

DETERMINISMO E PROBABILIDADE EM BIOLOGIA *

RUBENS MURILLO MARQUES ** e ELZA BERQUÓ ***

1. INTRODUÇÃO

Na aquisição de todo conhecimento o ponto de partida é o da consideração dos fatos concretos da realidade objetiva. O aspecto empírico, inicial do conhecimento, não passa senão de uma primeira etapa rudimentar da elaboração científica. É necessário explicar os fatos considerados, isto é, estabelecer um *sistema de idéias* ou *hipóteses* ou *modelos* tal que por via de inferência o pensamento possa, a partir dêle, chegar aos fatos objeto da teoria em questão. Caso se verifiquem as conseqüências, a hipótese é provisoriamente aceita como verdadeira, embora haja necessidade usualmente de modificações posteriores pela aquisição de novos conhecimentos ou descobertas.

Na formulação de um modelo matemático capaz de descrever o mecanismo segundo o qual um fenômeno se desenvolve no universo, dois caminhos podem ser seguidos, a saber: propor um modelo capaz de prever um único valor para a variável em estudo, ou empregar um modelo que leve em conta a flutuação a que estão sujeitos os valores da variável, associando a cada valor uma probabilidade de sua ocorrência. No primeiro caso, estaremos diante de um *modelo determinístico* e, no segundo, considerando um *modelo estocástico*.

A fim de ilustrar êstes dois procedimentos, consideraremos o fenômeno do crescimento de uma população de organismos. Para tanto, partiremos, em ambos os casos, de premissas que, embora não retratando totalmente a realidade, permitem de forma bastante simples a exposição das idéias fundamentais que norteiam êstes dois caminhos.

Seja então:

$X(t)$ = número de organismos na população no instante t .

Suporemos, por simplicidade, que $X(0) = 1$, isto é, que no instante inicial exista apenas um organismo na população. Admitiremos, ainda,

Recebido para publicação em 12-11-1964.

* Trabalho da Cadeira de Estatística Aplicada à Saúde Pública (Prof. Elza Berquó) da Faculdade de Higiene e Saúde Pública da USP.

** Instrutor da Cadeira.

*** Professor da Cadeira.

esta população sujeita a coeficientes de natalidade e mortalidade λ e μ , respectivamente, os quais suporemos constantes no decurso do tempo.

Seja:

$P_x(t) = P \{ X(t) = x \mid X(0) = 1 \}$ = Probabilidade de que tenhamos x organismos no instante t dado que no instante inicial tínhamos um organismo.

Nestas condições, o modelo determinístico nos conduzirá a uma equação matemática para $X(t)$, enquanto que o estocástico, a uma equação matemática para $P_x(t)$.

2. MODELO DETERMINÍSTICO

Além das pressuposições feitas na introdução, suporemos agora que as alterações sofridas pela população, quanto ao número de organismos, num intervalo de tempo $(t, t + \Delta t)$ sejam:

- proporcionais a Δt ;
- proporcionais ao número de organismos existentes na população no instante t , isto é, a $X(t)$.

Com isto, podemos escrever:

$$\begin{aligned} X(t + \Delta t) - X(t) &= \text{alteração sofrida no montante populacional no} \\ &\quad \text{intervalo de tempo } (t, t + \Delta t) \\ &= \lambda X(t) \Delta t - \mu X(t) \Delta t \\ &= \text{Número de organismos que nascem no intervalo} \\ &\quad \text{de tempo } (t, t + \Delta t), \text{ provenientes dos } X(t) \text{ existentes} \\ &\quad \text{no instante } t \text{ menos o número de mortes} \\ &\quad \text{no intervalo de tempo } (t, t + \Delta t), \text{ provenientes} \\ &\quad \text{dos } X(t) \text{ existentes no instante } t. \end{aligned}$$

Ou seja:

$$\frac{X(t + \Delta t) - X(t)}{\Delta t} = (\lambda - \mu) X(t)$$

Passando ao limite quando Δt tende a zero, resulta a seguinte equação diferencial:

$$\frac{d X(t)}{dt} = (\lambda - \mu) X(t)$$

cuja solução é:

$$X(t) = X(0) \cdot e^{(\lambda - \mu) t}$$

Mas, desde que $X(0) = 1$, vem:

$$X(t) = e^{(\lambda - \mu) t} \quad (1)$$

A equação (1) que constitui o modelo determinístico para o caso em aprêço, pode ser interpretada da seguinte maneira: tôda vez que partirmos de uma população que no instante inicial tem apenas um organismo e que está sujeita a coeficientes de natalidade e mortalidade constantes no decurso do tempo, chegaremos *certamente* no instante t a uma população constituída por $X(t)$ organismos. Esta interpretação justifica a denominação de modelo determinístico dada à equação (1), pois, uma vez fixadas as condições iniciais, existirá sempre uma imutabilidade nos resultados finais.

Assim, por exemplo, a título de motivação, suponhamos uma população de organismos nas seguintes condições:

$$X(0) = 1, \lambda = 0,70, \mu = 0,20$$

O número de organismos na população após uma, duas, três e quatro unidades de tempo, seria, de acôrdo com a (1), dado respectivamente por:

$$X(1) = 1,6487 \cong 2 \text{ organismos}$$

$$X(2) = 2,7183 \cong 3 \text{ organismos}$$

$$X(3) = 4,4817 \cong 4 \text{ organismos}$$

$$X(4) = 7,3891 \cong 7 \text{ organismos}$$

Os resultados acima mostram que, para as condições propostas, deveríamos ter *invariavelmente* 2, 3, 4 e 7 organismos na população após uma, duas, três e quatro unidades de tempo, respectivamente. No entanto, todo pesquisador no campo biológico sabe que isto não acontece com tal regularidade. Os resultados acima se devem ao fato de estarmos supondo uma total estabilidade nas condições e leis que governam o crescimento de populações. Ainda que os coeficientes de mortalidade e natalidade, como características individuais, fôssem mantidos constantes no decurso do tempo, existe um conjunto de causas cujo efeito acumulado pode produzir perturbações no sistema em aprêço.

Atentando para esta deficiência característica dos modelos determinísticos, e visto que é possível a obtenção de resultados que se afastam daqueles por eles previstos, melhor seria a construção de um modelo que permitisse a determinação das probabilidades de ocorrência de $X(t)$. Estes são os chamados modelos estocásticos.

3. MODELO ESTOCÁSTICO

Além das pressuposições feitas na introdução, suporemos que se a população no instante t tiver x organismos, num instante $t + \Delta t$ ela só poderá ter:

$x + 1$ com probabilidade $\lambda x \Delta t$

$x - 1$ com probabilidade $\mu x \Delta t$

x com probabilidade $1 - (\lambda + \mu) x \Delta t$

Conseqüentemente, a fim de calcularmos $P_x(t + \Delta t)$, notemos que o evento — x indivíduos no instante $t + \Delta t$ — ocorrerá por uma das três maneiras seguintes:

ou tínhamos $x - 1$ organismos no instante t e nasceu um organismo no intervalo de tempo $(t, t + \Delta t)$, o que acontece com probabilidade $\lambda (x - 1) \Delta t \cdot P_{x-1}(t)$;

ou tínhamos $x + 1$ organismos no instante t e morreu um organismo no intervalo de tempo $(t, t + \Delta t)$, o que acontece com probabilidade $\mu (x + 1) \Delta t \cdot P_{x+1}(t)$;

ou tínhamos x organismos no instante t e nada sucedeu no intervalo de tempo $(t, t + \Delta t)$, o que acontece com probabilidade $\{1 - \lambda (x - 1) \Delta t - \mu (x + 1) \Delta t\} P_x(t)$.

Nestas condições, podemos escrever:

$$P_x(t + \Delta t) = \lambda (x - 1) \Delta t \cdot P_{x-1}(t) + \mu (x + 1) \Delta t \cdot P_{x+1}(t) + \{1 - (\lambda + \mu) x \Delta t\} P_x(t)$$

isto é:

$$\frac{P_x(t + \Delta t) - P_x(t)}{\Delta t} = \lambda (x - 1) P_{x-1}(t) + \mu (x + 1) P_{x+1}(t) - (\lambda + \mu) x \cdot P_x(t)$$

Passando-se ao limite para Δt tendendo a zero, vem:

$$\frac{d P_x(t)}{dt} = \lambda(x-1) P_{x-1}(t) + \mu(x+1) P_{x+1}(t) - (\lambda + \mu) x P_x(t) \quad (2)$$

a qual vale para $x = 1, 2, \dots$

Para $x = 0$ tem-se:

$$\frac{d P_0(t)}{dt} = \mu \cdot P_1(t) \quad (3)$$

A solução da (2) é dada por:

$$P_x(t) = \{1 - \alpha(t)\} \{1 - \beta(t)\} \beta(t)^{x-1}, \text{ para } x = 1, 2, \dots \quad (4)$$

e da (3) por:

$$P_0(t) = \alpha(t) \quad (5)$$

onde:

$$\alpha(t) = \frac{\mu \{ e^{(\lambda-\mu)t} - 1 \}}{e^{(\lambda-\mu)t} - \mu}$$

$$\beta(t) = \frac{\lambda}{\mu} \alpha(t)$$

Como vemos, com as pressuposições apresentadas, podemos passar a calcular para cada um dos instantes considerados anteriormente, as probabilidades associadas a cada um dos valores de $X(t)$. Assim sendo, para $t = 1, 2$ e 3 teremos

$$\alpha(1) = 0,1360; \quad \alpha(2) = 0,2018; \quad \alpha(3) = 0,2371$$

$$\beta(1) = 0,4760; \quad \beta(2) = 0,7063; \quad \beta(3) = 0,8299$$

e portanto:

TABELA — Valores da probabilidade de têrmos x organismos após uma, duas e três unidades de tempo, segundo o número de organismos.

x	$P_x(1)$	$P_x(2)$	$P_x(3)$
0	0,1360	0,2018	0,2371
1	0,4527	0,2344	0,1298
2	0,2155	0,1655	0,1077
3	0,1026	0,1169	0,0894
4	0,0488	0,0826	0,0742
5	0,0232	0,0583	0,0616
6	0,0111	0,0412	0,0511
7	0,0053	0,0291	0,0424
8	0,0025	0,0205	0,0352
9	0,0012	0,0145	0,0292
10	0,0006	0,0102	0,0242
11	0,0003	0,0072	0,0201
12	0,0001	0,0051	0,0167
13		0,0036	0,0138
14		0,0025	0,0115
15		0,0018	0,0095
16		0,0013	0,0079
17		0,0009	0,0065
18		0,0006	0,0054
19		0,0004	0,0045
20		0,0003	0,0037
21		0,0002	0,0031
22		0,0001	0,0026
23			0,0021
24			0,0018
25			0,0014
26			0,0012
27			0,0010
28			0,0008
29			0,0007
30			0,0006
31			0,0005
32			0,0004
33			0,0003
34			0,0003
35			0,0002
36			0,0002
37			0,0001

Uma simples inspeção da tabela acima nos mostra que o valor previsto pelo modelo determinístico para $t = 1$, isto é, 2 organismos, ocorre com apenas 21,55% de probabilidade, sendo $x = 1$ o valor mais provável.

É de se esperar que, à medida que o tempo passa, os valores previstos pelo modelo determinístico tenham cada vez menos probabilidade de ocorrer, pois, torna-se cada vez mais difícil a manutenção da rigidez das condições sobre as quais se baseiam este modelo. De fato, para os instantes 2 e 3, os resultados previstos, com certeza pelo modelo determinístico eram, respectivamente, 3 e 4 organismos, aos quais o modelo estocástico correspondente associa probabilidades de 11,69% e 7,42%.

Vemos, pois, que às previsões feitas não levando em conta a flutuação do acaso estão associadas probabilidades cada vez menores à medida que nos afastamos do instante inicial. Note-se, no entanto, que embora isto aconteça, o número médio esperado de organismos previsto pelo modelo estocástico, $m(t)$, coincide com a previsão do modelo determinístico, pois demonstra-se facilmente que

$$m(t) = e^{(\lambda - \mu)t}$$

Vemos, portanto, que a melhor interpretação a ser dada ao modelo determinístico é que ele prevê o número médio esperado de organismos em cada instante.

Baseados, intuitivamente, neste fato é que os pesquisadores justificam o emprêgo do modelo determinístico. Entretanto, a introdução de um modelo estocástico permite a obtenção de informações que jamais poderiam ser obtidas à base de um modelo determinístico.

Assim, por exemplo, embora no exemplo considerado, a natalidade seja maior do que a mortalidade é possível uma eventual extinção da população num instante qualquer, com probabilidade, como vemos pela (5), $\alpha(t)$.

Assim as probabilidades de extinção para $t = 1, 2, 3, \dots$ são, respectivamente: 13,6%, 20,18% e 23,71%, ... convergindo para o valor $\frac{\mu}{\lambda} = 0,2857$ ou 28,57% se $\mu < \lambda$ e para o valor 1 se $\mu > \lambda$. Observe-se inclusive que em termos de resultados individuais, a partir do instante $t = 3$, a extinção é o resultado mais provável! Êstes fatos por si só mostram que não devemos fazer previsões a respeito de resultados individuais, mas, para um dado instante, calcular a probabilidade de que o número de organismos na população esteja compreendido entre dois valores, digamos, $x_1(t)$ e $x_2(t)$. Por exemplo, para $t = 3$, a probabilidade de que o número de organismos na população esteja compreendido entre 2 e 6 vale 41,40%, a probabilidade de que tenhamos mais do que 20 organismos será 1,77%, etc.

Outro fato digno de realce no modelo estocástico em questão é que a probabilidade de extinção não depende apenas da diferença $(\lambda - \mu)$, mas também do quociente $\frac{\mu}{\lambda}$. Assim, para $\lambda = 0,70$ e $\mu = 0,20$, $t = 2$, a probabilidade de extinção é 20,18%. Para $\lambda = 0,90$ e $\mu = 0,40$ e $t = 2$, isto é, para a mesma diferença $\lambda - \mu = 0,50$, a probabilidade de extinção é de 28,56%. Isto se justifica porque a variabilidade do número de

organismos no instante t depende não só de $(\lambda - \mu)$ mas também de $(\lambda + \mu)$ e vale:

$$\text{Variância de } X(t) = V \{ X(t) \} = \frac{\lambda + \mu}{\lambda - \mu} e^{(\lambda - \mu)t} \{ e^{(\lambda - \mu)t} - 1 \}$$

e conceitualmente pelo fato de que quanto maiores forem λ e μ maior será a dinâmica populacional e conseqüentemente maior será a probabilidade de uma eventual extinção.

Conforme salientamos no início dêste trabalho, o modelo estocástico desenvolvido para efeito ilustrativo é um caso particular do chamado "processo de renovação" (Birth and Death Process). O aprimoramento dêste tipo de modelo no sentido de melhor se ajustar à realidade, é conseguido através da consideração de fatores tais como: λ e μ variando em função do tempo, mutações, etc.

No caso do crescimento de populações humanas, o modelo pode ser melhorado pela consideração da variação dos coeficientes gerais de natalidade e mortalidade em função do tempo, bem como de suas especificações quanto ao sexo e à idade, podendo-se considerar ainda a população sujeita à ação dos movimentos migratórios.

Todavia, isto conduz a equações diferenciais para cujas soluções nem sempre se dispõe de uma teoria matemática já construída, sendo êste um dos sérios óbices para construção de modelos cada vez mais realistas.

RESUMO

A crescente importância que nos últimos anos vem tendo em Estatística os Processos Estocásticos e as suas inúmeras aplicações em Biologia, tais como as teorias sobre o crescimento de populações, sobre a propagação de epidemias, sobre genética de populações, etc., inspiraram êste trabalho que tem como objetivo divulgar as idéias fundamentais que norteiam êste campo. Para tanto utilizou-se um "processo de renovação".

SUMMARY

In the last twenty years, the Stochastic Process are assuming an increasing importance in Statistics and find multiple applications in Biology such as in the theory of population growth, stochastic models in the theory of epidemics, in the theory of gene frequencies and so on. This paper has the purpose of divulgating the basic ideas in this field, the motivation being the birth-and-death process.