\right)^{-1}(1-p)$$

where $0 \le p \le 1$. Notice in this case that the bound at p is added to the opposite bound at $1-p$. The formulas for multiplication under perfect and opposite dependence are analogous to those for addition. For subtraction and division, however, the operations are performed not on corresponding bounds, but on the opposite bounds. For instance, the p-box for the difference $Z = X - Y$ under opposite dependence is defined by the inverse of the functions

$$\left(\underline{F}_Z\right)^{-1}(p) = \left(\underline{F}_X\right)^{-1}(p) - \left(\overline{F}_Y\right)^{-1}(1-p)$$
$$\left(\overline{F}_Z\right)^{-1}(p) = \left(\overline{F}_X\right)^{-1}(p) - \left(\underline{F}_Y\right)^{-1}(1-p).$$

Example: Suppose that $X \sim$ weibull(1,2), where weibull(d,c) denotes a Weibull distribution with scale parameter d and shape parameter c. Now suppose that Y is distributed according to a distribution function about which all that is known is its possible range is [0, 1] and its mean = 0.2. We want to characterize the product XY under the assumption that X and Y are perfectly dependent. The probability distribution for X can be characterized by a degenerate p-box whose upper and lower bounds are coincident. This p-box is depicted in the left graph of Figure 13. For the sake of plotting convenience, the Weibull distribution was truncated at its 99.5% percentile, slightly above the value 2.3. The p-box for the variable Y is depicted in the middle graph of Figure 13. Consult Ferson et al. (2003) for details about how the best possible bounds for the unknown distribution function can be obtained from constraints on the range and mean. The multiplicative convolution under perfect dependence between X and Y is symbolized as X /×/ Y. The p-box for the product X /×/ Y can be obtained by computing the left bound at any probability level from the left bounds of X and Y at the same probability level. The right bound is computed similarly from the right bounds. In principle, these bounds for X and Y could be read from the left and middle graphs of the p-boxes shown in Figure 13. In a computer implementation, the bounds on X would be obtained by evaluating the quantile function (i.e., the inverse of the distribution function) of the Weibull distribution. The values of the bounds and their products are given for several different probability levels p in the table below.

	Left bounds			**Right bounds**		
p	X	Y	X /×/ Y	X	Y	X /×/ Y
0	0	0	0	0	0.202	0
0.01	0.1	0	0	0.1	0.202	0.0203
0.02	0.142	0	0	0.142	0.204	0.029
0.03	0.175	0	0	0.175	0.206	0.036

0.04	0.202	0	0	0.202	0.208	0.0421
0.05	0.226	0	0	0.226	0.211	0.0477
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.95	1.73	0.149	0.258	1.73	1	1.73
0.96	1.79	0.158	0.283	1.79	1	1.79
0.97	1.87	0.167	0.312	1.87	1	1.87
0.98	1.98	0.175	0.347	1.98	1	1.98
0.99	2.15	0.184	0.394	2.15	1	2.15
1	∞	0.191	∞	∞	1	∞

The resulting p-box for X /×/ Y consists of the left and right bounds in the fourth and seventh columns of the table as functions of the probability level in the first column. This p-box is shown in Figure 13 as the graph on the right.

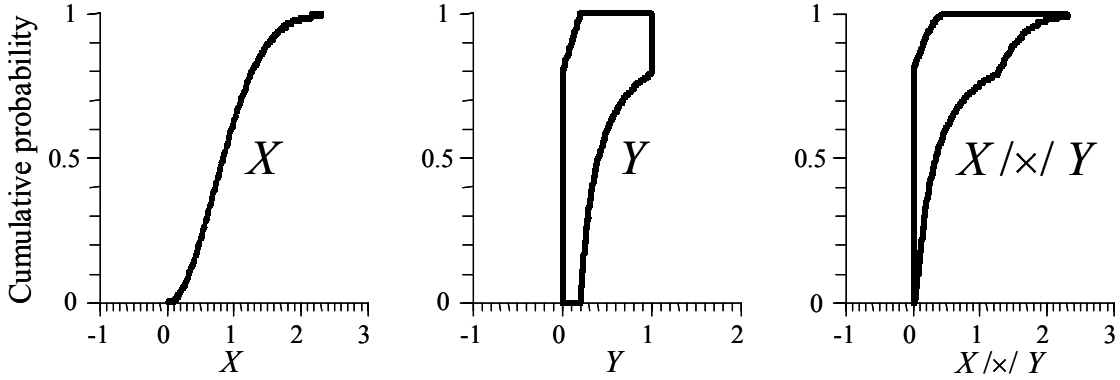

Figure 13: Probability boxes for factors and their product under perfect dependence.

3.5.2 Extreme dependence with Dempster-Shafer structures

Convolutions under extreme dependence assumptions are also possible between Dempster-Shafer structures. Berleant and Goodman-Strauss (1998) described algorithms for such calculations based on linear programming. Although they did not refer to the mathematical objects as Dempster-Shafer structures and they allowed multiple mass assignments to a single focal element, a simple normalization that condenses multiple masses would convert Berleant's objects into traditional Dempster-Shafer structures. One can often avoid the necessity of using linear programming to compute convolutions under extreme dependence by exploiting the transformations between p-boxes and Dempster-Shafer structures. The following three examples illustrate convolution of Dempster-Shafer structures under extremal dependence.

Example: Suppose that an uncertain number X is characterized by the Dempster-Shafer structure given by three equiprobable intervals {([1,3], 1/3), ([2,3],

1/3), ([3,5], 1/3)}* and that another uncertain number Y is characterized by a similar Dempster-Shafer structure {([2,8], 1/3), ([6,10], 1/3), ([8,12], 1/3)}. The lists within curly braces, consisting of pairs of intervals and masses, are Dempster-Shafer structures because the masses sum to one in both cases. Figure 14 depicts these two Dempster-Shafer structures (in the left and middle graphs) in terms of their cumulative plausibility and belief functions. Further suppose that X and Y are perfectly dependent and that we want to propagate them through an addition operation. The Dempster-Shafer structure of the convolution of X and Y assuming they are perfectly dependent, which we symbolize as X /+/ Y, can be computed with the Cartesian product depicted below.

	X			
Y	X /+/ Y	[1,3] 1/3	[2,3] 1/3	[3,5] 1/3
	[2,8] 1/3	[3, 11] 1/3	[4,11] 0	[5,13] 0
	[6,10] 1/3	[7,13] 0	[8, 13] 1/3	[9,15] 0
	[8,12] 1/3	[9,15] 0	[10,15] 0	[11, 17] 1/3

The focal elements of X and their associated masses are arrayed in the top row, and the focal elements and masses of Y are arrayed in the leftmost column. Each cell of the table consists of an interval and its associated mass. The marginal focal elements are used to compute the intervals in the interior of the table via interval arithmetic. For instance, the first focal element of X is added to the first focal element of Y to obtain [1,3] + [2,8] = [3,11]. Nine such interval additions are needed to fill up the table. If X and Y were independent, the masses associated with these intervals in the interior would be computed as the products of the masses of the marginal elements. In this case, because the quantities are perfectly dependent, only the *diagonal* elements of the Cartesian product get nonzero masses. In this example, the elements have a convenient order that makes computing the sum under perfect dependence very easy. Here, the first element of the resulting Dempster-Shafer structure arises from the addition of the first element of X with the first element of Y, the second element comes from adding the second intervals, and the third comes from adding the third intervals. In each of these three cases, the associated mass is just the same as the mass of X and Y elements. The result of this convolution is thus the Dempster-Shafer structure {([3, 11], 1/3), ([8, 13], 1/3), ([11, 17], 1/3)}. Figure 14 also shows (on the right) the cumulative belief and plausibility functions for the Dempster-Shafer structure characterizing the sum X /+/ Y.

*In this report, Dempster-Shafer structures are often specified in this format. Each pair in the list has the form (focal element, associated mass), where the masses in the list sum to one.

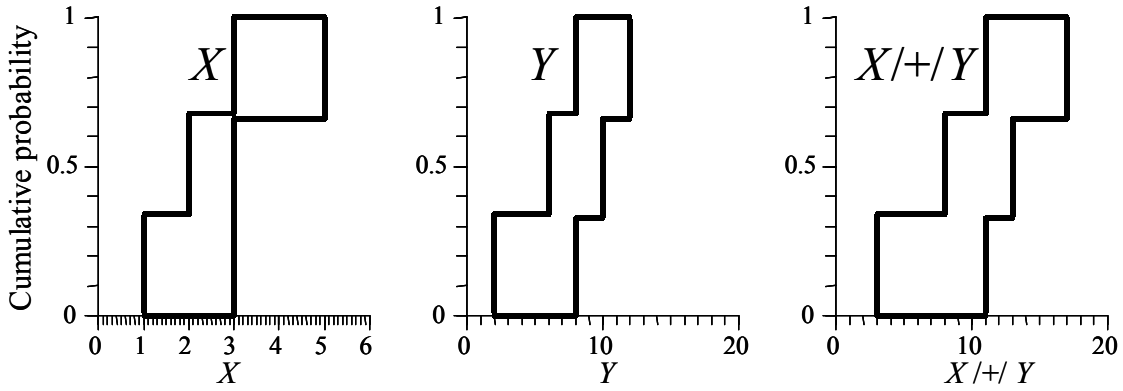

Figure 14: Cumulative plausibility and belief functions for addends and sum of Dempster-Shafer structures under perfect dependence.

Example: Another example involves a more complicated situation in which the focal elements and masses of the two Dempster-Shafer structures do not match quite so nicely. Let *X* be characterized as before by the Dempster-Shafer structure {([1,3], 1/3), ([2,3], 1/3), ([3,5], 1/3)}, but let *Y* be characterized by {([2,8], 1/4), ([3,10], 1/4), ([6,12], 1/4), ([8,15], 1/4)}. These two structures are depicted in terms of their cumulative plausibilities and beliefs in Figure 15. We seek to compute the *X* /+/ *Y*. Because *X* has three focal elements and *Y* has four, they cannot be combined using the method used in the previous example. However, that method could be used if we express the Dempster-Shafer structures in compatible forms. To do this, we expand the Dempster-Shafer structure for *X* to the twelve-element list {([1,3], 1/12), ([1,3], 1/12), ([1,3], 1/12), ([1,3], 1/12), ([2,3], 1/12), ([2,3], 1/12), ([2,3], 1/12), ([2,3], 1/12), ([3,5], 1/12), ([3,5], 1/12), ([3,5], 1/12), ([3,5], 1/12)}. This list is like a Dempster-Shafer structure in that the masses sum to unity, but it has repeated intervals. The Dempster-Shafer structure for *Y* is likewise expanded to a twelve-element list {([2,8], 1/12), ([2,8], 1/12), ([2,8], 1/12), ([3,10], 1/12), ([3,10], 1/12), ([3,10], 1/12), ([6,12], 1/12), ([6,12], 1/12), ([6,12], 1/12), ([8,15], 1/12), ([8,15], 1/12), ([8,15], 1/12)}. It is clear that these transformations neither lose information nor create additional structure beyond that in the original specifications for *X* and *Y*. The twelve-by-twelve Cartesian product can now be formed between these expanded lists. To reflect the perfect dependence between *X* and *Y*, all the mass is assigned to the diagonal elements of the matrix. When the masses are integrated for the focal elements, the result of this convolution is the Dempster-Shafer structure {([3,11], 3/12), ([4,13], 1/12), ([5,13], 2/12), ([8,15], 2/12), ([9,17], 1/12), ([11,20], 3/12)}. The cumulative plausibility and belief functions for this structure are depicted on the rightmost graph in Figure 15.

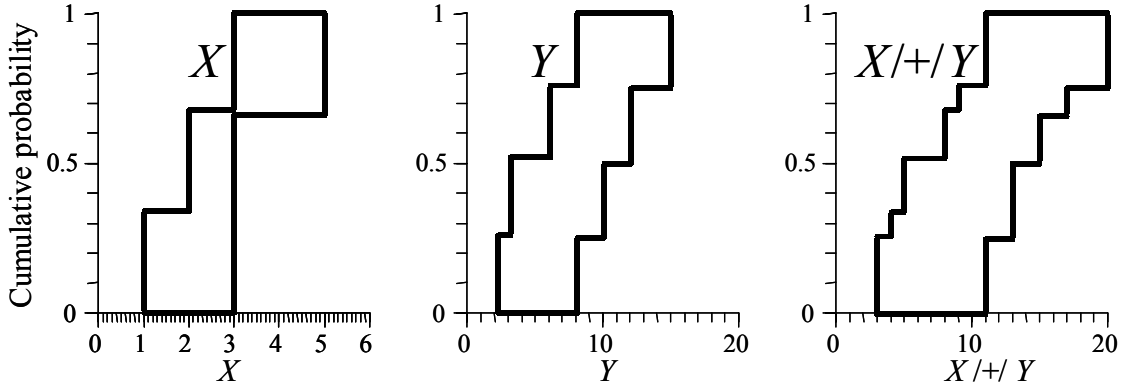

Figure 15: Addition of general Dempster-Shafer structures under perfect dependence.

Example: The focal elements of the Dempster-Shafer structures in the previous examples had a convenient order that enabled us to use a simple Cartesian product to effect convolutions under perfect and opposite dependence. Suppose we want to compute the sum X /+/ Y where X is {([1,4], 1/2), ([2,3], 1/2)} and Y is {([11,13], 1/2), ([12,14], 1/2)}. In this case, the focal elements of X cannot be arranged into a comparable Cartesian product. The equivalence relations explored in Ferson et al. (2003) suggest that a transformation of X into a related Dempster-Shafer structure could be used to do the calculation. In particular, we might replace X with the Dempster-Shafer structure {([1,3], 1/2), ([2,4], 1/2)}, which does allow the convenient ordering of its focal elements. This transformation of X leaves unchanged its cumulative plausibility and belief functions. If we then apply the method used in the previous examples, we will obtain the result {([12,16], 1/2), ([14,18], 1/2)}.

The algorithms based on linear programming described by Berleant and Goodman-Strauss (1998) give the same results in numerical experiments as the methods based on p-boxes, and Regan et al. (2004) argued that this equivalence is general. Jim Hall (pers. comm.) has pointed out that the transformation used in the third example corresponds to weak inclusion (sensu Dubois and Prade 1990) between the original and transformed Dempster-Shafer structures, and that this may have implications for what can be assumed about their relationship.

3.6 Simulating correlations

When the correlations among variables can be estimated empirically or assigned by theoretical considerations, several standard techniques are used to simulate the dependence relationships among variables in a Monte Carlo analysis. This would be useful in any sort of risk assessment or uncertainty propagation problem in which analysts would like to use their knowledge about the interactions or dependencies between the input variables to tighten or reduce the bias of the results. This section considers a few of the methods that represent *specified* correlation coefficients between

uncertain quantities. It is important to understand at the outset that merely giving a correlation coefficient, whether it be Pearson, Spearman, Kendall or some other index, is usually not sufficient to fully specify the dependence model. There must, therefore, be some further implicit or explicit constraint that makes the problem well defined.

Scheuer and Stoller (1962) described a numerical method in the general multivariate case (i.e., for two or more variables) that is still widely used to generate normal deviates with a specified variance-covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1k} \\ \vdots & \ddots & \vdots \\ \sigma_{k1} & \cdots & \sigma_{kk} \end{bmatrix}$$

where σ_{ii} is the variance of the i^{th} variable and σ_{ij}, for $i \neq j$, is the covariance between the i^{th} and j^{th} variables. Correlated normal random deviates are computed as weighted linear combinations of independent normal random deviates

$$Z_i = \sum_{j=1}^{i} c_{ij} Y_j, \quad i = 1,...,k,$$

where $Y_j \sim$ normal(0,1) are independent standard normal random deviates, and c_{ij} are the elements of a lower triangular matrix C solving $\Sigma = CC^{\text{T}}$, which can be obtained numerically by Cholesky decomposition (Gentle 1998; Press et al. 1992). Recursive formulas for the elements of C are given by Scheuer and Stoller (1962). The resulting Z_i are normally distributed, each with zero mean and unit variance, and they have the desired cross correlations. The final step is to rescale the values so they have the variance-covariance matrix Σ and translate them to the desired central tendencies. The Scheuer and Stoller method can generate extremal correlations (Section 3.5) by setting the appropriate correlations to ± 1. Because all the marginals are normal, correlations can be varied over the entire range $[-1, +1]$.

In the bivariate case $k = 2$, these equations reduce to

$$X_1 = Y_1 \sigma_1 + \mu_1$$
$$X_2 = \left(r Y_1 + \sqrt{1 - r^2}\, Y_2 \right) \sigma_2 + \mu_2$$

where r is the desired (Pearson product-moment) correlation between the two random variables, μ_1 and μ_2 are their respective desired means, σ_1 and σ_2 are their standard deviations, and Y_1 and Y_2 are independent standard normal deviates.

Example: Suppose we wish to estimate the distribution function of $X+Y$ where $X \sim$ normal(10,2) and $Y \sim$ normal(5,1) and X and Y have Pearson correlation 0.5. In this problem, we can use the formulas for the bivariate case with $\mu_1 = 10$, $\sigma_1 = 2$, $\mu_2 = 5$, σ_1

= 1, and $r = 0.5$. The value of $\sqrt{(1-r^2)} = 0.866$. Figure 16 depicts the distribution of the sum $X+Y$ as computed by a Monte Carlo simulation involving 1000 replications. The observed correlation between the variables was 0.510, compared to the planned value of 0.5. The observed mean of the resulting normal distribution was 15.11, compared to the theoretical expectation of 15 (the mean of a sum is not affected by correlation). The observed variance was 7.12. The theoretical variance can be computed from the covariance. The Pearson correlation is defined to be the covariance of the variables divided by their standard deviations. Thus, for the correlation of 0.5, the covariance would be 1. This allows us to compute the theoretical variance of the sum of these correlated normals as $V(X) + V(Y) + 2\ cov(X,Y) = 7$, rather larger than the variance under independence, which would be $V(X)+V(Y) = 5$.

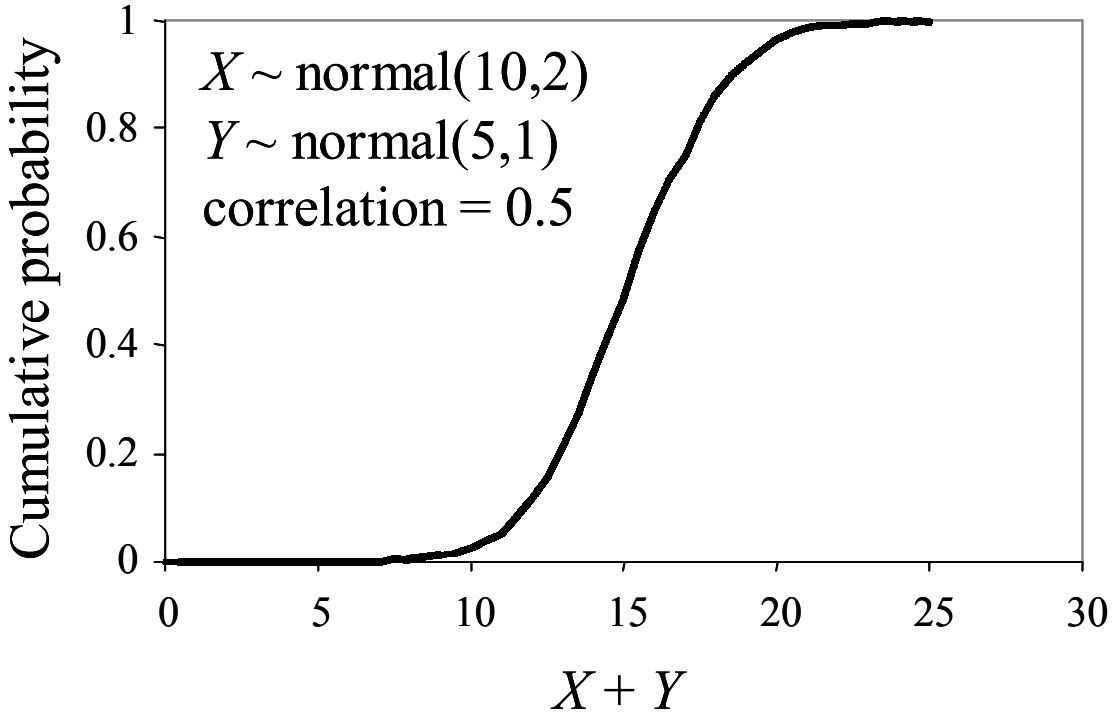

Figure 16: Distribution of the sum of correlated normals.

The Scheuer and Stoller method is accurate, numerically inexpensive and suitable for use in multivariate settings, but it is restricted to normally distributed variables. Although analogous methods can be derived for other distribution shapes (e.g. Song 1997), such a parametric approach seems impractical given the great variety

of distribution shapes* encountered in practical risk assessments. But, most importantly for this report, it does not seem possible that this method can be easily generalized for use with Dempster-Shafer structures and probability boxes. The reason for this is not because the methodology is based on sampling; one can 'sample' from a Dempster-Shafer structure or a p-box by selecting a random deviate $u \sim$ uniform(0,1) and taking as the sample the interval formed by the u^{th} quantiles of the cumulative plausibility and cumulative belief functions. One can also compose Dempster-Shafer structures and p-boxes from sets of such interval samples (see Ferson et al. 2003). The reason that the Scheuer and Stoller approach does not extend easily to the context of imprecise probabilities is that it requires that the marginal distributions be decomposed into distribution shape and parameters. The epistemic uncertainty embodied in Dempster-Shafer structures and p-boxes cannot be partitioned in this way. Their epistemic uncertainty is distributional too. Although one can specify a p-box by giving the shape and, say, interval bounds on the parameters such as the mean and standard deviation, it is not possible to characterize every p-box in this way. Dempster-Shafer structures can also be complicated structures that cannot be decomposed quite so handily. The method is not modeling the dependencies per se, but rather sidestepping the problem by inducing a superficial conformance of the simulated deviates with the specified dependencies.

Iman and Conover (1982) described another, more robust technique for simulating deviates from distributions of general distribution shapes and Spearman rank correlation structure. Iman and Davenport (1980; 1982a) gave many example scatterplots illustrating the results of the technique on assorted combinations of marginal distributions. Helton (1993, section 3.2; Helton and Davis 2002, section 5) reviewed this method and its compatibility with Latin hypercube sampling techniques, giving an overview of the algorithm, illustrative examples and an argument for the naturalness of Spearman correlation as how most people intuitively think about correlation. There are, nevertheless, some caveats (see Section 3.6.2) that an analyst should keep in mind about how the traditional Pearson correlation coefficient and rank correlation can differ substantially. The approach assumes that the joint distribution is a transform of a joint normal distribution and it uses a variant of the Scheuer and Stoller method. Interestingly, however, it does not seem to create extremal dependencies by setting the input correlations to ±1. The Iman and Conover method was criticized as ad hoc by Clemen and Reilly (1999) and considered dismissively by Cooke (1997), but it is fairly simple to implement and it can be applied to arbitrary marginal distributions in the multivariate context. Because it was the first method with these important features, it has been widely adopted. For instance, the Iman and Conover method is used in the Crystal Ball software package (Eric Wainwright, pers. comm.; Decisioneering 1996; Burmaster and Udell 1990; Metzger et al. 1998) and is probably the most widely used method for inducing correlations in Monte Carlo simulations. However, like the

*Lurie and Goldberg (1994) described an iterative approach for obtaining a desired pattern of Pearson correlations matching specified marginal distributions, but it is essentially a trial-and-error approach that can be computationally intensive.

Scheuer and Stoller method, the Iman and Conover method does not seem to generalize easily to Dempster-Shafer structures or p-boxes.

Nelsen (1986, 1987) gave methods to simulate bivariate deviates from distributions having arbitrary marginal shapes and arbitrary rank correlation (measured with either Spearman's ρ correlation or Kendall's τ correlation). Clemen and Reilly (1999) described another method based on the dependence model expressed in multivariate normal distributions. When the marginals are normal, the correlation can be specified with the Pearson coefficient, but because the dependence function can be freely wedded to arbitrary marginals, the approach immediately applies to all other distribution shapes too. The resulting correlations are no longer the specified Pearson's coefficients, but the transformation leaves rank correlations unchanged. Cario and Nelson (1997) described yet another very general analytical approach to the problem and spell out how it too can be applied in the multivariate case.

These recent approaches all rely on transformations that are usually expressed in terms of the theory of "copulas" (Schweizer 1991; Nelsen 1999). Copulas are simply the dependence functions that knit together marginal distributions to form their joint distribution. In fact, the copula is also called the "dependence function". Copulas have very simple structures and have many useful properties. They greatly simplify the generation of correlated random numbers. Consider, for example, the method of Clemen and Reilly (1999). This method can be described in three steps. Given k marginal distributions characterizing k random variables and a $k \times k$ positive semi-definite matrix of correlation coefficients (considered to be either Spearman correlations or Kendall correlations) that characterize the dependence between these random variables, the first step is to translate the given correlations into Pearson correlations. If the given values are Spearman correlations ρ_{ij}, compute $r_{ij} = 2\sin(\pi\rho_{ij}/6)$, $i,j = 1,\ldots, k$. If the given values are Kendall correlations τ_{ij}, compute instead $r_{ij} = \sin(\pi\tau_{ij}/2)$. These formulas depend on the family of copulas employed. The Clemen and Reilly method uses the normal family of copulas, so named because they are exactly the dependence functions embodied in all multivariate normal distributions. The second step is to generate k correlated unit normal random deviates Z_i using an approach such as that of Scheuer and Stoller (1962) according to the computed Pearson correlations. The third and last step is to compute

$$X_i = F_i^{-1}\left(\Phi(Z_i)\right)$$

for each i = 1, …, k, where Φ is the standard normal distribution function, and F_i^{-1}. is the inverse function of the desired marginal distribution for variable i. This process is repeated as many times as the analyst needs correlated deviates. Because the Spearman and Kendall correlations are invariant to monotone transformations, the X_i will have the same correlations as the Y_i. However, the Φ transformation changes the unit normal deviates into unit uniform deviates, and the F_i^{-1} functions changes unit uniform deviates into deviates that have the desired marginal distributions. Therefore the X_i have the

correct marginal distributions and the desired nonparametric correlations. Cario and Nelson (1997) call the strategy of transforming correlated normals to arbitrary marginal distributions "NORTA", an acronym for "normal to anything". This strategy is what gives the Iman and Conover method its flexibility too. Henderson et al. (2000) noted that, although this approach is quite flexible, there are possible variance-covariance structures and sets of marginals distributions that it cannot generate. They described an adjustment to the algorithm that will produce results that approximate the desired covariance.

Very similar strategies can be mounted for other copula families if one has access to an algorithm for creating correlated deviates within that family and a way to express the intended correlation in terms of the parameter(s) of the family. What had previously been the difficult part of accounting for arbitrary marginals in a multivariate context has been made straightforward. This copula approach must be considered the state of the art. It has proven so simple and fruitful that it has lately spawned something of a cottage industry in developing methods for generating correlated random numbers to specification. Nelsen (1986) used the Frank family of copulas. Nelsen (1987) used a family of copulas formed as convex linear combination of the Fréchet limiting cases. Clemen and Reilly (1999) used the normal family, and Kurowicka et al. (2001) used the elliptic family. These different families and parameterizations have various advantages and disadvantages, although these details are beyond the scope of this report. The normal family is actually among the most computationally intensive algorithms because it involves evaluating the normal distribution function. This function, which has no closed form expression, is of course widely available (e.g., Abramowitz and Stegun 1964). It is, for instance, the NORMSDIST function in Microsoft Excel. The operations needed for the Frank family, in contrast, are very convenient to compute, although the functions to translate the correlations are given as lookup tables.

Example: Suppose that X is lognormally distributed such that the mean μ of $\ln(X)$ is 5 and its geometric standard deviation $\sigma = 1$, and that Y has a beta distribution with parameters 10 and 2. Further suppose that the Spearman correlation between X and Y is 0.8. The complete calculations to compute the correlated random variables can be encapsulated in the following pseudocode (when we cannot specify values because they are random, we give the Microsoft Excel expressions that could be used to generate them):

$$
\begin{aligned}
r &= 2\sin(\pi\,\rho / 6) = 0.8135 \\
W_1 &= \Phi(U_1) = \text{NORMSINV(RAND())} \\
W_2 &= \Phi(U_2) = \text{NORMSINV(RAND())} \\
Z_1 &= W_1 \\
Z_2 &= r\,W_1 + \text{sqrt}(1-r^2)\,W_2 = 0.8135\,W_1 + 0.5816\,W_2 \\
X &= F_1^{-1}(\Phi(Z_1)) = \text{LOGINV(NORMSDIST}(Z_1),5,1) \\
Y &= F_2^{-1}(\Phi(Z_2)) = \text{BETAINV(NORMSDIST}(Z_2),10,2)
\end{aligned}
$$

where U_i are independent unit uniform deviates and words in all capital letters are functions in Excel syntax. Figure 17 shows 200 hundred random pairs of (X, Y) resulting from these calculations. Iman and Davenport (1980; 1982a) show a variety of comparable scatterplots for various different marginal distributions and Spearman correlations.

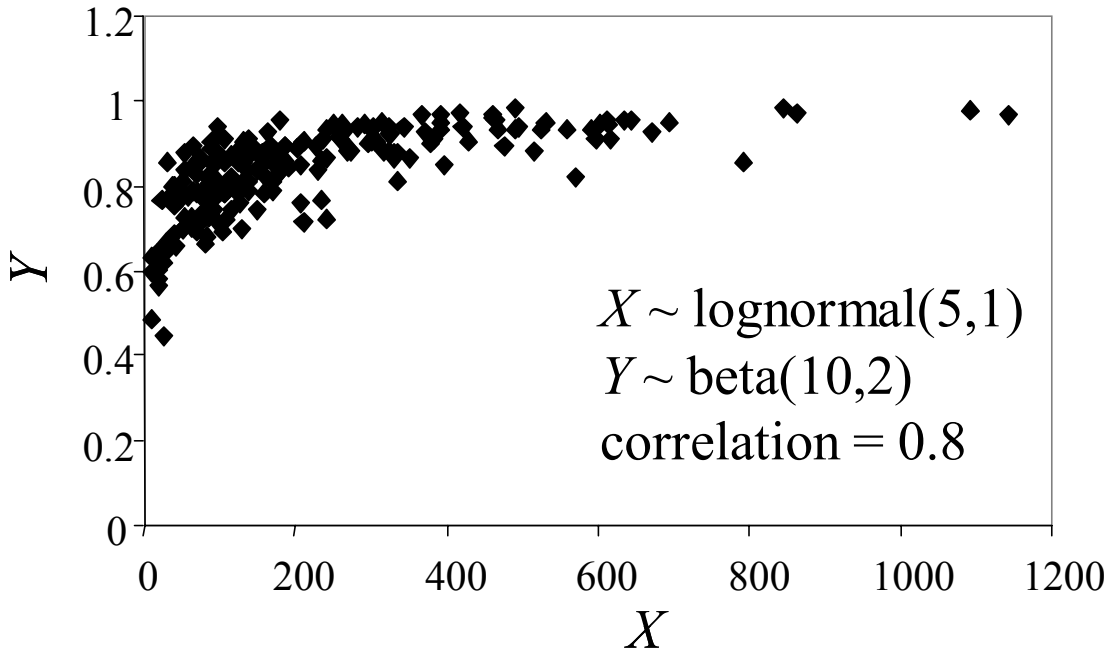

Figure 17: Scattergram depicting correlation (ρ=0.8) for non-normal marginals.

3.6.1 Feasibility of correlation matrix

There are mathematical constraints associated with correlations. For instance, one variable cannot be strongly positively correlated with each of two variables that are themselves strongly negatively correlated. Such constraints can be summarized by saying the matrix of correlations must always be a positive semi-definite matrix. Checking for positive semi-definiteness requires a special algorithm (Iman and Davenport 1982b). An infeasible correlation matrix amounts to gibberish for the model that uses it. If the input correlations are the result of coherent empirical studies, this will never be a problem. However, if the inputs are mixed results from different studies, or if they are based on hypothetical values or best professional judgments about correlations, infeasible configurations may be specified.

Many specially developed computer codes and even some commercially available software packages for Monte Carlo simulation do not check that the input correlation matrix satisfies the positive semi-definiteness condition. It is not clear what

their algorithms for generating correlated deviates would do if the feasibility condition is not satisfied. It is therefore important for analysts always to check that the input corresponds to a feasible correlation matrix.

If the matrix is positive semi-definite, then it is a possible correlation matrix. If it is not positive semi-definite, then it cannot be a correlation matrix in the first place and certainly should not be used in modeling dependencies in a risk analysis. This is the result of some impossible scenario such as variable X being strongly positively correlated with both Y and Z, but Y and Z being strongly negatively correlated with each other. It may even be possible and useful to employ the positive semi-definiteness of correlation matrices to tighten some interval estimate of correlation. For instance, knowing the correlations between X and Y and between X and Z can constrain the correlation between Y and Z to an interval smaller than $[-1, +1]$.

Unfortunately, this strategy of using available information about the relationships among some variables to inform us about the relationships among others does not extend to feasibility constraints on the qualitative (sign) information about dependencies, which is especially weak. Even information about how some variables are perfectly or oppositly dependent does not induce constraints that can be used to make inferences about unknown dependencies. Indeed, seemingly self-evident inferences involving extremal dependencies are demonstrably false. For instance, suppose A and B are oppositely dependent and that A and C are oppositely dependent. Thinking something like "the enemy of my enemy is my friend", one might expect that it would be possible to infer from this that B and C are perfect dependent. However, this is not a correct inference. Although one can infer that B and C could not be oppositely dependent, they may be independent. Here is a simple example. Consider discrete distributions such that there are four possible configurations as given in the following table.

A	B	C
1	3	3
2	1	3
2	3	1
3	1	1

It is easy to see by plotting these three variables against each other in various combinations, that A and B are oppositely dependent on one another, as are A and C. (Their Pearson correlation is - 0.707, but their Spearman correlation is - 1.) Nevertheless, B and C are independent. Likewise, it is very easy to construct examples in which a variable X is perfectly associated with both Y and Z, and yet the variables Y and Z themselves are independent. Thus, one cannot use information that some variables are maximally dependent to infer very much about other variables.

3.6.2 Caveats

There are four major caveats to consider when applying any of the methods discussed above to generate correlated variates.

(1) The first caveat is that there are many measures of correlation, and an analyst must specify which correlation is intended. The correlation measures are sensitive to different aspects of the possible dependence between variables. For instance, consider Figure 18, which depicts two dissimilar patterns of bivariate dependence. The Pearson correlation for the data in the graph labeled A is 0.786. The Spearman correlation is a comparable value of 0.794. These values agree that there is a moderately strong positive relationship between the two variables. In contrast, the Pearson correlation for the data depicted in graph B is 0.655, suggesting a somewhat weaker relationship. However, the Spearman correlation for the data in graph B is 1.0 because the points are monotonically increasing. The discrepancy arises, of course, from the fact that this perfect monotonic relationship is still strongly nonlinear.

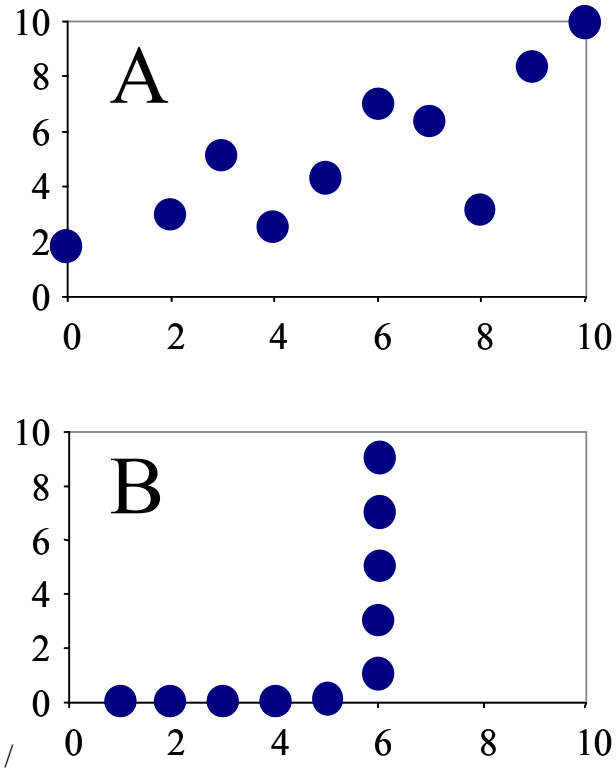

Figure 18: Two patterns of bivariate dependence.

(2) The second caveat about generating correlated variates is that one cannot reconstruct complex or idiosyncratic nonlinear dependencies. It is essential to understand, for instance, that some of the algorithms mentioned above that generate correlated random

deviates cannot produce a pattern of bivariate dependence that looks anything like the pattern shown in graph B of Figure 18, no matter what correlation coefficient is used as input.

(3) Although the methods are fairly well developed for simulating correlations and other dependencies in probabilistic risk analysis, these methods are not very useful if the nature or degree of correlation is unknown. This is the third caveat about generating correlated variates. If the analyst believes the variables may be correlated or dependent in some way, but does not know the magnitude or details of the interaction, the methods described above cannot be used because the methods need these details. Estimating correlations usually requires paired or joint observations that can often be difficult or sometimes even impossible to collect. Section 4 describes some approaches to accounting for partial or even total ignorance about dependence.

(4) The presence of correlated variables in a regression-based sensitivity analysis can produce strange and potentially misleading results (Helton and Davis 2000; Helton 1997). This is not a good reason to neglect correlations when they exist, but care should be exercised.

3.7 Parameterized copulas

As discussed in Section 3.6, specifying a correlation coefficient is insufficient to fully determine the dependence between two variables. Methods that seem to do this are implicitly assuming some copula family that takes the correlation and specifies a particular dependence function. This section shows how the theory of copula can be used directly and explicitly to compute arithmetic functions between the variables with arbitrarily complex dependencies.

It is important for an analyst to be able to model dependencies (rather than merely correlations) because complex dependencies routinely appear among variables in risk assessments and these dependencies can have profound impacts on the numerical results of risk calculations. Complex dependencies are certainly not rare. They are perhaps as common as nonlinearity generally in physical systems. Vesely et al. (1981) argued that dependencies may often dominate system performance. Hickman et al. (1983, their section 3.7) discuss a variety of reasons that dependencies can arise in nuclear power plants, including common-cause initiating events, functional dependencies, shared-equipment dependencies, physical interactions, human-interaction dependencies, and inter-component dependencies. Hart et al. (2004) discuss a variety of mechanisms that lead to complex dependencies.

Although there is a vast literature in risk analysis and uncertainty modeling on selecting marginal distributions (see Haimes et al. 1994; Morgan and Henrion 1990; Cullen and Frey 1999; inter alia), there has been considerably less consideration of how an analyst should select and implement models of dependence (but see Cooke 2002; Haas 1999; Hutchinson and Lai 1990). How would a risk analyst make use of detailed information about dependence in an assessment? How can we make use of our

knowledge about the constraints that prohibit or favor certain combinations of variable values in order to obtain better calculations and reduce our uncertainty about them? It turns out that it is relatively easy to do these things by making fuller use of the theory of copulas. The rest of this section gives a sketch of this theory.

Given marginal distributions F_X and F_Y characterizing random variables X and Y respectively, the Lesbesgue-Steiltjes integral

$$F_Z(z) = \int_{z=x+y} dC\big(F_X(x), F_Y(y)\big)$$

(which always exists) gives the distribution of the sum $Z = X+Y$ in terms of the dependence (copula) function $C(u,v)$ between X and Y. If the variables are independent, so that $C(u,v)=uv$, and the integral for the distribution function of Z reduces to

$$\int_{z=x+y} F_X(x)dF_Y(y) = \int_{-\infty}^{z} F_X(z-t)dF_Y(t)\,.$$

There are very similar formulas for the distribution of differences, products, quotients, etc., where the plus sign in the condition specifying the integral is replaced by a minus, multiplication sign, division sign, etc. We mention these formulas only in passing. In practice, they are very rarely actually used in risk assessments or other applications of probabilistic modeling, except in a few very simple cases (simulation methods are used instead). We mention these formulas only to say that their assumption of independence makes them far less useful for general problems than might be supposed.

The probability that a random variable is within a closed region can be estimated by integrating its probability density function over that region. In many cases, however, it can be far easier to compute it directly from the joint distribution function at the corners of the region. For example, the probability associated with the region $x_1 \le X \le x_2$, $y_1 \le Y \le y_2$ can be computed as the diagonal difference

$$H(x_1, y_1) - H(x_1, y_2) - H(x_2, y_1) + H(x_2, y_2),$$

where H is the cumulative bivariate joint distribution function for X and Y. Figure 19 shows how this diagonal difference finds the mass associated with a rectangle. The masses that are added together are depicted with hatching that slants upward; masses that are subtracted have hatching that slants downward. Because H is the joint cumulative distribution, the result is the cumulative mass that is associated with the rectangle marked b, that is, the rectangle limited by $x_1 \le X \le x_2$, $y_1 \le Y \le y_2$. Using this fact, the probability distribution function for $Z = g(X, Y)$, for some function g, can be estimated by the discretization

$$F_Z(z) = \sum_{z=g(x,y)} H(x+\Delta x, y+\Delta y) - H(x+\Delta x, y) - H(x, y+\Delta y) + H(x,y)\,.$$

The summation is needed because there could be several combinations of X and Y that produce the same value of Z, so the probability of each should be added together. Sklar's (1959) theorem tells us how to compute the joint distribution function $H(x, y)$ from specified marginal distributions and a dependence function represented by a copula. For any (univariate) distribution functions F_X and F_Y and any copula C, the function

$$H(x, y) = C(F_X(x), F_Y(y))$$

is a two-dimensional distribution function having marginals F_X and F_Y. If F_X and F_Y are continuous, then C is unique. This decomposition of a joint distribution into its marginals and the copula that knits them together is a very general approach. The marginal distributions can be specified arbitrarily. The discretization can be made arbitrarily fine to achieve any desired precision. Sklar's theorem generalizes to dimensions higher than two (Nelsen 1999; Cossette et al. 2001; cf. Bernat et al. 2004).

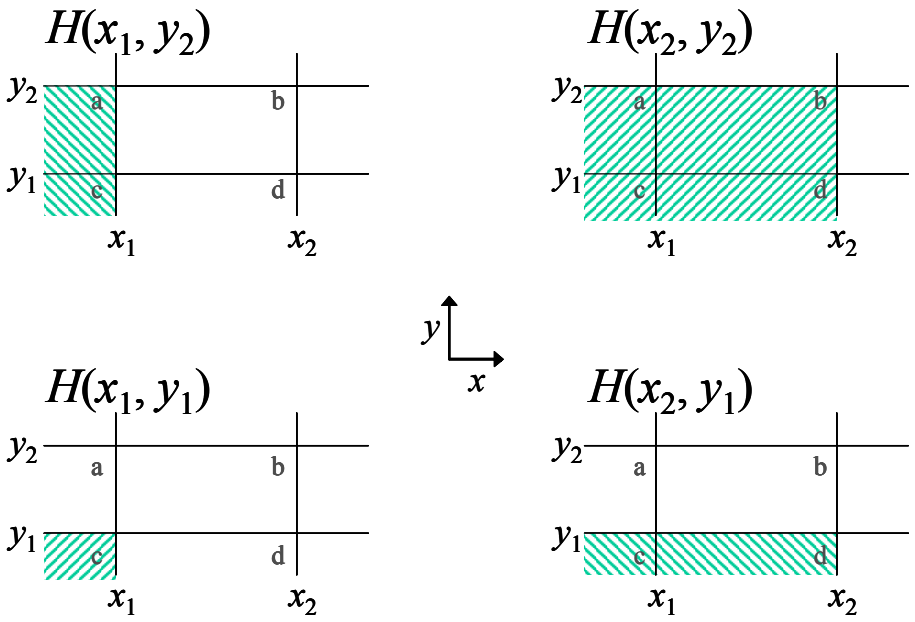

Figure 19: Diagonal difference estimate of the mass associated with a rectangle.

Hutchinson and Lai (1990) and Haas (1999) give accessible introductions to copulas and their use in modeling dependence in risk and uncertainty modeling. A copula is simply the dependence function between random variables. Figure 20 depicts the copulas corresponding to three special cases of dependence. Irrespective of what

the marginal distributions are, every perfectly dependent pair of random variables has the dependence function represented by the copula $M(u, v) = \min(u, v)$. Moreover, any pair of random variables characterized by given marginals can be made to be perfectly dependent on one another by combining their marginal distribution with the copula M as specified by Sklar's theorem. This function is shown on the left graph of Figure 20. Similar statements are possible with opposite dependence and the copula $W(u, v) = \max(u + v - 1, 0)$, which is depicted in the right graph. Finally, the middle graph, depicting the product copula $\Pi(u, v) = uv$, has the same intimate connection with independence. Fréchet (1951) and Hoeffding (1940) showed that, in fact, all dependencies between any two variables correspond to a copula between M and W, that is, for any copula C,

$$W(u,v) \le C(u,v) \le M(u,v)$$

(see Whitt 1976). These special cases are called the Fréchet-Hoeffding limits, or simply, the Fréchet limits.

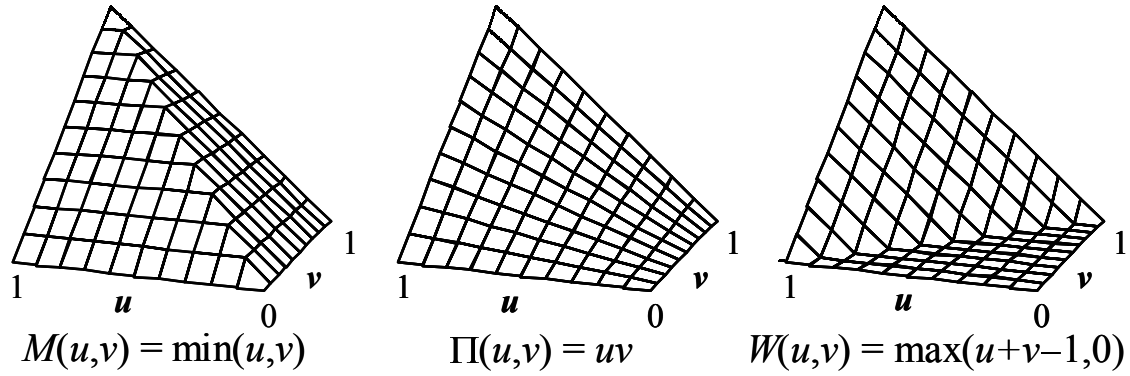

Figure 20: Three special cases of dependence functions.

Nelsen (1999) reviews many copula families that have been proposed. A family of copulas is a *model* of the dependence between random variables. A copula family that includes the special cases of perfect dependence, opposite dependence and independence is called comprehensive. There are several families that have this property. For example, the Frank family of copulas (Frank 1979; Nelsen 1999), which is defined by

$$C_{\text{Frank}}(u,v) = \log_s \left[1 + (s^u - 1)(s^v - 1)/(s - 1)\right]$$

where $0 \le s$. Perfect dependence arises in the limit as s tends to zero. Opposite dependence arises when s goes to infinity, and independence corresponds to an s of one. The Clayton family is

$$C_{\text{Clayton}}(u,v) = \max\left((u^{-\theta} + v^{-\theta} - 1)^{-1/\theta},\ 0\right)$$

where $-1 \le \theta$. Perfect dependence corresponds to $\theta = \infty$. Setting θ to -1 yields opposite dependence, and zero corresponds to independence. The Mardia family is defined as a linear combination of the three special cases

$$C_{\text{Mardia}}(u,v) = \frac{\theta^2(1+\theta)}{2} M(u,v) + (1-\theta^2)\Pi(u,v) + \frac{\theta^2(1-\theta)}{2} W(u,v)$$

where the functions M, Π and W are the functions defined above. Clemen and Reilly (1999) argued that the normal copula is especially useful. It is the dependence between variables that are multivariately normal. The normal copula is

$$C_{\text{normal}}(u,v) = \mathbf{\Phi}_r\left(\Phi^{-1}(u), \Phi^{-1}(v)\right)$$

where

$$\mathbf{\Phi}_r(x,y) = \frac{1}{2\pi\sqrt{1-r^2}} \int_{-\infty}^{x}\int_{-\infty}^{y} \exp\left(\frac{-\left(s^2 - 2rst + t^2\right)}{2(1-r^2)}\right) dsdt$$

and Φ^{-1} is the quantile function of the standard univariate normal distribution function, and $\mathbf{\Phi}_r$ is the joint cumulative distribution function for the standard bivariate normal distribution with Pearson correlation r, $0 < r < 1$.

Not all copula families are comprehensive. Indeed, most of the families described by Nelsen (1999) that have arisen in the statistical and engineering literature do not include all three special cases of perfect dependence, opposite dependence and independence. Elliptic copulas, which were recommended by Kurowicka et al. (2001), include all correlations between -1 and $+1$, but, interestingly, they do not include the independent copula Π when the correlation is zero.

As we discussed in Section 3.6.2, there are several measures of correlation that could be used to index the strength of dependence. Hutchinson and Lai (1990) describe a host of possible measures of correlation/dependence, including not only Pearson, Spearman and Kendall, but also medial correlation, maximal (sup) correlation, monotone correlation indices, and several indices of concordance, such as Blomqvist's (1950) quadrant measure. These various measures are scales that summarize dependence in different ways (comparable in this sense to the mph scale and kph scales on a speedometer, or perhaps to linear and log scales describing the same phenomenon). In principal, one could match a copula family with one of these measures, which would specify how to select a particular copula from the family.

The literature about dependence has exploded over the last two decades. Many disparate ideas and approaches have been proposed about how to measure and model dependence. Future research in this area may witness some pruning or weeding among

the many ideas that have developed to select a few that are most generally useful. At least for the foreseeable future, however, none of the copula families is likely to emerge as "the" family to use for all applications. Like the various measures of correlation, the different families are useful in different circumstances to represent different sorts of phenomena. This methodological embarrassment of riches need not, however, lead to confusion or anxiety among risk analysts. The algorithmic tools described in this and the following sections to compute convolutions can be used with essentially *all* of the copula families irrespective of their origin or complexity, so long as they admit the numerical specification of the dependence function. This means that, even though it may not always be clear which copula family or correlation measure we should be using, we will always be able to compute convolutions with our choices.

Examples: Figure 21 shows how the distribution of a sum of random variables $X \sim$ normal(5,1) and $Y \sim$ uniform(2,5) can vary with the choice of copula family and correlation measure that are used to model the dependence between X and Y. The top graph represents the Frank copula family, parameterized by the medial correlation coefficient. The middle graph represents the Mardia family parameterized by Kendall correlation, and the bottom graph represents the Clayton copula family (which specifies its own index of correlation). Each of the three graphs shows six distributions, corresponding to the correlations +1, +0.6, +0.2, –0.2, –0.6 and –1. The correlation of +1 represents perfect dependence and it produces the shallowest distribution function for the sum. The correlation of –1 likewise represents opposite dependence, which produces the steepest distribution function for the sum. These two distributions are identical among the three graphs. The other correlations yield distributions that are intermediate in slope to these two extreme cases. These intermediate distributions vary markedly in shape across the three graphs. Notice, for instance that all the distributions produced by the Frank copula go through a single point at the median. The distributions from the Mardia copula, on the other hand, share even greater commonality. They variously trace along a common set of percentiles in the middle of the distributions. The distributions from the Clayton copula don't have any common points for all correlations except at the extreme tails.

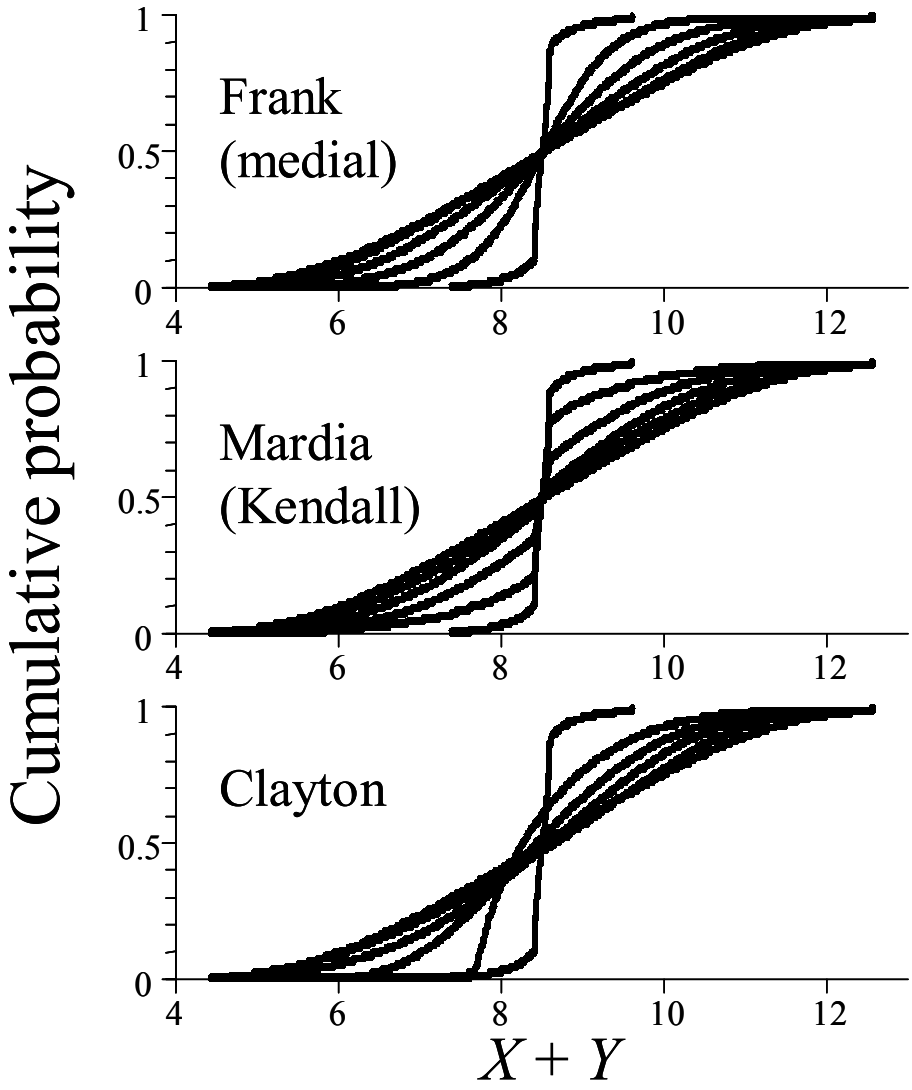

Figure 21: Variation in the distribution of a sum of correlated variables where correlation is interpreted according to three copula families.

3.7.1 Using copulas with Dempster-Shafer structures

The computational approach using a parametric copula that was described in Section 3.7 for computing convolutions among random variables characterized by probability distributions extends immediately to Dempster-Shafer structures so long as these structures have a natural order for their focal elements. In this case, as before, the probability associated with a closed region is computed directly from the joint distribution function at the corners of the region as the diagonal difference.

The first step of the calculation is to compute the focal elements. Suppose the first Dempster-Shafer structure has elements $\{([a_1, b_1], p_1), ([a_2, b_2], p_2), \ldots, ([a_n, b_n], p_n)\}$ where $a_1 \le a_2 \le \ldots \le a_n$ and $b_1 \le b_2 \le \ldots \le b_n$, and the second has the elements $\{([c_1, d_1], q_1), ([c_2, d_2], q_2), \ldots, ([c_m, d_m], q_m)\}$ where $c_1 \le c_2 \le \ldots \le c_m$ and $d_1 \le d_2 \le \ldots \le d_m$, and we wish to combine these two Dempster-Shafer structures with the binary function f. The $(i, j)^{\text{th}}$ elements in Yager's (1986) Cartesian product is $f([a_i, b_i], [c_j, d_j])$.

The second step in the calculation is to compute the masses associated with each of the focal elements. The probability mass associated with any closed region $x_1 \le X \le x_2, y_1 \le Y \le y_2$ can be computed from the joint distribution H as the diagonal difference

$$H(x_1, y_1) - H(x_1, y_2) - H(x_2, y_1) + H(x_2, y_2),$$

where each of these H's can be evaluated in terms of the copula applied to the marginal cumulative probabilities $H(x,y) = C(F(x), G(y))$. This means that the mass associated with the $(i, j)^{\text{th}}$ focal element is

$$C(P_i, Q_j) - C(P_{i-1}, Q_j) - C(P_i, Q_{j-1}) + C(P_{i-1}, Q_{j-1})$$

where P_i and Q_j are the cumulative masses

$$P_i = \sum_{k=1}^{i} p_k,$$

$$Q_j = \sum_{k=1}^{j} q_k.$$

Whenever the focal elements have a natural order, these cumulative masses have the interpretation as the probability that the underlying variable is smaller than the indicated focal element. The last step is to sum the masses for any elements of the Cartesian product that happen to be identical. The sum of all of the probability masses in the Cartesian product will be one (except for possible discretization error).

Example: Suppose we want to use the Frank family of copulas parameterized by Spearman correlation to compute the sum of $X+Y$. Further suppose that X is characterized by the Dempster-Shafer structure $\{([1,6], 0.25), ([5,9], 0.5), ([7,10], 0.25)\}$ and that Y is characterized by the Dempster-Shafer structure $\{([2,4], 0.5), ([3,5], 0.5)\}$, and their Spearman correlation is 0.6. The cumulative plausibility and belief functions for these two inputs are shown in the left and middle graphs of Figure 22. The focal elements for the answer can be obtained as the Cartesian product of the input focal elements, so, for instance, the first focal element is just [1,6]+[2,4] = [3,10]. The focal elements are thus

	$j = 1$	$j = 2$
$i = 1$	[3,10]	[4,11]
$i = 2$	[7,13]	[8,14]
$i = 3$	[9,14]	[10,15]

where i indexes the elements of the X structure, and j indexes the elements from the Y structure. The mass to be associated with the i,j focal element in the Cartesian product is the diagonal difference

$$C(P_i, Q_j) - C(P_{i-1}, Q_j) - C(P_i, Q_{j-1}) + C(P_{i-1}, Q_{j-1})$$

where C is the dependence function, P_i is the cumulated probability for the ith focal element of X, and Q_j is the cumulated probability for the jth focal element of Y. The value of P_1 is 0.25; the value P_2 is 0.75; the value of P_3 is 1. The value of Q_1 is 0.5, and the value of Q_2 is 1. (P_0 and Q_0 are of course zero.) For the dependence function we are using the Frank copula family

$$C(u, v) = \log_s(1 + (s^u - 1)(s^v - 1) / (s - 1))$$

where the parameter s is chosen to yield the correct Spearman correlation. From a table given by Nelsen (1986, table I*), the value of the Frank parameter s corresponding to a Spearman correlation of 0.6 is about 0.01. The matrix of the copula values is

	$j = 1$	$j = 2$
$i = 1$	0.211	0.25
$i = 2$	0.461	0.75
$i = 3$	0.5	1.0.

The mass for each focal element is obtained from the diagonal difference, so, for instance, the mass for the focal element when i=2 and j=1 is just

$$C(P_2, Q_1) - C(P_1, Q_1) - C(P_2, Q_0) + C(P_1, Q_0),$$

which is 0.461 – 0.211 – 0 – 0 = 0.25. Other elements are evaluated similarly. The matrix of masses to be associated with the focal elements is therefore

	$j = 1$	$j = 2$
$i = 1$	0.211	0.039
$i = 2$	0.25	0.25
$i = 3$	0.039	0.211.

Consequently, the Dempster-Shafer structure associated with the sum $X+Y$ assuming X and Y are correlated according to a Frank copula with correlation 0.6 is {([3,10], 0.211),

*s = exp(–exp(–z)), where z is Nelsen's tabled value.

([4,11], 0.039),([7,13], 0.25), ([8,14], 0.25), ([9,14], 0.039), ([10,15], 0.211)}. The cumulative plausibility and belief functions for this structure are displayed on the right-most graph of Figure 22.

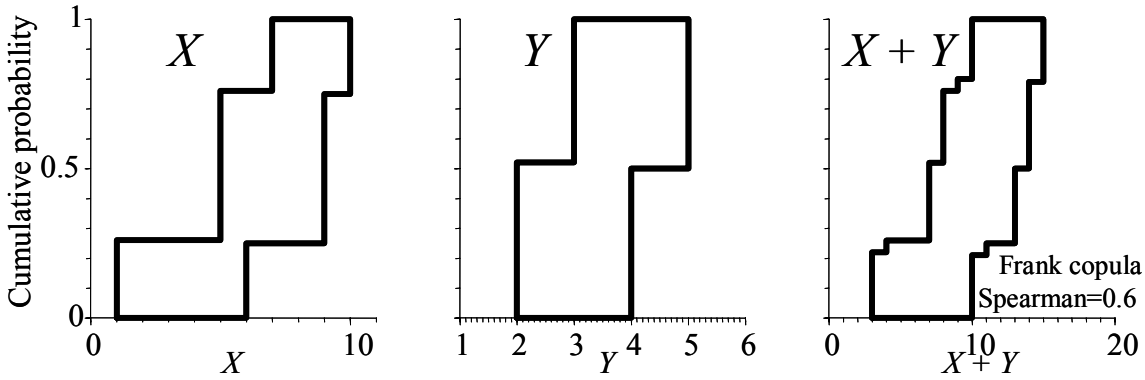

Figure 22: Convolution between Dempster-Shafer structures with a specified copula (Frank family, parameterized by Spearman correlation).

Examples: If the elements of a Dempster-Shafer structure do not have the convenient ordering, then the Dempster-Shafer structure should first be transformed to a related structure that is so ordered. For instance, if X were characterized by {([5,9], 0.5), ([1,6], 0.25), ([7,10], 0.25)}, it would first be rewritten as {([1,6], 0.25), ([5,9], 0.5), ([7,10], 0.25)}. This is merely a reordering of the focal elements, so it does not alter the Dempster-Shafer structure at all. If, on the other hand, X were characterized by {([1,6], 0.25), ([5,10], 0.5), ([7,9], 0.25)}, it would first be replaced by the related Dempster-Shafer structure {([1,6], 0.25), ([5,9], 0.25), ([5,10], 0.25), ([7,10], 0.25)}. This is necessary because [7,9] is inside [5,10] so there can be no convenient ordering of the focal elements. This might seem like a substantial change because it alters the number of focal elements from 3 to 4, but it does not change the cumulative plausibility and belief functions. Once both input structures have the requisite ordering, then they may be convolved using the method outlined in the previous example.

3.7.2 Using copulas with p-boxes

Using the machinery developed for Dempster-Shafer structures in the previous section, it is straightforward to calculate convolutions between p-boxes under the assumption that the variables represented by the p-boxes are dependent and their dependence is expressed by some parametric copula. The p-boxes are first discretized into Dempster-Shafer structures. The calculation then proceeds exactly as it did for Dempster-Shafer structures. In the case of p-boxes, however, the requisite natural ordering of the focal elements is always present because of the way the discretization is done. The resulting Dempster-Shafer structure is then reassembled into a p-box.

Discretization of p-boxes is described by Ferson et al. (2003, their section 2.3). If a p-box is specified by the bounding functions $\underline{F}(x)$ and $\overline{F}(x)$ such that $\underline{F}(x) \le \overline{F}(x)$, then the discretization will be have focal elements $[\overline{F}^{-1}(p), \underline{F}^{-1}(p)]$ where the superscript denotes the inverse function (or some quasi-inverse if the function is not strictly invertible), and $0 \le p \le 1$, for as many levels of p as are required to make an adequate discretization. The mass associated with the focal elements is Δp, which can be made arbitrarily fine. When the bounds of the p-box are step functions, the discretization can be exact; when they are curves, the discretization will be an approximation to the p-box.

Example: Suppose we want to compute the product XY, where the random variable X is a proportion whose median is no larger than 0.1 and whose 95th percentile is no larger than 0.3, and the random variable Y is a proportion with a mean of 0.2. These two marginal inputs are depicted in as the left and middle graphs of Figure 23. Suppose that the dependence between X and Y is to be modeled with a normal copula with a Kendall correlation of –0.5. As Clemen and Reilly (1999) explain, this Kendall correlation corresponds to a Pearson correlation of $2 \sin(\pi(-0.5)/6) = -0.51764$ for a normal copula. The discretization for X is the Dempster-Shafer structure {([0,0.1],0.5), ([0,0.3],0.45), ([0,1],0.05)}. The Dempster-Shafer structure for Y is only approximate because its bounds are not step functions. With 100 discretization levels, it is {([0, 0.202], 0.01), ([0, 0.204], 0.01),..., ([0.192, 1], 0.01)}. The 300 focal elements in the Cartesian product between X and Y consequently are the intervals {[0, 0.0202], [0, 0.0204],..., [0, 0.1],..., [0, 0.0606], [0, 0.0612],..., [0, 0.3],..., [0, 0.202], [0, 0.204],..., [0, 1]}. The associated mass for each of these intervals is computed using the diagonal difference with $C = C_{\text{normal}}$, the normal copula with $r = -0.51764$. The cumulated masses for X are $P_1 = 0.5$, $P_2 = 0.95$, and $P_3 = 1$. The cumulated masses for Y are $Q_1 = 0.01$, $Q_2 = 0.02$, …, and $Q_{99} = 0.99$ and $Q_{100} = 1$. The calculated Dempster-Shafer structure is then reintegrated to obtain the p-box shown in the right graph of Figure 23.

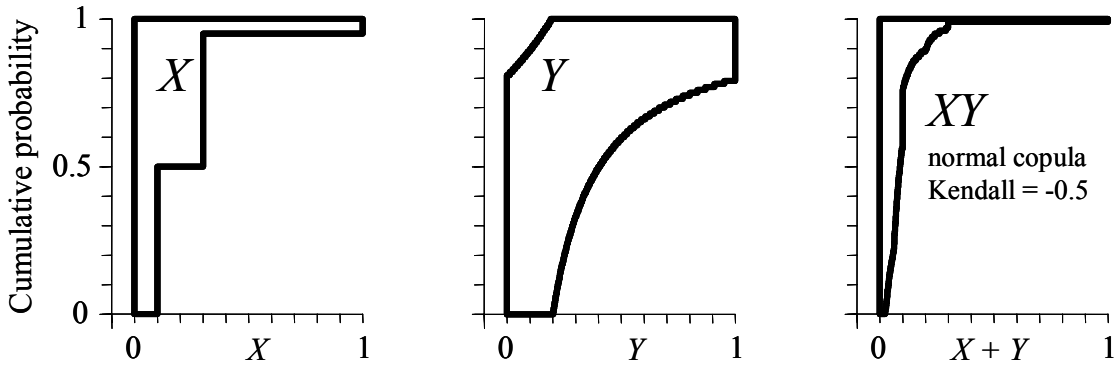

Figure 23: Convolution between p-boxes under a particular dependence function (normal copula, parameterized by Kendall correlation).

It is worth noting that the approach based on parameterized copulas developed here is compatible with, and provides a unifying justification for, the approaches used in Section 3.5.1 and 3.5.2 to convolve probability boxes or Dempster-Shafer structures under perfect or opposite dependence. We can conveniently illustrate this for the case of perfect dependence, for which the copula is $M(u,v) = \min(u,v)$. In this case, the joint distribution is

$$H(x, y) = M(F(x), G(y)) = \min(F(x), G(y)),$$

where $F(x)$ and $G(y)$ are the cumulative masses associated with values x and y respectively. Consider two uncertain numbers X and Y that have been discretized into n equiprobable levels. The cumulative mass associated with the i^{th} slice of the discretization is i/n. Such discretizations therefore imply that a particular diagonal difference

$$H(x, y) - H(x, y + \Delta y) - H(x + \Delta x, y) + H(x + \Delta x, y + \Delta y)$$

would be evaluated under perfect dependence as

$$\min\left(\frac{i}{n}, \frac{j}{n}\right) - \min\left(\frac{i}{n}, \frac{j+1}{n}\right) - \min\left(\frac{i+1}{n}, \frac{j}{n}\right) + \min\left(\frac{i+1}{n}, \frac{j+1}{n}\right)$$

for some i and j. Across the Cartesian product, there are three situations. If $i < j$, then $i+1 \le j$, and the diagonal difference would simplify to

$$\frac{1}{n}\big(i - i - (i+1) + (i+1)\big) = 0.$$

If, on the other hand, $j < i$, then $j+1 \le i$, and the diagonal difference simplifies to

$$\frac{1}{n}\big(j - (j+1) - j + (j+1)\big) = 0.$$

The quantity is nonzero only when $i = j$, in which case it is

$$\frac{1}{n}\big(i - i - j + (i+1)\big) = \frac{1}{n}.$$

This is why only the diagonal elements in the Cartesian product get mass when dependence is perfect. A similar argument can be constructed from the copula $W(u,v) = \max(u+v-1, 0)$ that shows why it is the anti-diagonal elements that get the mass when the dependence is opposite.

3.7.3 Caveat: incompleteness of correlation

As will be evident to the reader, modeling dependence between variables with a scalar measure of the strength of correlation is likely to be a superficial description of their relationship. *Any* model that characterizes dependence well enough to, say, allow the generation of sample variates is employing some particular dependence function (copula), although its details may be implicit and not obvious to the analyst. The set of such dependence functions parameterized by the correlation constitutes a copula family which is a model of the dependence. When such characterizations are used in risk assessments, analysts are relying on the assumption that this copula family is an appropriate one, and that it is faithful to the true dependence resulting from the mathematical or physical relationship between the variables both in terms of its summary correlation and its consequences for whatever arithmetic functions involving the dependent variables are to be computed. Sections 3.8 and 3.9 review the use of copulas that are not members of parametrically prescribed families. Section 4 considers methods that will be useful when little or no information is available to select the dependence function.

3.8 Empirical copulas

An empirical copula, also known as an empirical dependence function (Deheuvels 1979; Nelsen 1999), is a characterization of the dependence function between variables based on observational data. It is the analog of the empirical distribution function for the question of dependence. (In fact, if the marginal distributions are first transformed to standard uniforms, the empirical copula is just the empirical distribution function for the joint distribution.) An empirical copula can be computed from bivariate sample data (x_i, y_i), $i = 1,\ldots, n$, with

$$C(u,v) = C\left(\frac{j}{n}, \frac{k}{n}\right) = \frac{\#\left\{(x,y) \mid x \le x_{(j)},\ \ y \le y_{(k)}\right\}}{n}$$

where # denotes the cardinality of a set, and $x_{(j)}$ and $y_{(k)}$, for $1 \le j, k \le n$, denote order statistics from the sample. The numerator is just the number of points in a scattergram of the data that are lower and to the left of the point $(x_{(j)}, y_{(k)})$, where $x_{(j)}$ is the j^{th} smallest value of the sample x's and $y_{(k)}$ is the k^{th} smallest value of the sample y's.

Example: The leftmost graph in Figure 24 is the scattergram of 300 hypothetical data points exhibiting an interesting nonlinear dependence between its variables X and Y. The marginal distribution for X was uniform over the unit interval. Y was a randomized function of X. Its values were computed as

$$Y = \begin{cases} \dfrac{X}{3} + \dfrac{Z}{6}, & \text{if } W < \dfrac{1}{2} \\ \dfrac{1}{2} + \dfrac{X}{3} + \dfrac{Z}{6}, & \text{otherwise} \end{cases}$$

where W and Z were independent random uniform deviates also from [0,1]. This formula results in a distribution for Y that is close to (but not exactly) uniform over the unit interval. The 300 X- and Y-values were sorted independently. At each of 300×300 evenly spaced points in the unit square [0,1]×[0,1], the number of data points were tallied that were smaller (in the X direction) than the i^{th} sorted X-value and simultaneously smaller (in the Y direction) than the j^{th} sorted Y-value. These tallies were all normalized by dividing by 300. Plots of these normalized tallies are shown in the middle and right graphs of Figure 24. The function is the empirical copula associated with the scattergram. The vertical axis of the middle graph is the value $C(u,v)$ for the empirical copula at given values for u and v. The rightmost graph shows the same function, in a contour plot with gray shading.

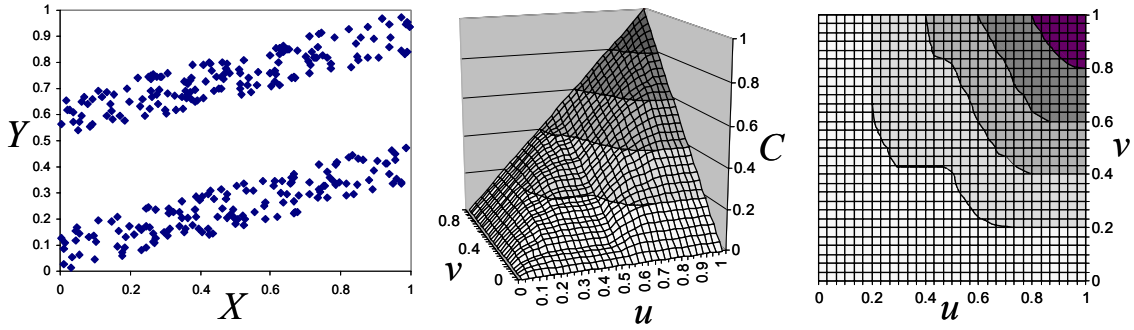

Figure 24: Empirical copula (middle and right graphs) from observational data (left graph).

By construction, the copula is zero along the u- and v-axes. It is intersects with the forty-five degree line in the v,C-plane at u = 1, and similarly in the u,C-plane at v = 1. All copulas have these characteristics; copulas differ only in the internal details of the 2-increasing function that connects these edges together. The observed scattergram creates, and is captured by, the wrinkles in the surface of the copula.

Empirical copulas may be used in the algorithms described in Section 3.7 and its subsections to make arithmetic calculations with probability distributions, Dempster-Shafer structures and probability boxes. This use is completely straightforward. The empirical copula simply replaces the C's (the parameterized copula) previously used. For any pair of values from the unit interval, the empirical copula returns another value from the unit interval. Rather than computing the copula from some parameterized

expression, its value is simply read from the (discretized) table storing the normalized tallies.

Some analysts find it hard to understand why one would bother with an empirical compula if computing it implies one already has empirical data for a joint distribution from which the model could be constructed directly. Does not this joint information allow one to sidestep the question of modeling dependence altogether? If the data are abundant and reliable, the answer might be yes. But in many cases, the empirical information is sparse, or based on imperfect samples, or associated with surrogate variables similar not identical to those of interest. In such situations, an analyst might prefer to build a model of the dependence that makes use the available data, but that can go beyond those data. It is important to understand that the calculations or simulations based on an empirical copula are *not* simply reconstructing the precise details* of the scattergram on which the copula is based. Instead, what is reconstructed is the overall statistical pattern of the dependence in the joint distribution. This is the sense in which the empirical copula goes beyond the joint data. It means that one can easly apply the dependence observed for one set of marginal distributions to another set of marginal distributions for which no joint data may even be available.

One important caveat in this use of empirical copulas is that they can only reflect the variation exhibited in the data sets used to construct them. Just as an analyst who uses an empirical distribution function (EDF) to model some univariate marginal distribution takes a risk that the available data may not reflect the entire range of possible values for that variable, the analyst who employs an empirical copula to model the bivariate dependence takes a risk the combinations observed in the available data may not reflect the full variety of combinations that are actually possible. These risks grow very large when empirical data are sparse, but, for reasons of dimension, they are likely to be much worse for empirical copulas than for (univariate) empirical distribution functions.

3.9 Constructed arbitrary copulas

Nelsen (1999) considers the problem of constructing copulas to suit one's modeling purposes. One application of constructed copulas might be to reflect hypothetical or planned constraints in the dependencies among variables. For instance, if a new safety release system is designed to activate when two variables become jointly extreme, then such conditions would not, in principle, be able to propagate through the system. The resulting dependence pattern would permit large values for both variables but wouldn't allow them to be simultaneously large. How could this dependence pattern be modeled and represented in calculations? One way to obtain a copula that respects such constraints is to manually edit an observed scattergram. This editing would consist of removing all or a proportion of the points in disallowed regions. In the absence of paired empirical observations, a bivariate scattergram developed using independent

*To reconstruct the precise locations of the data in the scattergram, one could employ a permutation or randomization strategy such as bootstrapping.

uniform marginals could be edited. The resulting "empirical" copula obtained from this edited scattergram would express the nonlinear dependence represented by the constraints. Another use of constructed copulas might be to posit the potential existence of combinations of variable values that have never been observed but are thought to be possible. This might be effected by *adding* points in underpopulated regions of the scattergram. The resulting edited scattergram could then be used to compute an empirical copula using the formula just discussed in Section 3.8 and incorporated into calculations using the methods described in Sections 3.7, 3.7.1, and 3.7.2.

4 Accounting for incertitude about dependence

Empirical information about dependencies among variables in an assessment is often sparse. In multivariate problems, it is commonly not practical to specify the entire variance-covariance matrix (Cooke 1997), much less the full multidimensional character of the dependence among the variables. In such an environment, analysts sometimes express a desire to relax some of the independence assumptions in their assessments. By this, what they may often mean is to *alter* those assumptions away from independence to some other precisely specified dependence assumption. But they may also actually intend to *relax* the assumption in the sense of making fewer or no assumptions about the dependence. The methods described in this section allow analysts to do this. The study of how assumptions about dependence affect the quantitative results of an assessment might be called dependency bounds analysis (Williamson and Downs 1990; Ferson 1994; Ferson and Long 1995).

4.1 Sensitivity analyses and dispersive Monte Carlo

In some situations, it may be reasonable to assume that some or all of the statistical relationships among the variables are simple* and do not harbor cryptic nonlinearities so that they can be adequately characterized by correlation coefficients. Even with such an assumption, the magnitude of the correlations may be unknown, or the correlation coefficients may be known only to within intervals because empirical information relating the variables is sparse. In these cases, straightforward sensitivity studies that vary correlations or some more integrated approach would be useful.

Whitt (1976) described a scheme to simulate bivariate distributions from specified marginals with the largest (or smallest) possible correlation. This technique has often been used in Monte Carlo simulations to estimate random quantities with minimal variance (see Bratley et al. 1983). The approach can be extended to the multivariate case to compute conservative estimates of risks. Suppose that X and Y are random variables with marginal distributions F and G respectively. The distribution with greatest dispersion that could arise as the sum $X+Y$ is that obtained when the variables have the largest possible correlation between them. This is because the variance of a sum is the sum of the variances of the addends plus twice their covariance, which depends directly on their correlation (Mood et al. 1975). The exact opposite is true for subtraction. The greatest dispersion in a distribution of differences comes from

*As might by now be apparent to the reader, there is no simple definition of what a simple dependence is. Simplicity might be characterized by linearity or ellipticity, symmetry, an absence of unexpected constraints or interactions, or any of several other qualitative features. A person's definition of simplicity is perhaps clear until he views real data with their idiosyncrasies and subtleties. All it means in this context is that a scalar correlation coefficient is a robust and appropriate descriptor for the pattern of dependence.

the case in which the correlation between the subtrahend and the minuend is most negative. Similar statements can be made about multiplication and division of numbers whose distributions are non-negative.

The following scheme allows one to sample Monte Carlo variates from specified distributions such that they exhibit the greatest possible correlation. Sample variates $X \sim F$ and $Y \sim G$ by randomly and repeatedly selecting u from a uniform random distribution between 0 and 1 and mapping the value through the inverse functions $X=F^{-1}(u)$ and $Y=G^{-1}(u)$. The resulting set of variate pairs has the maximal correlation possible subject to the specified marginal distributions F and G. To obtain the smallest (maximally negative) correlation possible between the variates, the mapping should be $X=F^{-1}(u)$ and $Y=G^{-1}(1-u)$. Since the most dispersed distribution arising from the addition of two random variables occurs when their correlation is maximal (Müller 1997; Goovaerts et al. 2000), a conservative estimate of the extreme values of the distribution of the sum $X+Y$ might be estimated with the quantity $F^{-1}(u) + G^{-1}(u)$ where u is sampled from a uniform distribution on [0,1]. This can be referred to as the (u,u) strategy. The most dispersed distribution arising from subtraction of two random variables is that obtained when there is minimal (that is, most negative) correlation between them. This can be estimated by $F^{-1}(u) - G^{-1}(1-u)$, which can be called the $(u,1-u)$ strategy. So long as the distributions are of like sign, then the dispersive sampling strategy for multiplication is like that for addition, and the one for division is like that for subtraction. That is, the products $F^{-1}(u)G^{-1}(u)$ form the most dispersed distribution that can result from multiplication of random variables having marginals F and G. Likewise, $F^{-1}(u)/G^{-1}(1-u)$ represents the most dispersed distribution that could result from division.

To compute conservative estimates of a distribution resulting from more complex arithmetic combinations of positive random numbers, one can use the (u,u) strategy among variables that are added or multiplied together and the $(u,1-u)$ strategy when they are subtracted or divided. A mixed expression such as $A/(1-B)$ requires, again, the (u,u) strategy, and in arbitrary mathematical expression involving multiple variables, assigning u and $1-u$ will necessitate a symbolic analysis of the mathematical expression. This approach will work in situations where the original distributions and intermediate results are strictly positive, and it could be generalized in a software implementation for general distributions. It is compatible with ordinary Monte Carlo analyses that assume independence or a particular correlation structure among other variables. This approach can be called dispersive Monte Carlo sampling because it yields simulated distributions with extremal dispersion, i.e., the largest possible variance, given the specified marginals. Note however that there is no mathematical guarantee that the tail probabilities computed will be upper bounds irrespective of the dependencies among the variables. However, in many situations the maximally dispersed result is likely to be an appropriately conservative estimate that may be useful for risk analysis for practical cases (Burgman et al. 1993, page 154; Ferson 1994; Bukowski et al. 1995; Müller 1997; Vyncke et al. 2000; Goovaerts et al. 2000; Kaas et al. 2000).

Example: Suppose that we need to estimate the distribution of $V=WX/(Y-Z)$, where $W\sim$ normal(5,1), $X\sim$ lognormal(1.2,0.3), $Y\sim$ uniform(2,14), and $Z\sim$ beta(2,3), and we want to use dispersive Monte Carlo sampling to conservatively account for ignorance about the magnitude of the correlations among the variables. In a Microsoft Excel spreadsheet, this simulation could be implemented with the expression =(NORMINV(U,5,1)*LOGINV(U,1.2,0.3))/(((1−U)*12+2)−BETAINV(U,2,3)) where U is (a single value of) a uniformly distributed random variate between zero and one. Figure 25 shows the distribution function (estimated from 1000 random samples) of V resulting from the dispersive Monte Carlo sampling as a black curve. The gray curve is the corresponding distribution of V assuming that all of the inputs were mutually independent. The differences between the two curves are striking. The observed mean of the distribution obtained from dispersive sampling was 4.9 and the observed standard deviation was 7.7. The observed mean of the distribution obtained from ordinary Monte Carlo sampling assuming independence was 3.1 and the observed standard deviation was 2.4. The tail weights are considerably greater for the dispersive distribution. For instance, the 90th percentile of the independent distribution is about 6. The 90th percentile of the dispersive distribution is more than twice that value. It is also clear that the left-tail percentiles for the dispersive distribution are lower than for the independent distribution.

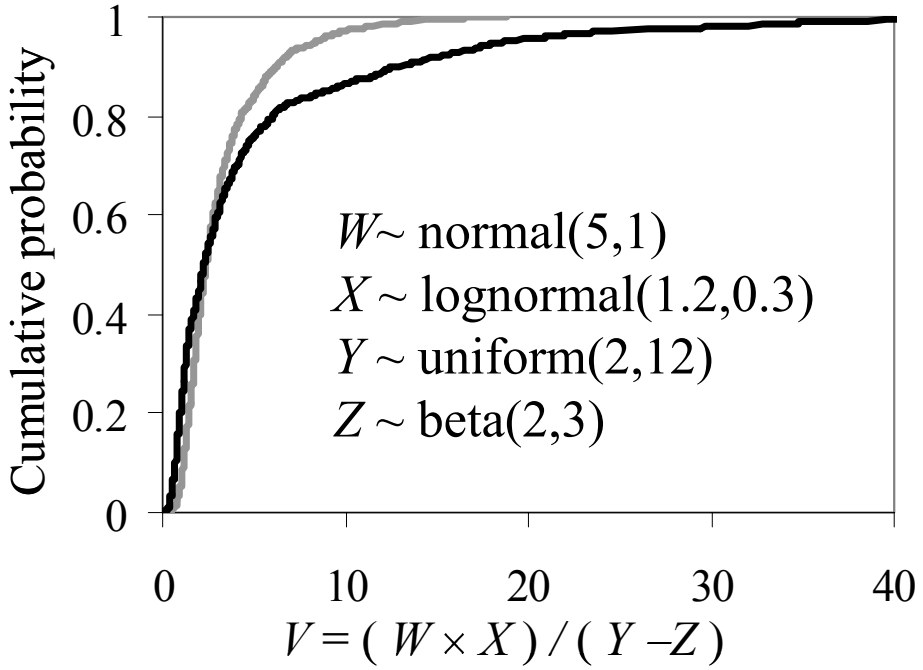

Figure 25: Arithmetic function of random variables estimated by dispersive Monte Carlo sampling (black) and ordinary Monte Carlo sampling assuming independence (gray)

A modification of the dispersive Monte Carlo approach is to make the correlations as large as *plausible*, rather than as large as possible. For instance, one might know that the correlation between two variables could not be larger than some degree, and feel it improper to assume counterfactually that the correlation is bigger than this maximum. In this approach, any pair of variables that would have been treated with the (u,u) strategy would now be assigned the largest reasonable correlation. And any pair of variables that would have been treated with the $(u,1-u)$ strategy would now be assigned the smallest (most negative) plausible correlation. The simulation would then employ the methods discussed in Section 3.6 to generate variables correlated as prescribed.

Example: This example merely illustrates the monotonicity of the influence of correlation on a convolution. Figure 26 shows the variation in the distributions of products XY where $X \sim$ normal(5,1) and $Y \sim$ normal(10,2) and X and Y are correlated to various degrees as computed in a Monte Carlo simulation involving 1000 replications using the method of Scheuer and Stoller (1962). There are nine distributions depicted, corresponding to Pearson correlations of -1, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75 and 1. The correlation of -1 yields the steepest distribution, and the correlation of $+1$ yields the shallowest. Intermediate correlations yield distributions intermediate in slope. The effect on the tails of the product distribution can be seen in the figure. The effect on the mean of the distribution is fairly minor and ranges roughly linearly from 48 to 52 as correlation increases. The effect on the standard deviation, on the other hand, is more substantial. For the correlation of -1, the standard deviation is 2.88, but for the correlation of $+1$, it is 20.5. Figure 27 depicts the effects on these parameters.

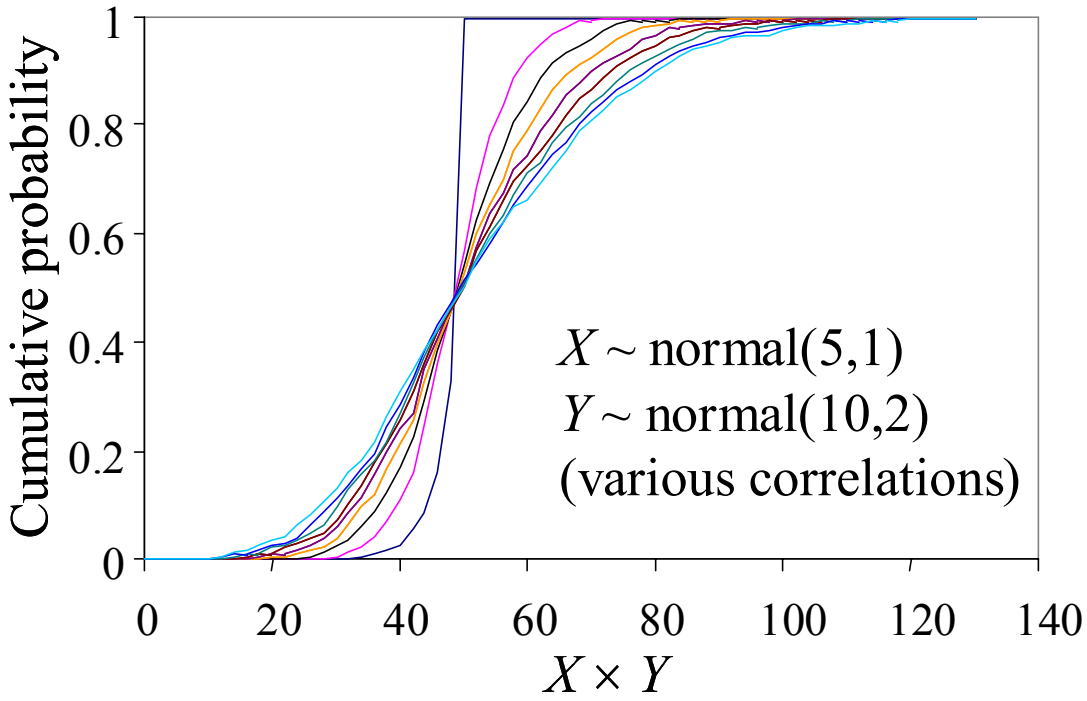

Figure 26: Distributions of products of normal factors under varying correlations.

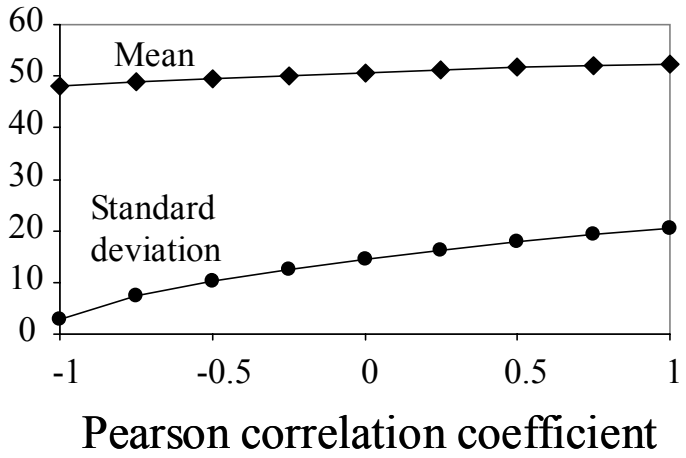

Figure 27: Effect of correlation on parameters of the product distributions.

There are some significant caveats associated with dispersive Monte Carlo sampling. The most important caveat is that it is incomplete. It cannot assess the possible consequences of nonlinear dependencies. Some authors (e.g., Burgman et al.

1993; Goovaerts et al. 2000; Kaas et al. 2000) seem to suggest that the dispersive answer is an "upper bound" on the unknown true distribution, but this is certainly not the case, as we shall discuss in Section 4.2. If an analyst uses this approach, it should be with a full understanding of the nature of the assumption that excludes so many potential dependencies. Another disadvantage of this approach is that it usually cannot be used to estimate the how narrow the final probability distribution might be. Although it can find the most dispersed (i.e., the widest) distribution, it cannot be used to get the least dispersed (narrowest) one. This is apparent as soon as one tries to estimate the narrowest distribution for the sum $A+B+C$. If u is assigned to A and $1-u$ to B, there isn't anything that can be assigned to C. (Assuming independence certainly does *not* yield the narrowest distribution. This is clear in the very simplest case. For instance, if A and B are uniformly distributed on [0,1], an assumption of independence yields the triangular distribution [0,1,2]. The narrowest possible distribution, of course, is the invariant distribution at 1.)

4.2 Fréchet bounds

When empirical information is lacking so that an analyst cannot be confident about the nature of the dependencies among the variables, it may often be useful to compute bounds on the risk result without making any assumption at all about one, some or all of the dependencies. Dependency bounds analysis (Makarov 1981; Frank et al. 1987; Williamson 1989; Williamson and Downs 1990; Ferson and Long 1995) produces such bounds, which are often also mathematically best possible in the sense of being as tight as possible given the stated information. This strategy generalizes the assumption of multivariate linearity of the relationships described in Section 4.1. Dependency bounds are sure to enclose the result, no matter what correlation or nonlinear dependency may exist between or among any of the variables. Thus, the results of a dependency bounds analysis are generally bounds on a cumulative distribution function, rather than an approximation of one such as might be given by Monte Carlo simulation. The strategy is flexible enough to model independence among some variables while making no assumption about the dependencies among others.

Kolmogorov posed the question: what are bounds on the distribution function of a sum given fixed marginal distributions? Only quite recently did Makarov (1981) solve Kolmogorov's problem by finding the best possible bounds for the distribution. Frank et al. (1987) showed that the solution to the Kolmogorov problem is a special case of a simple inequality and generalized the result to arbitrary functions increasing in each place such as multiplication of positive variables. Williamson and Downs (1990) articulated the strategies for how these bounding formulas could be used in practice, extended the proof of the best possible nature of the bounds, and described software code to implement the calculations. Here we briefly describe the main result of this body of work and give a numerical example to illustrate the calculation.

Let us start with the main theorem of Frank et al. (1987). Let X and Y be random variables on $\Re^+$ with distribution functions F and G respectively. For the

distribution function of $Z = L(X, Y)$ where L is a binary operation onto $\Re^+$ that is non-decreasing in each place and continuous (except possibly at infinity), $\tau_{W,L}(F, G)$ is an upper bound and $\rho_{W,L}(F, G)$ is a lower bound. These are pointwise best possible bounds.

The bounds $\tau_{W,L}(F, G)$ and $\rho_{W,L}(F, G)$ mentioned in the theorem are the supremal and infimal convolutions respectively, which are defined by

$$\tau_{C,L}(F,G)(z) = \sup_{z=L(x,y)} C\big(F(x), G(y)\big)$$
$$\rho_{C,L}(F,G)(z) = \inf_{z=L(x,y)} C^d\big(F(x), G(y)\big)$$

where $C^d(u, v) = u + v - C(u, v)$, as applied to the copula $W(u, v) = \max(u + v - 1, 0)$, which is the lower Fréchet-Hoeffding limit on all copulas. It is interesting that the lower copula bound determines both the lower and the upper bound on the distribution function of Z. The upper bound has no role in the calculation.

Setting $L(X, Y)$ to addition (which is increasing in each place) and simplifying, the upper and lower bounds on the distribution of $Z=X+Y$ are

$$\tau_{W,+}(F,G)(z) = \sup_{z=x+y} \max\big(F(x) + G(y) - 1, 0\big)$$
$$\rho_{W,+}(F,G)(z) = \inf_{z=x+y} \min\big(F(x) + G(y), 1\big)$$

respectively. Actually, these bounds are valid for all real values because addition is monotonic over all $\Re$. Comparable bounds on the product $Z=XY$ are almost identical except that the + in the conditions for the supremum and infimum is replaced with $\times$. These bounds are valid so long as X and Y are almost surely* positive. Transformations may be employed to extend the approach to multiplication between distributions that are either entirely non-negative or entirely non-positive, but distributions that straddle zero remain problematic in this approach.

Because subtraction is not increasing in each place, we require a substitution replacing Y with its image $-Y$, whose distribution is just $1-G(-y)$. Thus, the upper and lower bounds for the difference $Z=X-Y=X+(-Y)$ are

$$\sup_{z=x-y} \max\big(F(x) - G(y), 0\big)$$
$$1 + \inf_{z=x-y} \min\big(F(x) - G(y), 0\big)$$

respectively. With a similar substitution, the bounds for quotients of positive numbers can also be obtained.

*The phrase "almost surely" means "except possibly for a set of measure zero".

Note that the bounds obtained rigorously contain the distribution for *Z no matter what the correlation or dependency between* X *and* Y. They are also the pointwise best possible such bounds, which means they could not be any tighter without excluding some possible distributions. Note also that these bounds are often considerably wider that those obtained by wiggling the correlation coefficient in a sensitivity study.

Example: Suppose that $X \sim$ weibull(1.5,3.5) and $Y \sim$ uniform(2, 9). These input distributions are depicted on the left and middle graphs of Figure 28. (The Weibull distribution is truncated at the 99.5th percentile for convenience.) The best possible bounds on the distribution of the quotient without making any assumption about the dependence between X and Y are shown in the rightmost graph. The distribution resulting from convolution under an independence assumption is also shown on the same graph in gray for comparison.

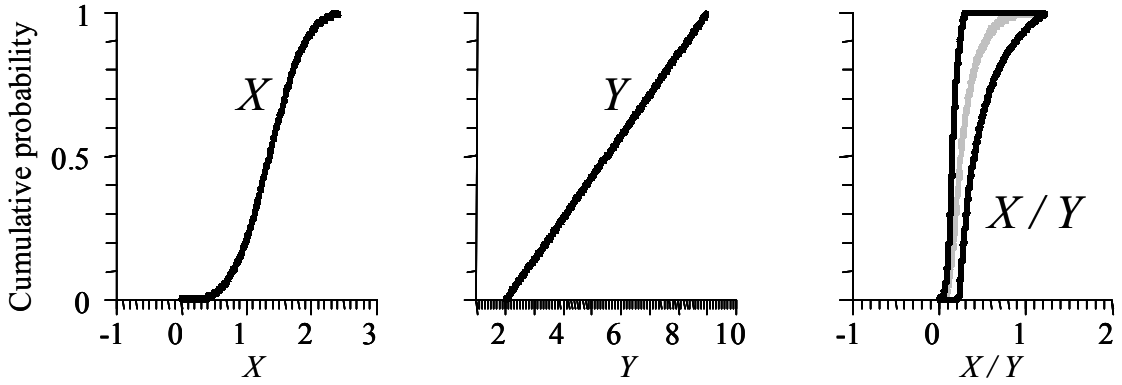

Figure 28: Dependency bounds for the quotient of random variables.

4.2.1 Fréchet bounds with Dempster-Shafer structures

This section explains how one can project Dempster-Shafer structures through arithmetic operations such as addition, multiplication, etc., without making any assumption about the dependence between the operands. We will first try to extend the approach we've used in previous sections but will discover that it does not produce good results. We will then introduce two methods that work well which are based on entirely different approaches.

The computational approach based on the Cartesian product described by Yager (1986; Ferson et al. 2003) apparently does not readily generalize to this case. To see why, consider the following numerical example. Suppose that X is the Dempster-Shafer structure {([0,1], 1/4), ([0,2], 1/4), ([2,3], 1/4), ([3,4], 1/4)} and that Y is the Dempster-Shafer structure {([2,5], 1/3), ([4,9], 1/3), ([8,9], 1/3)}. And suppose that we want to compute the Dempster-Shafer structure that characterizes the quotient X/Y without making any assumption about the dependence between the two quantities. One approach is to fashion the Cartesian product:

X / Y *X* *Y*	[0,1] 1/4	[0,2] 1/4	[2,3] 1/4	[3,4] 1/4
[2,5] 1/3	[0, 0.5] [0,0.25]	[0,1] [0,0.25]	[0.4,1.5] [0,0.25]	[0.6,2] [0,0.25]
[4,9] 1/3	[0,0.25] [0,0.25]	[0,0.5] [0,0.25]	[0.222,0.75] [0,0.25]	[0.333,1] [0,0.25]
[8,9] 1/3	[0,0.125] [0,0.25]	[0,0.25] [0,0.25]	[0.222,0.375] [0,0.25]	[0.333,0.5] [0,0.25]

where the elements of *X* are along the top row and the elements of *Y* are down the leftmost column. The first lines in the cells of the Cartesian product are the focal elements of the quotient. They are obtained by division of the marginal intervals, i.e., the focal elements of *X* and *Y* using standard interval arithmetic (Moore 1966). (Note, for instance, that [2,3] / [4,9] is [0.222,0.75] and *not* [0.5,0.333].) The second line in each cell is supposed to be the probability mass associated with that focal element. In this situation where we make no assumption about dependence between *X* and *Y*, this probability mass must be given as an interval. Following the rules described in Section 2.3 for computing probabilities of events when their dependence is unknown, the bounds of each interval are obtained from the Fréchet inequalities on the conjunction of two events: (*i*) the numerator is in that column's marginal focal element, and (*ii*) the denominator is in that row's marginal focal element. The operation is conjunction because we want the probability that *X* is in its focal element *and* that *Y* is in its focal element. In this case, and$_{\text{Fréchet}}$(1/3, 1/4) = [0, 0.25] for each cell.

This Cartesian product is unlike that suggested by Yager (1986) or the one considered in Section 3.5.2, because the probability masses do not seem to sum to unity. The sum of the twelve interval probability masses is [0, 3], which includes but does not equal 1. Note that it would be clearly wrong to arbitrarily assign equal masses to each focal element or scale them so that they sum to one. The only correct interpretation of this Cartesian product is that each cell could have as little as zero or as much as one quarter of the mass. Because we are making no assumption about dependence, we're not sure exactly where the mass might go. Figure 29 depicts the focal elements associated with the quotient that were computed in the Cartesian product. (In this graph, the vertical scale is not meaningful; the intervals are simply arranged in an arbitrary vertical order so that their horizontal locations will be clear). Note that the expression 'focal elements' for these intervals could be considered an abuse of terminology because their masses don't strictly sum to one. The cumulative plausibility function of a Dempster-Shafer structure characterizes the most leftward disposition of mass consistent with that structure. For the array shown in Figure 29, the cumulative plausibility function computed by this approach based on the Cartesian product must be the unit step function at zero $H_0(x/y)$. It would have to be there because 6 of the focal elements have left endpoints at zero. Any four of them could be such that all of the

mass is assigned to this value. The cumulative belief function, on the other hand, reflects the most rightward disposition of mass. If each focal element contains at most a quarter of the mass, we need only examine the largest four right endpoints. These are 2, 1.5, 1 and 1. Thus, the cumulative belief function computed by this approach is a discrete distribution with half its mass at 1, a quarter of its mass at 1.5 and the remaining quarter of its mass at 2. See the gray step functions in Figure 30.

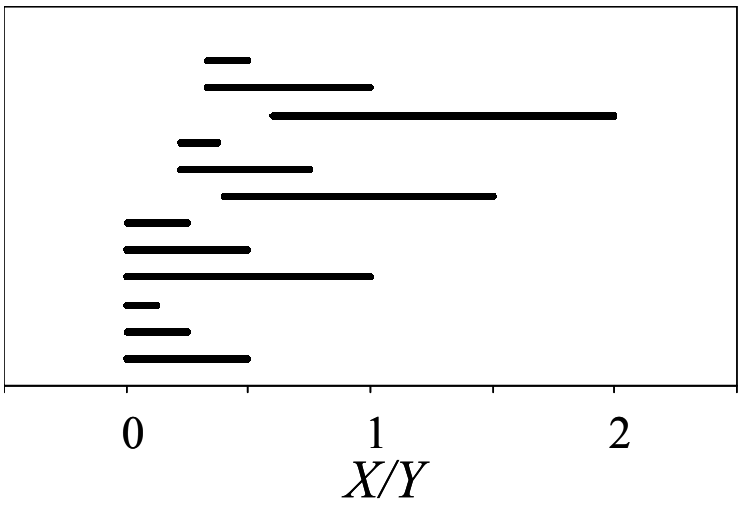

Figure 29: 'Focal elements' from the Cartesian product for the quotient *X/Y* under no dependence assumptions (only the horizontal position is meaningful).

Although the computational approach based on the Cartesian product respects the constraint that probabilities have to sum to unity, it does not take a full proper account of this fact to obtain the best possible bounds on the quotient. The probability interval given for each cell in the Cartesian product represents the best possible bounds on the mass associated with that interval, but there are necessary relationships between the elements that are not captured in this computational approach. For instance, notice that the focal elements in 6 cells of the Cartesian product overlap zero. According to the mass calculations in these 6 cells, the total mass that could be at zero must be between 0 and 1.5. Of course, we know that the mass cannot possibly be more than 1! Indeed, we can see by inspection that, at most, only half of the mass for *X* could be at the value zero because two columns have focal elements containing zero, and each has a probability mass of ¼ according to the top margin.

Berleant and Goodman-Strauss (1998; Berleant and Cheng 1998; Berleant and Zhang 2004a; 2004b) described a mathematical programming approach to the computation that produces a substantially better answer because it takes account of the information and constraints in a fully comprehensive way. Their approach maximizes the probability represented by the cumulative plausibility function (or left bound of a p-

box) for every value of the abscissa, subject to the constraints that the masses in each row must add up to 1/3, and the masses in each column must add up to 1/4, and all masses must be non-negative. The approach likewise minimizes the probability represented by the cumulative belief function at every value of the abscissa subject to the same constraints. More generally, a linear program is formulated for each abscissa value for each bound that takes as constraints the fact that the probability masses in the cells of each interior (non-marginal) row must sum to the marginal probability of that row, and likewise for each interior column. This approach neatly sidesteps the problem of determining exact joint distributions of probability masses and finds the maximum (or minimum) value of the p-box bounds given the set of all joint probability mass distributions satisfying the constraints. The output from these mathematical programming exercises yields the best-possible cumulative plausibility and belief functions for arithmetic operations and other binary functions that are monotonic in all directions (including diagonally; see Section 8 for an example of a function that s not monotonic in the diagonal direction) over the range of the input arguments. Because linear programming is required to implement this approach, they must generally be done by computer, even for fairly simple problems.

Example: Consider again the numerical example with which this section began. Evaluate the quotient X/Y without making any dependence assumption, where X is the Dempster-Shafer structure {([0,1], 1/4), ([0,2], 1/4), ([2,3], 1/4), ([3,4], 1/4)} and that Y is the Dempster-Shafer structure {([2,5], 1/3), ([4,9], 1/3), ([8,9], 1/3)}. The cumulative plausibility and believe functions for the two inputs are shown in the left two graphs of Figure 30. The calculations implied by the Berleant-Goodman-Strauss algorithm are too complex to show, but they yield the Dempster-Shafer structure ([0, 0.375], 0.08), ([0, 0.5], 0.25), ([0, 0.75], 0.09), ([0, 1], 0.11), ([0.222, 1], 0.14), ([0.222, 1.5], 0.09), ([0.222, 2], 0.01), ([0.333, 2], 0.25)}. The cumulative plausibility and belief functions for this structure are displayed as black bounds on the right-most graph of Figure 30. For contrast, the analogous functions that would have been obtained from the approach based on the Cartesian product are displayed as the gray bounds. As expected, the black bounds are considerably tighter than the gray bounds. The optimal bounds produced by the this calculation involving linear programming are identical to the results that would be obtained from the approach that will be described next in Section 4.2.2.

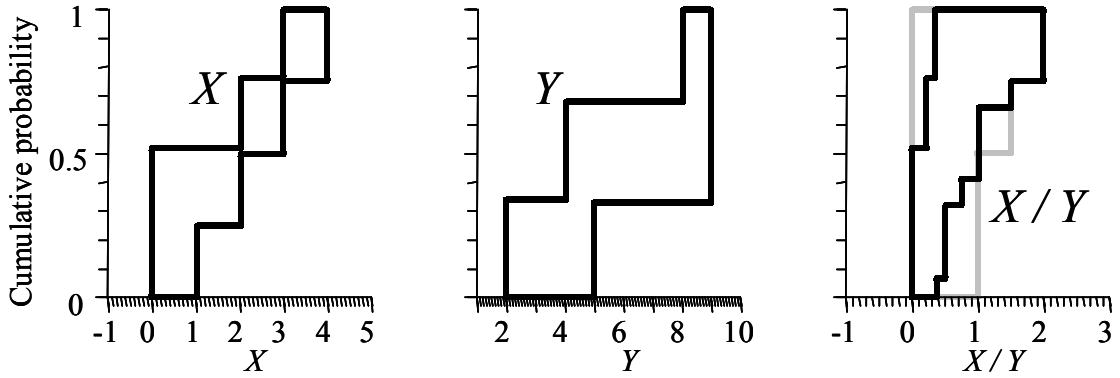

Figure 30: Quotient of Dempster-Shafer structures without any dependence assumptions.

4.2.2 Fréchet bounds with p-boxes

Williamson (1989; Williamson and Downs 1990, page 112) explained how the Fréchet bounds could be used with probability boxes. If the p-box for the quantity X is $[\overline{F}_X, \underline{F}_X]$ and the p-box for the quantity Y is $[\overline{F}_Y, \underline{F}_Y]$, then the p-box for the quantity $X+Y$ computed without dependence assumptions is $[\overline{F}_{X+Y}, \underline{F}_{X+Y}]$ where

$$\overline{F}_{X+Y}(z) = \inf_{z=x+y} \min\left(\overline{F}_X(x) + \overline{F}_Y(y),\ 1\right),$$
$$\underline{F}_{X+Y}(z) = \sup_{z=x+y} \max\left(\underline{F}_X(x) + \underline{F}_Y(y) - 1,\ 0\right).$$

In contrast, the analogous formulas for the p-box of the difference $X-Y$ are

$$\overline{F}_{X-Y}(z) = 1 + \inf_{z=x-y} \min\left(\overline{F}_X(x) - \underline{F}_Y(y),\ 0\right),$$
$$\underline{F}_{X-Y}(z) = \sup_{z=x-y} \max\left(\underline{F}_X(x) - \overline{F}_Y(y),\ 0\right).$$

Note that in these formulas, the upper bound is combined with the lower bound and the lower bound is combined with the upper bound. This is reminiscent of interval subtraction (Moore 1966). The formulas for multiplication and division of p-boxes characterizing non-negative quantities are very similar to the formulas for addition and subtraction respectively. The only differences are in the conditions for the supremums and infimums ($z=x+y$ becomes $z=xy$ for multiplication and $z=x-y$ becomes $z=x/y$ for division). Although the formulas given here are rather simple, they are not well suited to calculation via computer because they involve finding the largest and smallest values

over an infinite number of cases, but Williamson and Downs (1990) gave explicit algorithms that are convenient for such calculations.

Example: Suppose that $X \sim$ weibull([1,2], [3,4]) and $Y \sim$ uniform([2,5], [8,9]). (This means that X is distributed as a Weibull distribution characteristic life between 1 and 2, and shape parameter between 3 and 4, and Y is uniformly distributed but we are unsure about what the endpoints of this distribution are. See Ferson et al. 2003 for details about how p-boxes are generated from such specifications.) These inputs, along with the resulting p-box for the quotient X/Y are depicted in Figure 31. The result was computed without appealing to any assumption about the dependence between X and Y. It was obtained using the Williamson and Downs (1990) formulations described above. For instance, the upper bound on the distribution of the quotient was computed as

$$\overline{F}_{X/Y}(z) = 1 + \inf_{z=x/y} \min\left(\overline{F}_X(x) - \underline{F}_Y(y), 0\right)$$

where $\overline{F}_X(x)$ is the left bound on the X (which is the s-shaped curve from 0 to about 1.5) and $\underline{F}_Y(y)$ is the right bound of Y (which is the straight line going from 5 to 9). This operation to obtain the upper bound on the quotient is itself a convolution because the value of the bound at any value z is computed from all possible values of x and all possible values of y that combine to give that $z = x / y$.

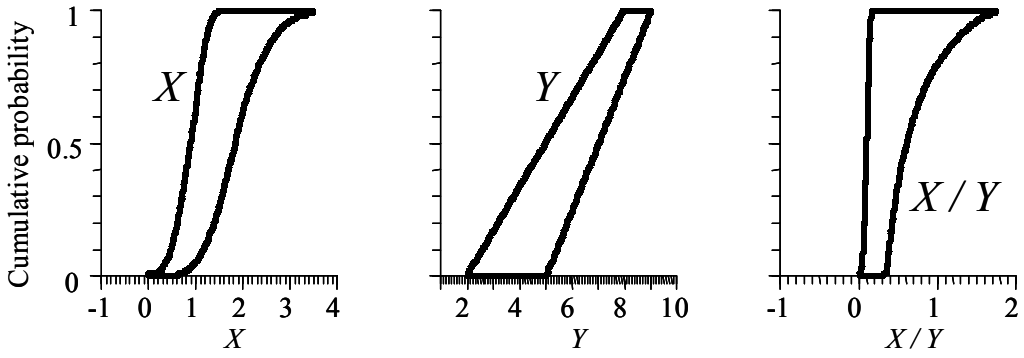

Figure 31: Quotient of probability boxes without dependence assumptions.

4.2.3 Dependence may not always matter

It is certainly true that the dependencies between random variables can often make a substantial difference to the results of risk assessments that depend on them. In such cases, knowing the dependency may provide considerable improvement (tightening) of the p-box or Dempster-Shafer structure that characterizes the answer. But there are situations involving epistemic uncertainty when knowing the dependence between random variables is irrelevant to the risk result that depends jointly on these variables. In particular, suppose a random variable X is characterized only by the interval $[x_1, x_2]$.

(This interval is equivalent to the degenerate Dempster-Shafer structure $\{([x_1, x_2], 1)\}$. It is also equivalent to the degenerate probability box $(H_{x1}(x), H_{x2}(x))$, consisting of spike bounds at the endpoints of the range.) Then, under many practical conditions, convolutions between X and any other probability distribution, Dempster-Shafer structure or p-box under the Fréchet assumption will yield the same result as the convolution under an independence assumption, or, indeed, any assumption about correlation or dependence between X and Y. We might write this as

$$X + Y = X \left|+\right| Y = X \,/+/\, Y = X \,\backslash+\backslash\, Y = convolve(X, Y, r) = convolve(X, Y, C)$$

where + denotes addition under no assumption about the dependence between X and Y, |+| denotes addition under independence, /+/ denotes addition under perfect dependence, \+\ denotes addition under opposite dependence, r is any correlation and C is any copula. Similar equivalences can hold for subtraction, multiplication, etc. This means that, if one of the operands is an interval or something equivalent to an interval, the knowing the dependence does not allow one to tighten or improve the estimate of a convolution involving it. In this sense, having only an interval estimate for a quantity means that the analyst is freed from any concern about its dependence with other variables with which it is to be combined in the assessment. Fetz and Oberguggenberger (2004) review the conditions under which these equivalences hold.

There are two caveats about this. The first caveat is that this applies only to convolutions of distributions. Dependence will generally still be relevant and important for tightening the result from the Fréchet case if the interval represents a probability of an event. The second caveat is that there are exceptions to this even for convolutions of distributions. For instance, suppose we want to compute the product XY where X is the interval $[-1, +1]$ and Y is the Dempster-Shafer structure $\{([-1, 0], 0.5), ([0, 1], 0.5)\}$. The convolution via Yager's (1986) Cartesian product, which assumes random-set independence, produces the same answer as the Fréchet case, but it is not* best possible if X and Y are strongly dependent (see Section 3.1.3).

4.3 Using partial information about the dependency

The algorithms of Yager (1986), Williamson (1989; Williamson and Downs 1990) and Berleant (1993; 1996; Berleant and Goodman-Strauss 1998) provide bounds for convolutions (*i*) under an assumption of independence, and (*ii*) in the Fréchet case with no assumption at all about dependence between the variables. The bounds for the Fréchet case are generally much wider than for they are for independence. It would be helpful to be able to tighten the bounds when partial information is available about dependencies among the input variables. Ideally, for instance, an analyst might like to be able to translate an empirical statement like "the correlation coefficient is between 0.5 and 0.6" into bounds on the convolution. It might likewise be desirable to account for the qualitative information about dependency to tighten the uncertainty in a risk

*It is worth understanding why it isn't because it highlights a difference between random-set independence and strong independence. See Section 8 for the explanation and proofs.

assessment. The result would be a tightening over the Fréchet case but a relaxation of the result based on a precise dependency model.

The algorithms currently used to compute convolutions between dependent variables, or between variables with unknown dependence, are of two distinct kinds. The first approach is fundamentally mathematical programming (e.g., Berleant and Zhang 2003; Berleant and Goodman-Strauss 1998). The second way is an analytical approach based on sets of copulas (e.g., Williamson and Downs 1990). Consequently, there are two avenues on which to seek ways to use partial information about dependencies.

4.3.1 Lower bound on copula

This section describes a general way that the copula approach of Williamson and Downs (1990) can be made to produce tighter bound on convolutions from partial knowledge about the dependency between the variables involved.

The Fréchet bounds (Section 4.2) limit the distributions of convolutions of variables when their dependence is totally unknown. The computation of these bounds depends on the marginal distributions of the variables, which operation ($+$, $-$, $\times$, $\div$) characterizes the convolution, and the Fréchet-Hoeffding lower limit W, which is the lower bound on all copulas. The bounds on the distribution functions could be tighter if the copulas could be restricted to a smaller set than "all possible dependencies". For instance, if C is a different lower bound on the copula connecting random variables X and Y with distribution functions F and G respectively, then $\tau_{C,+}(F, G)$ is the pointwise best possible upper bound and $\rho_{C,+}(F, G)$ is the pointwise best possible lower bound on the distribution of $Z=X+Y$, where

$$\tau_{C,+}(F,G)(z) = \sup_{z=x+y} C\big(F(x), G(y)\big)$$

$$\rho_{C,+}(F,G)(z) = \inf_{z=x+y} C^d\big(F(x), G(y)\big)$$

where $C^d(u, v) = u + v - C(u, v)$. The lower bound on the copulas need not be described parametrically or even have a closed-form expression. It could, for instance, be represented in a lookup table which could be computationally efficient.

There are some situations in which partial knowledge about the dependence could be used to obtain substantially tighter bounds on convolutions. For instance, knowing the value of the copula at one or more points allows us to improve the Fréchet-Hoeffding bounds (Nelsen 1999, page 62). (Such points should be in the interior of the unit square, because all the points on its edge are already specified for any copula.) Suppose the value of the copula at the point (a, b) is θ. This value must of course be in the interval $[\max(a+b-1, 0), \min(a, b)]$, which is the elementary Fréchet limit. Then the lower bound on the copula is the function

$$C_{lower}(u,v) = \begin{cases} \max(0, u-a+v-b+\theta), & u \le a,\ v \le b \\ \max(0, u+v-1, u-a+\theta), & u \le a,\ b \le v \\ \max(0, u+v-1, v-b+\theta), & a \le u,\ v \le b \\ \max(\theta, u+v-1), & a \le u,\ b \le v. \end{cases}$$

Knowing multiple points allow one to create a synthetic bound as the maximum at each point (u,v) of all these functions. A similar upper bound can be specified for the copula, but our interest focuses on the lower bound because it determines both upper and lower bounds on convolutions.

Nelsen et al. (2001; 2004) studied the problem of bounding copulas. Cossette et al. (2001) also reviewed ways to improve the Fréchet bounds given partial information about correlation and covariance among the inputs.

4.3.2 Sign of dependence

A promising approach to tighten risk calculations is to make use of information about the sign of the dependence between the variables. There are several notions of sign dependence. In this section, we consider the most important and common one.

Two random variables are "positive quadrant dependent" (PQD, Lehmann 1966; Hutchinson and Lai 1990; Nelsen 1991; 1995) if the probability that the random variables are both small (or large) together is at least as great as if they were independent. For random variables, $X \sim F(x)$ and $Y \sim G(y)$ whose joint distribution is H and whose copula is C, the following definitions of positive quadrant dependence are equivalent in the context of precise probabilities:

i) $P(X \le x,\ Y \le y) \ge P(X \le x)\ P(Y \le y)$ for all x and y,
ii) $H(x, y) \ge F(x)\ G(y)$ for all x and y,
iii) $C(u,v) \ge uv$ for all u and v, and
iv) $P(Y \le y \mid X \le x) \ge P(Y \le y)$.

Thus, random variables are PQD if their joint distribution function is no smaller than the product of their respective marginal distribution functions at every point in the space of the two variables. There are several conditions that imply variables will be PQD, including when each is a stochastically increasing function of the other, i.e., $P(Y>y \mid X=x)$ is a non-decreasing function of x for all y, and $P(X>x \mid Y=y)$ is a non-decreasing function of y for all x. Positive quadrant dependence implies non-negative Pearson, Spearman and Kendall correlations. However, the observation that a correlation is positive does not imply the variables are PQD. Even knowing a measure of correlation is 0.999 is not enough to conclude that two variables are positively quadrant dependent. This notion of dependence is quite a bit stronger than that measured by the traditional coefficients. Nevertheless, it seems that is often still a plausible one that may reasonably be expected to hold in nature in many circumstances. This idea has been used in many statistical and engineering settings (see the references in Hutchinson and

Lai 1990; Nelsen 1999), and seems to capture one sense analysts have in mind when they use the phrase 'positively depends'.

In the context of imprecise probabilities, however, it seems clear that, like the concept of independence (Section 3.1.3), the concept of positive quadrant dependence will furcate into multiple notions (Peter Walley, pers. comm.) so that the various definitions above could lead to different properties. Numerical discrepancies in calculations arising from the different possible definitions have not yet, however, been observed in practice.

Bounds on the convolution of PQD variables can be computed by extending the theorem of Frank et al. (1987) simply by replacing the lower copula bound W with Π (see Section 4.3.1 and 4.2). For example, bounds for a sum of PQD variables whose marginals F and G are

$$\tau_{\Pi,+}(F,G)(z) = \sup_{z=x+y} \big(F(x)G(y)\big),$$
$$\rho_{\Pi,+}(F,G)(z) = \inf_{z=x+y} \big(1-(1-F(x))(1-G(y))\big).$$

Note that these formulas give bounds that are *not* the same as an envelope of the perfect and independent convolutions (which would be narrower and pinch to a point that may be untenable; see Figure 41). The differences in the formulas for the other arithmetic operations are similar to those for the original Fréchet bounds. Straightforward derivations allow these formulas to be extended to p-boxes. There is no known analogous algorithm based on mathematical programming for Dempster-Shafer structures, but they can be handled by first converting them to probability boxes (Ferson et al. 2003).

There is of course a complementary notion of negative quadrant dependence. Random variables X and Y are negative quadrant dependent (NQD) if and only if X and $-Y$ are positively quadrant dependent. Convolutions with NQD variables can be calculated similarly. The intersection of the convolution bounds for positive and negative quadrant dependencies is *not* the same as the bounds obtained under independence.

One significant caveat about using qualitative information about signs of dependencies is that the feasibility checks (Section 3.6.1) cannot be used to check the reasonableness of an analyst's specifications. Because there is no analog of the positive semi-definiteness constraint on correlation matrices for sign information about dependencies, analysts are free to specify almost any pattern of intervariable dependencies. This means that there are very few consistency checks that can be used by reviewers or automatically applied by software to help to ensure that the assessment is reasonable. One can, however, infer from the fact that X and Y are PQD that X and $-Y$ are NQD, and that $-X$ and $-Y$ are PQD.

Figure 32 shows the lattice of dependencies arranged by sign considerations. It characterizes the relationship between the qualitative sign dependencies considered in

this section and the quantitative correlation constraints considered in Section 4.3.3. Going up on the lattice relaxes a dependence assumption. This would tend to widen the uncertainty of the output of an analysis. Going down on the lattice specifies a dependence more fully, and would generally result in tighter results. The figure visually emphasizes that positive quadrant dependence and negative quadrant dependence are subsets, respectively, of positive and negative correlations. Each node of this lattice contains any lower node in its conceptual definition of dependency. It also 'contains' the lower node in the sense that a convolution based on the lower node's dependency will always be tighter than a convolution based on the higher node's dependency.

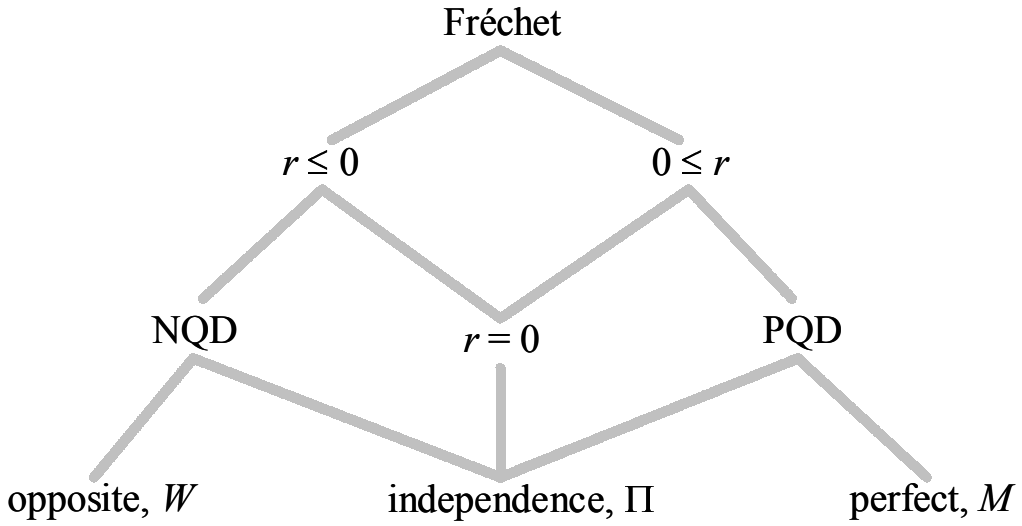

Figure 32: Lattice of sign dependencies.

Example: Consider two discrete random variables, X and Y, identically and uniformly distributed on the integers 1, …, 25. The Fréchet bounds for the convolution of these random variables are displayed on Figure 33 as gray step functions. If the variables are held to be PQD, these bounds contract to those shown in black on the figure. Note that, although there is a noticeable tightening of uncertainty, it is weakest at the tails. In assessments where focal interest is in extremal values of the output, the advantage of using knowledge about sign of the dependence may be less than might have otherwise been expected. The quantitative effects of using such knowledge will vary from problem to problem and it would premature to dismiss it without careful specific study.

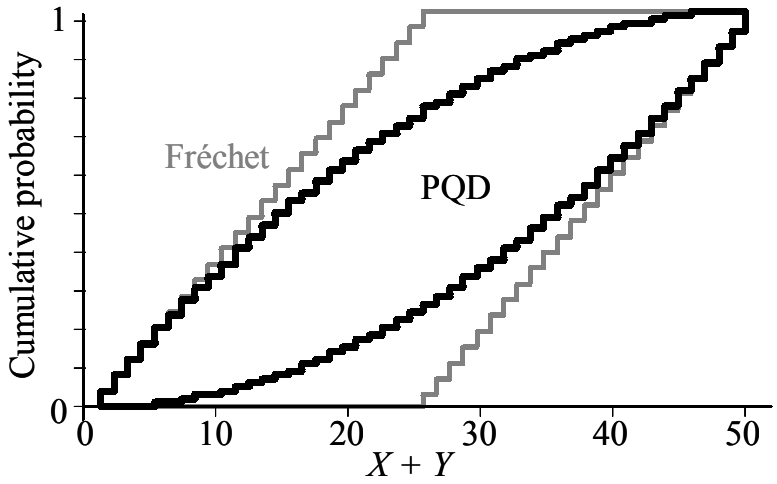

Figure 33: Convolution of positive quadrant dependent random variables (black) compared to the Fréchet bounds for the same variables (gray).

Example: The leftmost graph of Figure 34 depicts a p-box for a variable X that we know to be unimodal and to have as its minimum, maximum and modal values, 0, 1 and 0.2, respectively. The middle graph depicts the p-box for a variable Y which we know to have uniform distribution but only know the intervals [1,2] and [3,4] for its minimum and maximum respectively. Suppose that we wish to compute the bounds on the distribution of the sum $X+Y$ under the assumption that X and Y are positive quadrant dependent. According to the formulas described above, the upper and lower bounds on the distribution of the sum z are

$$\sup_{z=x+y}\left(\overline{F}(x)\overline{G}(y)\right),$$
$$\inf_{z=x+y}\left(1-(1-\underline{F}(x))(1-\underline{G}(y))\right),$$

where $\overline{F}(x) = \max(0, \min(1, x/0.2))$ is the left bound of X (which is the increasing straight line from 0 to 0.2), $\overline{G}(y) = \max(0, \min(1, (y-1)/2))$ is the left bound for Y (the increasing straight line from 1 to 3), $\underline{F}(x) = \max(0, \min(1, (x-0.2)/0.8))$ is the right bound of X (the straight line from 0.2 to 1), and $\underline{G}(y) = \max(0, \min(1, (y-2)/2))$ (the straight line from 2 to 4). To evaluate the supremum and infimums, the calculations require convolutions that consider all possible combinations of x and y that sum to the given z. The result in this case is very broadly uncertain. The uncertainty is almost as wide as the Fréchet case where we make no assumption whatever about dependence between X and Y. Knowing that the variables X and Y are positively dependent only reduces uncertainty by shaving a part off the lower right corner of the output p-box (and a much smaller part off the upper left corner).

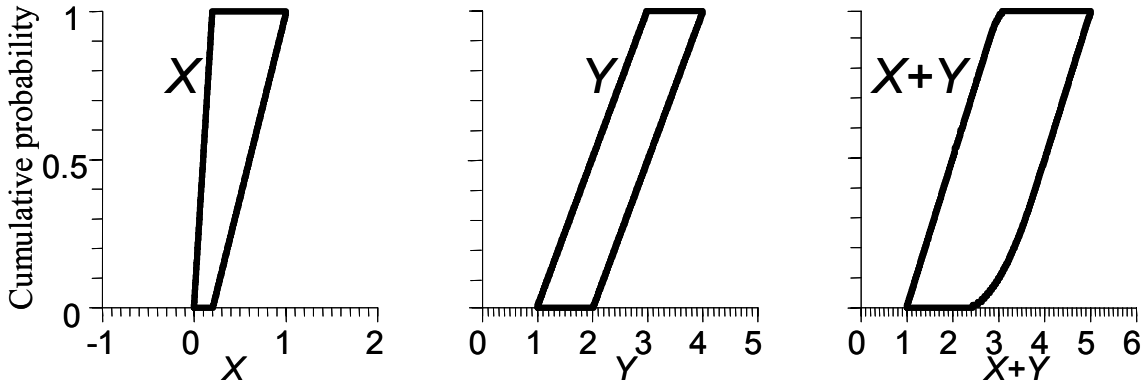

Figure 34: Convolution of positive quadrant dependent probability boxes.

4.3.3 Specified correlation coefficient

As already mentioned, it would be desirable to be able use quantitative information about dependencies among the input variables to improve the risk estimates. For instance, an analyst might like to be able to translate a statement like "the correlation coefficient is between 0.5 and 0.6" into improved bounds on the convolution. This section shows some example calculations that suggest that such information is not likely to be nearly as important as might have otherwise been thought.

Example: Suppose X and Y are both integers uniformly distributed between 1 and 4. What are bounds on the distribution function for $X+Y$ given that X and Y are uncorrelated, i.e., their Pearson correlation is zero? There is no known copula-based approach to solve this question (but see Nelsen et al. 2001; 2004). The question can, however, be framed and solved as a mathematical programming problem similar to those addressed in Berleant and Goodman-Strauss (1998) with the additional constraint that correlation is zero (Berleant and Zhang 2003a; 2003b). The result of this numerical exercise is shown in Figure 35. The gray cumulative plausibility and belief functions circumscribe the sum $X+Y$ given their uniform marginals and their lack of correlation. Also shown on the same figure in black are the bounds for the Fréchet case in which no assumption at all was made about dependence between X and Y. The surprising finding of this exercise is that most of the Fréchet case's incertitude already exists in the uncorrelated case. This means that the analyst's knowing the variables are uncorrelated does not help much to tighten the uncertainty of the result. We see that assessments in which an analyst has interpreted an empirical correlation near to zero as evidence of *independence* between the variables without ancillary evidence or argument are actually extremely weak and unreliable analyses.

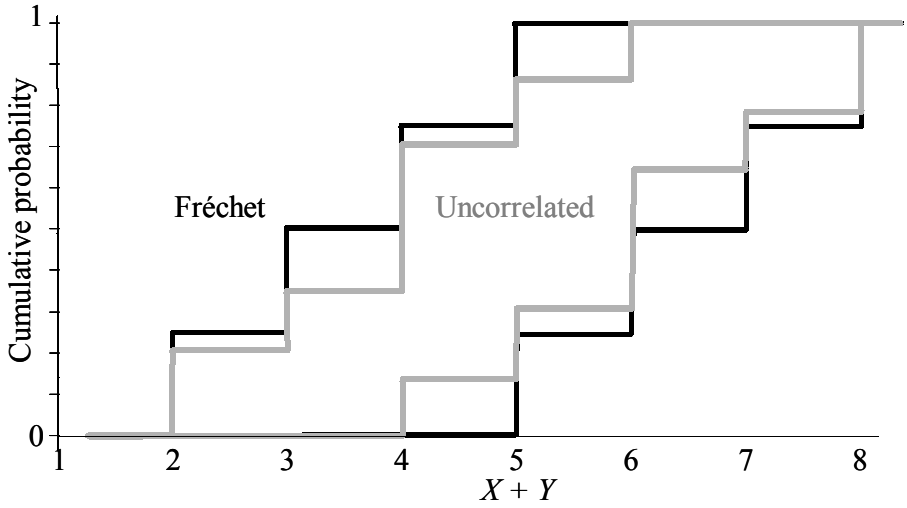

Figure 35: Bounds on convolution assuming inputs are uncorrelated compared to bounds under the Fréchet case for *X,Y* ~ discreteuniform(1,4)

Example: To dispel the notion that this finding could be due to having so few possible values in the marginal distribution, we repeated the exercise but allowed more values in the marginals. Suppose now that X and Y are integers, each uniformly distributed over the integers 1 to 25. The bounds on the convolution $X+Y$ in the Fréchet case for which no assumption is made about the dependence between X and Y are shown as black step functions in Figure 36 The bounds on the same convolution assuming that the variables have zero correlation are shown as gray step functions. Clearly, it is still the case that almost all of the uncertainty of the Fréchet case remains even when we add the information that the variables are uncorrelated. Of special interest are the distribution tails, where knowing the correlation is zero allows no improvement at all over the Fréchet case.

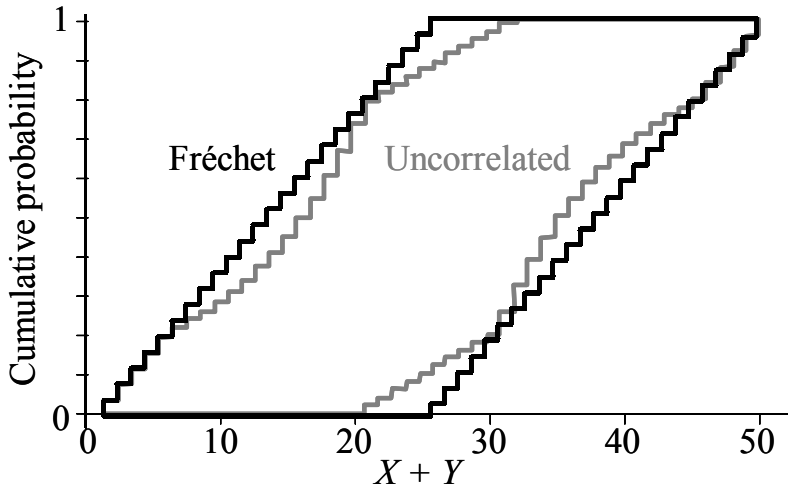

Figure 36: Bounds on convolution assuming inputs are uncorrelated compared to bounds under the Fréchet case for $X,Y \sim$ discreteuniform(1,25)

Example: We extend the previous example by specifying the correlation as 0.5 rather than zero. Again, mathematical programming is used to compute the bounds on the distribution (Berleant and Zhang 2003a; 2003b). The result shows fairly wide bounds on the sum $X+Y$, which is shown as the gray step functions in Figure 37. These bounds are slightly tighter overall than the comparable bounds for the case when correlation is zero shown in the previous figure, but they are still virtually as wide as the bounds for the Fréchet case in the tails where interest is usually focused.

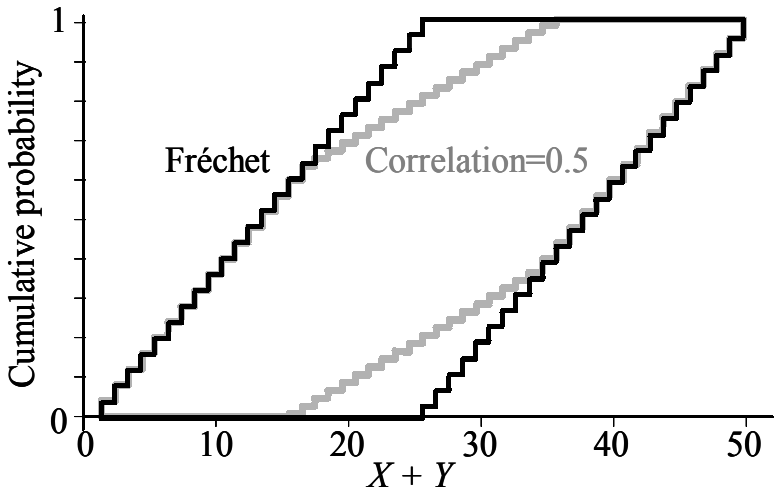

Figure 37: Bounds on convolution assuming inputs have correlation 0.5 (gray) compared to bounds for the Fréchet case (black).

We saw in Section 3.1.1 that uncorrelatedness does not imply independence, and that the variety of dependencies each yielding a correlation of zero could lead to wide variation in the resulting convolutions. The last example shows that, to a large extent, the same is true for other magnitudes of correlation. So what does quantitative information correlation tell us about convolutions? It turns out that the answer is, often, not very much. The reason for this is that many* copulas produce the same correlation. This is especially true of correlation near zero. Uncorrelatedness, when correlation is zero, entails very few constraints on the copula and therefore there is more freedom for the distribution of the convolution. There are very many ways a joint distribution can have zero correlation. The consequence of this is that, the closer a correlation gets to zero, the less it says about a dependency. This is an important fact that may seem counterintuitive. Although independence is a very strong assumption, uncorrelatedness is a very weak assumption, at least with respect to the bounds it implies for convolutions. It is interesting to note that saying a correlation is, say, 0.8, tells more about the convolution result than saying the correlation is exactly 0. As the specified correlation gets stronger and stronger (whether positively or negatively), in fact, it constrains the convolution more and more, until the extremal correlation is reached which corresponds to perfect or opposite dependence.

It is unlikely that this idea of using quantitative information about dependence can be rescued by switching to a different measure of correlation. The fact that Kendall's τ is 0, for instance, implies no nontrivial constraints on a copula *at all*. If Spearman's ρ is 0, there may be some weak constraints, but it is not entirely clear how to make use of this fact to tighten the bounds on the resulting distribution (but see Nelsen et al. 1999). The Pearson product moment correlation coefficient is itself not a function solely of the copula, but also depends on the marginals as well. Basically, the central problem is that these single scalar measures of dependency are *averages* over a two-dimensional function characterizing the interaction (the copula). But the extreme behaviors of the convolution function depend on *local* features of the copula. Thus, even if general constraint relationships could be found they are likely to be rather weak, so that knowing any of the correlation coefficients would not substantially improve or tighten the uncertainty about a calculation.

Despite the modestness of its quantitative effect on calculations, it would be misguided for analysts to neglect the accounting of all available information about correlations or dependencies in their assessments and uncertainty models. Making use of all such information, whether it be qualitative or quantitative, enhances the credibility of the resulting analysis. It is often very important to risk managers, decision makers and other consumers of uncertainty analyses that this information is incorporated and properly accounted for in the analysis, especially if the information was hard won by special empirical effort. Without assurances that the available information has been

*Recall that Figure 7 showed ten disparate dependence patterns associated with the same marginal distributions. Flury (1986) also gave examples of different copulas associated with the same marginals.

included in the analysis, they will find it easier to reject or ignore its conclusions. In situations when an analytical result can be shown to be best possible in the sense that it has the smallest uncertainty given the uncertainty present, the force of the conclusions becomes considerably stronger.

5 Myths about correlations and dependencies

This section summarizes several of the most pervasive and pernicious myths in risk analysis about correlations and dependencies. It addresses again many ideas that have already been introduced elsewhere in this report, but structures them in terms of the myths that interfere with conducting good assessments. Some of these myths are obvious and widely recognized but still often perpetuated for the sake of mathematical convenience. Other myths are more subtle or even rather esoteric and may not be widely appreciated by risk analysts. Several of the myths about correlations and dependencies outlined here can lead to profound errors in risk assessments.

Myth 1

All variables are mutually independent.

Many variables in complex natural and engineered systems are, in fact, correlated or have some nonlinear interdependence. Although most risk analysts recognized that it is improper to assume that variables are independent in the face of evidence that they are not, many do so anyway as a shortcut or mathematical convenience. In some cases, these counterfactual assumptions are laughable, as in the case of assuming mass of some component and its surface area are independent. In some cases, assuming a perfect or opposite dependence would be a better default assumption than independence. In general, it is incumbent on the analyst to model the dependence if only approximately.

There is also impropriety in cases where independence is routinely assumed when there are no observations or other evidence available about the dependence between variables one way or the other. The lack of evidence about dependence does not by itself justify an assumption of independence, although many analysts argue as though it does. Fact: wishing variables were independent so the analysis is easier doesn't make them so. In cases when the dependence is partially or completely unknown, appropriate methods to account for this epistemic uncertainty such as those described in Section 4 should be employed.

Myth 2

If *X* and *Y* are independent and *Y* and *Z* are independent, then *X* and *Z* are too.

Mutual independence between X and Y and between and Y and Z doesn't guarantee that X and Z are also independent. In other words, independence is not transitive. This fact should perhaps be obvious to risk analysts. We have nevertheless observed the corresponding faulty reasoning applied in actual risk assessments. Fact: independence between X and Y and between and Y and Z implies nothing at all about the dependence

between X and Z. Consider the following very simple example. Let (X, Y, Z) be a discrete distribution consisting of the four points (1,1,1), (1,2,1), (3,2,3) and (3,1,3), each with probability ¼. As depicted in Figure 38, plotting the three bivariate scattergrams (X versus Y, Y versus Z, and X versus Z) reveals that X and Y are independent, as are Y and Z, but that X and Z are (perfectly) positively dependent on each other. If the four equiprobable points of the discrete distribution are instead (1,1,3), (1,2,3), (3,2,1) and (3,1,1), then the first two graphs are unchanged, but the third graph would show an oppositely dependent relationship between X and Z.

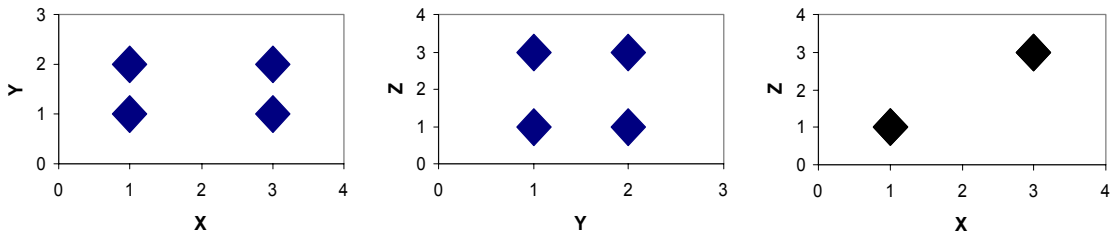

Figure 38. Discrete example of non-transitivity of independence.

Myth 3

Variables X and Y are independent if and only if they are uncorrelated.

Whenever correlation is introduced in beginning statistics courses, a counterexample to this myth such as that shown in Figure 39 is immediately presented. The variables X and Y in this graph are uncorrelated, i.e., they have a Pearson correlation of zero. However, they are clearly not independent. Despite widespread attempts to disabuse students of the difference between uncorrelatedness and independence, this myth or the consequences of the myth nevertheless pervade risk assessment. Uncorrelatedness does not generally* imply independence. Fact: independence implies that the correlation will be zero, but not vice versa.

---

*There are exceptions where uncorrelatedness actually does imply independence. One exception, for instance, is when X and Y both have Bernoulli distributions such that $P(X = 0) = P(X = 1) = 0.5$ and $P(Y = 0) = P(Y = 1) = 0.5$. Let $h(x, y)$ denote the joint mass function for X and Y. Let $a = h(0,0)$, $b = h(0,1)$, $c = h(1,0)$, $d = h(1,1)$, so $0 \le a,b,c,d \le 1$ and $a+b+c+d = 1$. Because the marginals are Bernoulli distributions, we know that $a+b = c+d = a+c = b+d = 0.5$. If X and Y are uncorrelated, then $r = E(XY) = E(X)\,E(Y)) / \text{sqrt}(V(X)\,V(Y)) = 0$, which implies $E(XY) = E(X)\,E(Y)$. But $E(XY) = \sum xy\,h(x,y) = 0\times0\times a + 0\times1\times b + 1\times0\times c + 1\times1\times d = d$. At the same time, $E(X)\,E(Y) = 0.5 \times 0.5 = 0.25$, so $d = 0.25$. But this means that b has to also equal 0.25 (because $b+d = 0.5$), and, in fact, the Bernoulli constraints cascade so that $a = b = c = d = 0.25$, which means that h is necessarily the independence copula. Thus, in this exceptional case, uncorrelatedness implies independence.

Sometimes the myth appears as "If two variables are normally distributed, then a zero correlation between them implies independence." Actually, normality of the marginal distributions is not sufficient. Melnick and Tenenbein (1982) provides counterexamples (see also Flury 1986; Kowalski 1973). Fact: if variables are *bivariately* normal, then zero correlation implies independence.

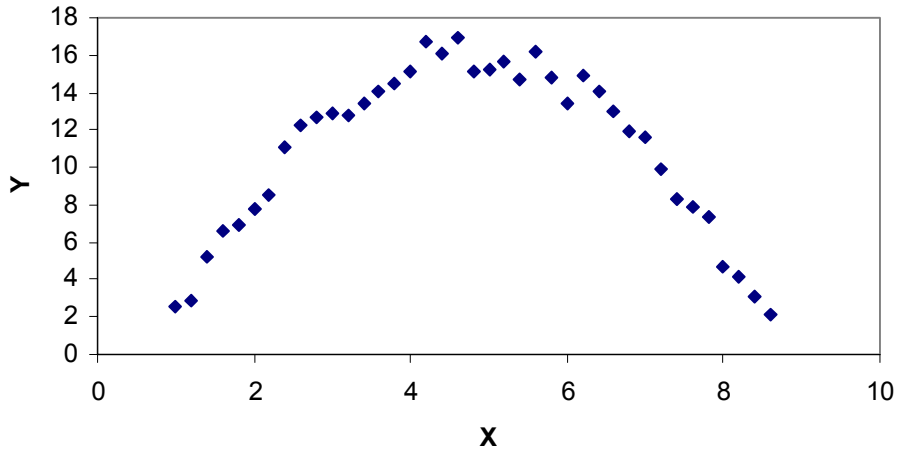

Figure 39. Variables that are uncorrelated but obviously dependent.

Myth 4

Zero correlation between *X* and *Y* means there's no relationship between *X* and *Y*.

This myth is closely related to the previous one. The phrase "no relationship" is really just another way of saying that knowing the value of either variable doesn't help in any way to establish the value of the other variable. (For precise probabilities, this is equivalent to saying the variables are independent; see Section 3.1.3.) Figure 39 also provides a counterexample to this myth. Fact: uncorrelatedness does not imply there is no relationship between the variables. *X* and *Y* are uncorrelated, but they clearly have a very strong relationship. Knowing that *X* is 3 tells us that *Y* is around 12. Knowing that *X* is 5 tells us that *Y* is around 15. Knowing that *Y* is 8 tells us that *X* is either around 2 or around 7.5. There is an immense amount of information embodied in the relationship between the two variables even though they have zero correlation. There are examples known in which random variables are uncorrelated yet are mutually completely dependent, that is, each is a function of the other.

Myth 5

Small correlations imply weak dependence.

Figure 39 also disproves this myth, which is closely related to the previous two myths. The falsity of this one is just as obvious, and yet it appears surprisingly often in multivariate data analyses and risk assessments.

Myth 6

Small correlations can be "safely ignored" in risk assessments.

In an important paper, Smith et al. (1992) suggested that small-magnitude correlations could be "safely ignored" in risk assessments seeking estimates of means of linear arithmetic functions of random variables. This is possible because *means* of sums and products are often similar to means for the independent case if a simple dependence with small correlation is introduced between the inputs. In the real world, however, there are three complications that prevent us from ignoring dependence among variables. First, many of the functions we need to evaluate are nonlinear. Second, the dependencies involved are more complicated than can be captured with simple correlation coefficients. Third, and probably most important, risk analysts are usually more concerned about the distributions' *tails* rather than their means. As has been illustrated several times in this report, tail risks can be radically influenced by dependencies even if correlation is zero. The Smith et al. (1992) paper has been widely overextended and abused, and risk analysts should generally try to account for all dependencies that relate their input variables to one another even if they might happen to yield correlations of small magnitude.

Myth 7

Different measures of correlation are similar.

Some risk analysts suggest that it doesn't make much difference which measure of correlation is employed and that the various measures are pretty much interchangeable. This view is false, however, as even cursory inspection of examples will easily reveal. There are many different measures of correlation that are in common use and many more that have been proposed. The most commonly used measures are Pearson's product moment correlation and Spearman's rank correlation, but there are a host of other measures that also arise in various engineering contexts, including Kendall's rank correlation, concordance, Blomqvist's coefficient, etc. Hutchinson and Lai (1990) review many of these. The choice of the measure can strongly influence the numerical characterization of a scattergram. Figure 40 shows a variety of bivariate relationships as scattergrams. Note that the units of the abscissa and ordinate are not shown because they are irrelevant and do not affect the magnitudes of the correlations. Each of the six scattergrams displayed has the same Spearman rank correlation, which is one, corresponding to perfect dependence or comonotonicity. But the scattergrams have

widely different Pearson correlation coefficients. For instance, the Pearson correlation for the scattergram in the upper, left-hand graph is one, but the Pearson correlation for the scattergram below it in the lower, left-hand graph is about 0.6. Fact: the various measures of correlation are sensitive to different features of the scattergram.

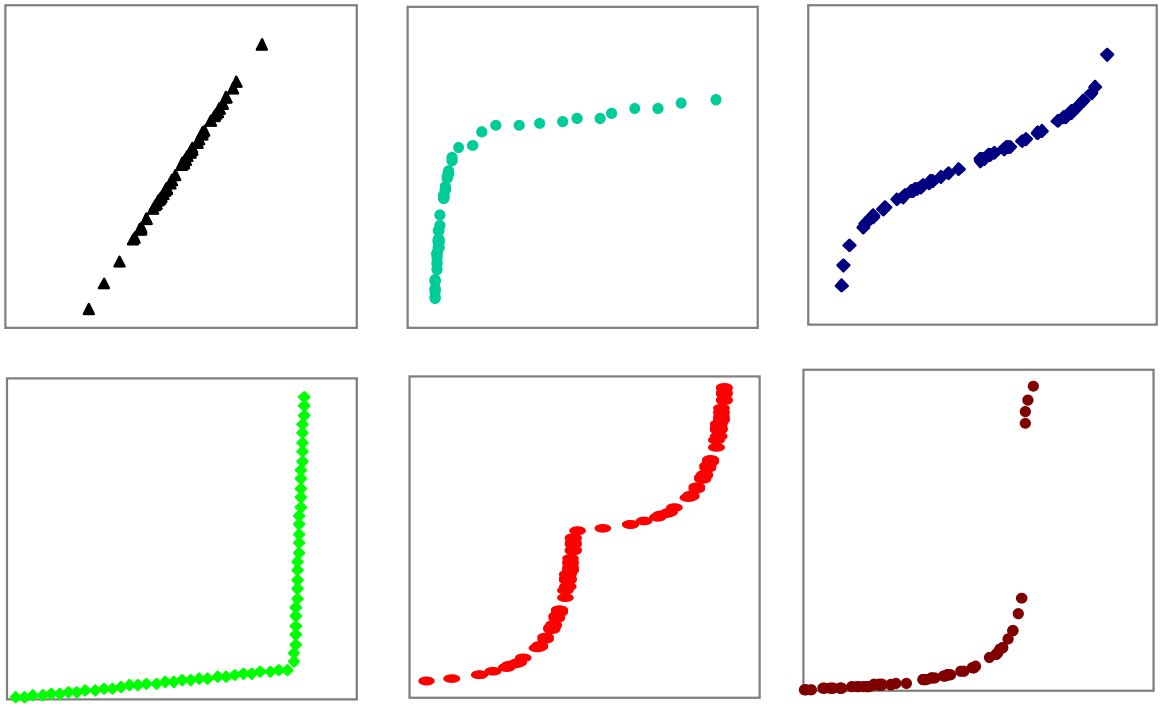

Figure 40. Different bivariate relationships with the same Spearman rank correlation (unity) but widely different Pearson correlation coefficients.

Myth 8

A correlation coefficient specifies the dependence between two random variables.

In fact, it takes a copula (dependence function) to fully specify the dependence between two random variables. A correlation coefficient is often a very poor summary of the dependence; it generally does *not* specify or determine the dependence. Instead, it determines only a class of such dependencies. In other words, many dependence functions have the same correlation. See Section 3.7 for an introduction to copulas.

Despite all appearances, Figure 40 is actually not an illustration of the fact that many dependence functions have the same correlation. All of the scattergrams in this figure have exactly the same dependence function (which is the copula M associated with perfect or comonotonic dependence). The differences among the scattergrams in Figure 40 are due entirely to the differences in the marginal distributions for the abscissa and ordinate variables.

In the assertion above that the correlation does not generally specify the dependence, the adverb "generally" was necessary because there some exceptions when the correlation does completely determine the dependence. One exception is when the correlation is extreme, that is, when the dependence is perfect or opposite. In this case, the Spearman rank correlation and Kendall correlation are either +1 or −1 (and the Pearson correlation is as large or as small as it can get given the marginal distributions). When one of these correlation measures is ±1, the dependence function is determined to be *M* or *W* respectively (see Section 3.7). Interestingly, as the correlation gets closer to zero, the family of dependence functions having that correlation gets larger and larger in its diversity. (This fact tends to explain why Myth 6 is not true.) See Figure 7 for examples of scattergrams corresponding to fixed marginal distributions with different dependence functions that all have the same Pearson correlation.

Myth 9

Pearson correlation varies between −1 and +1.

By convention, most measures of correlation are scaled so that they range in the interval [−1, +1]. Some measures, such as Spearman correlation, can always range over this entire interval. But not all correlation coefficients can vary across this range for arbitrary marginal distributions. The Pearson correlation, in particular, often cannot achieve either −1 or +1. For instance, if *X* is uniformly distributed over the unit interval [0,1] and *Y* is a lognormal distribution with underlying $\mu = 0$ and $\sigma = 1$, then the correlation between *X* and *Y* cannot be any larger than about 0.7. Depending on the marginal distributions involved, the largest possible Pearson correlation could in fact be arbitrarily close to zero. Fact: the Pearson correlation coefficient ranges within [−1, +1], but it may not reach all possible values in the interval for some marginal distributions.

Myth 10

Any patterns of correlations can be specified between multiple input variables.

This myth is unrelated to the previous one, where constraints arose because of marginals. Here we consider constraints on correlation that are irrespective of marginals. If variable *X* is strongly correlated to both variables *Y* and *Z*, then it may not be possible that *Y* and *Z* are strongly negatively correlated to each other. Fact: the pairwise correlations for a set of variables must satisfy certain feasibility constraints, so not all sets of correlations that one might specify are possible. These constraints are rather complicated, but can be summarized by saying that a correlation matrix must be positive semi-definite (see Section 3.6.1). Early versions of the software package @Risk (Palisade Corporation 1996; Salmento et al. 1989; Barton 1989; Metzger et al. 1998) did not account for this constraint, and consequently would have produced nonsensical results whenever users would specify an infeasible set of correlations.

Myth 11

Perfect dependencies between *X* and *Y* and between *X* and *Z* imply perfect dependence between *Y* and *Z*.

Extending the ideas discussed above about constraints on the correlation matrix, one might have expected that if a variable X is maximally correlated to variable Y, and variable Y is maximally correlated to variable Z, then we might be able to conclude that and X and Z are also maximally correlated. Expressed in other terms, comonotonicity between both X and Y and between Y and Z would seem to imply there should likewise be comonotonicty between X and Z. Furthermore, one might expect that if X and Y are maximally correlated (comonotonic) and Y and Z are minimally correlated (countermonotonic) then X and Z should be minimally correlated (countermonotonic) too.

Let // denote perfect dependence, i.e., maximal correlation and comonotonicity, and let \\ denote opposite dependence, i.e., minimal correlation and countermonotonicity. Below are facts that correct some of the mistaken ideas:

Fact: $X // Y$, and $Y // Z$ do not generally imply $X // Z$.
Fact: $X // Y$, and $Y \backslash\backslash Z$ do not generally imply $X \backslash\backslash Z$.
Fact: $X \backslash\backslash Y$, and $Y \backslash\backslash Z$ do not generally imply $X // Z$.

Perhaps even more surprising is that $X // Y$ and $Y // Z$ together don't even imply that X and Z can't be independent. A counterexample is easy to construct. Let (X, Y, Z) be discrete, taking on of the four values (1,1,1), (1,2,3), (3,2,1), and (3,3,3), each with probability ¼. Sketching the three bivariate plots reveals that $X // Y$ and $Y // Z$, but X and Z are independent. It is possible to conclude from perfect dependence between X and Y and between Y and Z that X and Z cannot be oppositely dependent, but that is a fantastically weaker conclusion that will rarely matter in a practical risk assessment.

If Y and Z are independent, then $f(Y)$ and $g(Z)$ are also independent, where f and g are arbitrary measurable functions (Roussas 1997, page 166). One might expect this fact could be extended to comonotonic or countermonotonic variables, but this is not the case. Let $\perp$ denote independence.

Fact: $X // Y$, and $Y \perp Z$ do not imply $X \perp Z$.
Fact: $X \backslash\backslash Y$, and $Y \perp Z$ do not imply $X \perp Z$.

The combination of perfect dependence with independence is subtle, and the mistakes that analysts make are understandable. In fact, however, assuming perfect or opposite dependence between X and Y and independence between Y and Z doesn't allow any conclusion at all about the dependence between X and Z. Any relationship between them is possible. One example would be where (X, Y, Z) take on the four values (1,1,3), (2,1,1), (2,3,3) and (3,3,1), each with probability ¼. Bivariate sketches show that $X // Y$

and $Y \perp Z$, but $X \setminus\setminus Z$. If the equiprobable points were instead (1,1,1), (2,1,3), (2,3,1) and (3,3,3), then still $X /\!/ Y$ and $Y \perp Z$, but now $X /\!/ Z$.

This flexibility about dependencies might be surprising because it seems to contradict the strictures on correlations mentioned in the discussion of the previous myth. In fact, the constraint of positive semi-definiteness that correlations must observe does not generalize to the case of dependencies, even in the extreme cases where correlations are minimal or maximal. However, this flexibility disappears if the analyst makes the assumption that the random variables have continuous distributions. In this case, $X /\!/ Y$, and $Y /\!/ Z$ *does* imply that $X /\!/ Z$, and similarly for the other displayed facts on the previous page. This highlights the strength of assumptions about the continuity of distributions, which themselves would need specific justification.

Myth 12

Monte Carlo simulations can account for dependencies between variables.

Cullen and Frey (1999, page 202) complain that critics of Monte Carlo simulation unfairly accuse it of "ignoring correlations". They point out that restricted pairing methods developed by Iman and his colleagues allow analysts to construct deviates in Monte Carlo simulations that have a prescribed correlation (Section 3.6; Iman and Conover 1982; Helton 1993; Helton and Davis 2002; 2003). However, what Cullen and Frey don't mention is that these algorithms pick a *particular* dependency function with the prescribed correlation, and that this is only one of infinitely many possible dependencies having this correlation. Fact: Monte Carlo methods can simulate correlations, but they do so by making unstated assumptions about the nature of the copula representing the dependence function. Monte Carlo methods cannot truly account for correlations in the sense of computing how low or high risks might be without making such assumptions. As discussed in Section 3.1.1, the effect on numerical results of these different dependence functions can be substantial, even though they may all have the same correlation coefficient.

The origin of Myth 7 discussed above seems to be due to the mistaken impression that Monte Carlo simulation can account for correlations. Given a magnitude of the correlation, one observes scattergrams from Monte Carlo simulations that are fairly similar to one another whichever measure of correlation is used. For instance, it may be hard to distinguish scattergrams generated in simulations using thei different measures. Their similarity, however, may mostly be a consequence of the very myopic selection of the copulas used in Monte Carlo simulations to generate correlated deviates. There are generally other possible choices for the dependence function that have the same correlation and yet produce substantially different scattergrams and result in considerably different answers.

Myth 13

Varying correlation coefficients constitutes a sensitivity analysis for uncertainty about dependence.

Fact: varying correlations is insufficient to explore the possible nonlinear dependencies between variables. For this reason, a sensitivity analysis based on varying correlation gives an incomplete picture of uncertainty that is far too tight, even if we vary correlation between −1 and +1. As an example, consider the problem of estimating the distribution of the sum $A+B$, where A and B are both uniform random numbers over the interval [2,13] but the dependence between A and B is unknown. The range of distributions that would be seen in Monte Carlo simulations by varying the correlation between A and B over its possible range of [−1, +1] is shown in Figure 41 as a gray slanted hourglass. The simulation strategies that could be used to obtain the hourglass are described in Section 3.6 and, in particular, in Section 4.1. This hourglass can be compared with the best possible limits on these distributions with no assumptions about dependence. These limits form a black parallelogram underneath the hourglass shape. The limits can be computed using the methods of Frank et al. (1987; Williamson and Downs 1990; Berleant and Goodman Strauss 1998) described in Section 4.2. The black parallelogram is substantially larger than the gray hourglass, and, although the bulk of the difference is about the central parts of the distribution rather than the extreme tails, the potential tails risks are underestimated by the Monte Carlo sensitivity analysis strategy.

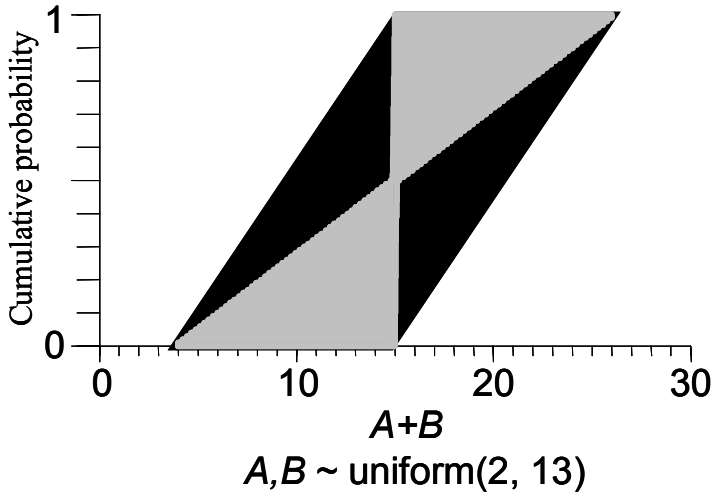

Figure 41. The envelope of distributions of sums $A+B$ obtainable from Monte Carlo simulations varying the A,B correlation between −1 and +1 (gray slanted hourglass) and best possible limits on these distributions making no assumptions about dependence (black parallelogram).

Myth 14

A model should be expressed in terms of *independent* variables only.

The idea behind this myth is discussed in Section 3.2. Strictly speaking, it might better be labeled an "opinion" rather than a "myth", but the idea is so clearly unworkable in general that it seems fair to list it here with other ideas that are impediments to conducting good risk assessments. The essential problem is that modeling all the underlying sources of the dependencies will often quickly become unwieldy and may be recursively complex. For instance, it would probably be unrealistic to ask a dam safety engineer to deconstruct his model's dependence on rainfall patterns in terms of independent variables (because even meteorologists can't do this). Fact: a statistical approach may be needed to handle dependencies in risk assessment models.

Myth 15

You have to know the dependence to model it.

Recent algorithmic advances allow the calculation of bounds on risks (1) under only partially specified dependence functions, or even (2) *without any assumption whatever about dependence*. As explained in Section 4.2, even if there is no information at all available about the dependence function relating variable X and Y for which we know the respective marginal distributions F and G, it is still possible to compute upper and lower bounds on the distribution function for $Z = X+Y$ as

$$\left[\sup_{z=x+y} \max\left(F(x)+G(y)-1, 0\right), \inf_{z=x+y} \min\left(F(x)+G(y), 1\right) \right].$$

These limits are bounds on the distribution function of the sum for every possible value z it might take. The limits are based on the classical Fréchet-Hoeffding limits for the dependence (copula) function. There are similar formulas for the distribution of differences, products, quotients, etc. Similar methods can be applied to probability boxes and Dempster-Shafer structures.

When there is partial information about the dependence function, such as that the relationship between X and Y is certainly positive (positive-quadrant dependent), then bounds on the distribution for Z can be computed with a formula like

$$\left[\sup_{z=x+y} \left(F(x)G(y)\right), \inf_{z=x+y} \left(1-\left(1-F(x)\right)\left(1-G(y)\right)\right) \right].$$

There are similar formulas for differences, etc., and for dependence functions that are surely negative. Similar methods can be applied to probability boxes and Dempster-Shafer structures. See Section 4.3 for a full discussion of these formulas.

Myth 16

The notion of independence generalizes to imprecise probabilities.

In fact, there are several distinct concepts that might deserve to be called independence in the context of imprecise probabilities (Couso et al 1999). These concepts are not equivalent to one another and they can lead to numerical differences in convolution results (Fetz and Oberguggenberger 2004). The issue is discussed in Section 3.1.3.

6 Choosing a strategy for modeling dependence

Figure 42 depicts the relationships among the various possible dependence assumptions that have been described in this report. It illustrates the fundamental tradeoff in modeling dependencies. One can have a result that makes few or no assumptions about dependence, but this result will be wider than it could be if the analyst had made use of available knowledge about dependence, or one can have a result that will be narrower but makes strong and perhaps unjustifiable assumptions about dependence.

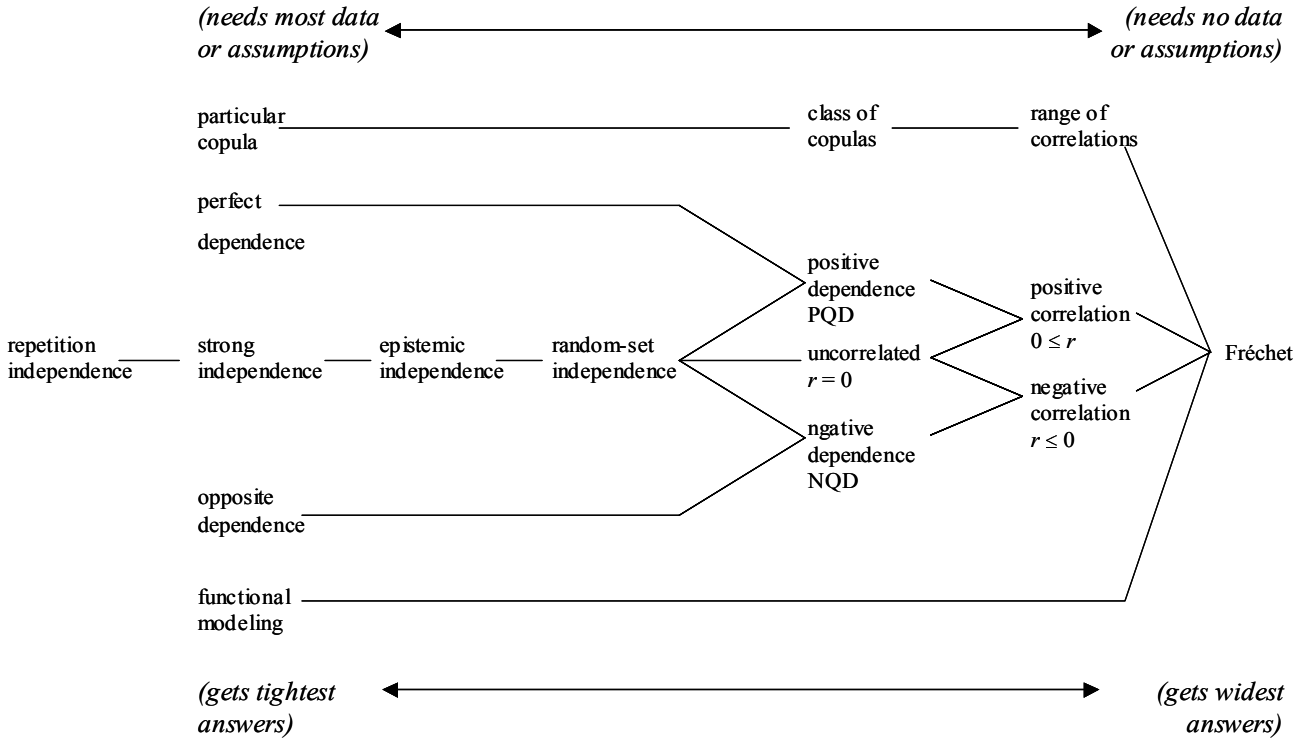

Figure 42: Relationships among dependency assumptions.

What is not emphasized in the figure is that, once one abandons the notion that the dependence function is known completely (i.e., that the particular copula can be specified precisely), then the resulting breadth of uncertainty in subsequent convolutions can be fairly large compared to the total possible breadth obtained in the Fréchet case. The uncertainty is often not greatly diminished even with additional qualitative information such as the sign of dependence or even with quantitative information such as the magnitude of the correlation coefficient. Consequently, once one admits uncertainty about the precise nature of the dependency function, there may

be little difference between the best possible bounds that account for what is known and the Fréchet bounds which make no assumption whatever about dependence. Further research is needed to discover the circumstances when partial information about intervariable dependencies can have a significant effect on the results of a calculation. In any case, however, such issues can be localized to considerations of particular pairs of variables or to small groups of variables; using the Fréchet case for some variables does not necessitate its use for other variables. The Fréchet case can be used in the same model with the independent case and, indeed, other assumptions about dependence.

The traditional approach to modeling dependencies in probabilistic assessments has been to start with the default assumption of independence and then consider evidence or scenarios that suggest some correlation or perhaps a functional dependence. Essentially, an analyst would start on the left side of the middle line of the diagram in Figure 42 at independence and move either up or down. We believe that a more appropriate strategy for risk analysts is to start with no assumptions, at the Fréchet case at the far right of the figure. As information in the form of empirical evidence or theoretical arguments is considered, one moves leftward. This difference of what assumptions are made by default can have a tremendous impact on the resulting assessments. If results are seen to be too uncertain to support decisions of concern there would be impetus to review more evidence, collect new data, and devote more empirical and theoretical attention to the matter. We view such an outcome as salutary for the discipline. The present state of affairs is too often one in which a risk analysis based on convenient assumptions is simply a regurgitation of preconceptions that adds little to the assessment process.

The following list is a synopsis of the various assumptions and modeling strategies that one could employ to account for dependence among random variables. Relevant sections in this report are given for each in parentheses.

Imprecise dependence

- Fréchet case (no assumption about dependence) (Section 4.2)
- Known or interval-bounded correlation (Section 4.3.3)
- Sign of dependence (Section 4.3.2)
  - Positive dependence (Section 4.3.2)
  - Negative dependence (Section 4.3.2)
- Linear dependence
  - Known correlation (Section 3.6)
  - Interval-bounded correlation (Section 4.1)

Precise dependence

- Specified dependence model (copula family) and correlation (Section 3.7)
- Known dependence function (copula) (Section 3.7)
- Empirical dependence function (Section 3.8)
- Perfect dependence (Section 3.5)
- Opposite dependence (Section 3.5)

Independence (Section 3.1)
Random-set independence (Sections 3.1, 3.1.3)
Epistemic independence (Section 3.1.3)
Strong independence (Section 3.1.3)
Repetition independence (Section 3.1.3)

Functional modeling

Known functional relationship (Section 3.2)
Stratification (Section 3.3)
Conditioning (Section 3.4)

6.1 Modeling dependence through a black box

Many of the techniques outlined in this report can be applied to real-world risk assessment or uncertainty modeling problems in which the evaluated function is a black box that cannot be decomposed into a sequence of binary arithmetic operations. In particular, any method that can be implemented via a sampling strategy can be used. This includes models that specify a dependence model (copula family) and correlation (Section 3.7) or an empirical copula (Section 3.8). Also amenable to sampling strategies are models of known functional relationships (Section 3.2), stratification approaches (Section 3.3), models assuming independence, or perfect or opposite dependence based onWhitt's sampling strategy (Section 3.5).

When dependence is not well parameterized, then it may be possible to use dispersive sampling within a black box assessment. Such use would require the analyst to specify the slope of the partial gradient with respect to each input variable through some sort of prior sensitivity analysis of the black box function. In fact, if the black box function is known to be monotonic in each input, then, in principle, the methods to account for the Fréchet case (making no assumption about dependence) (Section 4.2), or sign of dependence (Section 4.3.2) could be theoretically employed on the black box itself, although doing so would require enough samples to evaluate the required infimal and supremal convolutions. In most cases, it would probably be more reasonable to model the black box with a computationally simpler function such as a response surface. When the response surface is simple and decomposable, the dependency methods can be applied to it directly. If it too cannot be decomposed but it is monotonic, then an intensive sampling strategy to evaluate the infimal and supremal convolutions could be employed. As an alternative strategy, it might also be useful to use constructed copulas (Section 3.9) to study the putative response of the function to especially important features of the dependence.

7 Conclusions and future research

We make six major conclusions.

i) **Dependence considerations and special strategies to model dependence are not necessarily needed for each and every operation in a risk analysis.** There are many situations in which independence is clearly warranted. There are other situations in which dependence is irrelevant (such as for unary operations negation, square root, log, etc.). And there are cases in which the dependence between variables is moot because knowing the dependence has no effect on the result of the analysis. For instance, it is typically the case that if one of the operands is an interval (or a Dempster-Shafer structure or a p-box equivalent to an interval) knowing dependence will not allow one to reduce the uncertainty in the result compared to that obtained without making any assumption about dependence. Similarly, if one of the operands is a point value, then there can be no dependency between it and any variables with which it might be combined.

ii) **A sensitivity study in which correlation coefficients are varied among possible values to represent incertitude about inter-variable dependence is generally insufficient to reveal the true breadth of uncertainty.** If the dependence between variables can include subtle nonlinearities, then convolutions between these variables can yield tail risks that are more extreme at some percentile than the cases represented by perfect or opposite dependence. It is, however, possible to compute rigorous bounds on the convolution irrespective of the dependence between the inputs, and these bounds will also be best possible in many cases. Several convenient computational methods are currently available that can be used to calculate appropriate bounds on the distribution tails whenever there is partial or even complete ignorance about correlation and dependency among the variables. Dispersive Monte Carlo simulation is based on variance maximization/minimization. It is compatible with existing Monte Carlo methods, but it accounts only for linear dependencies. Dependency bounds analysis, based on Fréchet bounds, guarantees conservative estimates of tail probabilities no matter what dependence structure exists among the variables in the analysis.

iii) **Feasibility checks can be helpful.** When correlations are simulated, an analyst should ensure that the matrix of planned correlation coefficients is feasible by checking that it is positive semi-definite (Section 3.6.1). If the matrix is positive semi-definite, then it is a possible correlation matrix. If it is not positive semi-definite, then it cannot be a correlation matrix in the first place. It would be a mistake akin to specifying a negative variance and certainly should not be used in modeling dependencies in a risk analysis. It may even be possible and useful to employ this constraint on correlation matrices to tighten some interval

estimate of correlation. For instance, knowing the correlations between X and Y and between *X* and *Z* may constrain the correlation between *Y* and *Z* to an interval smaller than [−1, +1].

iv) **Using unjustified or false independence assumptions can be dangerous.** Extreme-event probabilities (i.e., the tails) of a statistical distribution resulting from probabilistic risk analysis can depend strongly on dependencies among the variables involved in the calculation. Ignoring dependencies that are present can result in substantial underestimate of tail risks (Section 3.1.2). Although well known techniques exist for incorporating correlation into analyses, unfortunately in practice they are often neglected on account of a paucity of information about joint distributions. Furthermore, certain forms of dependency that are not adequately measured by simple correlation must by necessity be omitted from such assessments.

v) **Knowing only that some correlation is zero is hardly better than knowing nothing at all about the dependence.** The bounds on convolutions between distributions, p-boxes or Dempster-Shafer structures that enclose all dependencies that create a correlation of a particular magnitude are surprisingly wide and, in the tails, essentially indistinguishable from the Fréchet bounds (Section 4.3.3).

vi) **Dependence is considerably more complicated in the context of imprecise probabilities than it is in classical probability theory.** For instance, the unique notion of independence in probability degenerates into many distinct notions when Dempster-Shafer structures, p-boxes or comparable structures are employed in an analysis (Section 3.1.3). Nevertheless, using random-set independence as implemented in the calculations with the Cartesian product will provide a conservative approach that never underestimates uncertainty about the result. If it is important to account for a more delicate notion of independence to reduce the uncertainty in the results, special ad hoc analysis may be required.

7.1 Future research

Once an analyst switches from a precise model of independence (in which the dependence function is completely specified) to an imprecise model of dependence (e.g., Fréchet or a sign dependence), there is often a marked increase in the overall uncertainty of the resulting convolution. As mentioned in Section 4.3, research is needed to develop intuition about when and under what circumstances partial information about dependencies between variables can be helpful in significantly reducing the uncertainty about the calculations in an analysis. Research is also needed on other ways to decrease the uncertainty about convolutions based on further empirical information or reasonable assumptions about dependence. Having a variety of approaches and arguments for reducing uncertainty would make the technology more flexible and appealing to analysts and modelers.

Another avenue of potentially useful research would be the development of methods to constrain copulas to reflect features of empirical joint distribution functions.

Even if they are not entirely reliable as descriptions of dependence because they are based on sparse information and data that were measured with error, such structures can suggest models of dependence for analysis. For instance, if the patterns in a scattergram of a joint distribution tend to suggest that a certain region is impossible by an absence of data points there, it could be useful to be able to reflect the impossibility of such variable pairings in specifying the family of copulas and thus the resulting bounds on convolutions. In particular, it would be useful to be able to construct copulas that embody several disparate features that are observed separately.

8 Appendix: an exceptional case

This appendix describes a numerical example that is of interest because it

i) illustrates a practical difference between random-set independence and strong independence (Section 3.1.3),
ii) constitutes an exception to the typical situation in which the convolution of uncertain numbers A and B is not affected by dependence assumptions if at least one of them is an interval (Section 4.2.3), and
iii) tempers the claim about the best-possible nature of the simple convolutions with probability boxes and Dempster-Shafer structures (see Ferson et al. 2003).

Suppose X is the interval $[-1, +1]$ and Y is the Dempster-Shafer structure $\{([-1, 0], 0.5), ([0, 1], 0.5)\}$. The cumulative plausibility and belief functions for X are shown on the top, left graph in Figure 43. The same functions for Y are shown on the top, middle graph. What can be said about the product $Z = XY$ assuming X and Y are independent? What can be said without any assumption about their dependence?

Let us address the second question first. We cannot use the methods of Frank et al. (1987) or Williamson and Downs (1990) to obtain the bounds on $Z = XY$ for the Fréchet case in which we make no assumption about dependence because the problem involves multiplication between distributions that have both positive and negative values. The method doesn't work for this case. But we can obtain the best possible bounds for the Fréchet case by ad hoc analysis. First, interval analysis tells us that the product is $Z = XY$ is totally constrained to the interval $[-1, +1]$. So we know that, whatever else is true, this interval forms a probability box about the product. Denote by B the discrete distribution with half its mass at -1 and half at $+1$. The Dempster-Shafer structure for B is $\{(-1, 0.5), (+1, 0.5)\}$. B is also depicted in the top, right graph of Figure 43. B is clearly consistent with X because it lies entirely within the range $[-1, +1]$. It is also consistent with Y because the locations for the point masses of B are within the two focal elements of Y and consistent with its partition of mass. Now suppose that the distributions for X and Y are both identical to B, and that they are perfectly dependent so that when X is -1, Y is too, and when X is $+1$, Y is too. In such a case, all the mass of Z will concentrate at the value $+1$. In that case, the cumulative belief and plausibility functions for Z are coincident spikes at $+1$. If, on the other hand, X and Y are oppositely dependent, then X and Y always have opposite signs and all the mass of Z concentrates at -1, and there is a spike at -1. Note that these two spikes are at the edges of the permissible range for Z as obtained by interval analysis. Therefore, the best possible p-box for the Fréchet case of the product $Z = XY$ is identical to the interval $[-1, +1]$ and the corresponding Dempster-Shafer structure is $\{[-1, +1], 1\}$.

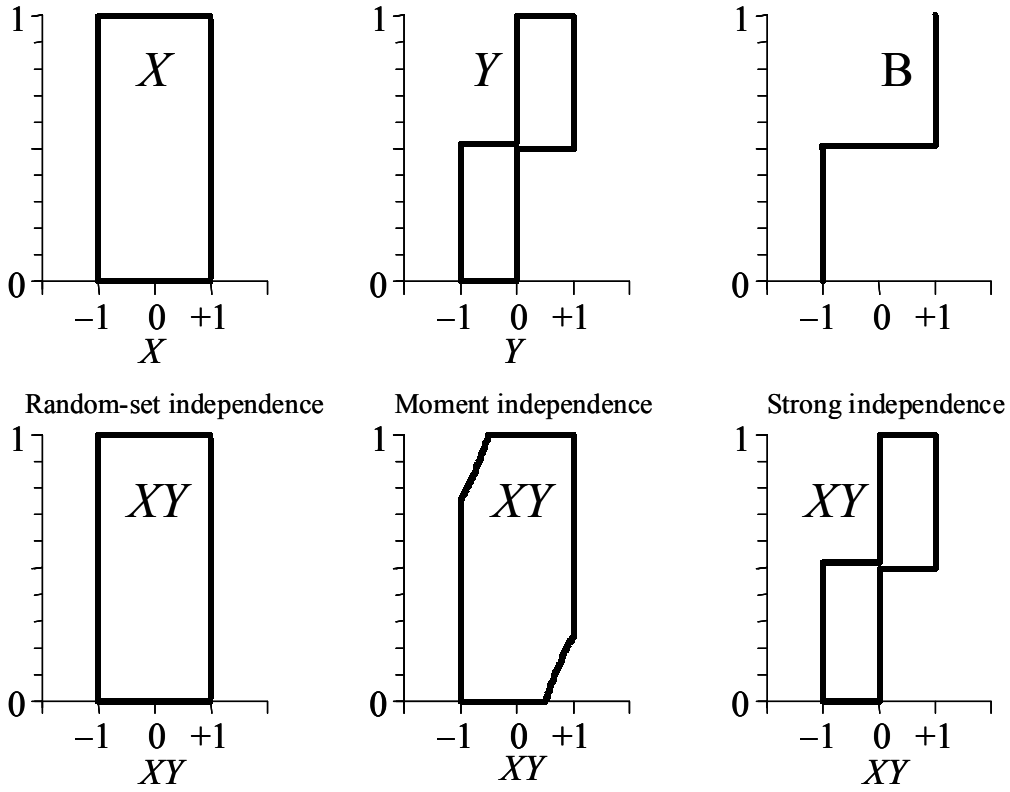

Figure 43: Inputs and outputs for the counterexample.

Now let us consider the problem of characterizing the product XY assuming that X and Y are independent. If we interpret the problem to be saying that X and Y have random-set independence, then we can compute the result with a convolution via Yager's (1986) Cartesian product with two rows and one column

XY *X* / *Y*	[−1, +1] 1
[−1, 0] 0.5	[−1, +1] 0.5
[0, +1] 0.5	[−1, +1] 0.5

where X is given in the right cell of the top row, and Y is given in the lower two cells of the first column. The variable X has only a single focal element; the variable Y has two. Each cell of the Cartesian product consists of an interval focal element on the first line and a probability mass on the second line. The interior of the Cartesian product has just two focal elements. They are the lower two cells in the right column. Because the

two intervals produced in the Cartesian product are the same, they are combined to produce a Dempster-Shafer structure with one focal element, {([−1, +1], 1)}, which is equivalent to the interval [−1, +1]. The algorithms given by Williamson and Downs (1990) and Berleant (1993) are equivalent to that of Yager for this problem and give the same answer.

So far so good, because the Fréchet case and the random-set-independent case have both given the answer as the interval [−1, +1]. But is this answer best possible if X and Y are independent? An easy argument shows that it is not. The mean of a product of independent random variables is the product of their means (Mood et al. 1974, page 180). The bounds on the mean of X are itself, that is, the interval [−1, +1]. The mean of Y is somewhere in the interval [−0.5, +0.5]. Interval analysis tells us that the product of these is just [−0.5, +0.5]. Therefore, $Z = XY$ where X and Y are independent has a range of [−1, +1] and a mean of [−0.5, +0.5]. But probability bounds analysis tell us that the distribution functions for any random variable with such a range and mean is constrained by the probability box shown in the bottom, middle graph of Figure 43 (Ferson et al. 2003). Clearly the interval shown on the bottom, left graph cannot be best possible if all distributions must lie inside the bounds in the bottom, middle graph. The discrepancy arises from the fact that we have used random-set independence when computing the earlier result but were interpreting the assertion that X and Y were independent to mean that they were strongly independent (Section 3.1.3).

So what are the best possible bounds on $Z = XY$ in case X and Y are strongly independent? The distribution for Z is a mixture of four cases: (*i*) when X and Y are both positive, (*ii*) when X is positive and Y is negative, (*iii*) when X is negative and Y is positive, and (*iv*) when both X and Y are negative. The probability that Z is positive comes only from the cases (*i*) and (*iv*). If we let $P_X = \mathrm{P}(0 \le X)$ and $P_Y = \mathrm{P}(0 \le Y)$, then $\mathrm{P}(0 \le Z) = P_X P_Y + (1 - P_X)(1 - P_Y)$. Because $P_Y = 1/2$ by construction, this simplifies to $\mathrm{P}(0 \le Z) = P_X/2 + (1 - P_X)/2 = 1/2$. This proves, then, that the median for Z is a pinch point at zero. Because the scalar zero is a possible value for X, the bounds on the product $Z = XY$ must include the scalar zero too. Because the scalar value one is a possible realization for X, and B is a possible realization for Y, the bounds on the product $Z = XY$ must also include the distribution B as well. Since both the scalar value zero and B are solutions under strong independence, their envelope (which is the largest p-box on the permissible interval with a medial pinch point at zero) is the best possible p-box on $Z=XY$ under strong independence between X and Y. This result is also depicted as the bottom, right graph in Figure 43.

This example shows that the convolution via Yager's (1986) Cartesian product, which assumes random-set independence, produces the same answer as the Fréchet case, but it is not best possible if X and Y are strongly independent. Further research is needed to fully understand the conditions under which the straightforward calculation involving the Cartesian product will be best possible and when it will at least be conservative in the sense of not underestimating uncertainty. Analysts must learn when it is reasonable to assume strong independence, and when it is not.

9 Glossary

2-increasing The property of a bivariate function Z such that $Z(a_2, b_2) - Z(a_1, b_2) - Z(a_2, b_1) + Z(a_1, b_1) \geq 0$ whenever $a_1 \leq a_2$ and $b_1 \leq b_2$. This property is distinct from and neither implies nor is implied by monotonicity. Copulas (q.v.) are 2-increasing functions.

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, or objective uncertainty.

almost surely A property holds almost surely if it holds always except possibly for a set of measure zero.

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

CDF Cumulative distribution function (see distribution function).

comprehensive A family of copulas is called comprehensive if it includes the special cases of perfect dependence, opposite dependence and independence.

comonotonicity Perfect dependence (q.v.) between random variables.

conjunction The truth-functional operation that yields true if all of its arguments are true and false otherwise.

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 A joint distribution function, each of the marginal distributions for which is uniform over the interval [0,1]. In the bivariate case, a copula is a function C:$[0,1]\times[0,1]\rightarrow[0,1]$ such that $C(0, a) = C(a, 0) = 0$ and $C(1, a) = C(a, 1) = a$ 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 $b_1 \leq b_2$. A copula is the function that expresses the dependence between variables and knits together their respective

marginal distribution functions to form their joint distribution function. For instance, the copula of continuous random variables $X \sim F(x)$ and $Y \sim G(y)$ is the joint distribution of $(F(X), G(Y))$. Every joint distribution can be decomposed into, and reconstructed from, its marginal distributions and a copula. Every possible dependence among random variables, even including functional relationships, is expressed by some copula. For any copula C, $\max(a+b-0) \leq C(a, b) \leq \min(a, b)$.

correlation The tendency of two paired variables to vary in the same direction. Correlation can be measured by several coefficients, conventionally constrained to the interval $[-1, +1]$, such as Pearson correlation, Spearman correlation, Kendall correlation, among others. If used without qualification, correlation usually refers to Pearson correlation.

countermonotonicity Opposite dependence (q.v.) between random variables.

covariance The first product moment of two variables about their means. For random variable X and Y, their covariance is $\mathrm{cov}(X, Y) = \mathrm{E}((X - \mathrm{E}(X))\,(Y - \mathrm{E}(Y))) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)$.

cumulative distribution function A distribution function (q.v.).

Dempster-Shafer structure A set of focal elements (in this report, closed intervals of the real line), each of which is associated with a non-negative value m such that the sum of all the m's is one.

Dempster-Shafer theory A variant of probability theory in which the elements of the probability space to which nonzero mass is attributed, called focal elements, are not singletons but rather sets which represent the indistinguishability of alternatives within bodies of evidence. Dempster-Shafer theory is sometimes called evidence theory.

dependence The relationship between events or between random variables. If one event (random variable) is unrelated to another event (random number), they are said to be independent. Otherwise, they are said to be dependent. Hutchinson and Lai (1990; section 11.4) review sets of axioms for a measure of dependence. Interestingly, the traditional measure, Pearson correlation, does not satisfy many of them.

disjunction The truth-functional operation that yields true if any of its arguments is true and false if they are all false.

dispersive Monte Carlo simulation A Monte Carlo simulation in which unknown correlations are set to their most extreme plausible values in order to obtain results that conservatively estimate variances and tail probabilities.

distribution function The function F, associated with a random variable X, that describes the probability $F(x)$ that X will take on a value not greater than x, which is often denoted as $\mathrm{Prob}(X \leq x)$. 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 independence The property that an analyst's uncertainty about either of two outcomes of a random experiment does not change when some information

about the outcome of one of them becomes known. For random variables X and Y, X and Y are epistemically independent if the conditional probability of each given the other is equal to its unconditional probability,

$$P(X|Y) = P(X),$$
$$P(Y|X) = P(Y).$$

In the context of imprecise probabilities, epistemic independence is defined in terms of lower bounds on expectations such that $\underline{E}(f(X)|Y) = \underline{E}(f(X))$ and $\underline{E}(f(Y)|X) = \underline{E}(f(Y))$ for all functions f where $\underline{E}(Z)$ denotes the infimum of all expectations of Z over all possible probability distributions that could characterize Z. Epistemic independence does not imply strong independence.

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.

event A subset of the sample space, which is the set of all possible outcomes of a random experiment. If the outcome of the random experiment is a member of an event, then the event is said to have occurred. In probability theory, an event is a collection of outcomes for which a probability has been assigned.

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.

Fréchet case The strategy of making no assumption about dependence.

Fréchet bounds Bounds on a joint distribution $H(x,y)$, specified by having marginal distributions $F(x)$ and $G(y)$, given by

$$\max\big(F(x) + G(y) - 1,\ 0\big) \le H(x, y) \le \min\big(F(x),\ G(y)\big).$$

These bounds are also known as the Fréchet-Hoeffding limits (Fréchet 1951; Hoeffding 1940). They are the distributional analogs of the bounds in the Fréchet inequalities.

Fréchet inequalities Inequalities due to Fréchet (1935) on the probabilities of conjunctions and disjunctions of events A_i given by

$$\max(0, a_1 + a_2 + \ldots + a_n - (n-1)) \le P(A_1 \,\&\, A_2 \,\&\, \ldots \,\&\, A_n) \le \min(a_1, a_2, \ldots, a_n),$$
$$\max(a_1, a_2, \ldots, a_n) \le P(A_1 \vee A_2 \vee \ldots \vee A_n) \le \min(1, a_1 + a_2 + \ldots + a_n),$$

where $a_i = P(A_i)$.

imprecise probabilities The subject of 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). The lower probability $\underline{P}(A)$ for event A is the maximum rate one would be willing to pay for a gamble that pays 1 unit of utility if A occurs and nothing otherwise. The upper probability $\overline{P}(A)$ for event A is $1 - \underline{P}(\text{not } A)$, i.e., one minus the lower probability of A not occurring. 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. Interval probabilities, Dempster-Shafer structures and probability boxes can be regarded as special-cases of imprecise probabilities.

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.

independence The unrelatedness between events or between random variables. In the context of probability theory, the concept of independence is unique for events and unique for random variables. Events are said to be independent (in the probabilistic sense) if the probability that both occur is the product of the probability of either occurring. Random variables X and Y are said to be independent if their joint distribution function H is equal to the product of their respective marginal distributions, in the bivariate case, $H(x, y) = F(x)\ G(y)$. In the context of imprecise probabilities, however, there are several concepts that could be called independence. See epistemic independence, random-set independence, repetition independence and strong independence.

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 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 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.

joint distribution A distribution function in two (or more) variables. In the bivariate case, a joint distribution $H(x, y)$ gives the probability that $X \leq x$ and, jointly, $Y \leq y$. A joint distribution can be decomposed into, and reconstructed from its marginal distributions and a copula that characterizes the dependence between the variables.

Kendall correlation The index named for M.G. Kendall (Huchinson and Lai 1990; Nelsen 1999) that measures the strength of the association between two variables X and Y. It is defined by

$$\tau = \mathrm{P}[0 < (X_1 - X_2)(Y_1 - Y_2)] - \mathrm{P}[(X_1 - X_2)(Y_1 - Y_2) < 0]$$

where (X_1, Y_1) and (X_2, Y_2) are independent realizations from a joint distribution. It can also be expressed as $\tau = \mathrm{cov}(\mathrm{sgn}(X_1 - X_2), \mathrm{sgn}(Y_1 - Y_2))$. The Kendall correlation coefficient measures monotonicity of the relationship between X and Y by considering the preponderance of pairs of bivariate data points (X_i, Y_i) and

(X_j, Y_j) that are concordant in the sense that Y_j is larger than Y_i is if X_j is larger than X_i. Unlike the Pearson correlation, any value in [−1, +1] is a possible Kendall correlation between any pair of marginal distributions. Independence between X and Y produces a value of zero for τ, although τ=0 does not conversely imply independence. Perfect correlation leads to a value of +1; opposite correlation leads to a value of −1.

Lucas correlation A measure of association between two events given by the Pearson correlation between the indicator functions of the events. See Section 2.2.

marginal distribution (or margin) Given two random variables X and Y and their (bivariate) joint cumulative distribution $H(x,y)$, $F(x,\infty)$, which is the limit of $F(x,y)$ as y approaches infinity, is marginal distribution for X, and $F(\infty, y)$ is the marginal distribution for Y. The marginal distribution gives the unconditional probability for one of the variables, so $F(x_i) = \text{Prob}(X_i \leq x_i)$, irrespective of the other variable and ignoring any information about it. Marginal distributions may also be defined in terms of more than one of the random variables. There are in general many joint distributions having specified marginal distributions.

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

negation The truth-functional operation that yields true if its argument is false and false if its argument is true. Extended to probabilities of events, the probability of a negation of an event is one minus the probability of the event.

negative quadrant dependence A pattern of dependence between random variables X and Y such that $P(X \leq x, Y \leq y) \leq P(X \leq x)\, P(Y \leq y)$. In this case, X and Y have non-positive (Spearman and Kendall and Pearson) correlations. The copula associated with this pattern of dependence is everywhere smaller than the copula associated with independence. Variables having negative quadrant dependence are said to be negatively quadrant dependent or NQD.

NQD Negatively quadrant dependent.

opposite dependence A pattern of dependence between events A and B such that $P(A \,\&\, B) = \max(0, P(A) + P(B) - 1)$, or between random variables X and Y such that $P(X \leq x, Y \leq y) = \max(0, P(X \leq x) + P(Y \leq y) - 1)$. In the case of random variables, opposite dependence is also called countermonotonicity and the related variables said to be countermonotone. In this case, X and Y have Spearman and Kendall correlations equal to −1 and the smallest Pearson correlation they could possibly have given their marginal distributions (Whitt 1976). If X and Y are oppositely dependent, then X is almost surely a non-increasing function of Y, and vice versa, and the graph of the support of the joint distribution function is non-increasing in the plane (Nelsen 1999, page 27).

p-box A probability box (q.v.).

Pearson correlation The statistic due to Karl Pearson that measures the strength of the association between two variables X and Y. It is defined by

$$r = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{V}(X)\,\mathrm{V}(Y)}} = \frac{\mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)}{\sqrt{\mathrm{V}(X)\,\mathrm{V}(Y)}}$$

where cov denotes the covariance of random variables, and V and E denote the variance and expectation (mean) of a random variable, respectively. The Pearson correlation measures the similarity of the association between X and Y with a straight line. It is only one of several measures of correlation that have been proposed. Independence between X and Y produces a value of zero for r. Perfect correlation leads to the largest possible value for r given the marginal distributions, although this value may be less than +1. Likewise, opposite correlation leads to the smallest possible value of r for the marginal distributions, but this may sometimes be greater than −1. Pearson correlation is also called product-moment correlation, and sometimes simply linear correlation.

perfect dependence A pattern of dependence between events A and B such that $\mathrm{P}(A \,\&\, B) = \min(\mathrm{P}(A), \mathrm{P}(B))$, or between random variables X and Y such that $\mathrm{P}(X \le x,\ Y \le y) = \min(\mathrm{P}(X \le x), \mathrm{P}(Y \le y))$. In the case of random variables, perfect dependence is also called comonotonicity and the related variables said to be comonotone or comonotonous. In this case, X and Y have Spearman and Kendall correlations equal to +1 and the largest correlation they could possibly have given their marginal distributions (Whitt 1976). If X and Y are perfectly dependent, then X is almost surely a non-decreasing function of Y, and vice versa, and the graph of the support of the joint distribution function is non-decreasing in the plane (Nelsen 1999, page 27).

positive quadrant dependence A pattern of dependence between random variables X and Y such that $\mathrm{P}(X \le x)\ \mathrm{P}(Y \le y) \le \mathrm{P}(X \le x,\ Y \le y)$. In this case, X and Y have non-negative (Spearman and Kendall and Pearson) correlations. The copula associated with this pattern of dependence is everywhere larger than the copula associated with independence. Variables having positive quadrant dependence are said to be positively quadrant dependent or PQD.

positive semi-definiteness Property of a matrix A by which A is symmetric and $0 \le x^{\mathrm{T}}Ax$ for all x, and by which all the principal minors of A are non-negative, and by which there exists a matrix C of rank r such that $A = C^{\mathrm{T}}C$.

PQD Positively quadrant dependent.

probability bounds analysis An analysis or calculation involving interval probabilities or probability boxes.

probability box A class of distribution functions $F(x)$ specified by a bounding pair of distribution functions $\underline{F}(x)$ and $\overline{F}(x)$ such that $\underline{F}(x) \le F(x) \le \overline{F}(x)$ for all x values.

quadrant dependence Dependence that is either positive quadrant dependence or negative quadrant dependence.

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-set independence The dependence between Dempster-Shafer structures X and Y, which have mass functions m_X and m_Y respectively, such that the Dempster-Shafer structure for the joint distribution has mass function $m(A_1 \times A_2) = m_X(A_1)\, m_Y(A_2)$ when A_1 is a focal element of X and A_2 is a focal element of Y, with $m(A)=0$ for all subsets not of the form $A = A_1 \times A_2$. This is the weakest definition of independence in the context of imprecise probabilities (cf. Couso et al. 2000). It therefore leads to the broadest uncertainty in results compared to other definitions of independence such as strong independence or epistemic independence.

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.

rank correlation Any measure of correlation based on the (within-variable) ranks of random variables rather than their absolute magnitudes. An unqualified reference to rank correlation usually refers to Spearman's rank correlation.

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.

repetition independence Independence between random variables that are identically distributed, although their distribution may be imprecisely known. Repetition independence is the analog in the context of imprecise probabilities of the constraint in probability theory that variables are independent and identically distributed (iid). Repetition independence implies a class of joint distribution functions that is smaller than from assuming strong independence, but repetition independence does not *imply* strong independence because the marginal distributions, whatever they are, must be identical, which precludes all other combinations of possible marginal distributions. This kind of independence corresponds to the smallest set of joint distribution functions of all the definitions of independence identified by Couso et al. (2000).

rigorous Exact or sure, as opposed to merely approximate. Usually said of bounds which can be rigorous without being best possible.

Sklar's theorem A result in the study of probabilistic metric spaces due to Sklar (1959) that states, if $H(x,y)$ is a joint distribution function with marginal distribution functions F and G, then there exists a copula C such that

$$H(x, y) = C(F(x), G(y)).$$

Conversely, for any univariate distribution functions F and G and any copula C, the function H is a two-dimensional distribution function having marginals F and G. If F and G are continuous, then C is unique; otherwise it can be uniquely

determined on the product of the ranges of F and G. The theorem generalizes to dimensions higher than two.

Spearman correlation The nonparametric index due to Spearman (1904; Hutchinson and Lai 1990; Nelsen 1999) that measures the strength of the association between two variables X and Y. It is defined by

$$\rho = P(X_1 < X_2,\ Y_1 < Y_3) + P(X_2 < X_1,\ Y_3 < Y_1)$$

which can also be expressed as $\rho = \text{cov}(\text{sgn}(X_2 - X_1), \text{sgn}(Y_3 - Y_1))$. Note that this formulation compares a vector (X_1, Y_1) with another vector (X_2, Y_3) with the same margins, but whose elements are independent. The Spearman correlation is identical to the Pearson correlation computed between the ranks of X and Y, or between their grades $F(X)$ and $G(Y)$, where $X \sim F$ and $Y \sim G$. The Spearman correlation measures the monotone association between the variables and is often considered a more appropriate measure of correlation for nonlinear relationships or non-normal variables than the traditional Pearson correlation. Unlike the Pearson correlation, any value in $[-1, +1]$ is a possible Spearman correlation between any pair of marginal distributions. Independence between X and Y produces a value of zero for ρ, although the fact that $\rho = 0$ does not imply independence. Perfect correlation leads to a value of $+1$; opposite correlation leads to a value of -1. The Spearman correlation is sometimes known as the grade correlation.

strong independence The complete absence of any relationship between variables. Variables X and Y are strongly independent if (*i*) X and Y result from random experiments, each governed a unique but possibly unknown probability distribution, (*ii*) the random experiments are stochastically independent (in the traditional sense), and (*iii*) there is no known relationship between the variables that would preclude some possible combinations of the possible marginal distributions. Variables X and Y are strongly independent if the set of possible joint distributions is the largest set such that each joint distribution $H(x, y) = F(x)\ G(y)$, where F is one of the possible distribution functions characterizing X and G is one of the possible distribution functions characterizing Y. Strong independence implies epistemic independence of the marginal experiments.

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.

total probability The probability of a single event.

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. This approach is also called second-order Monte Carlo simulation.

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).

uncorrelated Having a (Pearson) correlation of zero magnitude. Uncorrelatedness does not imply independence.

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.

10 References

Abramowitz, M. and I.A. Stegun. (1964). *Handbook of Mahematical Functions with Formulas, Graphs and Mathematical Tables*. National Bureau of Standards, Washington, DC.

Alefeld, G. and J. Herzberger. (1983). *Introduction to Interval Computations*. Academic Press, New York.

Barton W.T. (1989). Response from Palisade Corporation. *Risk Analysis* 9: 259-260.

Bedford, T.J., and R.M. Cooke. (2002). Vines —a new graphical model for dependent random variables. *Annals of Statistics* 30(4): 1031-1068. Available on-line at http://ssor.twi.tudelft.nl/~risk/files/mv3.pdf.

Berleant, D. (1993). Automatically verified reasoning with both intervals and probability density functions. *Interval Computations* 1993 (2): 48-70.

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

Berleant, D. and H. Cheng. (1998). A software tool for automatically verified operations on intervals and probability distributions. *Reliable Computing* 4: 71-82.

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 J. Zhang. (2004a). Representation and problem solving with Distribution Envelope Determination (DEnv). *Reliability Engineering and System Safety* 153-168. Available on-line at http://class.ee.iastate.edu/berleant/home/me/cv/papers/challengeProblems.pdf.

Berleant, D. and J. Zhang. (2004b). Using Pearson correlation to improve envelopes around the distributions of functions. *Reliable Computing* 10: 139-161. Available on-line at http://class.ee.iastate.edu/berleant/home/me/cv/papers/correlationpaper.pdf.

Bernat, G., A. Burns and M. Newby. (2004). Probabilistic timing analysis: an approach using copulas. *Journal of Embedded Computing* [to appear].

Blomqvist, N. (1950). On a measure of dependence between two random variables. *Annals of Mathematical Statistics* 21: 593-600.

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. Available in a 1951 reprinting from Dover Publications, http://store.doverpublications.com/.

Bratley, P., B.L. Fox and L.E. Schrage (1983). *Guide to Simulations*. Springer-Verlag, New York.

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

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

Burmaster, D.E. and P.D. Anderson. (1994). Principles of good practice for the use of Monte Carlo techniques in human health and ecological risk assessments. *Risk Analysis* 14: 477-481.

Burmaster D.E. and E.C. Udell. (1990). A review of Crystal Ball. *Risk Analysis* 10: 343-345.

Calabrese, E.J. and C.E. Gilbert. (1993). Lack of total independence of uncertainty factors (UFs): implications for the size of the total uncertainty factor. *Regulatory Toxicology and Pharmacology* 17: 44-51.

Cario, M.C. and B. L. Nelson. (1997). Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. http://www.iems.nwu.edu/%7Enelsonb/norta4.ps

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

Cheng, W.-Y. (2003). Default correlations (a Microsoft Power Point presentation file). http://personal.cityu.edu.hk/~efcheng/teaching/ef5154/defaultcorrelation.ppt.

Clemen, R. and T. Reilly (1999). Correlations and copulas for decision and risk analysis. *Management Science* 45: 208-224.

Cooke, R.M. (1991). *Experts in Uncertainty*. Oxford University Press, Cambridge.

Cooke, R.M. (1997). Markov and entropy properties of tree and vine-dependent variables. {variant title: Markov trees and tree dependent random variables] *Proceedings of the ASA Section on Bayesian Statistical Science*. Available on-line at http://ssor.twi.tudelft.nl/~risk/files/marktre.pdf.

Cooke, R. (2002). *Unicorn - Uncertainty Analysis with Correlations.* Available from http://ssor.twi.tudelft.nl/~risk/software/unicorn.html.

Cossette, H., M. Denuit and É. Marceau. (2001). Distributional bounds for functions of dependent risks. Discussion paper 0128, Institut de Statistique, Université Catholique de Louvain, http://www.stat.ucl.ac.be/ISpub/dp/2001/dp0128.ps.

Couso, I., S. Moral and P. Walley. (1999). Examples of independence for imprecise probabilities. Pages 121-130 in *Proceedings of the First International Symposium on Imprecise Probability and Their Applications*. G. de Cooman, F.F. Cozman, S. Moral and P. Walley (eds.), Imprecise Probabilities Project, Universiteit Gent.

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

Cozman, F.G. and P. Walley. (2001). Graphoid properties of epistemic irrelevance and independence. Pages 112-121 in *Proceedings of the Second International Symposium on Imprecise Probability and Their Applications*. G. de Cooman, T.L. Fine and T. Seidenfeld (eds.), Shaker Publishing, Maastricht.

Cui, W.C. and D.I. Blockley. (1990). Interval probability theory for evidential support. *Inernational Journal of Intelligent Systems* 5 (part2):183-192.

Cullen, A.C., and H.C. Frey. (1999). *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*. Plenum Press: New York.

Dall'Aglio, G., Kotz, S. and G. Salinetti (eds.) (1991). *Advances in Probability Distributions with Given Marginals: Beyond the Copulas*, Kluwer Academic Publishers, London.

Davis, J.P. and J.W. Hall (2003). A software-supported process for assembling evidence and handling uncertainty in decision-making. *Decision Support Systems* 35: 415-433.

Deheuvels, P. (1979). La fonction de dépendence empirique et ses propriétés: un test non paramétrique d'indépendence. *Bulletin de la Classe des Sciences de l'Academie Royale de Belgique* (ser. 5) 65: 274-292.

Decisioneering, Inc. (1996). *Crystal Ball: Forecasting and Risk Analysis for Spreadsheet Users*, Aurora, Colorado.

Dwyer, P. (1951). *Linear Computations*. John Wiley, New York.

Evan, M., N. Hastings and B. Peacock. (1993). *Statistical Distributions*. John Wiley & Sons, New York.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications. Volume 1.* John Wiley & Sons, New York.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications. Volume 2.* John Wiley & Sons, New York.

Ferson, S. (1994). Naive Monte Carlo methods yield dangerous underestimates of tail probabilities. *Proceedings of the High Consequence Operations Safety Symposium*, Sandia National Laboratories, SAND94–2364, J.A. Cooper (ed.), pp. 507–514.

Ferson, S. (1995). Quality assurance for Monte Carlo risk assessments. *Proceedings of the 1995 Joint ISUMA/NAFIPS Symposium on Uncertainty Modeling and Analysis*, IEEE Computer Society Press, Los Alamitos, California, pp. 14–19.

Ferson, S. (1996a). Automated quality assurance checks on model structure in ecological risk assessments. *Human and Environmental Risk Assessment* 2:558–569.

Ferson, S. (1996b). 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 M. Burgman. (1995). Correlations, dependency bounds and extinction risks. *Biological Conservation* 73:101-105.

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. 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.

Ferson, S., V. Kreinovich, L. Ginzburg, K. Sentz and D.S. Myers. (2003). Constructing probability boxes and Dempster-Shafer structures. Sandia National Laboratories, Technical Report SAND2002-4015, Albuquerque, New Mexico. Available at

http://www.sandia.gov/epistemic/Reports/SAND2002-4015.pdf. An extended version is also available at http://www.ramas.com/unabridged.zip.
Fetz, T. (2001). Sets of joint probability measures generated by weighted marginal focal sets. Pages 171-178 in *Proceedings of the Second International Symposium on Imprecise Probability and Their Applications*. G. de Cooman, T.L. Fine and T. Seidenfeld (eds.), Shaker Publishing, Maastricht.
Fetz, T. and M. Oberguggenberger. (2004). Propagation of uncertainty through multivariate functions in the framework of sets of probability measures. *Reliability Engineering and System Safety* 85: 73-88.
Finkel, A. (1990). *Confronting Uncertainty in Risk Management*. Center for Risk Management, Resources for the Future, Washington, D.C.
Finkel, A. M. (1994). Stepping out of your own shadow: A didactic example of how facing uncertainty can improve decision-making. *Risk Analysis* 14: 751-761.
Finkel, A.M. (1995). A second opinion on an environmental misdiagnosis: the risky prescription of *Breaking the Vicious Circle*. *New York University Environmental Law Journal* 3: 295-381.
Finley, B., D. Proctor, P. Scott, N. Harrington, D. Pasutenbach and P. Price. (1994). Recommended distributions for exposure factors frequently used in health risk assessment. *Risk Analysis* 14:533-553.
Fisher, R.A. (1957). The underworld of probability. *Sankhyā* 18 201-210.
Flury, B.K. (1986). On sums of random variables and independence. *The American Statistician* 40: 214-215.
Frank, M. (1979). On the simultaneous associativity of and $F(x, y)$and x+y–$F(x, y)$. *Aequationes Mathematicae* 19: 194-226.
Frank, M.J., Nelsen, R.B., and Schweizer, B. (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.
Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon, Sect. A* 9: 53-77.
Frey, H.C. (1992). Quantitative analysis of uncertainty and variability in environmental policy making. Environmental Science and Engineering Fellows Program, American Association for the Advancement of Science, Washington, DC. Available on-line at http://www4.ncsu.edu/~frey/frey_92.pdf.
Genest, C. and J. MacKay (1986). The joy of copulas: bivariate distributions with uniform marginals. *The American Statistician* 40: 280-283.
Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika* 74: 549-555.
Gentle, J.E. (1998). *Numerical Linear Algebra for Applications in Statistics*. Springer-Verlag, Berlin.
Good, I.J. (1965). *The Estimation of Probabilities*. MIT Press, Cambridge, Massachusetts.

Goovaerts, M.J., J. Dhaene and A. De Schepper. (2000). Stochastic upper bounds for present value functions. *Journal of Risk and Insurance* 67: 1-14.

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

Haas, C.N. (1999). On modelling correlated random variables in risk assessment. *Risk Analysis* 19: 1205-1214.

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

Haimes, Y.Y., T. Barry and J.H. Lambert. (1994). When and how can you specify a probability distribution when you don't know much? *Risk Analysis* 14:661-706.

Hansen, F. (1997). *Policy for Use of Probabilistic Analysis in Risk Assessment at the U.S. Environmental Protection Agency.* EPA/ORD/NCEA May 15:1-4. Available on-line at http://www.epa.gov/osp/spc/probpol.htm.

Hart, A., S. Ferson and J. Shaw. 2004 [tentative]. Problem formulation for probabilistic ecological risk assessments. *Proceedings of the SETAC Pellston Workshop on the Application of Uncertainty Analysis to Ecological Risks of Pesticides*. SETAC Press, Pensacola, Florida.

Hayes, B. (1993). The wheel of fortune. *American Scientist* 81:114-118.

Helton, J.C. (1993). Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering and System Safety* 42: 327-367.

Helton, J.C. (1994). Treatment of uncertainty in performance assessments for complex systems. *Risk Analysis* 14: 483-511.

Helton, J.C. (1997). Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. *Journal of Statistical Computing and Simulation* 57: 3-76.

Helton, J.C. and R.J. Breeding. (1993). Calculation of reactor accident safety goals. *Reliability Engineering and System Safety* 39: 129-158.

Helton, J.C. and F.J. Davis. (2000). Sampling-based methods for uncertainty and sensitivity analysis. SAND99-2240, Sandia National Laboratories, Albuquerque, New Mexico.

Helton, J.C. and F.J. Davis. (2002). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. SAND2001-0417, Sandia National Laboratories, Albuquerque, New Mexico.

Helton, J.C. and F.J. Davis. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering and System Safety* 81: 23-69.

Helton, J.C., J.D. Johnson and W.L. Oberkampf. (2004). An exploration of alternative approaches to the representation of uncertainty in model predictions. *Reliability Engineering and System Safety* 85: 39-72.

Henderson, S.G., B.A. Chiera and R.M. Cooke. (2000). Generating "dependent" quasi-random numbers. Pages 527-536 in *Proceedings of the 2000 Winter Simulation Conference*. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick (eds.), IEEE. Available on-line at http://ssor.twi.tudelft.nl/~risk/files/wsc2000hendersons-1i.pdf.

Hickman, J.W., et al. (1983). 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.

Hoeffding, W. (1940). Masstabinvariante Korrelationstheorie. *Schriften des Matematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin* 5 (Heft 3): 179-233 [translated in Hoeffding, W. (1940). Scale-invariant corelation theory. *Collected works of Wassily Hoeffding*, N.I. Fisher and P.K. Sen (eds.), Springer-Verlag, New York].

Hoffman, F. O., Hammonds, J.S. (1994). Propagation of uncertainty in risk assessments: The need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. *Risk Analysis* 14(5):707-712.

Hutchinson, T.P. and C.D. Lai. (1990). *Continuous Bivariate Distributions, Emphasizing Applications.* Rumsby Scientific Publishing, Adelaide 5000, Australia.

Iman, R.L. and W.J. Conover. (1980). Small sample sensitivity analysis techniques for computer models, with an application to risk assessment. *Communications in Statistics A9*: 1749-1842.

Iman, R.L. and W.J. Conover. (1982). A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics B11*: 311-334.

Iman, R.L. and J.M. Davenport (1980). Rank correlation plots for use with correlated input variables in simulation studies. SAND80-1903, Sandia National Laboratories, Albuquerque.

Iman, R.L. and J.M. Davenport (1982a). Rank correlation plots for use with correlated input variables. *Communications in Statistics (Simulation and Computation)* 11: 335-360.

Iman, R.L. and J.M. Davenport (1982b). An interactive algorithm to produce a positive-definite correlation matrix from an approximate correlation matrix (with a program users' guide). SAND81-1376, Sandia National Laboratories, Albuquerque.

Iman, R.L. and J.C. Helton (1985). A comparison of uncertainty and sensitivity analysis techniques for computer models. NUREG/CR-3904. National Technical Information Service, Springfield, Virginia.

Iman, R.L and M.J. Shortencarier. (1984). *A Fortran 77 Program and User's Guide for the Generation of Latin Hypercube and Random Samples for Use with Computer Models*. Technical Report NUREG/CR-3624, SAND83-2365, Sandia National Laboratories, Albuquerque, New Mexico.

Iman, R.L., J.C. Helton, J.E. Campbell. (1981a). An approach to sensitivity analysis of computer models, Part 1. Introduction, input variable selection and preliminary variable assessment. *Journal of Quality Technology* 13:174-183.

Iman, R.L., J.C. Helton, J.E. Campbell. (1981b). An approach to sensitivity analysis of computer models, Part 2. Ranking of input variables, response surface validation, distribution effect and technique synopsis. *Journal of Quality Technology* 13:232-240.

Jaynes, E.T. (1957). Information theory and statistical mechanics. *Physical Review* 106:620-630.
Johnson, N.L. S., Kotz, and N. Balakrishnan. (1994). *Continuous Univariate Distributions*. (Vols. 1 and 2). John Wiley & Sons, New York.
Joslyn, C. and S. Ferson. (2004). Approximate representations of random intervals for hybrid uncertainty quantification in engineering modeling. *Proceedings of the Fourth International Conference on Sensitivity Analysis of Model Output.* Los Alamos National Laboratory, Los Alamos, NM.
Kaas, R., J. Dhaene and M.J. Goovaerts. (2000). Upper and lower bounds for sums of random variables. *Insurance: Mathematics and Economics* 27: 151-168.
Kaplan, S. (1981). On the method of discrete probability distributions—application to seismic risk assessment. *Risk Analysis* 1:189-198.
Kaufmann, A. and M.M. Gupta. (1985). *Introduction to fuzzy arithmetic: theory and applications*. Van Nostrand Reinhold, New York.
Klir, G.J. and B. Yuan. (1994). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice-Hall, Upper Saddle River, New Jersey.
Kowalski, C.J. (1973). Non-normal bivariate distributions with normal marginals. *The American Statistician* 27: 103-106.
Kurowicka, D. and Cooke, R.M. (2002). The vine copula method for representing high dimensional dependent distributions: application to continuous belief nets. Proceedings of the 2002 Winter Simulation Conference E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes (eds.), IEEE. Available on-line at http://ssor.twi.tudelft.nl/~risk/files/kurowickad-2i.zip.
Kurowicka, D., J. Misiewicz and R.M. Cooke (2001). Elliptical copulae. Pages 209-214 in *Monte Carlo Simulation*, Schueller and Spanos (eds), Balkema, Rotterdam. Available on-line at http://ssor.twi.tudelft.nl/~risk/files/ellcop.pdf.
Kyburg, H.E., Jr. (1998). "Interval Valued Probabilities," on-line publication available at http://ippserv.rug.ac.be/documentation/interval_prob/interval_prob.html, The Imprecise Probabilities Project, edited by G. de Cooman and P. Walley, an electronic collection at http://ippserv.rug.ac.be/home/ipp.html.
Lehmann, E.L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics* 37: 1137-1153.
Levine, R.D. and M. Tribus (1976). *The Maximum Entropy Formalism*. MIT Press, Cambridge.
Li, D. (2000). On default correlation: a copula function approach. *Journal of Fixed Income* 9(4): 43-54.
Loui, R.P. (1986). Interval based decisions for reasoning systems. *Uncertainty in Artificial Intelligence*. L.N. Kanal and J.F. Lemmer (eds.), Elsevier Science Publishers, Amsterdam.
Lucas, D.J. (1995). Default correlation and credit analysis. *Journal of Fixed Income* 4(4): 76-87.

Lurie, P.M., M.S. Goldberg. (1994). A method for simulating correlated random variables from partially specified distributions. Institute for Defense Analysis IDA paper P-2998, Alexandria, VA.

Madan, D.P. and J.C. Owings (1988). Decision theory with complex uncertainties. *Synthese* 75: 25-44.

Makarov, G.D. (1981). 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.

Markov [Markoff], A. (1886). Sur une question de maximum et de minimum proposée par M. Tchebycheff. *Acta Mathematica* 9: 57-70.

Matheron, G. (1975). *Random Sets and Integral Geometry*. Wiley, New York.

Melnick, E.L. and A. Tenenbein. (1982). Misspecification of the normal distribution. *The American Statistician* 36: 372-373.

Metzger, J. N., Fjeld, R. A., Hammonds, J. S., Hoffman, F. O. (1998). Evaluation of software for propagating uncertainty through risk assessment models. *Human and Ecological Risk Assessment* 4(2): 263-290.

Meyer, M.A. and J.M Booker (1991). *Eliciting and Analyzing Expert Judgment: A Practical Guide*. Academic Press, New York.

Mood, A.M., F.M. Graybill, D.C. Boes. (1974). *Introduction to the Theory of Statistics*. McGraw Hill, Boston.

Moore, R.E. (1966). *Interval Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey.

Moore, R.E. (1979). *Methods and Applications of Interval Analysis*. SIAM Studies on Applied Mathematics, Volume 2, Philadelphia.

Morgan, J.T. (1984). *Elements of Simulation*. Chapman & Hall, London.

Morgan, M.G. and M. Henrion. (1990*). Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, Cambridge, UK.

Morgan, M.G. and M. Henrion (1998). Analytica: a software tool for uncertainty analysis and model communication. Chapter 10 of *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press, New York. Available on-line at http://www.lumina.com/software/ch10.9.PDF.

Müller, A. (1997). Stop-loss order for portfolios of dependent risks. *Insurance: Mathematics and Economics* 21: 219-223.

Nelsen, R.B. (1986). Properties of a one-parameter family of bivariate distributions with specified marginals. *Communications in Statistics (Theory and Methods)* A15:3277-3285.

Nelsen, R.B. (1987). Discrete bivariate distributions with given marginals and correlation. *Communications in Statistics (Simulation and Computation)* B16:199-208.

Nelsen, R.B. (1991). Copulas and association. Pages 51-75 in *Advances in Probability Distributions with Given Marginals: Beyond the Copulas*, G. Dll'Aglio, S. Kotz and G. Salinetti (eds.), Kluwer Academic Publishers, London.

Nelsen, R.B. (1995). Copulas, characterization, correlation and counterexamples. *Mathematics Magazine* 68: 193-198.
Nelsen, R.B. (1999). *An Introduction to Copulas*. Lecture Notes in Statistics 139, Springer-Verlag, New York.
Nelsen, R.B., J.J. Quesada-Molina, J.A.Rodrígues-Lallena and M. Úbeda-Flores. (2001). Bounds on bivariate distribution functions with given marginals and measures of association. *Communications in Statistics (Theory and Methods)* 30: 1155-1162.
Nelsen, R.B., J.J. Quesada-Molina, J.A.Rodrígues-Lallena and M. Úbeda-Flores. (2004). Best-possible bounds on sets of bivariate distribution functions.*Journal of Multivariate Analyses* [in press].
Neumaier, A. (1990). *Interval Methods for Systems of Equations*, Cambridge University Press, Cambridge.
Neumaier, A. (2002). Taylor forms—use and limits. *Reliable Computing* 9: 43-79.
Oberkampf, W.L. and J. C. Helton. (2002). Investigation of evidence theory for engineering applications. AIAA Non-Deterministic Approaches Forum, April 2002, Denver, Colorado, paper 2002-1569.
Oberkampf, W.L., J. C. Helton and K. Sentz. (2001). Mathematical Representation of Uncertainty," AIAA Non-Deterministic Approaches Forum, April 2001, Seattle, Washington, paper 2001-1645.
Palisade Corporation. (1996). *@Risk: Advanced Risk Analysis for Spreadsheets*. Newfield, New York.
Press, W.H. B.P. Flannery, S.A. Teukolsky and W.T. Vetterling. (1992). *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, Cambridge University Press.
Regan, H.M., S. Ferson, and D. Berleant. (2004). Equivalence of five methods for bounding uncertainty. *International Journal of Approximate Reasoning* 36: 1-30.
Ripley, B.D. (1987). *Stochastic Simulation*. John Wiley, New York.
Risk Assessment Forum. (1994). *Peer Review Workshop Report on Ecological Risk Assessment Issue Papers*. U.S. Environmental Protection Agency, Washington, D.C.
Roberts, L. (1990). Risk assessors taken to task. *Science* 248:1173.
Robbins, H.E. (1944). On the measure of a random set, I. *Annals of Mathematical Statistics* 15: 70-74.
Robbins, H.E. (1945). On the measure of a random set, II. *Annals of Mathematical Statistics* 16: 342-347.
Rosenthal, A. (1975). A computer scientist looks at reliability computations. *Reliability and Fault Tree Analysis: Theoretical and Applied Aspects of System Reliability and Safety Assesment*, R.E. Barlow, J.B. Fussell and N.D. Singpurwalla (eds.), SIAM, Philadelphia.
Roussas, G.G. (1997). *A Course in Mathematical Statistics*. Academic Press, San Diego.

Salmento, J.S., E.S. Rubin, and A.M. Finkel. (1989). A review of @Risk. *Risk Analysis* 9: 255-257.
Scarsini, M. (1984). On measures of concordance. *Stochastica* 8: 201-218.
Scheuer, E.M. and D.S. Stoller. (1962). On the generation of normal random vectors. *Technometrics* 4:278-281.
Schweizer, B. (1991). Thirty years of copulas.*Advances in Probability Distributions with Given Marginals: Beyond the Copulas*, G. Dll'Aglio, S. Kotz and G. Salinetti (eds.), Kluwer Academic Publishers, London.
Sentz, K. and S. Ferson. (2002). *Combination of Evidence in Dempster-Shafer Theory*. Sandia National Laboratories, Technical Report SAND 2002-0835, Albuquerque, New Mexico. Available at http://www.sandia.gov/epistemic/Reports/SAND2002-0835.pdf.
Sklar, A. (1959). Fonctions de répartition à *n* dimensions et leurs marges. *Publications de L'Institut de Statistiques de l'Université de Paris* 8: 229-231.
Skelboe S. (1974). Computation of rational interval functions. *BIT* 14: 87-95.
Smith, A.E., P.B. Ryan and J.S. Evans. (1992). The effect of neglecting correlations when propagating uncertainty and estimating the population distribution of risk. *Risk Analysis* 12:467-474.
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 and Company, 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.
Song, P. (1997). Generating dependent random numbers with given correlations and margins from exponential dispersion models. *Journal of Statistical Computing and Simulation* 56: 317-335.
Spearman, C. (1904). "General intelligence", objectively determined and measured. *American Journal of Psychology* 15: 201-293. See a republication at http://psychclassics.yorku.ca/Spearman/.
Springer, M.D. (1979). *The Algebra of Random Variables*. Wiley, New York.
Tessem, B. (1992). Interval probability propagation. *International Journal of Approximate Reasoning* 7: 95-120.
Tonon, F., A. Bernadini and I. Elishakoff. (1999). Concept of random sets as applied to the design of structures and anlysis of expert opinions for aircraft crash. *Chaos, Solitons and Fractals* 10: 1855-1868.
Tonon, F., A. Bernadini and A. Mammino. (2000a). Determination of parameters range in rock engineering by means of Random Set Theory. *Reliability Engineering and System Safety* 70: 241-261.
Tonon, F., A. Bernadini and A. Mammino. (2000a). Reliability analysis of rock mass respsonse by means of Random Set Theory. *Reliability Engineering and System Safety* 70: 263-282.

Vesely, W.E., F.F. Goldberg, N.H. Roberts, D.F. Haasl. (1981). *Fault Tree Handbook*. Nuclear Regulatory Commission, Washington, DC.

Voit, E.O., W.L. Balthis and R.A. Holster. (1995). Hierarchical MC/P modeling with s-distributions: concepts and illustrative analysis of mercury contamination in king mackerel. *Environmental International* 21: 627-635.

Vyncke, D., M.J. Goovaerts and J. Dhaene and (2000). Convex upper and lower bounds for present value functions. Research report 0025, Department of Applied Economics, KU Leuven, Belgium.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall, London.

Walley, P. and T.L. Fine. (1982). Towards a frequentist theory of upper and lower probability. *Annals of Statistics* 10: 741-761.

Whitt, W. (1976). Bivariate distributions with given marginals. *The Annals of Statistics* 4:1280-1289.

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.

Wise. B.P. and M. Henrion. (1986). A framework for comparing uncertain inference systems to probability. *Uncertainty in Artificial Intelligence*. L.N. Kanal and J.F. Lemmer (eds.), Elsevier Science Publishers, Amsterdam.

Yager, R.R. (1986). Arithmetic and other operations on Dempster-Shafer structures. *International Journal of Man-machine Studies* 25: 357-366.