

Das possíveis tendências da Matemática do Acaso

J. Tiago de Oliveira

Lisboa

Quem poderia enumerar as mudanças incontáveis que a atmosfera tem dia a dia, e disto predizer hoje o tempo que fará daqui a um mês ou a um ano?

JACOB BERNOULLI, *Ars Conjectandi* (1713)

Falar do impacto da Estatística no mundo moderno e do modo como ela está profundamente integrada no evoluir da civilização industrial dos nossos dias é já tarefa pouco convidativa. Mas, ainda, em pequeno toque de bruxaria, tentar prever a sua evolução, prognosticar do seu futuro, é decerto bem mais complexo. Se a predição é um tema estatístico, todo o estudioso deste sector sabe bem que a sua confiança, a segurança do prever, se esvae à medida que se avança no futuro, à medida que o futuro devém passado.

Com esta ressalva, poderíamos talvez transformar a questão: Quais são os projectos próximos da Estatística? em: Quais são as linhas de força actuais do seu desenvolvimento? Quais são os problemas em aberto?

Mas, mesmo assim, se nos voltarmos para o passado, quantas foram as linhas de evolução percebidas claramente pelos contemporâneos? Quantas vezes foram certas direcções inicialmente rejeitadas? É fácil, quando o modo da evolução já foi presente e é passado, encontrar que o que sucedeu «tinha de suceder», que «certos desenvolvimentos estavam até implícitos em trabalhos anteriores», etc. Estas análises «a-posteriori»,

bem úteis de resto!, são fáceis; difícil é sim o prever no presente. A história situa-se numa dialéctica entre acções pessoais e tendências sociológicas, sendo as primeiras cada vez mais prementes quanto mais se entra na história de um ramo cultural, embora o impulso socio-tecnológico mantenha sempre a sua pressão. A mostrar esta pressão, uma só nota: a tão consabida existência de «époças grávidas de invenção», em que a mesma descoberta surge simultaneamente em vários investigadores; a recordar a acção pessoal, uma só pergunta: Sem CANTOR, de quantos anos se teria atrasado a teoria dos conjuntos? Ou teria, por providencialismo ou inevitabilidade histórica, de acontecer «um CANTOR» naquela época? E a questão transmuda-se, de novo: Quais são, tão-só, os principais problemas em aberto?; já que o impacto, nas linhas de força, de investigadores presentes ou passados é desconhecido. Neste movimento browniano tendencial do pensar, impelido pela pressão histórico-tecnológica, vamos pois tentar diagnosticar algumas vias não percorridas ou pouco percorridas ainda, alguns caminhos por abrir.

E, de novo, uma dificuldade se nos antolha: Quem conhece todo um ramo do saber para dele fazer uma perspectiva total? A questão, dado que há, necessariamente, escolha pessoal, torna-se em: Quais *lhe* parecem ser os principais problemas em aberto? É a esta pergunta, tão restrita, que vamos tentar dar resposta. A auxiliá-la, vamos servir-

-nos, também, do notável estudo de P. AUGER «Les tendances actuelles de la recherche scientifique», UNESCO (1961) no que toca a evolução possível das Probabilidades e Estatística e do heterodoxo e desafiador ensaio de JOHN W. TUKEY «The future of data analysis», Ann. Math. Stat., vol. 33, 1962. Mas, recordemo-lo de novo, esta escolha tem sempre algo de aleatório (e não o teria em Estatística!) pois o movimento, que depois será «naturalíssimo», deverá transcender, pela sua dinâmica interna, os projectos aqui indicados.

No princípio do século eram quase só palavras, descrições. A Estatística é então, somente, a mera compilação de dados, a sua disposição em tabelas, uns tantos cálculos de médias e outras estatísticas simples... e pouco mais; a decisão estatística era, tantas vezes!, feita de modo intuitivo, vendo se o valor calculado a partir da amostra estava próximo ou distante daquele que teóricamente se esperava; um eco, ainda, encontra-se num tratadista das Probabilidades (J. USPENSKY, «Introduction to Mathematical Probability», 1937) ao avaliar a heterogeneidade, verificando se o coeficiente de divergência era próximo ou distante da unidade. A fixação da variabilidade estava ainda bem longe!

Mas a situação começa breve a mudar. FISHER, NEYMAN e WALD, após o esforço de KARL PEARSON e «STUDENT» lançam os fundamentos da Estatística moderna, a procura dos métodos óptimos da inferência, o estudo do comportamento indutivo, rigorizando a comparação intuitiva e vaga. E desde 1912 (o primeiro trabalho de FISHER), a Estatística é explosão, passando do estadió de descrição verbal ao nível de disciplina matemática. Por 1950 a situação pode descrever-se como segue: da estatística estática — isto é, dos processos em que o acaso intervém, digamos, sempre do mesmo modo,

sem evolução — pode supor-se (e poderá?!, quando se procuram projectos de futuro) que estão lançadas as traves-mestras fundamentais; está formulada a teoria da estimação, do teste de hipóteses, da discriminação, etc. e, acima de tudo, um método geral de inferência (ou decisão) estatística. O texto, ainda vivo, de H. CRAMÉR («Mathematical Methods of Statistics», 1945) dá uma excelente perspectiva do domínio em causa que o tempo, de certo!, alargou e aprofundou. Evidentemente que, traves-mestras lançadas, há muito ainda a fazer, um edifício a completar. E, adentro desse esquema, se vem trabalhando desde então. Ao mesmo tempo, vem-se desenvolvendo a teoria dos processos estocásticos, com primórdios no estudo da evolução económica, a futura Estatística dinâmica, que ainda tem imenso a formular. É esta, em duas breves linhas, a situação do dia de hoje, após meio século de vida intensa.

E agora, em breve parêntesis, uma nota sobre o impacto da Matemática do Acaso no sector das aplicações. FISHER, como se sabe, era estatístico na Estação Agrária Experimental de Rothamstead; daí que os métodos de análise estatística no domínio agro-pecuário tenham sofrido desde logo forte impulso, permitindo a melhoria da produtividade, o aumento da eficácia, o estudo cuidado e metódico das condições de produção, etc. As aplicações industriais surgem cerca da década de 30: as cartas de contróle e os métodos, sucessivamente alargados, de contróle de lotes (estes tão ligados ao desenvolvimento dos testes de hipóteses) são, talvez, os primeiros contributos da Estatística ao aperfeiçoamento tecnológico da sociedade industrial, em que a estabilidade de produção, o seu padronizar, desempenham papel relevante; depois, a análise sequencial, desenvolvida durante a II guerra mundial, a teoria de segurança (reliability), os testes de duração (life tests), etc. são novos e importantes ins-

trumentos colocados ao dispor da tecnologia industrial. E a lista poderia prosseguir; o multiplicar de artigos, e mesmo de novas revistas!, sobre este sector, é um índice efectivo da sua importância. No domínio das aplicações médicas, o estudo da eficácia dos fármacos, da qualidade dos tratamentos, a detecção de causas possíveis de doença, etc. são algumas das várias orientações. A Econometria e a Teoria das Comunicações, a Sociologia e a Biometria, etc. eis, ainda, alguns dos diversos sectores em que o impacto metodológico e analítico da Estatística se fez sentir.

E, agora, os projectos? Ah!, era tão fácil, qual Sancho!, escapar através da porta cómoda de que, traves-mestras lançadas, se estava apenas na fase dos pequenos problemas, das soluções locais, adequadas a cada questão científica, um pouco decretar a morte lenta da Estatística, uma «morte térmica», com aumento de produção (e entropia!), como ensina a Termodinâmica. Ah!, a via de salvação era simples... mas não parece justa. Está-se ainda longe, bem longe, da fase do bordado no tecido das relações, este ainda por tecer completamente.

Muitos problemas de tipo experimental necessitam ainda de solução específica. Topam-se a cada instante. E, tantas vezes, há para êles, tão-só, solução aproximada e ineficiente, tantas vezes difícil de manejar, embora solução. Daí que novos métodos de ataque sejam necessários. O que está formulado é *um sistema, uma teoria, uma abordagem teórica do comportamento indutivo, desse salto para o desconhecido*. Aqui e ali, surgem poucas tentativas esparsas de um novo fundamento da inferência.

Direções como, por exemplo, o fundamentar da decisão estatística em termos da distância (ou separação) entre os dados e as hipóteses estatísticas em jogo são possíveis e têm sido, um pouco, tentadas. Avaliações de eficiência em termos outros que a compa-

ração da variância são viáveis e, de certo, úteis; a introdução geral das funções de perda (por WALD) libertou um pouco a Estatística deste uso, embora muitas vias estejam inexploradas: o estudo da eficiência dos testes têm sido, somente, feito nesses termos.

Doutro lado, toda a aplicação da Estatística está baseada no presuposto de que dado modelo é duma descrição teórica suficiente do fenómeno aleatório em estudo. E se assim não for, se o modelo não for adequado? Poucos trabalhos existem neste sector do estudo da *adequação* ou *prova dos modelos*.

Aqui, a dialéctica é complexa: a um lado os modelos disponíveis não são muitos; a outro a aplicação irreflectida de modelos inadequados pode levar a grandes riscos na decisão em concreto; a prova dos modelos bem como o uso de direcções alternativas de decisão deverão não só fundar mais largamente a sua validade como também alargar o âmbito das aplicações.

Uma outra via, conexas com esta, é a do desenvolvimento dos *métodos não-paramétricos*, já agora em expansão larga, que permitem, em tantos casos, a necessária libertação dos modelos estreitos, que, muitas vezes, podem levar a conclusões erróneas. Faltam também métodos de análise da *heterogeneidade das amostras* que tão grandemente podem viciar um estudo estatístico, quando falsamente pressupostas. Conexo com as possibilidades de um novo fundamento da inferência e a adequação dos modelos está, também, o estudo da *robustez da inferência*, em especial do problema complexo das *misturas*, que podem levar à inversão de resultados clássicos (preferência não da média mas da mediana no caso das distribuições normais, como o mostra um trabalho de TUKEY) quando dois modelos (um contaminado e outro não) podem ser ambos descrição igualmente eficiente do problema empírico em causa. Eis pois alguns, mas só *alguns*, dos problemas em aberto no domínio da Estatística estática.

E para os processos estocásticos? Destes, os mais largamente estudados são os processos estacionários e os markovianos, extensões naturais da independência que tantas vezes aparece na Estatística estática; deve porém observar-se que há ainda muito a encontrar nos modelos markovianos, de tal modo que AUGER o aponta como um dos grandes problemas, embora de então para cá algo se tenha feito. Mas processos markovianos (que fundamentam *um* determinismo estatístico) e estacionários são apenas alguns, embora importantes, dos muitos que uma *fisiologia dos processos estocásticos* terá que estudar probabilisticamente, em primeira etapa. Está ainda em início a *estatística dos processos estocásticos* com os seus problemas já clássicos de estimação (ou extracção de sinal), de

teste de hipóteses (ou detecção de sinal), discriminação, etc, estudados alguns apenas para os processos de estrutura menos complexa. Fundamentada que foi a *teoria da predição* do ponto de vista probabilístico (WIENER-KOLMOGOROFF) (e sob poucas perspectivas), a *teoria estatística da predição* pode dizer-se não existir: o que há, tão-só, são arremedos baseados na teoria probabilística, aproximações.

Ora aí está! Após esta curta volta aqui ficam indicadas *algumas*, não *as!*, direcções possíveis do evoluir próximo que *me* parece natural para a Estatística. O tempo se encarregará de mostrar quão pouco se disse e quanto se falhou.