Modelo probabilístico para uma variável fuzzy

M. Missio¹,

Matemática-Unidade de Dourados – UEMS, 78.060-900, Dourados, MS.

L. C. Barros²,
Depto. de Matemática Aplicada, IMECC, – UNICAMP,
13.083-970, Campinas, SP.

Resumo. Neste trabalho, propomos a integração de duas ferramentas matemáticas - a Teoria de Probabilidades e a Teoria de Conjuntos Fuzzy - no tratamento de incertezas. A integração entre as duas teorias é realizada por meio de um modelo probabilístico, elaborado para tratar incertezas de variabilidade presentes no domínio das funções de pertinência da variável de entrada de um Sistema Baseado em Regras Fuzzy. Para o modelo probabilístico, fazemos uso do Método de Monte Carlo associado à distribuição Weibull.

Palavras-chave: Modelo Fuzzy; Modelo Probabilístico; Método de Monte Carlo; Distribuição Weibull.

1. Introdução

Em epidemiologia, as incertezas podem ser agrupadas segundo duas classes: a variabilidade, originada da heterogeneidade da população ou da estocasticidade e a do conhecimento parcial, que resulta de erros sistemáticos de

 $^{^{1}}$ maristela@uems.br

²laeciocb@ime.inicamp.br

medida ou do desconhecimento de parte do processo considerado, conforme Massad et al. (2008). Portanto, ainda segundo os autores, as duas classes de incertezas devem ser tratadas por métodos diferentes e mais apropriados a cada uma. Segundo Barros e Bassanezi (2006), a incerteza proveniente da aleatoriedade de eventos está bem desenvolvida, ocupando um lugar de destaque na galeria da Matemática por meio da Teoria da Probabilidade. Porém, Massad et al. (2008) reportam que essa teoria não consegue, na maioria das vezes, abordar o problema da ignorância e da subjetividade. Nesses casos, a lógica fuzzy pode ser considerada como uma das ferramentas matemáticas mais poderosas para lidar com tais incertezas, imprecisões e verdades parciais.

Em geral, os modelos matemáticos que tratam de fenômenos epidemiológicos, incorporando incertezas em suas variáveis, utilizam a teoria da Probabilidade ou a Teoria de Conjuntos Fuzzy, a depender do tipo de incerteza. No entanto, ao desenvolver um Sistema Baseado em Regras Fuzzy (SBRF), a variável de entrada do sistema pode apresentar, além da incerteza proveniente de conhecimentos parciais, tratada pela lógica fuzzy, a incerteza de variabilidade no domínio das suas funções de pertinência. A primeira está presente na definição do "evento", e a segunda na chance de ocorrência de "seus elementos". Ou seja, cada termo (ou evento fuzzy) tem fronteira não clara. Por outro lado, "seus elementos" (suporte) não são equiprováveis segundo uma distribuição aleatória, de acordo com Missio (2008).

Nesse sentido, uma vez detectadas características de incertezas de variabilidade em uma variável fuzzy, faz-se necessário integrar ao modelo fuzzy um modelo estocástico, para tratar esse tipo de incerteza.

O uso integrado de modelos fuzzy e estocásticos no tratamento das duas classes de incertezas pode ser encontrado em Missio (2008). Nesse trabalho, realizou-se um estudo da evolução espaço-temporal da febre aftosa em bovinos por meio de um modelo compartimental tipo SIR (Suscetível, Infectado, Recuperado), envolvendo o sistema de Equações Diferenciais Parciais (EDP) (1). As variáveis de estado do sistema são representadas por

 $S=S(x,y;t),\ I=I(x,y;t)$ e R=R(x,y;t), com $(x,y)\in\Omega\subset\mathbb{R}^2$ e $t\in(0,T)$ para algum T>0 fixo.

$$\begin{cases}
\frac{\partial S}{\partial t} = \operatorname{div}(D_S(\mathcal{C}, \mathcal{P}).\nabla S) - \lambda(\beta, I)S \\
\frac{\partial I}{\partial t} = \operatorname{div}(D_I(\mathcal{C}, \mathcal{P}, \mathcal{FD}).\nabla I) + \lambda(\beta, I)S - \mu I \\
\frac{\partial R}{\partial t} = \operatorname{div}(D_R(\mathcal{C}, \mathcal{P}).\nabla R) + \mu I,
\end{cases} \tag{1}$$

satisfazendo as condições iniciais $S(x,y;0) = S_0$, $I(x,y;0) = I_0(x,y)$ e R(x,y;0) = 0.

Os parâmetros $D_S = D_S(\mathcal{C}; \mathcal{P})$, $D_I = D_I(\mathcal{C}, \mathcal{P}, \mathcal{FD})$ e $D_R = D_R(\mathcal{C}, \mathcal{P})$ representam, respectivamente, os coeficientes de difusão das populações de suscetíveis, infectados e recuperados. São dependentes da população (\mathcal{P}) , das características ambientais (\mathcal{C}) e da fase da doença (\mathcal{FD}) , sendo as duas últimas consideradas variáveis incertas. A força de infecção da febre aftosa é representada pelo parâmetro $\lambda = \beta I$, com β indicando a taxa de transmissão da doença. O parâmetro constante μ representa a taxa de passagem de infectado para recuperado.

No modelo, foram consideradas incertezas do ponto de vista epidemiológico, socioeconômico e ambiental, incorporadas pelos parâmetros de dispersão e de transmissão do sistema de EDP. Os parâmetros incertos foram estimados fazendo-se uso de modelos fuzzy, mais especificamente, SBRF. O acoplamento de modelos fuzzy e probabilísticos foi necessário, uma vez que, as incertezas eram oriundas da subjetividade e da aleatoriedade.

Neste estudo, com base em Missio (2008), propomos um modelo estocástico para tratar aspectos de incertezas probabilísticas presentes no domínio das funções de pertinência da variável de entrada de um SBRF.

Para o modelo probabilístico, fazemos uso de uma distribuição de probabilidade e do método de Monte Carlo, a fim de se escolher um valor de entrada da variável fuzzy para alimentar o SBRF. Este método gera uma amostra de valores aleatórios por meio de uma distribuição de probabilidade

devidamente escolhida. A média amostral obtida é utilizada como entrada no SBRF. Adotamos esse método por ser uma técnica simples e confiável quanto à geração de números aleatórios, com a possibilidade de repetir o experimento em um número considerável de vezes, visando minimizar o erro na estimativa.

2. Objetivos

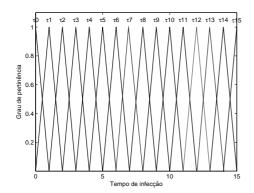
- 2.1 Desenvolver um modelo probabilístico para tratar incertezas de variabilidade presentes em uma variável fuzzy.
- 2.2 Fazer a integração entre SBRF e modelos probabilísticos, envolvendo o Método de Monte Carlo associado à função de distribuição Weibull.

3. Metodologia

Como motivação para realizar a integração da Teoria da Probabilidade com a Teoria de Conjuntos Fuzzy, utilizamos a variável fuzzy tempo de infecção da população, denotada por τ . Essa é a variável de entrada do SBRF elaborado por Missio (2008) para modelar a variável fuzzy fase da doença, denotada por \mathcal{FD} , de forma que $\mathcal{FD} = \mathcal{FD}(\tau)$.

A variável τ , conforme definida em Missio (2008), indica o período da infecção em que a população infectada pela febre aftosa se encontra. Esta variável divide o tempo máximo de aproximadamente 15 dias em que o indivíduo fica doente em pequenos intervalos, tornando possível determinar os diversos momentos que podem acontecer alterações dos sinais clínicos da população infectada. Ou seja, com relação aos sinais clínicos, a doença pode ser classificada qualitativamente em diferentes fases, dependendo do tempo de infecção. Por exemplo, no período entre 5 e 6 dias de infecção as lesões são bastantes severas, podendo-se dizer que a doença se apresenta numa fase aguda.

Neste sentido, Missio (2008) classificou qualitativamente a variável τ pelos termos lingüísticos tempo0 ($\tau 0$), tempo1 ($\tau 1$), tempo2 ($\tau 2$), tempo3 ($\tau 3$), ..., tempo14 ($\tau 14$) e tempo15 ($\tau 15$). A variável de saída \mathcal{FD} do SBRF é representada pelos conjuntos fuzzy inicial, moderada leve, moderada, moderada aguda e aguda. As funções de pertinência dos conjuntos fuzzy assumidos pelas variáveis τ e \mathcal{FD} estão ilustradas nas respectivas Figuras 1 e 2. O eixo horizontal das funções de τ varia entre 0 e 15 dias, tendo em vista que nesse intervalo de tempo ocorre a recuperação da população de infectados.



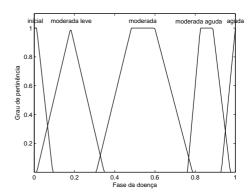


Figura 1: Funções de pertinência dos con-figura 2: Função de pertinência dos conjuntos fuzzy assumidos pela variável τ . juntos fuzzy assumidos pela variável \mathcal{FD}

Para compor a base de regras do SBRF, Missio (2008) fez a associação entre os termos subjetivos das variáveis τ e \mathcal{FD} , estabelecendo regras do tipo: Se $\tau = \tau 3$ então \mathcal{FD} é moderada leve. O SBRF utilizado para modelar a variável \mathcal{FD} é estruturado pelo esquema da Figura 3.

Ao realizarmos a leitura da variável lingüística τ , quando posto em ação o SBRF $\mathcal{FD} = \mathcal{FD}(\tau)$, temos a nítida característica de incertezas de natureza fuzzy e probabilística. O conjunto fuzzy tempo5 da população de infectados exemplifica mais precisamente um evento fuzzy, onde a incerteza fuzzy está presente na definição do evento (ver Figura 4). E para dar entrada ao SBRF é necessário obter a chance desses "eventos" ocorrerem, caracterizando uma incerteza de variabilidade (probabilística). Ou seja, os "eventos" não são

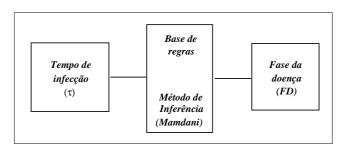


Figura 3: Estrutura do Sistema Baseados em Regras Fuzzy para modelar a variável \mathcal{FD} .

equiprováveis segundo uma distribuição aleatória.

Sendo assim, considerando o experimento "tempo de infecção da população", a chance de ocorrência de um dos eventos deste está intrinsecamente associada à probabilidade de que tal evento ocorra. Isto remete à idéia de que o tempo de infecção da população τ pode ser definido como uma variável aleatória. A Figura 4 ilustra o SBRF para modelar a variável \mathcal{FD} , destacando o experimento "tempo de infecção da população" e seus eventos "tempo0, tempo1, [...], tempo15", os quais envolvem incertezas que são tratadas por meio de ferramentas fuzzy e probabilísticas, respectivamente.

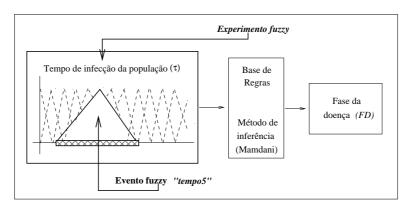


Figura 4: Esquema ilustrativo de um **experimento fuzzy** e seus **eventos**. Em destaque o evento fuzzy "tempo5".

Para qualquer simulação computacional de um sistema físico que envolva aleatoriedade, deve ser incluído um método para geração de sequências de números aleatórios. Um destes métodos amplamente utilizado em procedimentos de seleção amostral (para gerar aleatoriamente valores para variáveis incertas) é a simulação de Monte Carlo. Assim, no contexto desse artigo, o método consiste basicamente em gerar sucessivas amostras de valores aleatórios, de acordo com uma distribuição de probabilidade.

Levando-se em conta o fato relevante de que o tempo de infecção da população τ é limitado pelo intervalo [0,15], optou-se por uma função de distribuição de probabilidade específica que é usada extensivamente na análise de dados de vida: a distribuição Weibull truncada.

A distribuição Weibull, nomeada em 1939 pelo seu criador, o físico sueco Ernest Hjalmar Wallodi Weibull, é uma distribuição de probabilidade contínua e limitada, muito versátil para aplicações práticas e elimina os inconvenientes que outras distribuições podem provocar, dado que nunca ocorrem valores negativos para os tempos de vida. É uma generalização da distribuição exponencial usada em estudos de tempo de vida de equipamentos, análise de sobrevivência e em outras áreas, devido à sua versatilidade e simplicidade, segundo Dantas (2004).

Uma variável aleatória tem distribuição Weibull com parâmetros $\alpha>0$ e $\eta>0$ se sua densidade de probabilidade é descrita por:

$$f(v) = \begin{cases} \frac{\alpha}{\eta} \left(\frac{v}{\eta}\right)^{\alpha - 1} \exp\left\{-\left(\frac{v}{\eta}\right)^{\alpha}\right\}, & v \ge 0\\ 0, & \text{caso contrário} \end{cases}$$
 (2)

onde $\alpha>0$ é o parâmetro de forma, conhecido também como inclinação da distribuição Weibull, e $\eta>0$ é o parâmetro de escala da distribuição, e controla as dimensões que a curva assume. A função de distribuição é representada por

$$F(v) = 1 - \exp\left\{-\left(\frac{v}{\eta}\right)^{\alpha}\right\}. \tag{3}$$

O valor esperado do tempo de infecção da população τ , pela distribuição Weibull, é dado pela expressão:

$$E(\tau) = \eta \Gamma \left(1 + \frac{1}{\alpha} \right), \tag{4}$$

com $\Gamma(y) = \int_0^\infty v^{y-1} \exp\{-v\} dv$ representando a função gama.

A taxa de falha de um sistema cujo tempo de vida tem função de distribuição F(v), com densidade de probabilidade f(v), é dada por

$$h(v) = \frac{f(v)}{1 - F(v)}.$$

Essa taxa, que é uma função do tempo, é aproximadamente igual à probabilidade de que ocorra falha num intervalo de tempo $\triangle v$, após o instante v, dado que não houve falha até esse instante, conforme Dantas (2004).

No contexto aqui tratado, a expressão "taxa de falha" é vista de maneira positiva, indicando a taxa de sucesso em relação ao tempo em que o indivíduo deixa de ser doente. Assim, a partir de f(v) e de F(v), obtém-se a taxa de falha da distribuição Weibull:

$$h(v) = \frac{f(v)}{1 - F(v)} = \frac{\frac{\alpha}{\eta} \left(\frac{v}{\eta}\right)^{\alpha - 1} \exp\left\{-\left(\frac{v}{\eta}\right)^{\alpha}\right\}}{\exp\left\{-\left(\frac{v}{\eta}\right)^{\alpha}\right\}}.$$

Simplificando, tem-se

$$h(v) = \frac{\alpha}{\eta} \left(\frac{v}{\eta}\right)^{\alpha - 1}.$$
 (5)

A fim de estimar um valor razoável para o parâmetro de forma α , observa-se a partir do modelo (1), que $\frac{\partial R}{\partial t}\cong \mu I$. Como no início a população de infectados I cresce com o tempo t $(t\in[0,T])$, é de se esperar que $\frac{\partial R}{\partial t}$ também cresça com t. Consequentemente, a taxa de falha (em que há recuperação da doença) cresce com o tempo. Ou seja, $h(t)=\frac{\alpha}{\eta}\left(\frac{t}{\eta}\right)^{\alpha-1}$ é crescente. Portanto $\alpha\geq 1$.

Considerando que no início do processo da epidemia a população de recuperados, mensurada pela função R(t), apresenta uma pequena concavidade para cima (ver Capítulo 7 em Missio (2008)), adotamos $\frac{\partial R}{\partial t} \cong t$. Assim, partindo do pressuposto de que a taxa de falha deve ter um comportamento semelhante à curva dos recuperados, que cresce com o tempo t no início do processo de evolução da epidemia, admite-se que a taxa de falha pode ser aproximada por uma função linear da forma h(t) = lt. Então, de (5) tem-se $\alpha - 1 = 1$ e, portanto, $\alpha = 2$.

Com isso, o parâmetro de forma, indicando a inclinação da distribuição Weibull, é estimado em $\alpha=1$ quando t=1, e $\alpha=2$ quando t>1. Desta forma, tem-se que a geração de valores aleatórios segue uma distribuição exponencial no primeiro dia de evolução do sistema, ou seja, existe uma probabilidade maior de ocorrer valores aleatórios mais próximos de zero. Por outro lado, quando t>1, há maior probabilidade de ocorrer a geração de valores aleatórios mais próximos do limite superior de τ .

Lembrando que a variável tempo de infecção da população τ é definida no intervalo de 0 a 15 dias, τ assume um conjunto limitado. Desta forma, utilizando-se o Método de Monte Carlo, faz-se a geração de valores aleatórios v, por meio da distribuição uniforme e, em seguida, ocorre a geração de valores τ por meio de uma Weibull truncada, com base na seguinte transformação:

$$\tau = \eta \left(-\log \left((1 - v)e^{-\left(\frac{a}{\eta}\right)^{\alpha}} + ve^{-\left(\frac{b}{\eta}\right)^{\alpha}} \right) \right)^{1/\alpha}. \tag{6}$$

Os parâmetros a e b indicam, respectivamente, os limites inferior e superior da variação de $\tau \in [0,15]$. Assim, estimamos a=0, pois 0 é o limite inferior de τ . Por outro lado, considerando o tempo (em dias) de evolução do sistema representado por $t \in [0,15]$, ao estimar o parâmetro b é natural ter a variável tempo de infecção da população $\tau < 1$ no primeiro dia (t=1); $\tau < 2$ no segundo dia (t=2) e, assim sucessivamente até $\tau \leq 15$ no décimo quinto dia (t=15). O tempo de infecção de cada individuo é sempre limitado superiormente por $\tau = 15$, tendo em vista que os animais se recuperam da

doença nesse período. Face a isso, τ pode assumir qualquer valor inferior ou igual a 15. Assim, adotamos o parâmetro b=t, para $t\in[0,15]$.

Considerando que o tempo médio de doença é em torno de 10 dias e que η cresce com b, sendo b o valor máximo a ser assumido por τ , o valor esperado do tempo de infecção da população τ é estimado em $E(\tau) = 10 \left(\frac{b}{15}\right)$. Então

$$\eta\Gamma\left(1+\frac{1}{\alpha}\right) = 10\left(\frac{b}{15}\right).$$

Essa expressão possibilita estimar o parâmetro de escala de maneira que a distribuição esteja truncada em b. Ou seja, adotando-se η dessa forma, tem-se a probabilidade nula de estimar $\tau > t$.

4. Resultados

A fim de aferir a proposta de interligação da Teoria de Conjuntos Fuzzy e Teoria de Probabilidade realizamos, o acoplamento dos modelos probabilístico e fuzzy. Para tal, construímos um algoritmo, utilizando o *Matlab* - versão 7.0, que possibilita simular o Método de Monte Carlo, conectando-o com o Sistema Baseado em Regras Fuzzy, modelado no *Fuzzy Logic Toolbox* contido no ambiente *Matlab*.

Considerando que o tempo de infecção da população τ é limitado pelo intervalo [0,15] dias, o valor de entrada da variável τ para o SBRF é estimado diariamente a partir de t=0 ($t\in[0,15]$) de evolução da epidemia, pelo Método de Monte Carlo, por meio do seguinte algoritmo.

- 1. Gera-se uma amostra de m valores aleatórios, através da distribuição uniforme;
- 2. Faz-se uma transformação desta amostra, por meio da função de distribuição Weibull truncada (6), para se obter uma nova amostra de τ no intervalo [0,t], considerando os valores dos parâmetros a,b,α e η , a cada unidade de tempo t, conforme descrição anterior;

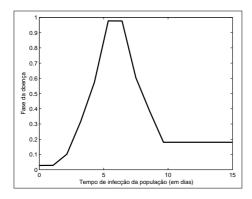
3. Finalmente, o valor de entrada para alimentar o SBRF é dado pela média amostral $\bar{\tau}$ da nova amostra obtida em 2.

O tamanho da amostra (m) gerada tem grande importância na determinação da qualidade dos resultados, uma vez que, quanto maior for a amostra, mais a distribuição amostral se aproxima das funções originais. Porém, se o processo requerer um tempo elevado de execução, poderemos fazer uso do Teorema Central do Limite, aproximando a distribuição amostral média pela distribuição normal.

Cada valor $\overline{\tau}$ representando o tempo de infecção da população, que é a média dos valores aleatórios simulados, obtido pelo Método de Monte Carlo, é enviado para o controlador fuzzy. Desta forma, a variável fase da doença \mathcal{FD} é inferida a partir da base de regras estabelecidas em Missio (2008). Utilizamos o método de inferência de Mamdani; este associa a t-norma min para o conectivo lógico "e" e a t-conorma max para o conectivo "ou", segundo Barros e Bassanezi (2006). A resposta clássica (não fuzzy) é obtida através da defuzzificação, utilizando-se o Método do Centro de Massa, que é o método mais utilizado, de acordo com Missio (2008). A Figura 5 apresenta a curva obtida pelo mapeamento realizado pelo modelo fuzzy, mais especificamente pelo SBRF, integrado ao modelo probabilístico.

Ao observar a curva que descreve a variável \mathcal{FD} , obtida pelo SBRF integrado ao modelo probabilístico e ilustrada na Figura 5, pode-se intuir que existe uma aproximação dessa com a curva que representa a evolução teórica da febre aftosa em um indivíduo infectado, representada pela curva do virus na Figura 6.

É sabido que para comparar duas curvas, deve-se conhecê-las em todos os pontos, e aqui está se supondo que a função que representa a curva da evolução teórica da febre aftosa em um indivíduo infectado não é conhecida. Todavia, Dias (2006) mostra que um SBRF tem a capacidade de aproximar funções contínuas, ou seja, que toda função contínua definida em um intervalo fechado pode ser aproximada por um sistema baseado em regras fuzzy à medida que o número de regras aumente indefinidamente.



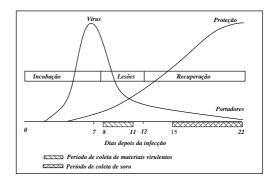


Figura 5: Curva obtida pelo mapeamento do modelo fuzzy integrado ao modelo probabilístico para a variável \mathcal{FD} .

Figura 6: Evolução teórica da febre aftosa em um indivíduo infectado, extraída de Fernãndez e Söndahl (1978).

5. Conclusões

Considerando a aproximação entre as curvas, ilustradas nas Figuras 5 e 6, suficientemente satisfatória, conclui-se que o SBRF integrado ao modelo probabilístico, por meio do Método de Monte Carlo via função de distribuição Weibull, demonstrou possuir boa capacidade para determinar a variável fase da doença dos indivíduos infectados, com base em características de evolução do vírus e das alterações dos sinais clínicos dos animais infectados.

Adotamos o Método de Monte Carlo no modelo probabilístico, por se caracterizar como uma técnica simples e confiável em gerar números aleatórios, com a possibilidade de repetir o experimento em um número considerável de vezes para minimizar o erro na estimativa. Para cada variável incerta do modelo (isto é, que tem uma faixa de valores possíveis), foi preciso optar por uma distribuição de probabilidade, e esta foi a principal dificuldade de utilização desta técnica. No entanto, ao selecionar aleatoriamente valores de forma independente, o Método de Monte Carlo contribui para uma visão também probabilística de uma variável de entrada fuzzy para um SBRF.

Portanto, mesmo percebendo diferenças significativas entre as teorias fuzzy e probabilística, pode-se comprovar a existência de grandes possibi-

lidades delas se complementarem. Esta junção resulta em uma ferramenta realmente útil a ser aplicada em modelos similares ao utilizado nesse trabalho.

Referências

- Barros, L. C. e Bassanezi, R. (2006). *Tópicos de Lógica Fuzzy e Biomatemática*, volume 5. Coleção IMECC Textos didáticos, Campinas. UNICAMP IMECC.
- Dantas, C. A. B. (2004). *Probabilidade: Um Curso Introdutório.*, volume 5 of *2.ed.* Editora da Universidade de São Paulo, São Paulo.
- Dias, M. (2006). Equações diferenciais ordinárias com campo de direções parcialmente conhecido. Dissertação de Mestrado, IMECC, Unicamp, Campinas, SP.
- Fernandez, A. A. e Söndahl, M. (1978). Diagnóstico das enfermidades vesiculares e classificação do vírus da febre aftosa, volume 5. Centro Pan-Americano de Febre Aftosa, Rio de Janeiro, RJ.
- Massad, E., Ortega, N., Barros, L., e Struchiner, C. (2008). Fuzzy Logic in Action: Applications in Epidemiology and Beyond. Ed. Springer, Berlin, Heidelberg,.
- Missio, M. (2008). Modelos de EDP integrados à Lógica Fuzzy e Métodos Probabilísticos no tratamento de incertezas: uma aplicação à febre aftosa em bovinos. Tese de Doutorado, IMECC, Unicamp, Campinas, SP.