

Diferentes Abordagens à Teoria de Possibilidade

Estevão Esmi and Laécio Carvalho de Barros *

Instituto de Matemática, Estatística e Computação Científica
Universidade de Cammpinas

Resumo Há na literatura matemática pelo menos duas abordagens distintas sobre teoria de possibilidade: uma “conjuntista” e outra “probabilista”. Tal fato torna-se evidente quando vemos que uma distribuição de possibilidade pode ser identificada com um conjunto fuzzy e, por outro lado, por meio de medida de probabilidade superior. Tal ambiguidade resulta da falta de consenso sobre a interpretação adequada para distribuição de possibilidade e como se deve computá-las. Neste trabalho, veremos que esse tipo de confusão é desnecessária pois cada uma dessas visões surgem de abordagens distintas no tratamento de incertezas.

Keywords: Distribuição de possibilidade, conjuntos fuzzy, probabilidade superior, distribuição condicional.

1 Introdução

Recentemente, os dois primeiros autores deste trabalho junto com outros pesquisadores submeteram para publicação um artigo que envolvia a noção de não interatividade entre distribuições de possibilidade introduzida por Zadeh [1,2]. Um dos revisores deste artigo questionou o uso de tal noção pois como ele apontou: medidas de possibilidade nada mais são do que casos particulares de medidas de probabilidade superiores [3,4,5,6] que surgem de certas extensões da teoria de probabilidade clássica e a noção de não interatividade, tal como proposta por Zadeh, parece não condizer com nenhuma das extensões da noção de independência estocástica encontrada na literatura. As observações do revisor em questão nos fez atentar que algumas confusões e mal-entendidos com respeito à teoria de possibilidades surge pelo fato de haver mais de uma abordagem à mesma. Motivados por este comentário, decidimos escrever esse artigo onde o nosso principal objetivo é mostrar pelo menos três abordagens distintas à teoria de possibilidade. Esperamos conseguir passar ao leitor que, por mais que essas três abordagens apresentem algumas coincidências em certos casos, elas não são equivalentes, sobretudo quando adentrarmos sobre conceitos como condicional, independência, etc, que são importantes na elaboração de métodos de inferência e raciocínio aproximado envolvendo possibilidade.

Neste trabalho não pretendemos apresentar conceitos e resultados novos ou fazer uma revisão aprofundada da literatura de teoria de possibilidade (mesmo

* Este trabalho foi parcialmente pelo CNPq.

porque não há espaço suficiente aqui para tal). Nosso objetivo neste trabalho é apresentar rapidamente três abordagens à teoria de possibilidade de maneira concisa sendo a primeira baseada em teoria de conjuntos fuzzy e as últimas duas em teoria de medidas. Embora as bases teóricas da primeira e das demais abordagens serem distintas, há diversas coindências entre elas sobretudo nas definições de distribuição (conjunta) de possibilidade e medida de possibilidade quando consideramos a σ -álgebra dada pelo conjunto das partes do universo em questão. Ao nosso ver, tais equivalências acabam por causar certa confusão entre alguns pesquisadores que ora adotam resultados oriundos de uma abordagem ora de outra sem atentar para respectivas bases teóricas.

Atualmente há na literatura diversas ferramentas matemáticas que lidam com a modelagem de incertezas. Em particular focaremos em três destas: teoria de probabilidade, teoria de conjuntos fuzzy e teoria de possibilidade. Dentre estas, a mais antiga e bem consolidada é a de probabilidades. A teoria de conjuntos fuzzy surge em 1965 com Zadeh [7] estendendo a teoria de conjuntos e subsequente a teoria de possibilidade [8,9,1] que surge como uma teoria “intermediária” entre as outras duas. Posteriormente, descobriram-se conexões da teoria de possibilidade com teoria de evidência e probabilidade intervalar [4,10,11,12]. As três trazem terminologias semelhantes, porém conceitualmente distintas. Por exemplo, há a noção de medida de probabilidade, medida de possibilidade e medida fuzzy, sendo essa última extensão das outras duas. Também há na literatura as denominações, distribuição de probabilidades e de possibilidades. No caso fuzzy esse termo não é usado. A conexão entre as teorias de conjuntos fuzzy e de possibilidades se dá por meio das chamadas distribuições, *i.e.*, funções definidas de um universo Ω no intervalo $[0, 1]$, que como veremos, são identificadas com funções de pertinências de certos conjuntos fuzzy. Já a conexão entre as teorias de possibilidades e de probabilidades se dá por meio de suas respectivas medidas, sem a necessidade explícita de distribuições.

A teoria de possibilidade proposta por Zadeh toma como base a teoria de conjuntos fuzzy [1], que por sua vez é uma extensão da teoria de conjuntos [7]. Esta abordagem, como já mencionamos, consiste essencialmente em identificar funções de pertinência de certos conjuntos fuzzy com funções de distribuições de possibilidade. Nesse caso, todo ferramental matemático empregado para se processar e inferir sobre distribuição de possibilidades é próprio da teoria de conjuntos fuzzy. Assim podemos interpretar a distribuição de possibilidade como uma função que avalia “a possibilidade de um elemento estar em um certo conjunto”.

Por outro lado, a exemplo do que acontece em probabilidade, a teoria de possibilidade também pode ser conduzida com base em teoria de medidas. Aqui, distinguimos dois casos de interesse. O primeiro consiste em desenvolver a teoria de possibilidade nos moldes (ou de maneira análoga) ao caso de probabilidade, “simplesmente” substituindo medida de probabilidade por possibilidade nas definições e, então, estudando as propriedades resultantes desta substituição. Tal estratégia é alinhada com aplicações de teorias de medidas gerais de incertezas que incluem medidas não-aditivas, medidas fuzzy, medidas de Choquet, etc. Dada a semelhança com o caso probabilístico, podemos arguir que esta aborda-

gem pode ser empregada para avaliar e modelar algo como “a possibilidade de um evento ocorrer”. O segundo caso de interesse surge da associação de medidas de possibilidades com casos particulares de certas medidas de probabilidade superior que é um conceito que aparece em diversos contextos como, por exemplo, do estudo de coleções de medidas de probabilidade [6]. Especificamente neste contexto, uma medida de probabilidade superior corresponde ao envelope superior de um dado conjunto de medidas de probabilidade em uma σ -álgebra. Associar medidas de possibilidade à medidas de probabilidade superior nos habilita a emprestar as ferramentas e operadores estabelecidos para probabilidade superior em cálculos envolvendo medidas de possibilidade. Assim, tal abordagem à teoria de possibilidade pode ser vista como um segmento da teoria de probabilidade superior. Neste caso, a interpretação que cabe à distribuição de possibilidade deve estar alinhada ao de probabilidade superior que envolve noções de medidas de probabilidade. Por exemplo, “a *possibilidade* de um evento A ” pode ser traduzida como “a maior probabilidade *possível* do evento A ocorrer”.

Este artigo está organizado como se segue. Iniciamos por apresentar uma abordagem à teoria de possibilidade baseada em teoria de conjuntos fuzzy, que para estabelece-la não é requerido o conceito de medida, apesar de estar em acordo com a noção de medida de possibilidade como veremos nas seções subsequentes. Em seguida, relembramos alguns conceitos básicos de probabilidade necessários para o que se segue nas Seções 4 e 5. Na Seção 4, apresentamos uma abordagem à teoria de possibilidade baseada em teoria de medidas cujo desenvolvimento se dá de maneira análoga ao da teoria de probabilidades. Na Seção 5 relembramos rapidamente a noção de medida de probabilidade superior e indicamos a sua conexão com medida de possibilidade. Por fim, encerramos este artigo com algumas considerações finais.

2 Teoria de Possibilidade Via Conjuntos Fuzzy

Um abordagem sobre teoria de possibilidade encontrada na literatura [1,13] pode ser vista como um segmento da teoria de conjuntos fuzzy. O ponto chave dessa abordagem consiste em associar a noção de distribuição de possibilidade dada na Definição 1 com a de conjuntos fuzzy tal como se segue.

Definition 1 (Distribuição de Possibilidade). *Seja Ω um conjunto não vazio qualquer. Uma distribuição de possibilidade em Ω é uma função $\pi : \Omega \rightarrow [0, 1]$ tal que $\sup_{\omega \in \Omega} \pi(\omega) = 1$.*

Um (sub)conjunto fuzzy A do universo Ω é identificado com uma função $\varphi_A : \Omega \rightarrow [0, 1]$, chamada de *função de pertinência de A* , onde o valor $\varphi_A(\omega)$ denota a pertinência do elemento $\omega \in \Omega$ ao conjunto fuzzy A . A fim de facilitar a notação, denotaremos $\varphi_A(\omega)$ simplesmente por $A(\omega)$. Também a classe dos conjuntos fuzzy em Ω pelo símbolo $\mathcal{F}(\Omega)$.

Um subconjunto fuzzy $R \in \mathcal{F}(\Omega_1 \times \dots \times \Omega_n)$ é denominado uma relação fuzzy (n-ária), onde o valor $R(\omega_1, \dots, \omega_n)$ denota o grau de relação entre $\omega_1, \dots, \omega_n$.

A projeção fuzzy da relação R em Ω_i é conjunto fuzzy $P_R^i \in \mathcal{F}(\Omega_i)$ dado para todo $y \in \Omega_i$ por

$$P_R^i(y) = \sup\{R(\omega_1, \dots, \omega_n) \mid \omega_i = y \text{ e } \omega_j \in \Omega_j \text{ para } j \neq i\}. \quad (1)$$

Dado uma relação fuzzy binária $R \in \mathcal{F}(\Omega_1 \times \Omega_2)$ definimos sua relação fuzzy inversa $R^{-1} \in \mathcal{F}(\Omega_2 \times \Omega_1)$ por $R^{-1}(y, x) = R(x, y)$ para todo $(y, x) \in \Omega_2 \times \Omega_1$.

Dizemos que um subconjunto fuzzy A de Ω é normal se $\sup_{\omega \in \Omega} A(\omega) = 1$ [14]. Note que da Definição 1, temos que uma distribuição de possibilidade pode ser associada biunivocamente a uma função de pertinência de um conjunto fuzzy normal. Tal identificação nos permite definir conceitos e operadores para distribuição de possibilidade através dos já estabelecidos na teoria de conjuntos fuzzy. Por exemplo, um distribuição de possibilidade conjunta π em $\Omega_1 \times \dots \times \Omega_n$ corresponde a função de pertinência de uma relação fuzzy normal no respectivo produto cartesiano, onde as distribuições marginais de π são dadas pelas funções de pertinências das respectivas projeções fuzzy. Repare que, uma vez que a relação fuzzy associada a π é normal, segue que suas projeções fuzzy também são normais e, portanto, as distribuições marginais de π representam distribuições de possibilidade.

Distribuições de possibilidade condicionais podem ser definidas em termos dos chamados métodos de inferência e raciocínio aproximado fuzzy baseados em processos dedutivos tais como *modus ponens* e *modus tollens* oriundos da lógica proposicional [15,16]. O resto desta seção dedica-se em apresentar tais métodos fuzzy e relaciona-los com cálculos de distribuições condicionais.

Sejam $A, B \in \mathcal{F}(\Omega)$, dizemos que $A \subseteq B$ sempre que $A(\omega) \leq B(\omega)$ para todo $\omega \in \Omega$. Além da relação de inclusão, também podemos definir operações de intersecção e união como se segue. Definimos respectivamente a intersecção e união entre os conjuntos fuzzy A e B como sendo os conjuntos fuzzy $C = A \cap B$ e $D = A \cup B$ dados para todo $\omega \in \Omega$ por $C(\omega) = \min\{A(\omega), B(\omega)\}$ e $D(\omega) = \max\{A(\omega), B(\omega)\}$. Note que as definições acima estendem a teoria de conjuntos clássica uma vez que um subconjunto Y de Ω pode associado biunivocamente a um conjunto fuzzy cuja função de pertinência é dada pela sua função indicadora (ou característica).

Diversas operações entre conjuntos podem ser definidas em termos de operadores binários denominados de *conectores lógico fuzzy* que nada mais são que extensões dos operadores booleanos para o intervalo unitário. Considere uma operador binário $\star : [0, 1]^2 \rightarrow [0, 1]$ comutativo, associativo e crescente tal que existe $e \in [0, 1]$ tal que $a \star e = a$ para todo $a \in [0, 1]$. Se $e = 1$, então a operação \star é dita uma *t-norma* e estende o conectivo lógico Booleano “e”. Agora, se $e = 0$, então a operação \star é dita uma *t-connorma* (ou *s-norma*) e estende o conectivo lógico Booleano “ou”. O produto cartesiano de conjuntos fuzzy $A_i \in \mathcal{F}(\Omega_i)$, $i = 1, \dots, n$, com respeito a uma t-norma t é uma relação fuzzy $R \in \mathbb{R}(\Omega_1 \times \dots \times \Omega_n)$ dada por $R(\omega_1, \dots, \omega_n) = A_1(\omega_1) t \dots t A_n(\omega_n)$.

Uma implicação fuzzy é um operador binário $\triangleright : [0, 1]^2 \rightarrow [0, 1]$ decrescente no primeiro argumento e crescente no segundo argumento que estende a implicação da lógica Booleana, isto é, $0 \triangleright 0 = 0 \triangleright 1 = 1 \triangleright 1 = 1$ e $1 \triangleright 0 = 0$.

Implicações fuzzy podem ser definidas em termos de outros operadores da lógicas fuzzy. Por exemplo, dado uma t-norma t , podemos definir uma implicação fuzzy \triangleright_t , chamada *residual* ou *R-implicação*, dada por

$$a \triangleright_t b = \sup\{c \in [0, 1] \mid a t c \leq b\}, \quad \forall a, b \in [0, 1].$$

Em particular, a R-implicação fuzzy com respeito a t-norma do mínimo ($t \equiv \cdot$) é dada por

$$a \triangleright \cdot b = \begin{cases} 1 & \text{se } a \leq b \\ \frac{b}{a} & \text{se } a > b \end{cases}, \quad \forall a, b \in [0, 1]. \quad (2)$$

Se t denota uma t-norma contínua à esquerda no primeiro argumento e se \triangleright_t denota a respectiva R-implicação fuzzy, então vale a seguinte desigualdade:

$$r t (r \triangleright_t y) \leq y, \quad \forall r, y \in [0, 1]. \quad (3)$$

Além disso, se $x \in [0, 1]$ é tal que $r t x \leq y$, então $x \leq (r \triangleright_t y)$, isto é, $(r \triangleright_t y)$ é única solução maximal da desigualdade (3) para r e y dados. Como consequência direta da última observação e da monotocidade de t , temos que $\exists x \in [0, 1]$ tal que $r t x = y \Leftrightarrow r t (r \triangleright_t y) = y$.

Baseados nas definições acima, definimos a noção de distribuição possibilidade condicional como se segue. Seja π uma distribuição de possibilidade conjunta em $\Omega_1 \times \Omega_2$ com marginais $\pi_i = P_\pi^i$ para $i = 1, 2$ e seja t uma t-norma contínua à esquerda no primeiro argumento. Dado $y \in \Omega_2$, o conjunto fuzzy $\pi_y \in \mathcal{F}(\Omega_1)$ cuja função de pertinência é dada por $\pi_y(x) = \pi_2(y) \triangleright_t \pi(x, y)$ é chamado de distribuição de possibilidade condicional em Ω_1 dado $y \in \Omega_2$ com respeito à t-norma t . Neste caso, denotamos o grau da pertinência x em π_y pelo símbolo $\pi(x|y)$, isto é, $\pi(x|y) = \pi_2(y) \triangleright_t \pi(x, y)$, $\forall x \in \Omega_1$. Repare que $\pi(\cdot|y)$ não necessariamente é normal e, portanto, pode não ser uma distribuição de possibilidade. Uma condição suficiente para que $\pi(\cdot|y)$ seja uma distribuição de possibilidade é de existir $x \in \Omega_1$ tal que $\pi(x, y) = \pi_2(y)$. Dizemos que π_1 é π_2 são independente se $\pi_1(x) = \pi(x|y)$ para todo $x \in \Omega_1$ e $y \in \Omega_2$. Adicionalmente, dizemos que π_1 e π_2 são não interativos se $\min\{\pi_1(x), \pi_2(y)\} = \pi(x, y)$ para todo $x \in \Omega_1$ e $y \in \Omega_2$.

Note que a definição de condicional acima requer que uma distribuição conjunta seja fornecida. Considere o caso onde ao invés da distribuição conjunta tenhamos disponível apenas as distribuições marginais π_1 e π_2 . Neste caso podemos tomar a distribuição conjunta $\pi_{\tilde{t}}$ dada pelo produto cartesiano de π_1 e π_2 com respeito a uma t-norma \tilde{t} qualquer e, então, aplicar a definição anterior. Especificamente, definimos a distribuição condicional com respeito à t e \tilde{t} de π_1 dado π_2 como $\pi(x|y) = \pi_2(y) \triangleright_t (\pi_1(x) \tilde{t} \pi_2(y))$ [15]. Repare que t e \tilde{t} podem não ser iguais na definição acima. Assumir que π_1 e π_2 são não interativos na definição anterior equivale a assumir que $\tilde{t} \equiv \wedge$ e, neste caso, temos que $\pi(x|y) = \pi_2(y) \triangleright_t (\pi_1(x) \wedge \pi_2(y)) = \pi_2(y) \triangleright_t \pi_1(x)$.

3 Conceitos Básico de Probabilidade

A teoria de probabilidade é embasada na teoria de medidas que nada mais é que uma função de conjuntos, cujo domínio é chamado de σ -álgebra, como veremos a

seguir. Seja Ω um conjunto não vazio e seja \mathcal{A} uma classe de subconjuntos de Ω . Dizemos que \mathcal{A} é uma σ -álgebra em Ω se satisfaz as seguintes três propriedades: $\Omega \in \mathcal{A}$; se $A \in \mathcal{A}$, então $A^c \in \mathcal{A}$; Se $A_i \in \mathcal{A}$ para $i > 0$, então $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$. O par (Ω, \mathcal{A}) é chamado de espaço \mathcal{A} -mensurável, ou simplesmente mensurável.

A maior σ -álgebra de Ω é o conjunto das partes de Ω denotado por $\mathcal{P}(\Omega) = \{Y \mid Y \subseteq \Omega\}$. Note que intersecção de σ -álgebras também são σ -álgebras. A menor σ -álgebra que contém $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ é dita σ -álgebra gerada por \mathcal{C} . Em particular, a σ -álgebra gerada por todos os subconjuntos aberto de \mathbb{R} é chamada σ -álgebra de *Borel*, denotada aqui pelo símbolo \mathcal{B} .

Definition 2 (Medida de Probabilidade). *Seja \mathcal{A} uma σ -álgebra de um conjunto Ω . Uma função P de \mathcal{A} para $\mathbb{R}^+ := [0, +\infty)$ é dita uma **medida de probabilidade** se P satisfizer as seguintes propriedades:*

P1: $P(A) \leq 1$ para todo $A \in \mathcal{A}$;

P2: $P(\Omega) = 1$;

P3: Se $A_1, A_2, \dots, A_n, \dots \in \mathcal{A}$ tal que $A_i \cap A_j = \emptyset$ para $i \neq j$, então

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n). \quad (\sigma\text{-aditividade})$$

Se P é uma medida de probabilidade definida em um espaço mensurável (Ω, \mathcal{A}) , então a tripla (Ω, \mathcal{A}, P) é dita espaço de probabilidade. Além disso, cada elemento $A \in \mathcal{A}$ é chamado de evento e $P(A)$ é a probabilidade do evento A . Aqui, denotaremos a classe das medidas de probabilidade definidas em um mesmo espaço mensurável (Ω, \mathcal{A}) pelo símbolo $\mathbb{P}(\Omega, \mathcal{A})$.

A partir das propriedades $P1 - P3$, pode-se mostrar que uma medida de probabilidade P em (Ω, \mathcal{A}) também satisfaz as seguintes propriedades:

P4: $P(\emptyset) = 0$;

P5: $P(A) \leq P(B)$ se $A \subseteq B$, $A, B \in \mathcal{A}$;

P6: $P(A) = 1 - P(A^c)$ para todo $A \in \mathcal{A}$.

Uma medida de probabilidade no espaço mensurável $(\mathbb{R}, \mathcal{B})$ pode ser definida em termos de uma *função de densidade de probabilidade*, que nada mais é do que uma função $f : \mathbb{R} \rightarrow \mathbb{R}^+$ tal que $\int_{-\infty}^{+\infty} f(x)dx = 1$. Em particular, definimos a medida de probabilidade P para $A \in \mathcal{B}$ por $P(A) = \int_A f(x)dx$. Apenas como registro, ressaltamos que essa é uma forma canônica de “construir” a medida P que “avalia” cada evento A .

Uma variável aleatória (v.a.) X em um espaço de probabilidade (Ω, \mathcal{A}, P) nada mais é do que uma função $X : \Omega \rightarrow \mathbb{R}$ tal que

$$X^{-1}(B) := \{\omega \in \Omega \mid X(\omega) \in B\} \in \mathcal{A}, \quad \forall B \in \mathcal{B}.$$

Note que a composição de X com uma função contínua $g : \mathbb{R} \rightarrow \mathbb{R}$ dá origem a outra v.a. denotada aqui simplesmente por $g(X)$.

Note que através da variável aleatória X podemos induzir uma medida de probabilidade \tilde{P} no espaço mensurável $(\mathbb{R}, \mathcal{B})$ dada por $\tilde{P}(B) = P(X^{-1}(B))$

para todo $B \in \mathcal{B}$. Assim, o contra-domínio da variável aleatória X passa a ter estrutura de espaço de probabilidade $(\mathbb{R}, \mathcal{B}, \bar{P})$ que é particularmente interessante pois permite associar, quando possível, a medida de probabilidade \bar{P} à uma função de densidade de probabilidade. Neste ponto, a teoria de probabilidade se bifurca em dois casos: discreto e contínuo. O primeiro ocorre quando a imagem de X , denotada aqui por $Im(X)$, é um conjunto enumerável e, neste caso, a função de densidade de probabilidade é uma função $f : Im(X) \rightarrow [0, 1]$ dada por $f(x) = \bar{P}(\{x\}) = P(X^{-1}(\{x\}))$ para todo $x \in Im(X)$. O segundo caso ocorre quando $Im(X)$ é não enumerável e existe uma função $f : \mathbb{R} \rightarrow \mathbb{R}^+$, dita densidade de probabilidade, tal que $\bar{P}(B) = \int_B f(x)dx = \int_{X^{-1}(B)} P(d\omega)$ para todo $B \in \mathcal{B}$.

Note que, em ambos os casos, discreto e contínuo, se soubermos a função de densidade de probabilidade f , então podemos determinar a respectiva medida \bar{P} que governa no espaço de probabilidade do contra-domínio da v.a. X , mesmo que não saibamos explicitamente a medida P que governa no espaço de probabilidade do domínio de X . Na prática, muitas vezes dispomos apenas de um conhecimento parcial da função de densidade probabilidade f e é exatamente neste contexto que surgem estudos por exemplo de probabilidade superior e inferior.

Sejam $X : (\Omega_X, \mathcal{A}_X, P_X) \rightarrow (\mathbb{R}, \mathcal{B}, \bar{P}_X)$ e $Y : (\Omega_Y, \mathcal{A}_Y, P_Y) \rightarrow (\mathbb{R}, \mathcal{B}, \bar{P}_Y)$ duas variáveis aleatórias, onde \bar{P}_X e \bar{P}_Y denotam as medidas de probabilidades induzidas pelas v.a. X e Y , respectivamente. Uma medida de probabilidade conjunta de (X, Y) é uma medida de probabilidade \bar{P}_{XY} definida no espaço mensurável $(\mathbb{R} \times \mathbb{R}, \mathcal{B} \times \mathcal{B})$ que satisfaz as seguintes condições:

$$\begin{aligned} \text{M1 } \bar{P}_{XY}(A \times \mathbb{R}) &= \bar{P}_X(A) = P_X(X^{-1}(A)), \quad \forall A \in \mathcal{B}; \\ \text{M2 } \bar{P}_{XY}(\mathbb{R} \times B) &= \bar{P}_Y(B) = P_Y(Y^{-1}(B)), \quad \forall B \in \mathcal{B}. \end{aligned}$$

Por convenção, denotamos $\bar{P}_{XY}(A \times B) = \bar{P}_{XY}(X \in A, Y \in B)$ para todo $A \times B \in \mathcal{B} \times \mathcal{B}$ e as medidas de probabilidade \bar{P}_X e \bar{P}_Y são ditas medidas de probabilidade marginais.

Considere o espaço de probabilidade (Ω, \mathcal{A}, P) e $B \in \mathcal{A}$ tal que $P(B) > 0$. A probabilidade condicional de um evento $A \in \mathcal{A}$ dado B , denotada por $P(A|B)$, é dada por

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (4)$$

Os eventos A e B são ditos *independentes* se $P(A|B) = P(A)$ ou, equivalentemente, se $P(A \cap B) = P(A)P(B)$.

Vejamos agora o conceito de variáveis aleatórias independentes. Seja \bar{P}_{XY} uma medida de probabilidade conjunta de (X, Y) , onde X e Y são duas variáveis aleatórias. Dado $A \times \mathbb{R}, \mathbb{R} \times B \in \mathcal{B} \times \mathcal{B}$ com $\bar{P}_{XY}(\mathbb{R} \times B) = \bar{P}_Y(B) > 0$, pelo visto acima, temos que a probabilidade condicional de $A \times \mathbb{R}$ dado $\mathbb{R} \times B$ é dada respectivamente por

$$\bar{P}_{XY}(A \times \mathbb{R} | \mathbb{R} \times B) = \frac{\bar{P}_{XY}(A \times \mathbb{R} \cap \mathbb{R} \times B)}{\bar{P}_{XY}(\mathbb{R} \times B)} = \frac{\bar{P}_{XY}(A \times B)}{\bar{P}_Y(B)}. \quad (5)$$

Alternativamente, também denotamos $\bar{P}_{XY}(A \times \mathbb{R} | \mathbb{R} \times B)$ pelo símbolo $\bar{P}_{XY}(X \in A | Y \in B)$. No caso onde $A \times \mathbb{R}$ e $\mathbb{R} \times B$ são independentes temos que

$$\bar{P}_X(A) = \bar{P}_{XY}(A \times \mathbb{R}) = \bar{P}_{XY}(A \times \mathbb{R} | \mathbb{R} \times B) \quad (6)$$

ou, similarmente, se $\bar{P}_{XY}(A \times B) = \bar{P}_X(A)\bar{P}_Y(B)$. Dizemos que X e Y são variáveis aleatórias *independente* se para todo $A, B \in \mathcal{B}$, com $\bar{P}_Y(B) > 0$, vale a Equação (6).

4 Teoria de Possibilidade Via Teoria de Medidas

Nesta seção abordaremos teoria de possibilidade via teoria de medidas nos moldes do que é feito na teoria de probabilidade. A idéia aqui consiste em substituir nas definições da seção anterior medidas de probabilidade por medidas de possibilidade.

Definition 3 (Medida de Possibilidade). *Seja \mathcal{A} uma σ -álgebra de Ω . Uma medida de possibilidade em \mathcal{A} é uma função $\Pi : \mathcal{A} \rightarrow [0, 1]$ que satisfaz $\Pi(\emptyset) = 0$ e $\Pi(\Omega) = 1$ e que para todo $A, B \in \mathcal{A}$ tem-se $\Pi(A \cup B) = \max\{\Pi(A), \Pi(B)\}$.*

Uma consequência direta da Definição 3 é que Π é monótona crescente, isto é, sejam $A, B \in \mathcal{A}$ tais que $A \subseteq B$, então $\Pi(A) \leq \Pi(B)$ pois $\Pi(B) = \max\{\Pi(A), \Pi(A^c \cap B)\}$.

Sejam $(\Omega_X, \mathcal{A}_X)$ e $(\Omega_Y, \mathcal{A}_Y)$ dois espaços mensuráveis. Dados duas medidas de possibilidade $\Pi_X : \mathcal{A}_X \rightarrow [0, 1]$ e $\Pi_Y : \mathcal{A}_Y \rightarrow [0, 1]$, uma medida de possibilidade conjunta em $\Omega_X \times \Omega_Y$ é uma medida de possibilidade $\Pi_{XY} : \mathcal{A}_X \times \mathcal{A}_Y \rightarrow [0, 1]$ tal que satisfaz as seguintes condições:

$$\begin{aligned} \text{N1 } \Pi_{XY}(A \times \Omega_Y) &= \Pi_X(A), \quad \forall A \in \mathcal{A}_X; \\ \text{N2 } \Pi_{XY}(\Omega_X \times B) &= \Pi_Y(B), \quad \forall B \in \mathcal{A}_Y. \end{aligned}$$

As medidas de possibilidade Π_X e Π_Y são ditas medidas de possibilidade marginais. Repare que as propriedades N1 e N2 e a monotocidade de Π_{XY} implicam que $\Pi_{XY}(A \times B) \leq \min\{\Pi_X(A), \Pi_Y(B)\}$.

Para definir condicional possibilística podemos seguir a estratégia de substituir a medida de probabilidade por possibilidade na Equação (4), o que resultaria na seguinte definição para a possibilidade de um evento $A \in \mathcal{A}$ dado um outro evento $B \in \mathcal{A}$ com $\Pi(B) > 0$ [16]:

$$\Pi(A|B) = \frac{\Pi(A \cap B)}{\Pi(B)}. \quad (7)$$

Com tudo, outras medidas condicionais possibilísticas podem também ser utilizadas se adotarmos como base outras igualdades para o caso probabilístico derivadas das propriedades das medidas de probabilidade. Por exemplo, para um espaço de probabilidades (Ω, \mathcal{A}, P) , podemos utilizar as propriedades P1 - P6 para obter as seguintes igualdades

$$P(A|B) = \frac{P(A \cap B)}{1 - P(B^c)} = \frac{P(A \cap B)}{P(A \cap B) + 1 - P((A^c \cap B)^c)}.$$

O que resultaria nas seguintes definições de possibilidade condicional [16]:

- (a) $\Pi_a(A|B) = \frac{\Pi(A \cap B)}{1 - \Pi(B^c)}$ para $\Pi(B) < 1$;
(b) se $(\Pi(A \cap B) + 1 - \Pi((A^c \cap B)^c)) > 0$, então

$$\Pi_b(A|B) = \frac{\Pi(A \cap B)}{\Pi(A \cap B) + 1 - \Pi((A^c \cap B)^c)}. \quad (8)$$

Na literatura também encontramos propostas axiomáticas envolvendo elementos da lógica fuzzy para definir medida de possibilidade condicional sem a noção de medida de possibilidade conjunta (o leitor interessado pode consultar [17]).

Se \mathcal{A} é a maior σ -álgebra de Ω , isto é, $\mathcal{A} = \mathcal{P}(\Omega)$, então uma medida de possibilidade Π pode ser associada a uma única função $\pi : \Omega \rightarrow [0, 1]$ e vice-versa através da seguinte relação:

$$\Pi(\{\omega\}) = \pi(\omega), \quad \forall \omega \in \Omega. \quad (9)$$

Note que a função π acima está de acordo com a noção de distribuição de possibilidade conforme a Definição 1 que é estabelecida via teoria de conjuntos fuzzy nas seções anteriores. Assim, a noção de distribuição de possibilidade surge em ambas abordagens, via teoria de conjuntos fuzzy e via teoria de medidas, de maneiras independente, porém, equivalente.

Fazendo um paralelo com teoria de probabilidades, temos que distribuição de possibilidade está para medida de possibilidade tal como distribuição de densidade de probabilidade está para medida de probabilidade uma vez que

$$\Pi(A) = \Pi\left(\bigcup_{\omega \in A} \{\omega\}\right) = \bigvee_{x \in A} \Pi(\{\omega\}) = \bigvee_{x \in A} \pi(\omega) \quad (10)$$

para todo $A \in \mathcal{A}$. Relembre que em probabilidade, quando há uma distribuição de densidade de probabilidade o cálculo da probabilidade de um certo evento resume-se ao cálculo de um somatório ou uma integral sobre a respectiva função de densidade. Contudo, diferente do que acontece na teoria de probabilidade onde a noção de distribuição de densidade de probabilidade é subjacente a de variável aleatória, a noção de distribuição de possibilidade requer apenas que $\mathcal{A} = \mathcal{P}(\Omega)$ e não necessita da definição de algo como “variável possibilística”. Vale também destacar que em probabilidade o conceito de distribuição de densidade de probabilidade se divide em dois casos, o discreto e o contínuo, porém, em possibilidade não há necessidade em dividir em casos.

5 Teoria de Possibilidade Via Probabilidade Superior

A noção de probabilidade superior aparece na literatura em diferentes contextos ligado a teoria clássica de probabilidades [4,16,3]. Seja \mathcal{M} uma coleção de medidas de probabilidade em um mesmo espaço mensurável (Ω, \mathcal{A}) , isto é, $\mathcal{M} \subseteq \mathbb{P}(\Omega, \mathcal{A})$, as funções $\bar{P}, \underline{P} : \mathcal{A} \rightarrow [0, 1]$ dadas por

$$\bar{P}(A) = \sup_{P \in \mathcal{M}} P(A) \quad \text{e} \quad \underline{P}(A) = \inf_{P \in \mathcal{M}} P(A), \quad \forall A \in \mathcal{A}, \quad (11)$$

são chamadas respectivamente de medidas (coerentes) de probabilidades superior e inferior (com respeito à \mathcal{M}). Alguns autores definem medida de probabilidade superior/inferior como sendo uma cota superior/inferior para \mathcal{M} e no caso particular onde esta corresponde a menor/maior das cotas superiores/inferiores, dadas na Equação (11), eles chamam a respectiva medida de “coerente” [6]. Portanto, neste texto, lidaremos apenas com medidas de probabilidade superior e inferior ditas coerentes por tais autores.

Outro ponto que vale apenas destacar é que a mesma medida de probabilidade superior/inferior pode ser gerada por diferentes coleções de medidas de probabilidade. Se \bar{P} e \underline{P} são medidas de probabilidades superior e inferior com respeito a uma coleção de medidas de probabilidade \mathcal{M} qualquer, então pode-se demonstrar que

$$\mathcal{M} \subseteq [\underline{P}, \bar{P}] := \{P \in \mathbb{P}(\Omega, \mathcal{A}) \mid \underline{P}(A) \leq P(A) \leq \bar{P}(A), \forall A \in \mathcal{A}\}.$$

Desse modo, segue que $[\underline{P}, \bar{P}]$ é a coleção maximal entre as coleções de medidas de probabilidade que produzem as mesmas medidas de probabilidade superior \bar{P} e inferior \underline{P} . A relação acima também é encontrada em uma área de pesquisa chamada probabilidade intervalar que, de maneira bem grosseira, destina-se a investigar pares de funções $P_1, P_2 : \mathcal{A} \rightarrow [0, 1]$ tais que $P_1(A) \leq P_2(A)$ para todo $A \in \mathcal{A}$. Uma formulação axiomática para medidas probabilidades intervalares pode ser encontrada em [18].

Note que a noção de probabilidade inferior é dual ao de probabilidade superior segundo a seguinte relação derivada da propriedade P6:

$$\underline{P}(A) = \inf_{P \in \mathcal{M}} P(A) = \inf_{P \in \mathcal{M}} 1 - P(A^c) = 1 - \sup_{P \in \mathcal{M}} P(A^c) = 1 - \bar{P}(A^c).$$

Assim, é suficiente focar apenas em uma delas e no nosso caso daremos atenção especial para a medida de probabilidade superior uma vez que medidas de possibilidade correspondem a casos particulares de medidas de probabilidade superior [11,10,6]. Seja Π uma medida de possibilidade em (Ω, \mathcal{A}) e seja \bar{P} a medida de probabilidade superior com respeito à $\mathcal{M}(\Pi) = \{P \in \mathbb{P}(\Omega, \mathcal{A}) \mid P(A) \leq \Pi(A) \forall A \in \mathcal{A}\}$, pode-se verificar que $\mathcal{M}(\Pi)$ é de fato não vazio e que $\Pi = \bar{P}$, onde \bar{P} denota a medida de probabilidade superior com respeito à $\mathcal{M}(\Pi)$.

Ao identificar uma medida de possibilidade com uma medida de probabilidade superior, podemos nos valer de toda teoria desenvolvida para probabilidade superior encontrada na área de probabilidade intervalar [4,3] ou mais geralmente na área de probabilidade imprecisa [6]. Por exemplo, podemos utilizar as noções de probabilidade condicional superior de um evento A dado outro evento B , dada como se segue para uma coleção de medidas de probabilidade \mathcal{M} , para definir possibilidade condicional. Mais precisamente, sejam $A, B \in \mathcal{A}$ e $\mathcal{M} \subseteq \mathbb{P}(\Omega, \mathcal{A})$, definimos a probabilidade condicional superior de A dado B por [5]

$$\bar{P}(A|B) = \sup_{P \in \mathcal{M}} P(A|B). \quad (12)$$

Se \mathcal{M} é tal que $\mathcal{M} = [\underline{P}, \overline{P}]$, então, podemos mostrar a seguinte igualdade

$$\overline{P}(A|B) = \frac{\overline{P}(A \cap B)}{\overline{P}(A \cap B) + \underline{P}(A^c \cap B)} = \frac{\overline{P}(A \cap B)}{\overline{P}(A \cap B) + 1 - \overline{P}(A \cup B^c)}. \quad (13)$$

Assim, temos que a possibilidade condicional de A dado B nesta abordagem é dada pela Equação (8) se considerarmos a coleção de medidas de probabilidade $\mathcal{M}(\Pi)$. Contudo, diferente do que argumentado na seção anterior onde a Equação (8) é obtida “macaqueando” as equações para probabilidade condicional clássica, nesta abordagem, a fórmula é obtida por considerar o limite superior de probabilidades condicionais, ou seja, por um processo de cálculo de um certo limite levando em conta uma certa família de medidas de probabilidade. Do ponto de vista matemático, esta abordagem é bem mais consistente pois define de maneira única a noção de condicional, evitando abordagens *ad hoc* tais como discutida na seção anterior.

Em teoria de probabilidades, se dois eventos $A, B \in \mathcal{A}$ de medidas não nulas são independentes, então, temos que $P(A) = P(A|B) = P(A|B^c)$ e $P(B) = P(B|A) = P(B|A^c)$. Baseado nessas observações, podemos definir independência entre dois eventos $A, B \in \mathcal{A}$ para uma medida de probabilidade superior como se segue [6]. Dois eventos $A, B \in \mathcal{A}$ são ditos independentes com respeito a uma medida de probabilidade superior $\overline{P} : \mathcal{A} \rightarrow [0, 1]$ se as seguintes igualdades $\overline{P}(A) = \overline{P}(A|B) = \overline{P}(A|B^c)$ e $\overline{P}(B) = \overline{P}(B|A) = \overline{P}(B|A^c)$ são satisfeitas.

6 Considerações Finais

Neste trabalho apresentamos de maneira muito sucinta três abordagens à teoria de possibilidade encontradas na literatura. Apesar delas apresentarem diversas coincidências não podemos deixar de atentar para o fato que cada uma delas tem estruturas próprias, como estruturas algébricas distintas. Por exemplo, a abordagem discutida na Seção 4 requer mensurabilidade no universo de discurso Ω enquanto na Seção 2 não. Além disso, se quisermos estabelecer algum tipo de equivalência entre as abordagens das Seções 2 e 4 devemos considerar a σ -álgebra $\mathcal{P}(\Omega)$. Mesmo as abordagens das Seções 4 e 5, que são desenvolvidas sob espaços mensuráveis, requerem desenvolvimentos bem diferentes, pois a Seção 4 considera um espaço mensurável equipada com medida não aditiva, que é a medida de possibilidade, enquanto que a Seção 5 considera uma família de espaços de probabilidades definidas sob o mesmo espaço mensurável.

Tal como indicamos no decorrer do texto, podemos obter interseções entre as três abordagens discutidas aqui, porém, isso exige certas particularizações tal como assumir que estamos lidando com a maior σ -álgebra $\mathcal{P}(\Omega)$ e com conjunto de medidas de probabilidade $\mathcal{M}(\Pi)$ tal como definida na Seção 5. Contudo, mesmo com todas essas particularizações, tais equivalências tornam-se mais difíceis de serem estabelecidas quando vamos caminhando para métodos de inferência nos quais envolve conceitos de condicional, independência, etc. Por isso, é importante ter sempre em mente as diferenças que surgem das distintas

abordagens, pois coisas que acontecem em uma poderão não acontecer na outra tal como é o caso da noção de não interatividade vista na Seção 2.

Referências

1. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems* **1**(1) (1978) 3–28
2. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning - i. *Information Sciences* **8** (1975) 199–249
3. De Campos, L.M., Huete, J.F., Moral, S.: Probability intervals: a tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **2**(02) (1994) 167–196
4. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. *The annals of mathematical statistics* (1967) 325–339
5. Tanaka, H., Sugihara, K., Maeda, Y.: Non-additive measures by interval probability functions. *Information Sciences* **164**(1) (2004) 209–227
6. Walley, P.: *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis (1991)
7. Zadeh, L.A.: Fuzzy sets. *Information and Control* **8**(3) (1965) 338–353
8. Dubois, D., Nguyen, H.T., Prade, H.: Possibility theory, probability and fuzzy sets misunderstandings, bridges and gaps. In: *Fundamentals of fuzzy sets*. Springer (2000) 343–438
9. Gaines, B.R., Kohout, L.: Possible automata. In: *Proc. Int. Symp. on Multiple-Valued Logic*, University of Indiana, Bloomington, IN (1975) 183–196
10. De Cooman, G., Aeyels, D.: On the coherence of supremum preserving upper previsions. In: *Proc. of the 6th Inter. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'96)*, Granada, Spain, July. (1996) 1–5
11. Dubois, D., Prade, H.: When upper probabilities are possibility measures. *Fuzzy sets and systems* **49**(1) (1992) 65–74
12. Shafer, G., et al.: *A mathematical theory of evidence*. Volume 1. Princeton university press Princeton (1976)
13. Carlsson, C., Fullér, R.: *Possibility for decision*. Springer (2011)
14. Klir, G.J., Yuan, B.: *Fuzzy Sets and Fuzzy Logic; Theory and Applications*. Prentice Hall, Upper Saddle River, N. Y. (1995)
15. Barros, L., Bassanezi, R.: *Tópicos de lógica fuzzy e biomatemática*. 3rd edn. G. P. Silveira, Campinas, SP, Brazil (2015)
16. Nguyen, H.T., Walker, E.A.: *A first course in fuzzy logic*. CRC press (2005)
17. Bouchon-Meunier, B., Coletti, G., Marsala, C.: Independence and possibilistic conditioning. *Annals of Mathematics and Artificial Intelligence* **35**(1-4) (2002) 107–123
18. Walley, P., Fine, T.L.: Towards a frequentist theory of upper and lower probability. *The Annals of Statistics* (1982) 741–761