Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 7, n. 1, 2020.

Trabalho apresentado no XXXIX CNMAC, Uberlândia - MG, 2019.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Estimativa da Condição Inicial Fuzzy para Modelos de Biomatemática via Métodos Estatísticos

Michael Macedo Diniz [1]
Instituto Federal de São Paulo - IFSP, Campus São José dos Campos

Resumo. O valor da condição inicial de um modelo normalmente está sujeito a fatores de incerteza, e portanto é adequado fazer uso de um número fuzzy para representá-lo. O objetivo deste trabalho é apresentar uma estratégia de modelagem da condição inicial fuzzy a partir de estimadores estatísticos. Para isso, vamos utilizar os estimadores oriundos do método dos quadrados e o estimador de Lincoln-Petersen para calcular intervalos de confiança para o tamanho populacional, com esses intervalos será determinado o conjunto fuzzy que representa a condição inicial. Como resultado, buscamos delinear um método que permita modelar condições iniciais fuzzy a partir de estimativas estatísticas da condição inicial determinística do modelo. Por fim, serão feitas simulações computacionais em um caso fictício para ilustrar a aplicação do método proposto.

Palavras-chave. Condição Inicial Fuzzy, Estimadores de Lincoln-Petersen, Método dos Quadrados

1 Introdução

Em situações reais, determinar a condição inicial e os parâmetros de um modelo nem sempre é uma tarefa simples. No caso específico da condição inicial de modelos de dinâmica populacional, a dificuldade normalmente está associada ao tamanho da população, ou à dificuldade de acesso ao habitat em que vive. Sendo assim, no uso prático deste modelos qualquer estimativa para condição inicial possui componentes de incerteza e aleatoriedade que precisam ser consideradas.

De acordo com Seber [10], o primeiro trabalho voltado para a estimativa do tamanho de uma população foi redigido por Pierre Simon Laplace em 1783 e intitulado por *Sur les naissances, les mariages, et les morts*. Neste trabalho Laplace buscou estimar a população francesa a partir de dados relacionados ao número de mortes e nascimentos num determinado período. Segundo Siqueira [7], em 1896, Carl G. J. Petersen apresentou um método similar ao de Laplace para estudar a população de peixes no Mar Báltico, e em 1930, Frederick Lincoln utilizou este mesmo método para estimar a densidade populacional de patos selvagens da América do Norte. O estimador proposto nos trabalhos citados é conhecido hoje como *Estimador de Lincoln-Petersen*, atuando como pilar principal dos

[1] michael.diniz@ifsp.edu.br

chamados *métodos de Captura-recaptura*, utilizados para estimar a densidade populacional em diversos contextos. Maiores detalhes sobre estes métodos podem ser encontrados em [6, 12].

Proposta por Lotfi A. Zadeh em 1965 [13], a teoria dos conjuntos fuzzy busca representar matematicamente grandezas incertas. Esta ideia já foi extensivamente adotada para representar a condição inicial em sistemas dinâmicos, por exemplo em [3, 11], ou especificamente para tratar da condição inicial de modelos voltados a fenômenos biológicos [1].

Normalmente, os trabalhos que envolvem condição inicial fuzzy não abordam detalhadamente como este parâmetro pode ser obtido em situações práticas. Sendo assim, o objetivo principal deste trabalho é apresentar um método para estimar a condição inicial fuzzy de um modelo de dinâmica populacional, buscando assim aproximar ainda mais os modelos teóricos da literatura à situações práticas da realidade.

2 Metodologia

2.1 Métodos estatísticos para estimativa do tamanho de uma população

Segundo Malajovich [8], na prática podemos utilizar dois métodos básicos para estimar o tamanho de uma população, o método da Captura-recaptura e o método dos quadrados. O primeiro é adequado para populações de grande mobilidade, já o segundo, é utilizado quando a população de interesse tem pouca mobilidade, como plantas ou larvas.

2.1.1 Método da Captura-recaptura

O tipo mais básico de método de Captura-recaptura pode ser descrito da seguinte forma: inicialmente é feita uma captura aleatória de n_1 elementos da população, estes elementos são marcados e devolvidos ao ambiente em que estavam, num segundo momento é realizada uma segunda captura de n_2 elementos e são identificados os m_2 elementos que foram capturados duas vezes, com isso, uma estimativa do tamanho N desta população pode ser dada pelo estimador de Lincoln-Petersen

$$\hat{N} = \frac{n_1.n_2}{m_2}. \tag{1}$$

Chapman [4] mostra que a distribuição do estimador de Lincoln-Petersen é assintoticamente normal, e portanto, para grandes amostras ($n_1 > 30 \quad \text{e} \quad n_2 > 30$), considera-se o intervalo de $100(1-\gamma)\%$ de confiança para N dado por

$$\hat{N} \pm z_{\frac{\gamma}{2}} S$$

onde $z_{\frac{\gamma}{2}}$ é o valor de $Z \sim N(0,1)$ que tem probabilidade acumulada de $1 - \frac{\gamma}{2}$ e S é o estimador para o desvio padrão de $\hat{N}$ [9].

Ainda segundo Sadinle [9], para contornar o viés do estimador de Lincoln-Petersen e do estimador da variância, o intervalo de confiança para população pode ser estimado de forma mais precisa por

DOI: 10.5540/03.2020.007.01.0378

$$\frac{(n_1+1)(n_2+1)}{m_2+0.5} - 1.5 \pm z_{\frac{\gamma}{2}} \sqrt{\frac{(n_1+1)(n_2+1)(n_{12}+0.5)(n_{21}+0.5)}{(m_2+0,5)^3}} \tag{2}$$

onde n_{12} é o número de elementos que foram capturados na primeira captura e não foram na segunda, e n_{21} é o número de elementos que não foram capturados na primeira captura, mas foram na segunda.

2.1.2 Método dos quadrados

Este método é adequado para se determinar o número N de indivíduos de uma população estática, isto é, os elementos não se locomovem ao longo do ambiente em que vivem. Neste método, divide-se a região de interesse em m subregiões de mesmo tamanho, nas quais seja possível fazer a contagem da população. Posteriormente é definida uma amostra aleatória de n subregiões e contado o número de indivíduos em cada uma delas. Com isso, é possível obter um intervalo de confiança para média μ do tamanho da população em cada sub-região, este intervalo é dado por

$$\overline{x} \pm \frac{z_{\frac{\gamma}{2}} \hat{S}}{\sqrt{n}}. \tag{3}$$

onde $\overline{x}$ é a média amostral. Por fim, como o número de subregiões é conhecido, basta multiplicar os extremos do intervalo de confiança para média pelo número de subregiões, e assim é obtido um intervalo de confiança para o tamanho total da população.

2.2 Intervalo de confiança fuzzy

Seja $\tilde{u}$ um subconjunto fuzzy de $\mathbb{R}$ e $\mu_{\hat{u}} : \mathbb{R} \to [0,1]$ sua *função de pertinência*. O *suporte* de $\hat{u}$ é o conjunto clássico $\operatorname{supp} \hat{u} = \{x \in \mathbb{R} : \mu_{\hat{u}}(x) > 0\}$. Os α-*níveis* são definidos por $[\hat{u}]^\alpha = \{x \in \mathbb{R} \,|\, \mu_{\hat{u}}(x) \geq \alpha\}$ para todo $\alpha \in (0,1]$ e $[\hat{u}]^0 = \overline{\operatorname{supp} \hat{u}}$.

Definição 2.1. *(Ver [1].) Um subconjunto fuzzy $\hat{u}$ é um* número fuzzy *se $\mu_{\hat{u}} : \mathbb{R} \to [0,1]$ e*

1. *todos α-níveis de $\hat{u}$ são conjuntos não vazios, com $0 \leq \alpha \leq 1$;*
2. *todos α-níveis de $\hat{u}$ são intervalos fechados de $\mathbb{R}$;*
3. *$\operatorname{supp} \hat{u} = \{x \in \mathbb{R} : \mu_{\hat{u}}(x) > 0\}$ é limitado.*

Neste texto trabalharemos com métodos de estimação fuzzy propostos em [2] e que serão descritos a seguir. Considerando X uma variável aleatória com função de densidade de probabilidade $f(x;\theta)$, com θ sendo um parâmetro da função. Seja $Y = g(X_1, ..., X_n)$ uma estatística usada para estimar θ a partir de uma amostra aleatória $X_1, ..., X_n$. A partir da distribuição de Y calculamos um intervalo de confiança para θ com nível de confiança de $1-\gamma$, que chamaremos simplesmente de $100(1-\gamma)\%$ intervalo de confiança.

Buckley [2] propõe encontrar o $100(1-\gamma)\%$ intervalo de confiança para todo $\epsilon \leq \gamma < 1$, onde ϵ é um número muito próximo de zero. Cada um desses intervalos pode ser denotado por

$$[\theta_1(\gamma), \theta_2(\gamma)] \tag{4}$$

Consideraremos o intervalo $[\theta^*, \theta^*]$ para $\gamma = 1$, onde θ^* é a estimativa pontual para o parâmetro θ, sendo assim, temos intervalos para $0,01 \leq \gamma \leq 1$. Estes intervalos definem o conjunto fuzzy $\widetilde{\theta}$ através de seus α-níveis da seguinte forma

$$[\widetilde{\theta}]^\alpha = [\theta_1(\alpha), \theta_2(\alpha)] \tag{5}$$

Para $0 \leq \alpha < \epsilon$, definimos $[\widetilde{\theta}]^\alpha = [\theta_1(0,01), \theta_2(0,01)]$.

Com isso, a partir dos intervalos de confiança para um parâmetro θ construímos um conjunto fuzzy que representa o "empilhamento" desses intervalos.

3 Resultados

Para estimarmos a condição inicial fuzzy de uma determinada população utilizaremos o intervalo de confiança fuzzy em conjunto com os estimadores estatísticos provenientes do método de Lincoln-Petersen e do método dos quadrados. Ilustraremos este resultado a partir de simulações computacionais de uma situação hipotética:

Problema: Suponha um habitat de $10000m^2$ onde vive uma determinada espécie de animal que se alimenta exclusivamente de um único tipo de planta presente neste habitat. Busca-se desenvolver um modelo para analisar a dinâmica populacional entre o animal e a planta. Para isso, é necessário saber o tamanho da população de animais e da quantidade de plantas na região, porém, determinar essas informações por simples contagem é inviável pelo tamanho das populações e pelo tamanho da região em estudo. Como podemos estimar estes valores e conhecer por completo as possibilidades de precisão e confiabilidade dessa estimativa ?

3.1 Estimativa da quantidade de plantas

No caso da quantidade de plantas é conveniente usar o método dos quadrados, pois as plantas não se deslocam dentro da região, para isso, a região de $10000m^2$ será dividida em 1000 sub-regiões de $10m^2$. Para fins de simulação, associaremos para cada sub-região uma quantidade de plantas que varia aleatoriamente entre 0 e 50, com isso, conhecemos a priori o tamanho real da população, que neste caso é 25126 plantas. Iremos verificar se o método dos quadrados, em conjunto com a proposta deste artigo faz uma estimativa razoável deste valor.

Foi sorteada aleatoriamente uma amostra de 50 regiões. A média amostral é de 26,04 plantas por região e o desvio padrão amostral de 14,43 plantas. Com isso, um intervalo com 95% de confiança para média de plantas em cada região e um intervalo com 95% de confiança para o tamanho da população é dado por [22,04 , 30,04] e [22039 , 30040] respectivamente.

Repetindo o cálculo do $100(1-\gamma)\%$ intervalo de confiança para $0,01 < \gamma \leq 1$, e considerando cada um destes intervalos como sendo o α-nível do conjunto fuzzy, obtemos o resultado apresentado na Figura 2:

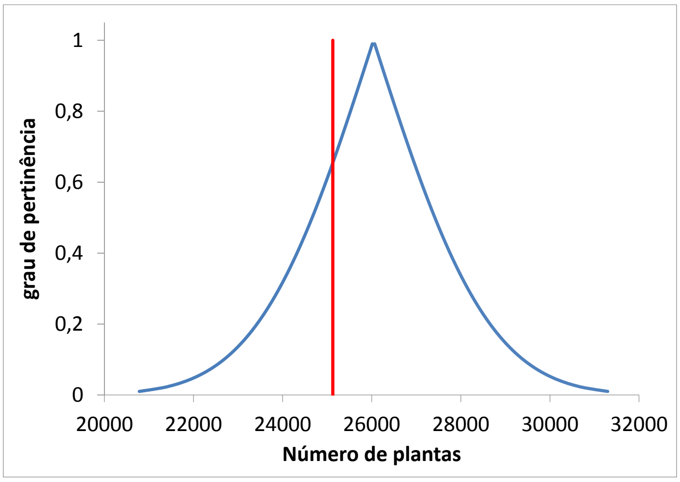

Figura 1: Estimativa fuzzy para condição o tamanho da população (condição inicial) de plantas. A linha vertical representa o N da população simulada.

O número fuzzy expresso na Figura 1 resume a relação entre precisão e confiabilidade para a estimativa do tamanho populacional. A medida que o α-nível aumenta, aumenta também a precisão e diminui a confiança da estimativa.

3.2 Estimativa da quantidade inicial de animais

Como os animais usualmente transitam no habitat em que vivem não é adequado utilizar o método dos quadrados, para isso, utilizaremos o método de Lincoln-Petersen apresentado na Subseção 2.1.1.

Iremos trabalhar com uma população de 500 animais, este número é o que desejamos estimar com o método. Inicialmente, simulamos a captura de 50 elementos da espécie, cada um deles foi marcado e foi feita uma nova captura de 50 animais. Verificou-se que o número de animais capturados nos dois processos de captura foi 6. Sendo assim, o estimador de Lincoln-Petersen é de aproximadamente 417 animais. Um intervalo com 95% de confiança, com base na equação 2, é $[94, 703]$.

Repetindo o cálculo do $100(1-\gamma)\%$ intervalo de confiança para $0,01 < \gamma \leq 1$, e considerando cada um destes intervalos como sendo o α-nível do conjunto fuzzy, obtemos o seguinte resultado:

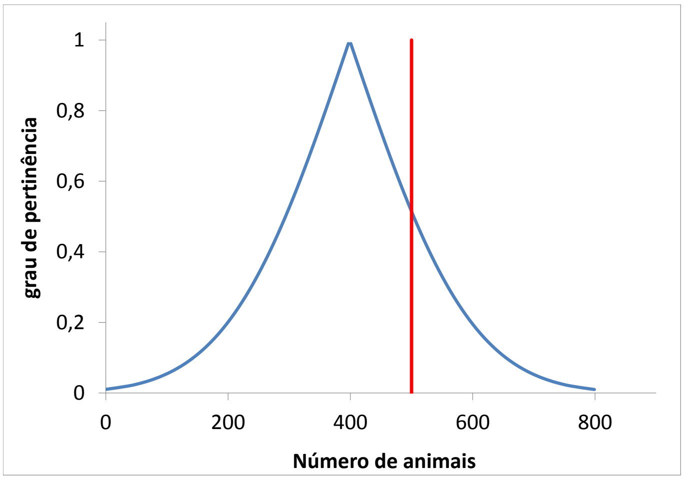

Figura 2: Estimativa fuzzy para condição o tamanho da população (condição inicial) de animais. A linha vertical representa o N da população simulada.

A estimativa para este caso é mais imprecisa, isso é devido ao estimador que estamos utilizando.

Ressaltamos que em ambos os exemplos, o conjunto fuzzy obtido não é uma representação linguística da condição inicial e sim, uma representação generalizada da estimativa amostral intervalar. Por exemplo, se 28000 tem grau de pertinência 0,4 ao conjunto fuzzy condição inicial, significa que 28000 pertencerá a todo intervalo com $(1 - 0{,}4)$ graus confiança. Claramente, esta estimativa fuzzy será diferente a cada nova amostra coletada.

4 Conclusões

Apresentamos neste trabalho uma alternativa para estimar a condição inicial fuzzy de modelos de biomatemática. No texto trabalhamos com o auxílio do estimador oriundo do método dos quadrados e do estimador de Lincoln-Petersen, entretanto, salientamos que a mesma estratégia pode ser aplicada fazendo uso de outros estimadores que porventura forem mais convenientes.

Referências

[1] L.C. Barros, R.C. Bassanezi, W.A. Lodwick, *A first Course in Fuzzy Logic, Fuzzy Dynamical Systems, and Biomathematics*. Springer-Verlag Berlin Heidelberg, 2016 .DOI:10.1007/978-3-662-53324-6

[2] J. J. Buckley. *Fuzzy probabilities: new approach and applications*. Springer-Verlag , Heidelberg, 2005.

[3] M. S. Cecconello et al. On the stability of fuzzy dynamical systems. *Fuzzy Sets and Systems*, 248:106-121,2014.

[4] C.J. Chapman. Some properties of the hypergeometric distibuition with applications to zoological censuses. *U California Public Stat*, 1:131-160, 1951.

[5] L. T. Gomes, L. C. de Barros, and B. Bede. *Fuzzy differential equation in various approaches*. In *SpringerBriefs in Mathematics*. SBMAC- Springer, 2015. ISSN: 2191-8198.

[6] E. B. Hook, R. R. Reagal. Capture-recapture methods in epidemiology: Methods and limitations. *Epidemiologic Reviews*, 17:243-264, 1995.

[7] E. L. Siqueira Jr., B. S. Lyra, C. M. L. Ferreira. Aplicação do método de captura-recaptura para estimar o rendimento do professor em sala de aula. In: Simpósio nacional de probabilidade e estatística, 19, 2010, Campinas. Resumos. Campinas: Unicamp, 2010.

[8] M. A. Malajovich. *Biologia e Meio Ambiente - 20 atividades selecionadas*. BTeduc. 2015.

[9] M. Sadinle. Transformed Logit Confidence Intervals for Small Populations in Single Capture–Recapture Estimation. *Communications in Statistics-Simulation and Computation*, 38:1909-1924, 2009.

[10] G. A. G. Seber. *The estimation of animal abundance and related parameters*. 2nd edition. London, England: Charles Griffin, 1982.

[11] Seikkala, S. On the fuzzy initial value problem. *Fuzzy Sets and Systems* 24:319-330, 1987.

[12] K. Tilling. Capture-recapture methods - useful or misleading. *International Journal of Epidemiology*. Internation Epidemiologic Association 30:12-14, 2001.

[13] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338-353, 1965.

DOI: 10.5540/03.2020.007.01.0378