

MEMÓRIAS
DA
ACADEMIA DAS CIÊNCIAS
DE
LISBOA

CLASSE DE CIÊNCIAS

TOMO XLV

**A matemática da evolução e a
evolução da matemática**

FÁBIO CHALUB



ACADEMIA DAS CIÊNCIAS
DE LISBOA

LISBOA • 2018

A matemática da evolução e a evolução da matemática

FÁBIO CHALUB

Inicialmente gostava de agradecer à Academia das Ciências de Lisboa, particularmente ao Prof. João Paulo Dias, o convite para falar num dos vários eventos comemorativos do ano Darwin. É bom notar que a matemática não foi esquecida nem se omitiu desta efeméride. De facto, há um consenso cada vez mais alargado da relevância da matemática nas ciências da vida, o que pode ser visto consultando algumas das referências bibliográficas [1,2].

O propósito destas notas é discutir, brevemente, o que poderia ser chamado de “a matemática da evolução”, ou seja, aqueles resultados, de cariz nomeadamente matemático cuja interpretação biológica é direta e, pelo outro lado, quais problemas originados na biologia precisam ainda de desenvolvimentos matemáticos que permitam sua melhor compreensão.

É um pouco difícil estabelecer uma data para o que poderia ser o início da biomatemática; por vezes arbitramos o problema dos coelhos de Fibonacci, de 1202 (“Um casal de coelhos demora um mês para o amadurecimento sexual; a partir deste momento, produz um novo casal por mês. Começando com um único casal, quanto casais estarão presente no mês n ?”) [3]. No entanto, apesar da inspiração biológica, este era apenas um problema perdido entre vários cujo objectivo era o de ilustrar algumas técnicas matemáticas.

Também Euler trabalhou em problemas de dinâmica populacional [4]; apesar de seu trabalho mostrar uma abordagem correcta e em larga medida moderna tanto do ponto de vista matemático como de modelação, sua motivação não era propriamente biológica: seu objectivo primário era mostrar que toda a população da Terra poderia ter sido gerada a partir apenas de um casal em alguns milhares de anos (“e não em centenas de milhares de anos como propõem alguns, em desacordo com as escrituras”).

Muitos trabalhos, no entanto, deveriam ser mencionados (e não serão) como os de Bernoulli [5], sobre a varíola (que poderia ser considerado, sem muita dificuldade, o primeiro trabalho que reúne de forma original as técnicas matemáticas da época na solução de um problema biológico relevante), Leslie e as populações estruturadas (ainda hoje chamamos as matrizes usadas naquele trabalho de “matrizes de Leslie”) [6] ou os de Verhulst sobre a equação logística [7]. Estes têm um perfeito balanço entre a motivação biológica, o uso da matemática na formulação e solução de um problema e a interpretação do resultado obtido, voltando ao ponto de partida: a realidade. Uma recente revisão destes trabalhos, combinando a perspectiva histórica com o conteúdo científico pode ser visto em [8].

Nosso ponto de partida será no entanto diferente: quando, no início do século XX um grupo de investigadores, muitos dos quais com formação em física e matemática se uniu para produzir o que ficou conhecido como “síntese evolutiva moderna” [9]. Nosso caminho, ao longo deste artigo não terá nem

ambições de ser historicamente rigoroso, menos ainda de ser exaustivo. Inicialmente enunciaremos alguns resultados de cariz essencialmente matemático que deve convencer o leitor que uma boa compreensão da dinâmica populacional envolve o estudo da matemática; posteriormente mostraremos alguns resultados que, na opinião do autor, ainda necessitam de novas técnicas rigorosas para sua melhor compreensão.

Quando Darwin formulou a teoria da evolução [10], em meados do século XIX, não havia a menor indicação do mecanismo da hereditariedade; de alguma forma imaginava-se que o filho era uma mistura mais ou menos homogénea das características de seus genitores (a teoria do “blending”). Desta forma, dado tempo suficiente uma população deveria ser homogenizada: ou seja, todos os indivíduos seriam idênticos. Isto estava em óbvio desacordo com a realidade.

A explicação estava já disponível. G. Mendel descobrira a genética mais ou menos a mesma época em que Darwin desenvolveu os conceitos centrais da evolução [11]. Mas, ao contrário do inglês que recebeu muita atenção ainda em vida, os experimentos de cruzamentos de ervilhas do austríaco foram largamente ignorados em seu tempo.

A sua redescoberta, já no início do século XX, permitiu resolver o *puzzle* da variabilidade genética, e esta não apenas era compatível com a evolução, como era a peça que faltava. Neste contexto, gostava de aqui destacar o primeiro resultado essencialmente matemático de interesse para a genética: a lei de Hardy-Weinberg, cujo nome refere a trabalhos independentes, mas quase simultâneos de G. Hardy, matemático inglês e W. Weinberg (médico alemão) [12,13].

Teorema: *Se: i) os acasalamentos forem aleatórios; ii) a população for de tamanho infinito; iii) não houver emeral natural; iv) não houver mutações; v) não houver migração; então não há evolução (ou seja, as frequências de cada gene são invariantes no tempo).*

A demonstração é, do ponto de vista matemático, bastante elementar, usando apenas alguns resultados clássicos de combinatória. No entanto, este teorema explicou algo que àquela época intrigava os biólogos: como a evolução não leva a eliminação dos genes recessivos. A partir da segunda geração, mostra a lei de Hardy-Weinberg, se não houver nenhuma força evolutiva presente, as frequências não mais mudam. Ademais, a remoção de cada uma das hipóteses acima, isoladamente, leva a alterações na composição da população, mostrando que cada condição acima é suficiente para vermos evolução em uma população.

Um segundo resultado fulcral na evolução matemática é o Teorema Fundamental de Fisher, formulado por R. Fisher em 1930 [14], envolvendo um dos conceitos centrais da biologia evolutiva: o de *fitness*, ou seja a probabilidade de um certo indivíduo deixar descendentes na geração seguinte:

Teorema: *A taxa de crescimento do fitness médio de uma população é idêntico a variância do fitness médio a cada momento.*

A formulação acima é uma das mais usuais do teorema de Fisher; no entanto sua boa compreensão exige um olhar mais apurado na demonstração. De facto, o que está por trás deste teorema é um lema que afirma que dada uma certa característica quantificável (traço) e que se mantenha de pai para filho,

então a diferença do valor médio deste traço entre duas gerações sucessivas é igual a covariância entre o traço e o *fitness* da geração antecessora. Aplicando o lema ao próprio *fitness*, temos o teorema acima. (A equação de Price generaliza o resultado de Fisher para ambientes variáveis, mas não veremos este resultado aqui [15].)

Como a variância é sempre não negativa, então uma consequência imediata é que o *fitness* médio é não decrescente, sendo constante apenas quando assume o mesmo valor em todos os indivíduos.

No entanto, a afirmação de que o *fitness* se mantém de pai para filho é extremamente forte: exclui uma grande categoria de características cuja modelação requer, geralmente, o uso da teoria de jogos.

Esta foi introduzida por Von Neumann e Morgenstern [16] no estudo de jogos e do comportamento económico. Posteriormente, Nash introduziu o conceito de equilíbrio que leva o seu nome. Nash era sobretudo um matemático, e apresentou duas demonstrações distintas da existência de equilíbrio, ambas baseadas centralmente em resultados gerais. No primeiro caso, no teorema do ponto fixo de Kakutani [17]; no segundo, no teorema de Brouwer [18], um resultado geral que garante a existência de um ponto fixo em qualquer aplicação contínua do disco de qualquer dimensão em si próprio.

O equilíbrio de Nash tornou-se um conceito central em economia: a existência, em condições muito gerais, de uma configuração onde cada agente económico maximiza a sua utilidade. Para isto, foi necessário introduzir, do ponto de vista da modelação do problema real, o conceito de “racionalidade”: cada um escolhe a estratégia que maximiza seu ganho.

Na biologia, evidentemente, não podemos usar tal conceito; no entanto, Maynard Smith e Price [18,19] pensaram na teoria de jogos de forma dinâmica: cada indivíduo faz aquilo para qual é geneticamente programado a fazer. Alguns destes comportamento serão bem sucedidos, outros nem por isto. O resultado é que a quantidade de descendentes deixada por cada tipo presente na população será variável, gerando uma dinâmica nas frequência. Os pontos de equilíbrio desta dinâmica serão, em algum sentido, o equivalente ao equilíbrio de Nash para os jogos evolutivos.

Assim, introduziram o conceito de estratégias evolutivamente estáveis (ESS). Dizemos que uma certa estratégia é um ESS se existir uma barreira de invasão positiva para qualquer grupo de indivíduos de estratégia distintas (mutantes). Como o conjunto de estratégias possíveis é compacto, existe uma barreira global. Isto significa que qualquer grupo pequeno de mutantes, ao entrar em uma população, terá um *fitness* médio inferior ao dos residentes. Desta forma sua frequência diminuirá na geração seguinte e tenderá a desaparecer com o tempo.

Apesar da ideia apelativa, um ESS não é necessariamente um atrator da dinâmica; antes pelo contrário é um estado que, ao ser atingido não é possível escapar (pelo menos com uma quantidade pequena de mutantes; quando permitimos grandes quantidades de indivíduos desviantes, a situação é distinta).

Claramente, as ideias dinâmicas percorrem a discussão anterior. Isto permitiu a Taylor e Jonker [20] definir a *dinâmica do replicador*. A taxa de variação (derivada temporal) da frequência relativa de um dado tipo é proporcional a diferença entre o *fitness* do tipo e o *fitness* médio da população. Imediatamente este é um sistema de $n-1$ equações diferenciais (onde n é o número de tipos presentes na população). As técnicas de estudo são aquelas clássicas da teoria de equações diferenciais ordinárias, e, evidentemente, de sistemas dinâmicos. Veja também [21].

Uma ideia que está sempre por trás do uso de equações diferenciais na dinâmica populacional é que a quantidade de indivíduos na população em estudo é infinita. Claro que esta hipótese é irreal; no entanto (e esta é a expectativa), deve servir como boa aproximação quando a população é, de facto, muito grande.

Para corresponder a expectativa acima, devemos modelar uma população finita (e há muitas maneiras de fazer isto) e verificar se os resultados de interesse obtidos neste caso convergem (em algum dos muitos sentidos que esta palavra adquire em uma teoria matemática) para aqueles obtidos com o uso de equações diferenciais.

Uma primeira maneira de modelar a evolução de uma população finita é usar o chamado processo de Moran [21] (os resultados abaixo mantêm-se mesmo quando usamos o mais realista processo de Wright-Fisher [9,22]; de facto, é expectativa do autor que eles valham em condições muito mais gerais do que aquelas exigidas por estes dois processos).

O processo de Moran é definido de uma forma muito simples: considera-se uma população de tamanho fixo N , composta por dois tipos de indivíduos. Calcula-se o *fitness* de cada um dos tipos (a forma exata de cálculo depende da modelação do problema, sendo o mais usual considerar todos os possíveis pares, calcular o *fitness* a partir de sua identificação, em cada encontro, com o ganho dado por uma certa matriz dada e usando os conceitos de teoria de jogos; ou então simplesmente considerar os *fitness* constantes – o que é um caso particular do anterior); depois escolhe-se um indivíduo para morrer com probabilidade $1/N$ e outro (possivelmente o mesmo) para reproduzir com probabilidade proporcional ao *fitness*.

Continuamos este processo até que um estado estacionário seja obtido. Não é difícil perceber que os únicos pontos fixos desta dinâmica são dados pelos estados homogêneos (ou seja, toda a população de um mesmo tipo). Portanto, estes devem ser o estado final, sendo possível calcular a probabilidade de fixação (ou extinção) de um dado tipo.

Parece, a primeira vista, que temos um resultado em desacordo com o uso de equações diferenciais no mesmo problema, pois neste caso é possível termos estados estacionários para populações não homogêneas (como é o caso do exemplo que motivou a criação da teoria evolutiva dos jogos, o jogo conhecido como de Pombas e Falcões, que modela a evolução da agressividade animal).

O ponto central é que os processos discretos possuem estados transientes que, grosso modo, duram um tempo crescente com N . Desta forma, quanto maior a população, mas relevante será este estado. Uma população infinita, quando modelada via equações diferenciais, deve ser comparada com o estado transiente e não com o estado final.

Do ponto de vista da álgebra linear, a modelação usando um processo discreto baseia-se em matrizes estocásticas. Estas têm sempre um valor próprio unitário, que descreve os estados estacionários finais. A equação diferencial surge a partir do estudo do maior valor próprio menor do que um, ou seja, do segundo valor próprio (que descreve o estado transiente mais relevante).

Assim, considerando que uma população é, na verdade, sempre finita, temos que a validade de uma equação diferencial como modelo para a dinâmica de uma população finita será necessariamente limitada no tempo (tão maior será a validade quanto maior for a população). Para estendermos a validade além deste limite devemos usar um modelo baseado em equações com derivadas parciais.

Mais especificamente, partindo da matriz de transição do processo de Moran, considerando populações progressivamente maiores e intervalos entre as sucessivas gerações Δt cada vez mais pequenos (mais

geralmente, relacionando ambos por $\Delta t=1/N^2$, e assumindo uma hipótese conhecida como “*weak selection principle*” (que não será discutido) obtemos que a $p(t,x)$, que descreve a densidade de probabilidade de encontrarmos uma eneral x de mutantes no instante t satisfaz uma equação a derivadas parciais

$$\partial_t p = \frac{1}{2} \partial_x^2 (x(1-x)p) - \partial_x (x(1-x)(\psi^{(A)}(x) - \psi^{(B)}(x))p)$$

definida para x no intervalo fechado de 0 a 1 [23].

A equação acima é a adjunta (num sentido clássico, chamado backward/forward de Kolmogorov) de uma importantíssima equação de genética populacional, chamada a equação de Kimura [24,25].

Esta equação tem coeficiente de difusão degenerado (igual a zero na fronteira do intervalo) o que gera alguma dificuldade em sua análise. Desta forma, é impossível fixar condições de fronteira (as fronteiras são absorventes, e portanto nada do que aconteça na fronteira pode afetar o interior do intervalo); assim a unicidade da solução (que deve ser o limite da solução do caso discreto) é garantida por outra via. Temos que obter leis de conservação do caso discreto (que são dadas pelos enerals próprios da matriz de transição adjunta) e eneralso-las para o caso limite.

Uma boa parte da matemática necessária para o estudo da equação acima teve que ser desenvolvida pelos autores. Infelizmente, vários resultados são específicos, e a situação geral ainda não está clara. De qualquer forma, como muitos outros exemplos de limites contínuos de processos discretos estão a ser trabalhados, temos a expectativa de que através de uma profusão de exemplos, a situação global comece a ser clarificada.

A análise toda acima omite um fator que dia após dia vem sendo considerado fundamental para a análise da evolução. Esta é, naturalmente, um processo em multi-escala [26]. Ou seja, genes organizam-se em indivíduos, indivíduos em sociedade, com todo tipo de organização intermédia. A evolução atua em todos estes níveis. No entanto, apesar de ser possível simular a evolução em vários níveis há uma carência clara na formulação de conceitos (como equilíbrio de Nash, por exemplo) para jogos com este grau de sofisticação.

Jogos de vários níveis tem sido usado com sucesso para o estudo por exemplo da evolução da ética, da cooperação e das normas sociais [27,28,29]. De facto, é genericamente aceite que a evolução em multi-nível é mais relevante para o estudo da evolução social do que da evolução biológica (é importante que haja um fluxo migratório relevante das unidades básicas de evolução – os genes no caso da biologia e os memes no caso da evolução social – de forma que haja alguma forma de generalização entre os vários níveis). Tais estudos tem sido feito, quase sem exceção, de forma numérica, havendo uma enorme lacuna de conhecimento matemático no estudo do que alguns chamam de “metapopulações”. Este é apenas mais um exemplo onde mais trabalho rigoroso se mostra fortemente necessário.

Gostava finalmente de terminar chamando a atenção dos estudantes que a relação entre matemática e biologia vai muito além das “aplicações da matemática a biologia”. Apesar de a história não ser tão longa quanto a das relações com a física, não há dúvida que este campo de trabalho não é mais uma promessa para o futuro, mas já um dos assuntos mais ativos da ciência moderna.

BIBLIOGRAFIA

- 1 – J. E. Cohen. Mathematics Is Biology's Next Microscope, Only Better; Biology Is Mathematics' Next Physics, Only Better. *PLoS Biol* 2(12): e439 (2004).
- 2 – B. Sturmfeld. Can Biology Lead to New Theorems? Annual Report 2005 of the Clay Mathematics Institute (2005).
- 3 – Leonardo Pisano Bogollo (Fibonacci). Liber Abaci (1202).
- 4 – L. Euler. Recherches eneralis sur la mortalite et la multiplication du genre humain. *Mem. Acad. R. Sci. Belles Lett. (Belg.)* 16, pp. 144–146 (1760).
- 5 – D. Bernoulli. Essai d'une nouvelle analyse de la mortalité causée par la petite vérole. *Mem. Math. Phy. Acad. Roy. Sci. Paris, anée 1760* (1766).
- 6 – Leslie, P.H. The use of matrices in certain population mathematics. *Biometrika*, 33(3), 183–212 (1945).
- 7 – P. F. Verhulst. Notice sur la loi que la population poursuit dans son accroissement. *Correspondance mathématique et physique* 10: pp. 113–121 (1838).
- 8 – N. Bacaer. Histoires de mathématiques et de populations, Editions Cassini (2009).
- 9 – W. J. Ewens. *Mathematical Population Genetics*, Springer (2004).
- 10 – C. Darwin. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, John Murray (1859).
- 11 – J. G. Mendel. Versuche über Pflanzenhybriden *Verhandlungen des naturforschenden. Vereines in Brünn, Bd. IV für das Jahr, 1865* (1866).
- 12 – G. H. Hardy. Mendelian proportions in a mixed population. *Science* 28: 49–50 (1908).
- 13 – W. Weinberg. Über den Nachweis der Vererbung beim Menschen. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg* 64: 368–382 (1908).
- 14 – R. A. Fisher. *The Genetical Theory of Natural Selection*, Clarendon Press, Oxford (1930)
- 15 – G. R. Price. Extension of covariance selection mathematics. *Annals of Human Genetics* 35: 485-490 (1972).
- 16 – J. v. Neumann, and O. Morgenstern, *Theory of Games and Economic Behavior*. Princeton, NJ. Princeton University Press (1944).
- 17 – J. Nash. Equilibrium points in n -person games *Proceedings of the National Academy of Sciences* 36(1):48-49 (1950).
- 18 – J. Nash. Non-Cooperative Games *The Annals of Mathematics* 54(2):286-295 (1951).
- 18 – J. Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press (1982).
- 19 – J. Maynard Smith. And G. R. Price. The logic of animal conflict, *Nature* 246:15-18 (1973).
- 20 – P. D. Taylor, P.D. and L. Jonker. Game dynamics and evolutionarily stable strategies, *Mathematical Biosciences* 40 145-156 (1978).
- 21 – J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press (1998).
- 21 – P. A. P. Moran *The Statistical Processes of Evolutionary Theory*, Clarendon Press (1962).
- 22 – M. Nowak. *Evolutionary Dynamics: exploring the equations of life*. Belknap/Harvard (2006).
- 23 – F. A. C. C. Chalub and M. O Souza. From discrete to continuous evolution models: A unifying approach to drift-diffusion and replicator dynamics *Theoretical Population Biology* 76 (4), 268-277 (2009).
- 24 – M. Kimura. Solution of a process of random genetic drift with a continuous model. *Proceedings of the National Academy of Sciences*, 41: 144-150 (1995).
- 25 – M. Kimura. On the probability of fixation of mutant genes in a population. *Genetics*, 47: 713-19 (1962).
- 26 – L. Keller (ed). *Levels of Selection in Evolution*. Princeton University Press (1999).
- 27 – H. Ohtsuki, Y. Iwasa. 2004. How should we define goodness?– reputation dynamics in indirect reciprocity. *J. Theor. Biol.* 231 (1), 107–120 (2004). Erratum in: *J. Theor. Biol.* 232(4), 451 (2005).
- 28 – J. M. Pacheco, F. C. Santos and F. A. C. C. Chalub, Stern-judging: A simple, successful norm which promotes cooperation under indirect reciprocity. *PLoS Comput. Biol.* 2 (12):e178 (2006).
- 29 – F. A. C. C. Chalub, F. C. Santos and J. M. Pacheco, The Evolution of Norms. *Journal of Theoretical Biology* 271 (2) (2006)

(COMUNICAÇÃO APRESENTADA À CLASSE DE CIÊNCIAS
NA SESSÃO DE 7 DE MAIO DE 2009)