10^{-2}m

À PROCURA DA ESC

A incerteza domina o mundo e a probabilidade é uma categoria inescapável.

DINIS DUARTE PESTANA
Departamento de Estatística e Investigação Operacional,
Faculdade de Ciências da Universidade de Lisboa
e Centro de Estatística e Aplicações da Universidade de Lisboa
dinis.pestana@fc.ul.pt

Além disso, os mesmos argumentos que originam a noção de Sorte podem, por outro lado, proporcionar noutros casos o cotejo adequado entre as noções de Acaso e Desígnio: Podemos imaginar Acaso e Desígnio como possíveis causas que entre si competem para a realização de alguns tipos de acontecimentos, e calcular a Probabilidade de que estes ocorram devido a um ou a outro.

A. de Moivre (1718). Prefácio de *The Doctrine of Chances.*

Lembro-me de, em jovem, um amigo de família ter comprado uma aparelhagem de alta fidelidade, um luxo que queria exibir a toda a gente. E lembro-me do ar aborrecido com que de vez em quando se franzia todo e exclamava: "*Mais um ruído!*" Fiquei com pena dele, por ter passado a ouvir os ruídos em vez de música.

O aforismo popular "ver as árvores e não ver a floresta" expressa esta perversão de dar mais atenção ao pormenor do que aos "pormaiores", que são os padrões que mais importa perceber. Felizmente esta perversão está limitada pelos nossos sentidos, que não percepcionam abaixo de um determinado limiar. Um dos sucessos da evolução foi barrar os canais da percepção abaixo e acima de determinados limites. Os nossos órgãos dos sentidos estão afinados para a escala apropriada: Se os nossos antepassados vissem nuvens de átomos em vez de um urso ou um tigre, porventura não teriam sido antepassados de ninguém...

ALA CERTA

Mas, como é do destino do homem estar insatisfeito, procurámos ampliar as nossas capacidades com tecnologias que nos permitam construir imagens de realidades que estão para além da resolução que os nossos olhos conseguem. Em certo sentido, o próprio paradigma da ciência tem mudado a nossa capacidade de ver (e medir) as outras escalas, quer seja ver, com telescópios e radiotelescópios, o que antes estava infinitamente longe, quer seja, com microscópios electrónicos ou de *scanning*, ver o que antes era infinitesimal, quer seja ainda ver a formação de imagens associadas a ondas de som ou a ondas electromagnéticas. No fundo, trata-se de formas sofisticadas de usar a ciência e a tecnologia para domesticar a escala, transformá-la para fazer surgir objectos antes invisíveis. Com a mudança para uma escala apropriada estamos constantemente a ampliar ou a diminuir a realidade.

SOB O SIGNO DA INCERTEZA

A evolução dotou-nos também da capacidade de abstrair os padrões que nos permitem avaliar o que é mais provável e usar esse conhecimento ao tomar decisões num regime de incerteza. Vivemos sob o signo da incerteza ("*Tempo de solidão e de incerteza*", num belo poema de Sophia de Mello Breyner), mas aprendemos a conviver com ela. Tão bem que os primeiros construtores da ciência moderna tiveram a genialidade de "aparar" a variabilidade que é incontornável nos dados concretos que recolhemos.

Galileu – um gigante do pensamento científico – foi capaz de vislumbrar a verdade revelada pelos dados a que tinha acesso, mas que estava *mascarada*. Com os instrumentos de que dispunha para medir o espaço e o tempo – e tinha que o fazer simultaneamente –, as relações parabólicas que deduziu eram uma fantasia genial. Muito antes de os conceitos apropriados terem sido formalmente inventados, e de a metodologia da investigação científica se ter estabelecido, ele já tinha percebido que não há ciência do concreto e que o papel do cientista é criar os modelos que transformam a informação em conhecimento.

De facto, o conhecimento deve assentar em factos e não em preconceitos. Mas nenhuma colecção de factos constitui, por si mesma, conhecimento. Os dados contêm um sinal, mas inevitavelmente o *sinal* está perturbado por *ruído*, que o distorce, esconde, confunde. Se a escala do ruído for desprezável relativamente à do sinal, construímos com relativa facilidade um modelo adequado, pelo menos operativamente. Mas, se a escala do ruído for grande, decifrar o código pode ser tarefa quase insolúvel. Como exemplo da primeira situação: a linguagem "secreta" infantil que consiste em duplicar cada sílaba, e mudar a primeira das sílabas gémeas começando-a por *p* (este último segmento de frase passaria a ser dita *pi i pumupardar pââ pripripeimeipara pasdas pisipalapasbas pégépeasmeas* …); como exemplo da segunda, a encriptação da frase com um código baseado na factorização de números primos, de grande dificuldade de decifração ¢{¡}˜Œ?{Œ{=?°¡¢?°Œ{ˆŒ=}=°\Œ±Œ=}-ˆ¡ˆÅ=. Uma área muito importante das aplicações de reconhecimento de padrões tem a ver com formas de ampliar a escala do sinal e apagar, na medida do possível, o ruído, levando a escala deste a ficar tão perto de zero quanto possível.

A PROBABILIDADE DOMESTICA A INCERTEZA

Galileu escreveu: "*Deus fez o Mundo em linguagem matemática, compete ao sábio decifrá-la*". A leitura dos números, quando se tornou evidente que a variabilidade é uma característica intrínseca dos dados, tornou-se o objecto da Estatística. Em certo sentido, a Estatística é uma arte marcial do intelecto: tal como o objectivo do judo não é destruir a força do adversário, mas antes procurar usá-la em proveito próprio, a glória da Estatística é ter evidenciado que sem variabilidade não há conhecimento e que o objectivo não é destruir a variabilidade: é delimitá-la, eventualmente constrangê-la, moderá-la – e usá-la como fonte de conhecimento.

Para chegar a este ponto, foram necessárias várias revoluções intelectuais. Não foi imediatamente evidente, mas, se olharmos para trás, ganhamos a convicção de que um dos maiores passos intelectuais dos últimos três séculos foi a descoberta da contingência, o reconhecimento de que a incerteza domina o mundo e que a probabilidade é uma categoria inescapável. Assim como a Estatística não procura anular a variabilidade, mas apenas pretende delimitá-la e usá-la como factor de conhecimento, o objectivo da Probabilidade é domesticar a incerteza, levá-la a uma escala que nos permita tomar decisões com riscos comportáveis.

Como os matemáticos Gnedenko e Kolmogoroff afirmaram no seu lúcido prefácio ao livro que escreveram sobre somas de variáveis aleatórias[1], por baixo do caos superficial que é a aparência dos fenómenos, há padrões perduráveis, que correspondem ao que é permanente e necessário. Sob esta perspectiva, os teoremas de caracterizações e resultados assimptóticos são o garante do valor gnoseológico da Teoria da Probabilidade.

Todos nós usamos médias. Porém, raramente nos questionamos sobre as razões por que elas são tão reveladoras. Dois resultados importantes da Teoria da Probabilidade justificam amplamente o seu uso, e alertam para os limites a partir dos quais ele se transforma em abuso.

A Lei dos Grandes Números afirma que a média de uma amostra converge para o valor médio populacional (no caso deste existir). Se pensarmos, em particular, na repetição de uma mesma experiência inúmeras vezes, e, se apenas registarmos se ocorre ou não o acontecimento A, podemos considerar o que se designa por variável indicatriz de A,

$$I_A = \begin{cases} A & \bar{A} \\ P(A) & 1 - P(A) \end{cases}$$

(como é usual, $\bar{A}$ designa o complementar do acontecimento A; na representação acima, na primeira linha escrevem-se os resultados possíveis da experiência, na segunda linha as correspondentes probabilidades) cujo valor médio é $E(I_A) = P(A)$. Assim, a noção de que a probabilidade pode ser conceptualizada como o limite da frequência relativa com que o acontecimento se realiza é uma consequência da Lei dos Grandes Números.

O Teorema do Limite Central, por outro lado, estabelece que, em condições muito gerais, se a variância populacional for σ^2 (e, consequentemente, a variância da média de uma amostra de dimensão n for $\frac{\sigma^2}{n}$), a distribuição assimptótica da média da amostra $\bar{X}$ é aproximadamente normal, com o mesmo valor médio μ que a população, mas com o erro padrão $\frac{\sigma^2}{\sqrt{n}}$. Este resultado mostra a enorme vantagem de usar médias: a escala do erro está a ser reduzido. Infelizmente, o ritmo a que o erro se reduz é baixa, vai com $\sqrt{n}$; para reduzir o erro a um décimo são necessárias 100 observações – e para reduzir, na mesma proporção relativa, o erro a um décimo de um décimo, são necessárias 10 000 observações!

O TEOREMA DO LIMITE CENTRAL É UM MICROSCÓPIO MENTAL

É pena que nem toda a gente conheça este resultado na sua plenitude, porque ele estabelece, de facto, um novo paradigma para obter resultados de grande precisão. É quase um microscópio electrónico mental: se for barato medir, mais vale medir grosseiramente e muitas vezes, porque ao fazer a média estamos a compensar erros por excesso com erros por defeito, e a obter resultados muito mais rigorosos com muito menos esforço. Há uma trintena de anos, para medir ângulos para preparar plantas e mapas usava-se um aparelho de precisão, caríssimo como todos os aparelhos de precisão, e de difícil manutenção, porque a peça fundamental era um disco que pesava alguns quilos e que tinha de estar muito polido e limpo por causa do atrito e que, por isso, tinha que ir constantemente para afinação. Taborda e Oliveira (1973)[3] aproveitaram a lição do Teorema do Limite Central: usaram um disco muito leve com uma descontinuidade magnética que permitia a aquisição de dados em tempo real pelo computador; a leveza do disco levava a que vibrasse e os erros se compensassem. Como adquiriam 600 dados por minuto (os computadores da altura não tinham a velocidade dos actuais), ao fim de dois ou três minutos tinham obtido uma média que indicava o ângulo desejado com uma precisão muito maior do que o aparelho clássico. Tornou-se, naturalmente, o *standard* de medição de ângulos em geodesia.

O Teorema do Limite Central Clássico diz respeito a populações bastante concentradas (por isso, com variância finita), em que a probabilidade de observações muito distantes da média decresce rapidamente para zero. A desigualdade de Chebycheff mostra que $P\left(|X - \mu| \geq k\sigma\right) \leq \frac{1}{k^2}$, $k>0$, no caso de haver variância. E o que acontece no caso de não haver variância, isto é, de a soma das caudas da distribuição, $P\left(|X| > x\right)$, ser elevada? Por exemplo, no caso de variáveis aleatórias de Cauchy, sabemos que a média tem exactamente a mesma distribuição que qualquer uma das parcelas — é a origem da famosa frase *"quem viu uma Cauchy viu todas"*. O esforço de amostragem não compensa.

Percebeu-se, enfim, que era tudo uma questão de escala:

se $$\frac{P\,(|X|> tx)}{P\,(|X|> x)} \rightarrow t^{-\alpha},\ \alpha \in (0,2), \quad t>0$$

(em vocabulário técnico, se a cauda da distribuição for uma função de variação regular de ordem α), então a escala apropriada para a sucessão de somas convergir é $n^{\frac{1}{\alpha}}$. O Teorema do Limite Central Clássico corresponde à situação em que a escala apropriada é $n^{\frac{1}{2}}$.

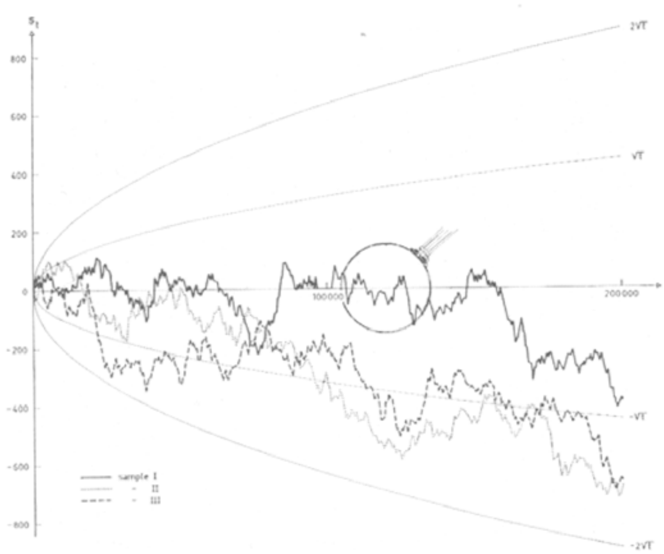

Fig .1. Simulação do movimento browniano.

Felizmente, depois de construir a teoria para somas, concluiu-se que para máximos e para mínimos a situação era muito semelhante: o peso das caudas ditava quais eram os modelos assimptóticos estáveis que se adaptam à realidade. Curiosamente, no caso de máximos, há uma situação em que a escala é irrelevante, apenas há que fazer translações para a origem (só a localização importa): nas populações de Gumbel, em que $P\,(X \leq x\,) = e^{-e^{-x}}$, prova-se que o máximo de n observações cresce tão moderadamente quanto $\ln(n)$, e que se, subtrairmos esta sucessão de valores à sucessão de máximos parciais (o que quer dizer que nos mantemos localizados na proximidade da origem), a distribuição mantém-se de Gumbel.

FRACTAIS, FIGURAS CUJAS PARTES REPRODUZEM O TODO

Recentemente a Matemática tem-se ocupado muito de objectos fractais, objectos que são iguais ou semelhantes a si mesmos a todas as escalas. O título de um trabalho pioneiro de Lèvy4 refere essa estranheza: "*as curvas e superfícies de que as partes são semelhantes ao todo*". O movimento browniano, que deve o seu nome ao botânico escocês Robert Brown, que julgou que os grãos de pólen que caíam na água tinham movimento próprio (o movimento é afinal devido aos choques das moléculas de água), tem essa característica. A fig. 1 mostra simulações do movimento browniano e uma ampliação que procura evidenciar essa convivência extraordinária de todas as escalas.

Para terminar com uma nota tão pessoal como a do início: às vezes penso que as imagens que a NASA gravou na sonda espacial que enviou a figura humana pelo Universo, e em que também gravou o número π por acreditar na universalidade da matemática, correm o perigo de não serem sequer vistas apenas por uma questão de escala. Mas, se a curva do movimento browniano tivesse sido incluída entre as imagens, essa sim, seria decerto vista, porque estaria sempre à escala certa...

REFERÊNCIAS

[1] Gnedenko, B. V. and Kolmogoroff, A.N. (1954). Limit Distributions for Sums of Independent Random Variables. Addison-Wesley, Reading, Mass.

[2] Taborda, J.R. and Ferraz de Oliveira, J. (1973). About a new device for measuring angles with high precision. Rev. Ciências Matemáticas, IV, Ser. A, 17-28.

[3] Lèvy, P. (1937). Les courbes planes ou gauches et les surfaces compostes de parties semblabes au tout, Jour. Ec. Pol. (3) 7, 227-247; 8, 249-291.