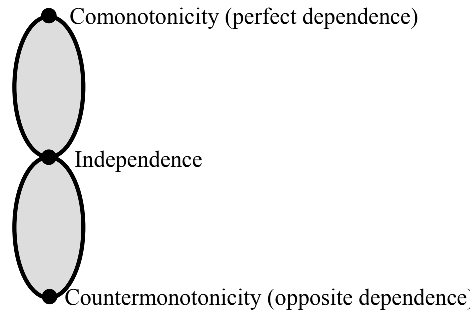
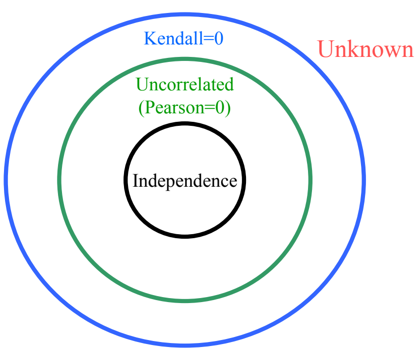
Independence when probabilities are imprecise

W. Troy Tucker, Scott Ferson and William Oberkampf
Applied Biomathematics and Sandia National Laboratories

9th ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability





Independence

- In the context of precise probabilities, there was a unique notion of independence
- In the context of imprecise probabilities, however, this notion disintegrates into several distinct concepts (Couso et al. 1999)
- The different kinds of independence behave differently in computations (Fetz 2001)

Equivalent definitions of independence

For precise probabilities, all these definitions are equivalent, so there's a *single concept*

- $H(x,y) = F(x)\ G(y)$, for all values x and y
- $P(X \in I,\ Y \in J) = P(X \in I)\ P(Y \in J)$, for any $I, J \subseteq \mathbf{R}$
- $h(x,y) = f(x)\ g(y)$, for all values x and y
- $E(w(X)\ z(Y)) = E(w(X))\ E(z(Y))$, for arbitrary w, z
- $\phi_{X,Y}(t,s) = \phi_X(t)\ \phi_Y(s)$, for arbitrary t and s

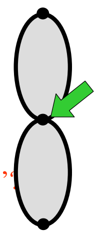
$P(X \le x) = F(x)$, $P(Y \le y) = G(y)$ and $P(X \le x,\ Y \le y) = H(x, y)$;
f, g and h are the density analogs of F,G and H; and
ϕ denotes the Fourier transform

Imprecise Probability

- DS structures
- Walley's "previsions"
- Coherent lower probability
- Probability bounds
- Interval probabilities
- Monte Carlo simulation with interval inputs
- Etc.

Imprecise probability independence

- Random-set independence
- Epistemic independence
- Strong independence
- Repetition independence
- Others?



Which should be called 'independence'?

Notation

- X and Y are random variables
- F_X and F_Y are their probability distributions
- F_X and F_Y aren't known precisely, but we know they're within classes $\mathcal{M}_X$ and $\mathcal{M}_Y$

$X \sim F_X \in \mathcal{M}_X$
$Y \sim F_Y \in \mathcal{M}_Y$

Repetition independence

- X and Y are random variables
- X and Y are independent (in the traditional sense)
- X and Y are identically distributed according to F
- F is unknown, but we know that $F \in \mathcal{M}$

$\Rightarrow$ X and Y are *repetition independent*

Analog of iid (independent and identically distributed)
$\mathcal{M}_{X,Y} = \{H : H(x, y) = F(x)\, F(y),\ F \in \mathcal{M}\}$

Strong independence

- $X \sim F_X \in \mathcal{M}_X$ and $Y \sim F_Y \in \mathcal{M}_Y$
- X and Y are stochastically independent
- All possible combinations of distributions from $\mathcal{M}_X$ and $\mathcal{M}_Y$ are allowed

$\Rightarrow$ X and Y are *strongly independent*

Complete absence of any relationship between X, Y
$\mathcal{M}_{X,Y} = \{H : H(x, y) = F_X(x)\, F_Y(y),$
$F_X \in \mathcal{M}_X,\ F_Y \in \mathcal{M}_Y\}$

Epistemic independence

- $X \sim F_X \in \mathcal{M}_X$ and $Y \sim F_Y \in \mathcal{M}_Y$
- $\underline{E}(f(X)|Y) = \underline{E}(f(X))$ and
  $\underline{E}(f(Y)|X) = \underline{E}(f(Y))$ for all functions f
  where $\underline{E}$ is the smallest mean over all possible probability distributions

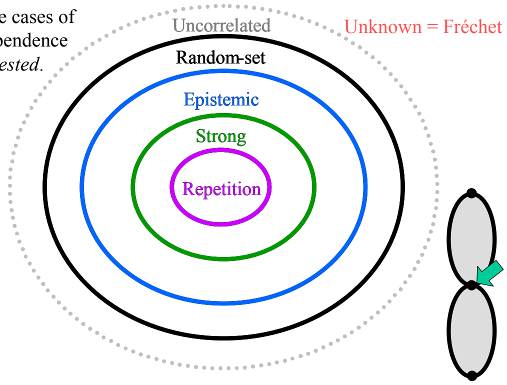
$\Rightarrow$ X and Y are *epistemically independent*

Lower bounds on expectations generalize the conditions $P(X|Y) = P(X)$ and $P(Y|X) = P(Y)$

Random-set independence

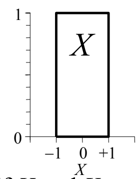
- Embodied in Cartesian products
- X and Y with mass functions m_X and m_Y are *random-set independent* if the Dempster-Shafer structure for their joint distribution has mass function $m(A_1 \times A_2) = m_X(A_1)\, m_Y(A_2)$ whenever A_1 is a focal element of X and A_2 is a focal element of Y, and $m(A) = 0$ otherwise
- Often easiest to compute

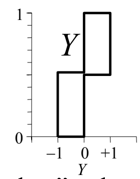
These cases of independence are *nested*.



Interesting example

- $X = [-1, +1]$, $Y = \{([-1, 0], ½), ([0, 1], ½)\}$



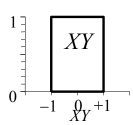


- If X and Y are "independent", what is $Z = XY$?

Compute via Yager's convolution

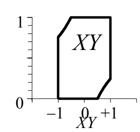
	Y	
	([−1, 0], ½)	([0, 1], ½)
X ([−1, +1], 1)	([−1, +1], ½)	([−1, +1], ½)

The Cartesian product with one row and two columns produces this p-box



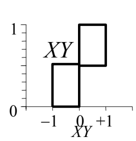
But consider the means

- Clearly, EX = [−1,+1] and EY=[−½, +½].
- Therefore, E(XY) = [−½, +½].
- But if this is the mean of the product, and its range is [−1,+1], then probability bounds analysis tells us better bounds on the CDF.



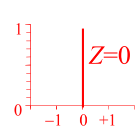
And consider the quantity signs

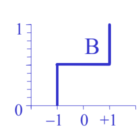
- What's the probability P_Z that $Z < 0$?
- $Z < 0$ only if $X < 0$ or $Y < 0$ (but not both)
- $P_Z = P_X(1-P_Y) + P_Y(1-P_X)$, where
  $P_X = P(X < 0)$, $P_Y = P(Y < 0)$
- But P_Y is ½ by construction
- So $P_Z = ½P_X + ½(1-P_X) = ½$
- Thus, zero is the median of Z
- Knowing median and range improves bounds

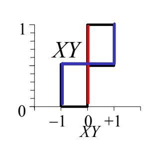


Best possible

- These bounds are realized by solutions
  If $X \equiv 0$, then $Z \equiv 0$
  If $X \sim Y \sim$ B = {(−1, ½),(+1, ½)}, then $Z \sim$ B

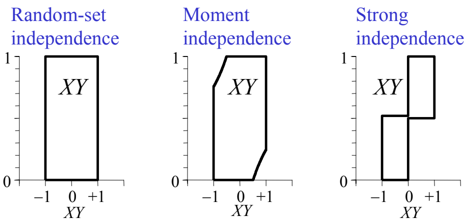






- So these bounds are also best possible since these examples live on the edges.

So which is correct?



The answer depends on what one meant by “independent”.

So what?

- The example illustrates a practical difference between random-set independence and strong independence
- It disproves the conjecture that the convolution of uncertain numbers is not affected by dependence assumptions if at least one of them is an interval
- It tempers the claim about the best-possible nature of convolutions with probability boxes and Dempster-Shafer structures

Strategy for risk analysts

- Random-set independence is conservative
- Using the Cartesian product approach is always rigorous, though may not be optimal
- Convenient methods to obtain tighter bounds under other kinds of independence await development

Uncertainty algebra

Operands and operation	***Answers under different dependence assumptions***
1) One interval ⇒	random-set = unknown
2) Two intervals ⇒	strong = epistemic = random-set = unknown
3) Interval and a function of an interval ⇒	strong = epistemic = random-set = unknown
4) One interval and monotone operation ⇒	strong = epistemic = random-set = unknown
5) Monotone operation ⇒	strong = epistemic = random-set
6) Two precise distributions ⇒	strong = epistemic = random-set
7) All cases ⇒	repetition ⊆ strong ⊆ epistemic ⊆ random-set ⊆ unknown

(after Fetz and Oberguggenberger 2004)

(Colored words denote the set of distributions that result from binary operations invoking the specified assumption about dependence)

Theoretical dependence cases

- Random-set independent
  - o Epistemically independent
  - o Strongly independent
  - o Repetition independent
- Perfectly associated (comonotonic)
- Oppositely associated (countermonotonic)
- Known copula
- Specified copula family and correlation
- Known functional relationship
- Positively quadrant dependent (PQD)
- Negatively quadrant dependent (NQD)
- Known or interval-bounded correlation
- Fréchet case

For precise probabilities, these are all the same.

These cases yield precise distributions from precise input distributions

Acknowledgments

- Peter Walley
- Thomas Fetz
- Gert de Cooman
- Sandia National Laboratories’ Epistemic Uncertainty Project

End