

# Modelagem matemática para o crescimento de tumores e uma breve análise da evolução do câncer de próstata

Natália F. Reis,<sup>1</sup> Laércio L. Vendite,<sup>2</sup>  
DMA, IMECC – Unicamp, 13.083-859, Campinas/SP.

**Resumo.** O câncer é uma das principais causas de morte em todo o mundo. Composto por uma variedade de doenças, sua complexa estrutura é objeto de muitos estudos e pesquisas atuais. A modelagem matemática é uma ferramenta útil para descrever o crescimento tumoral, com base em um rico acervo literário. Neste trabalho, utilizamos o modelo matemático clássico de Gompertz para descrever e comparar o crescimento tumoral através de dois modelos, com variações de parâmetros retirados da literatura. Em seguida, usamos a Lei Universal de Crescimento para ajustar os dados relacionados à evolução do câncer de próstata a um polinômio de grau 3, e validamos o modelo através do método *leave-one-out*. As simulações apresentadas foram realizadas no *software Matlab* e na linguagem de programação *Octave*.

**Palavras-chave:** Câncer de próstata; modelagem matemática; modelo de Gompertz.

## 1. Introdução

O câncer é um conjunto de mais de 100 doenças malignas, cujas características em comum são o crescimento desordenado das células que podem invadir tecidos adjacentes no organismo. Suas causas podem estar ligadas aos desequilíbrios do meio ambiente, exposição à radiação, hábitos pessoais e ao próprio envelhecimento (Naoum e Naoum, 2016). Ainda de acordo com Naoum e Naoum (2016), o adenocarcinoma de próstata, que é um dos tipos mais

---

<sup>1</sup>n262894@dac.unicamp.br

<sup>2</sup>lvendite@unicamp.br

comuns de câncer no mundo, demora cerca de 20 a 30 anos para aparecer no organismo e leva cerca de 3 a 15 anos para se transformar na doença. As diversas evoluções do câncer no organismo do ser humano podem ser descritas através da modelagem matemática.

O modelo matemático de Gompertz (1825) foi originalmente desenvolvido para descrever o crescimento tumoral em condições controladas. Sua função correspondente tem formato sigmoidal de crescimento, caracterizado como lento no início e no final de determinado período de tempo.

A lei universal de crescimento proposta por West (2018), descreve o crescimento de organismos vivos através de escalas, considerando fatores como a nutrição, a temperatura e a genética.

O foco deste trabalho é estudar a evolução tumoral e analisar estatisticamente um modelo ajustado a um restrito conjunto de dados, verificando se é um bom preditor para o modelo.

A metodologia utilizada consiste em testar duas situações diferentes na alteração de parâmetros do modelo de Gompertz, com condições iniciais e taxas de crescimento variadas, para avaliar seus desenvolvimentos até atingirem os limites máximos suportáveis. Além disso, utiliza-se o conjunto de dados obtidos de Guiot et al. (2003) e Yorke et al. (1993), para analisar 11 pontos que representam a evolução do carcinoma de próstata no tempo. Os dados serão ajustados a um polinômio de grau 3 e, em seguida, aplicaremos o método de validação cruzada *leave-one-out*. Para esta técnica, divide-se o conjunto de dados em conjuntos de treinamento e teste. Ao final, verifica-se o RMSE (erro quadrático médio) da análise.

O objetivo do manuscrito é comparar e analisar a dinâmica do crescimento tumoral através de dois modelos diferentes: o crescimento sigmoidal assimétrico, previamente modelado por Gompertz e o comportamento dos dados na evolução do câncer de próstata, que inicialmente foi modelada pela lei universal de crescimento e em seguida, estudada estatisticamente.

Utiliza-se como base teórica, modelos utilizados na evolução do crescimento tumoral e cálculos estatísticos que auxiliam na predição de modelos matemáticos com quantidades restrita de dados.

Para realizar as simulações computacionais utilizaremos o *software Matlab* (Quarteroni e Saleri, 2007) e a linguagem de programação *Octave* (Quarteroni e Saleri, 2007), que fornecerão os resultados e gráficos para a modelagem.

Este trabalho está organizado em quatro seções: a seção 2 está dividida

em cinco subseções, onde a primeira descreve de forma geral a modelagem matemática, a segunda retrata o modelo matemático clássico de Gompertz, a terceira fala sobre a Lei Universal de Crescimento no contexto de crescimento tumoral e a quarta e quinta focam sobre a validação cruzada *leave-one-out*. Na seção 3 os resultados obtidos são relatados e na seção 4 é feita a conclusão do estudo.

## 2. Fundamentos matemáticos

Este estudo está fundamentado nos três tópicos a seguir: revisão de modelos matemáticos, análises e comparações entre simulações que descrevem o comportamento de tumores e método de validação para prever a evolução do câncer de próstata em determinado período de tempo.

### 2.1. Modelagem matemática

A modelagem matemática, que se concentra no desenvolvimento de modelos, é considerada uma arte em si mesma, de acordo com Bassanezi (1999). Muito do que foi produzido na matemática tem sido direcionado para modelos e teorias, sendo justificado por suas aplicações.

A dinâmica das populações está ligada às variações numéricas ao longo do tempo, como o tamanho da população, a taxa de sobrevivência, a taxa de mortalidade e a taxa de reprodução, que são de destaque pela sua importância na descrição característica do meio.

### 2.2. Modelo de Gompertz

No ano de 1825, Benjamin Gompertz desenvolveu um modelo para o crescimento de uma população de determinada espécie, o qual se mostrou bastante eficaz quando comparado com os dados reais da mesma população (Gompertz, 1825).

Atualmente, este modelo é bastante utilizado no estudo do crescimento de células tumorais. É um modelo de séries temporais em que o crescimento é menor no começo e no fim do intervalo de tempo. A equação diferencial correspondente é:

$$\frac{dN}{dt} = rN(t)\ln\left(\frac{k}{N(t)}\right)$$

em que,  $N(t)$  é a população tumoral,  $r$  é a taxa de crescimento intrínseco das células ( $r > 0$ ) e  $k$  é a capacidade máxima que o tumor pode atingir com nutrientes disponíveis.

Sua solução é definida como:

$$N(t) = Ke^{-e^{rt} \ln\left(\frac{n_0}{k}\right)}$$

Com o desenvolvimento da angiogênese,  $N(t)$  aumenta cada vez mais, de modo a escrevermos:

$$\lim_{t \rightarrow \infty} N(t) = k$$

Com isso, é fácil verificar de maneira simples que o ponto de máximo global é  $N = \frac{k}{e}$  e seu valor máximo é alcançado quando:  $V_{max} = r \frac{k}{e}$ .

Por  $k$  ser intimamente ligado à quantidade de células tumorais  $N(t)$ , no instante  $t$ , consideramos que um tumor possua um limite que não ultrapasse o valor de  $10^{13}$ .

### 2.3. Lei universal de crescimento

Reconhecemos o crescimento de organismos através de características presentes na natureza. Pensar em escalas para analisar o crescimento de indivíduos não é algo tão familiar, no entanto, são fundamentais para descrever leis que regem o crescimento universal.

De acordo com West (2018), é possível redimensionar o tamanho de tumores utilizando essa lei. Usando a razão  $r = \left(\frac{m}{M}\right)^{0,25}$ , temos que  $m$  é a massa assintótica do táxon e  $r$ , a proporção de energia total, cujo objetivo é a manutenção celular.

Utilizando-se o tempo adimensional reescalado, temos:

$$\tau = \frac{1}{4} \alpha M^{-0,25} t - \ln \left( 1 - \left( \frac{m_0}{M} \right)^{0,25} \right) = \alpha r_0 t - \ln(1 - r_0) \quad (2.1)$$

Assim,  $m_0$  é a massa de nascimento tumoral e  $r_0 = \left(\frac{m_0}{M}\right)^{0,25}$ . Além disso,  $\alpha = 0,25 a m_0^{-0,25}$  e  $a$  é o parâmetro constante aproximado dentro do táxon.

A lei universal de crescimento é representada pela equação a seguir:

$$r = 1 - e^{-\tau}.$$

As variáveis  $m_0$  e  $M$  são as massas inicial e final do tumor e  $a$ , o parâmetro que associa as características tumorais à metástase e à invasão nos tecidos vizinhos.

## 2.4. Validação cruzada - *leave-one-out*

Um método simples e muito utilizado para estimar erros de previsão é o método de validação cruzada. Este método estima diretamente o erro extra-amostral esperado  $E' = E[L(Y, \hat{f}(X))]$  e o erro médio de generalização quando o método  $\hat{f}(X)$  é aplicado a uma amostra de teste independente da amostra utilizada para treinar o modelo, selecionada a partir da distribuição conjunta de  $X$  e  $Y$ . Geralmente, a validação cruzada estima bem o erro de previsão esperado do modelo.

## 2.5. *K-fold*

Quando um conjunto de dados é escasso, geralmente não é interessante separar um conjunto de validação cruzada e usá-lo para avaliar o modelo de previsão. Para refinar o problema, a validação cruzada *k-fold* usa parte dos dados disponíveis e ajusta o modelo a um método diferente para testá-lo (Hastie et al., 2009).

Como exemplo, imagine que alguns dados foram divididos em 5 partes.

Para a  $k$ -ésima parte (validação), ajustamos o modelo às outras  $k-1$  partes e calculamos o erro de previsão do modelo ajustado na previsão da  $k$ -ésima divisão dos dados. Com isso, para  $k = 1, 2, \dots, K$  combinou-se as estimativas  $K$  do erro de previsão.

Se  $\kappa : \{1, \dots, N\} \rightarrow \{1, \dots, K\}$  é uma função de indexação que indica a partição a qual a observação  $i$  é alocada pela randomização e denotamos por  $\hat{f}^{-k}(x)$  a função ajustada, calculada com a  $k$ -ésima parte dos dados removida, então, a estimativa da validação cruzada do erro de previsão é

$$CV(\hat{f}) = \frac{1}{N} \sum_N^{i=1} L(y_i, \hat{f}^{-k(i)}(x_i))$$

O caso  $K = N$  é conhecido como validação cruzada *leave-one-out*. Para  $\kappa(i) = i$ , e na  $i$ -ésima observação, o ajuste é calculado usando todos os dados exceto um deles. Para um conjunto de modelos  $f(x, \alpha)$  dado por um parâmetro de  $\alpha$ , denotamos por  $\hat{f}^{-k}(x, \alpha)$  o ajuste do  $\alpha$ -ésimo modelo com a  $k$ -ésima parte dos

dados removida. Então, para este conjunto de modelos, definimos:

$$CV(\hat{f}, \alpha) = \frac{1}{N} \sum_N^{i=1} L(y_i, \hat{f}^{-k(i)}(x_i, \alpha))$$

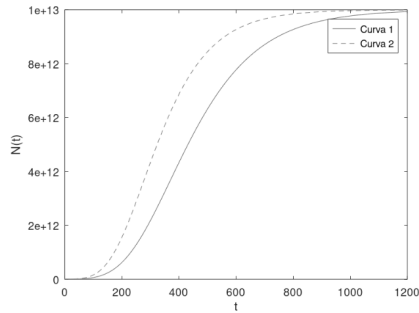
A função  $CV(\hat{f}, \alpha)$  fornece uma estimativa da curva de erro de teste, e encontramos o parâmetro  $\alpha$  que a minimiza. Nosso modelo final escolhido é  $f(x, \hat{\alpha})$ , que então é ajustado a todos os dados.

Para  $K = N$ , o estimador de validação cruzada é aproximadamente não enviesado para o erro de previsão verdadeiro (esperado), mas pode ter alta variância porque os  $N$  “conjuntos de treinamento” são muito similares entre si. A carga computacional também é considerável, exigindo  $N$  aplicações do método de aprendizado.

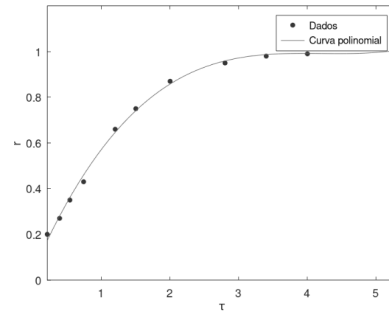
A curva de aprendizado tem uma inclinação considerável no tamanho do conjunto de treinamento de dado, a validação cruzada de cinco ou dez dobras subestimar o erro de previsão verdadeiro. O viés pode ser uma desvantagem na prática, dependendo do objetivo. Por outro lado, a validação cruzada *leave-one-out* tem baixo viés, mas pode ter alta variância (Hastie et al., 2009).

### 3. Metodologia e resultados

Através de uma breve revisão de literatura, foram retirados parâmetros de Schabel Jr (1975) para modelar o crescimento tumoral pelo modelo de Gompertz, como mostra a figura 1(a).



(a) Modelos de Gompertz



(b) Evolução do câncer

Figura 1: Comparação entre dois modelos de Gompertz que descrevem o crescimento tumoral em 1(a) e evolução do câncer de próstata no tempo em 1(b).

O modelo determinístico utilizado descreve as etapas de evolução do tumor. Ele possui a propriedade de crescimento exponencial que diminui à medida que a população celular aumenta. Os parâmetros usados foram: condições iniciais  $n_0 = 2^0 \cdot 10^0$  e  $n_0 = 2^{20} \cdot 10^{16}$ , representando o número de células tumorais no início, para as curvas 1 e 2, respectivamente, capacidade suporte  $k = 5 \cdot 10^{13}$ , taxa de crescimento  $r = 0,006$  referente à curva 1 e  $r = 0,008$  relacionada à curva 2 e tempo  $t = 5$  anos.

Ambas as curvas apresentam acelerados crescimentos no início e se diferenciam pelas condições iniciais. No entanto, as duas funções atingem uma mesma capacidade suporte, ainda possuindo taxas de crescimentos diferentes.

O modelo apresentado quantifica o número de células que se replicam na evolução tumoral por demonstrar precisão e permitir melhor previsão desse crescimento no futuro.

A simulação seguinte trabalha com escalas reduzidas e mostra também de maneira análoga o crescimento tumoral através do aumento exponencial.

Os dados utilizados neste estudo foram retirados de Guiot et al. (2003). São dados *in vitro* que descrevem a evolução do câncer de próstata e que foram ajustados à lei universal de crescimento, no estudo mencionado. Para melhorar o ajuste, utilizamos a ferramenta *WebPlotDigitizer* para extrair os dados da pesquisa original e aplicamos o método dos mínimos quadrados, encontrando o polinômio  $p(x) = 0,013x^3 - 0,164x^2 + 0,687x + 0,036$ .

O gráfico apresentado possui as variáveis  $\tau$ , como mostra a equação 2.1, que é o tempo adimensionalizado e  $r = 1 - e^{-\tau}$ , o valor relativo à proporção de energia total consumida pelo tumor durante sua evolução. Como unidade de medida de massa foi utilizada a grama ( $g$ ).

De fato, quando analisamos um modelo, é comum verificarmos que quanto maior o número de variáveis, mais o erro de estimativa tende a aumentar, podendo não ser bom preditor para grandes quantidades de instâncias. No nosso estudo, analisamos uma amostra de 11 dados para avaliar a validade e a eficiência do modelo que descreve o comportamento do carcinoma de próstata no tempo.

Foi realizada uma análise estatística das variáveis, como mostra a tabela 1, abaixo.

Tabela 1: Análise estatística dos dados de evolução do câncer de próstata.

| <b>x</b> | <b>y</b> | <b>Erro<sup>2</sup></b> | <b>SQReg</b> | <b>SQtot</b> |
|----------|----------|-------------------------|--------------|--------------|
| 0,21     | 0,20     | 0,2278                  | 0,2541       | 0,2278       |
| 0,39     | 0,27     | 0,1659                  | 0,1580       | 0,1659       |
| 0,54     | 0,35     | 0,1071                  | 0,0999       | 0,1071       |
| 0,74     | 0,43     | 0,0612                  | 0,0473       | 0,0612       |
| 1,2      | 0,66     | 0,0003                  | 0,0009       | 0,0003       |
| 1,5      | 0,75     | 0,0053                  | 0,0041       | 0,0053       |
| 2        | 0,87     | 0,0371                  | 0,0324       | 0,0371       |
| 2,8      | 0,95     | 0,0744                  | 0,0775       | 0,0744       |
| 3,4      | 0,98     | 0,0916                  | 0,0944       | 0,0916       |
| 4        | 0,99     | 0,0978                  | 0,0968       | 0,0978       |
| 5,3      | 1        | 0,1041                  | 0,1035       | 0,1041       |

Nela estão apresentados:

- **x**: valores referentes ao tempo de evolução do tumor.
- **y**: valores referentes ao volume do tumor.
- **Erro<sup>2</sup>**: erro ao quadrado.
- **SQreg**: soma dos quadrados da regressão.
- **SQtot**: soma total dos quadrados.

A princípio, analisamos o coeficiente de determinação ( $R^2$ ) para verificarmos a medida de ajuste do modelo estudado. Obtivemos um desempenho de aproximadamente 0,9799, indicando que o modelo explica 97,99% da variância da variável dependente a partir das variáveis independentes. Visualmente, o modelo também mostrou um bom ajuste, no entanto, isso levantou a hipótese de *overfitting* nos dados. Assim, utilizamos a validação cruzada *leave-one-out* no modelo.

Após isso, dividimos o conjunto de dados em um conjunto de treinamento, utilizando todos os valores numéricos, exceto um, que foi reservado para o conjunto de teste. Repetimos esse processo 11 vezes e comparamos a raiz quadrada média dos erros (RMSE) do modelo original com os RMSEs dos demais modelos.



O valor do RMSE total para os 11 dados originais foi de 0,01449, enquanto a média dos RMSE dos outros modelos foi de 0,01589. Isso significa que o erro aumentou em apenas 10% em relação ao modelo original, o que indica um pequeno *overfitting* em relação ao modelo inicial, mostrando boa generalização do modelo.

## 4. Conclusões

O modelo de Gompertz considerou taxas de crescimentos tumorais baixas e as comparou visualmente através de duas curvas com condições iniciais diferentes. Embora o comportamento de crescimento de ambas as curvas se diferenciasse e se tornassem mais sutis à medida que essa taxa diminuísse, ambas tendem a uma mesma capacidade suporte, mostrando que tumores possuem características diferentes em suas etapas de evolução.

Utilizando os dados experimentais, foi possível encontrar um bom ajuste para o modelo, pois polinômios de graus mais baixos facilitam o cálculo de operações mais complexas. Pela validação cruzada *leave-one-out*, obtivemos informações adicionais sobre o ajuste do modelo. Os resultados encontrados indicaram que não há *overfitting* significativo, já que o erro aumentou em um pequeno percentual dos modelos de teste para o modelo original.

Os dois modelos foram apresentados para mostrar que o crescimento exponencial descreve bem o crescimento tumoral, seja ele quantificando o número de células, ou reduzido proporcionalmente por escalas. De fato, na intenção de buscar melhores resultados, o aperfeiçoamento dos dados foi possível e mostrou um olhar diferente na generalização e busca de melhor predição do modelo.

## Referências

- Bassanezi, R. C. (1999). Modelagem matemática uma disciplina emergente nos programas de formação de professores. *Biomatemática*, 9:9–22.
- Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. in a letter to francis baily, esq. frs &c. *Philosophical Transactions of the Royal Society of London*, 115:513–583.
- Guiot, C., Degiorgis, P. G., Delsanto, P. P., Gabriele, P., e Deisboeck, T. S.

- (2003). Does tumor growth follow a “universal law”? *Journal of theoretical biology*, 225(2):147–151.
- Hastie, T., Tibshirani, R., Friedman, J. H., e Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.
- Naoum, P. C. e Naoum, F. A. (2016). *Biologia médica do câncer humano*. Editora Mc Will, São Paulo.
- Quarteroni, A. e Saleri, F. E. (2007). *Cálculo Científico com MATLAB y Octave*. Springer-Verlag. portoghese 1a ed. 2007; italiain 2a stampata, 2007.
- Schabel Jr, F. (1975). Concepts for systemic treatment of micrometastases. *Cancer*, 35(1):15–24.
- West, G. (2018). *Scale: The universal laws of life, growth, and death in organisms, cities, and companies*. Penguin.
- Yorke, E., Fuks, Z., Norton, L., Whitmore, W., e Ling, C. (1993). Modeling the development of metastases from primary and locally recurrent tumors: comparison with a clinical data base for prostatic cancer. *Cancer research*, 53(13):2987–2993.