

***o sinal
e o ruído***

o sinal e o ruído

nate silver

TRADUÇÃO DE
ANA BEATRIZ RODRIGUES E CLAUDIO FIGUEIREDO

REVISÃO TÉCNICA
EUCHERIO RODRIGUES



Copyright © Nate Silver, 2012

Todos os direitos reservados, incluindo o direito de reprodução no todo ou em parte em quaisquer meios. Publicado mediante acordo com The Penguin Press, membro do Penguin Group (USA), Inc.

TÍTULO ORIGINAL
The Signal and the Noise

PREPARAÇÃO
Elisa Nogueira

REVISÃO
Clara Diamant
Clarissa Peixoto

CAPA
Yentus & Booher

ADAPTAÇÃO DE CAPA
Julio Moreira

DIAGRAMAÇÃO
ô de casa

CRÉDITOS DAS ILUSTRAÇÕES
Figura 4.2: Cortesia do Dr. Tim Parker, da Universidade de Oxford

Figura 7.1: Taubenberger, Jeffery; Morens, David. "1918 Influenza: The Mother of All Pandemics". *Emerging Infectious Disease Journal*, vol. 12, n. 1, janeiro de 2006, Centers for Disease Control and Prevention

Figuras 9.2, 9.3A, 9.3C, 9.4, 9.5, 9.6 e 9.7: Cburnett, Wikimedia Commons

Figura 12.2: Cortesia do Dr. J. Scott Armstrong, The Wharton School, Universidade da Pensilvânia

CIP-BRASIL. CATALOGAÇÃO-NA-FONTE
SINDICATO NACIONAL DOS EDITORES DE LIVROS, RJ
S592s

Silver, Nate, 1978-

O sinal e o ruído: por que tantas previsões falham e outras não / Nate Silver; tradução Ana Beatriz Rodrigues, Claudio Figueiredo. – 1. ed. - Rio de Janeiro: Intrínseca, 2013.

544 p. : il. ; 23 cm.
Tradução de: The signal and the noise
ISBN 978-85-8057-346-6

1. Previsão. 2. Previsão - Metodologia. 3. Teoria bayesiana de decisão estatística. 4. Estatística. 5. Teoria do conhecimento. I. Título.

13-00329

CDD: 519.5

CDU: 519.2

[2013]

Todos os direitos desta edição reservados à

EDITORA INTRÍNSECA LTDA.
Rua Marquês de São Vicente, 99/3ª andar
22451-041 – Gávea
Rio de Janeiro – RJ
Tel./Fax: (21) 3206-7400
www.intrinseca.com.br

1ª edição

JUNHO DE 2013

impressão

LIS GRÁFICA

papel de miolo

PÓLEN SOFT 70 G/M²

papel de capa

CARTÃO SUPREMO ALTA ALVURA 250 G/M²

tipografia

ELECTRA

À minha mãe e ao meu pai

SUMÁRIO

Introdução 9

1. UM ERRO DE PREVISÃO CATASTRÓFICO 25
2. VOCÊ É MAIS ESPERTO QUE UM ANALISTA DE TV? 55
3. TUDO O QUE ME IMPORTA SÃO OS VS E OS DS 83
4. HÁ ANOS VOCÊS DIZEM QUE A CHUVA É VERDE 119
5. DESESPERADAMENTE EM BUSCA DE SINAL 153
6. COMO SE AFOGAR EM UM METRO DE ÁGUA 185
7. UM MODELO A SEGUIR 212
8. MENOS, MENOS E MENOS ERRADO 239
9. FÚRIA CONTRA AS MÁQUINAS 269
10. A BOLHA DO PÔQUER 302
11. SE NÃO PUDE VENCÊ-LOS... 339
12. UM CLIMA DE SAUDÁVEL CETICISMO 379
13. O QUE OS OLHOS NÃO VEEM O CORAÇÃO SENTE 420

Conclusão 454

Agradecimentos 463

Notas 467

Índice 527

INTRODUÇÃO

Este é um livro sobre informação, tecnologia e progresso científico. É um livro sobre competição, livres mercados e a evolução das ideias. É um livro sobre coisas que nos tornam mais inteligentes do que qualquer computador e sobre falhas humanas. É um livro sobre como aprendemos, passo a passo, a conhecer o mundo de forma objetiva e por que às vezes damos um passo atrás.

É um livro sobre previsão — algo que fica na interseção entre todas essas coisas. Um estudo sobre os motivos pelos quais algumas previsões vingam e outras não. Minha esperança é que possamos desenvolver um pouco mais de discernimento para o planejamento do futuro e, assim, diminuir a chance de repetir os erros.

Mais informações, mais problemas

O início da revolução da tecnologia da informação não ocorreu com a invenção do microchip; ocorreu, sim, com a invenção do tipo móvel para impressão. A criação de Johannes Gutenberg, em 1440, tornou a informação disponível às massas, e a explosão de ideias por ela produzida teve consequências inesperadas e efeitos imprevisíveis. Foi uma centelha para a Revolução Industrial em 1775,¹ um ponto de virada no qual a civilização, que praticamente não havia feito progresso científico ou econômico durante grande parte da sua existência, passou a ter as taxas exponenciais de crescimento e mudança que nos são familiares hoje. Desencadeou os eventos que produziram o Iluminismo europeu e a fundação da república americana.

Mas a invenção do tipo móvel produziria, antes, outra coisa: centenas de anos de guerra santa. Quando a humanidade passou a acreditar que poderia prever e escolher seu destino, iniciou-se a época mais sangrenta da história.²

Os livros já existiam antes de Gutenberg, mas a escrita e a leitura não eram muito difundidas. Ao contrário, eram itens de luxo da nobreza: os escribas produziam um exemplar de cada vez.³ Na época, pagava-se aproximadamente 1 florim (moeda de ouro que valeria cerca de 200 dólares atuais) por cada cinco páginas reproduzidas de um único manuscrito;⁴ desse modo, um livro como o que você está lendo custaria cerca de 20 mil dólares. É provável que também viesse com uma série de erros de transcrição, pois seria uma cópia da cópia de uma cópia, e os erros teriam se multiplicado e mudado a cada geração.

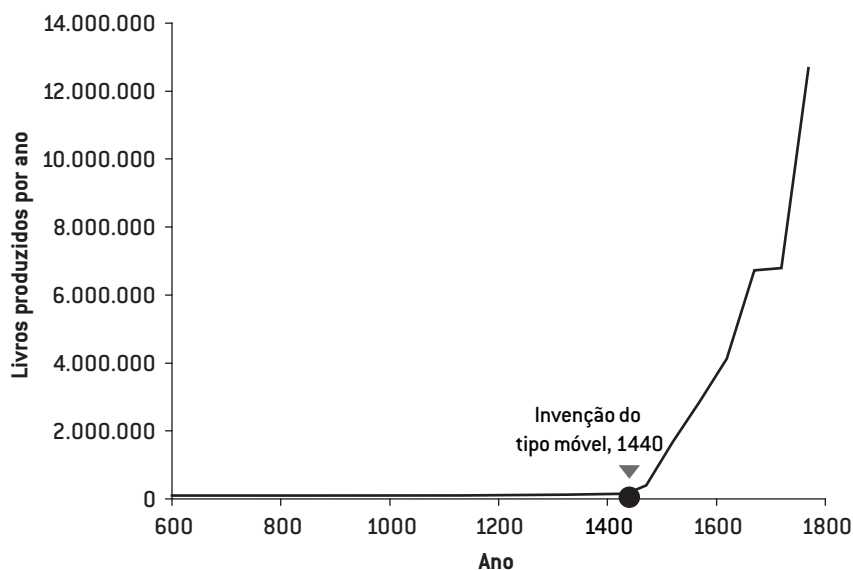
Isso dificultava muito o acúmulo de conhecimento. Era necessário um esforço heroico para impedir que esse volume *diminuísse*, uma vez que os livros poderiam se deteriorar muito antes de serem reproduzidos. Várias edições da Bíblia sobreviveram, junto com um pequeno número de textos de cânones, como Platão e Aristóteles. Contudo, uma quantidade inimaginável de sabedoria se perdeu ao longo dos anos,⁵ e havia pouco incentivo para registrá-la nas páginas.

A busca pelo conhecimento parecia inerentemente fútil, se não completamente vã. Se hoje temos uma sensação de efemeridade graças à rapidez com que as coisas estão mudando, essa era uma preocupação muito mais literal para as gerações que nos antecederam. Não havia “nada de novo sob o sol”, como dizem os belos versos no Eclesiastes, não tanto porque tudo fora descoberto, mas porque tudo seria esquecido.⁶

A invenção do tipo móvel mudou isso de maneira permanente e profunda. De uma hora para outra, o custo de produção de um livro diminuiu quase trezentas vezes;⁷ assim, um livro que poderia ter custado 20 mil dólares em valores atuais passou a custar 70. O tipo móvel disseminou-se com grande rapidez pela Europa; da Alemanha de Gutenberg a Roma, Sevilha, Paris e Basileia, até por volta de 1470, e, daí, para quase todas as outras principais cidades europeias nos dez anos seguintes.⁸ O número de livros em produção cresceu exponencialmente, aumentando cerca de trinta vezes no primeiro século depois da invenção.⁹ O conhecimento humano começou a se acumular, e com rapidez.

INTRODUÇÃO

FIGURA I.1: PRODUÇÃO DE LIVROS NA EUROPA



Entretanto, como ocorreu nos primórdios da World Wide Web, a qualidade da informação variava muito. Embora a invenção do tipo móvel tenha pagado dividendos quase imediatos com a produção de mapas de melhor qualidade,¹⁰ a lista de best-sellers logo passou a ser dominada por textos hereges e pseudocientíficos.¹¹ Agora, os erros podiam ser produzidos em massa, como na chamada “Wicked Bible” [A Bíblia maldita], que cometeu o mais infeliz erro tipográfico na história da página impressa: “Cobiçarás a mulher do próximo.”¹² Enquanto isso, a exposição a tantas novas ideias gerava confusão em massa. A quantidade de informações aumentava com uma rapidez muito maior do que nossa capacidade de saber o que fazer com ela ou nossa habilidade de diferenciar informações úteis ou equivocadas.¹³ Paradoxalmente, ter um volume tão maior de conhecimentos compartilhados estava ampliando o isolamento em termos nacionais e religiosos. O atalho instintivo que tomamos diante do “excesso de informação” é sua seleção; escolhemos aquelas que nos interessam e ignoramos as restantes, tornando aliados aqueles que fizeram as mesmas escolhas e inimigos os demais.

Os primeiros entusiastas do tipo móvel foram aqueles que o utilizaram para a evangelização. As 95 teses de Martinho Lutero não eram tão radicais; sentimentos semelhantes haviam sido discutidos muitas vezes. O que era revolucionário, como escreve Elizabeth Eisenstein, é que as teses de Lutero

“não ficaram afixadas à porta da igreja”.¹⁴ Ao contrário, foram reproduzidas pelo menos trezentas mil vezes pela prensa móvel de Gutenberg¹⁵ — um sucesso retumbante até para os padrões modernos.

O cisma produzido pela Reforma Protestante de Lutero não tardou a mergulhar a Europa em guerra. De 1524 a 1648, houve a Guerra dos Camponeses na Alemanha, a Guerra de Esmalcalda, a Guerra dos Oitenta Anos, a Guerra dos Trinta Anos, as guerras religiosas francesas, as guerras confederadas irlandesas, a Guerra Civil Escocesa e a Guerra Civil Inglesa — muitas das quais travadas ao mesmo tempo. Não podemos deixar de lado a Inquisição espanhola, que começou em 1480, ou a Guerra da Liga de Cambraia, de 1508 a 1516, embora não tivessem tanto a ver com a difusão do protestantismo. Só a Guerra dos Trinta Anos matou um terço da população da Alemanha,¹⁶ e o século XVII foi, possivelmente, o mais sangrento da história, seguido de perto pelo início do século XX.¹⁷

Entretanto, de alguma forma, em meio a tudo isso, o tipo móvel começava a gerar progresso científico e literário. Galileu compartilhava suas ideias (censuradas), e Shakespeare produzia suas peças.

As peças de Shakespeare muitas vezes ressaltam a ideia de destino, como faz a tragédia grega. O que as torna tão trágicas é a lacuna entre o que seus personagens gostariam de realizar e o que a sorte lhes reserva. A ideia de controlar o próprio destino parece ter se tornado parte da consciência humana nessa época — embora as competências necessárias para alcançar esse fim ainda não existissem. Ao contrário, quem desafiava o destino geralmente acabava morto.¹⁸

O tema é explorado vividamente na peça *Júlio César*. Na primeira metade da obra, César recebe vários sinais de advertência — que ele chama de predições¹⁹ (“cuidado com os idos de março”) — de que sua coroação poderia se transformar em uma carnificina. César, é claro, ignora-os, insistindo, muito orgulhoso, que os sinais se referem à morte de outra pessoa ou interpretando seletivamente os indícios. Eis então que César é assassinado.

“Os homens podem, porém, interpretar coisas ao seu modo/ Livre da finalidade da coisa propriamente dita”, adverte-nos Shakespeare pela voz de Cícero — um bom conselho para qualquer um que deseje mergulhar na riqueza de informações recém-descobertas. Era difícil distinguir sinal e ruído. A história que os dados nos contam muitas vezes é aquela que gostaríamos de ouvir, e geralmente fazemos de tudo para que ela tenha um final feliz.

No entanto, se *Júlio César* recorre ao antigo conceito de predição — associando-o a fatalismo, adivinhação e superstição —, ela também introduz

uma ideia mais moderna e mais radical: a de que poderíamos interpretar esses sinais para extrair deles alguma vantagem. “Há momentos em que os homens são donos do seu destino”, afirma Cássio, na esperança de persuadir Brutus a participar da conspiração contra César.

A ideia do homem como dono de seu destino começava a ganhar popularidade. Hoje, as palavras inglesas *predict* e *forecast* são usadas como sinônimas,^{*} mas, na época de Shakespeare, tinham significados diferentes. *Predict* era aquilo que faziam os adivinhos; *forecast*, por outro lado, era algo mais próximo à ideia de Cássio.

O termo *forecast* vem das raízes germânicas do inglês,²⁰ enquanto *predict* vem do latim.²¹ *Forecast* refletia a nova noção mundana dos protestantes, e não a outra visão do Sacro Império Romano. *Forecasting* implicava, em geral, planejar em condições de incerteza. Sugeria exercer prudência e esforçar-se, mais ou menos da mesma maneira que utilizamos nos dias de hoje a palavra *previsão*.²²

As implicações teológicas dessa ideia são complicadas,²³ mas eram menos complexas para quem desejava ter uma existência proveitosa no mundo terreno. Essas qualidades estavam fortemente associadas à ética protestante no trabalho, que, para Max Weber, propiciou o surgimento do capitalismo e a Revolução Industrial.²⁴ Essa noção de *forecast* estava muito ligada à ideia de progresso. Todas as informações contidas nos livros deveriam nos ajudar a planejar nossa vida e a prever, de forma proveitosa, o rumo do mundo.

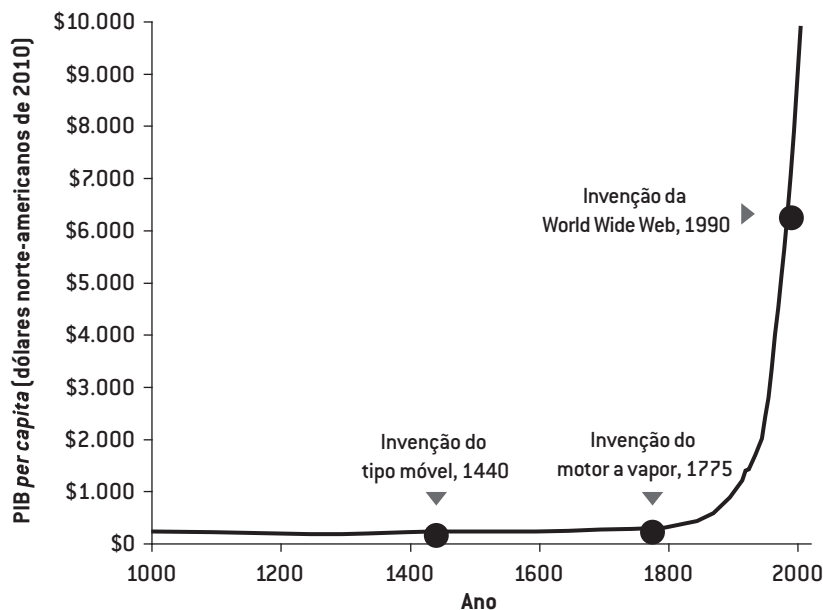
Os protestantes que conduziram séculos de guerra santa estavam aprendendo a usar o conhecimento acumulado para mudar a sociedade. A Revolução Industrial começou, em grande parte, nos países protestantes e naqueles em que havia liberdade de imprensa, onde as ideias religiosas e científicas podiam circular sem medo de censura.²⁵

Nunca é demais enfatizar a importância da Revolução Industrial. Durante boa parte da história do homem, o crescimento econômico ocorreu num ritmo de talvez 0,1% ao ano, suficiente para permitir um aumento gradativo da população, mas *nenhuma* elevação dos padrões de vida *per capita*.²⁶ Foi então que, de uma hora para outra, o progresso chegou onde antes não havia

^{*} Em português, as duas palavras podem ser traduzidas de formas diferentes, de acordo com o contexto: *weather forecast*, previsão do tempo; *prediction of the future*, previsão do futuro; *consumption forecast*, estimativa de consumo. As conclusões do autor quanto à etimologia de cada termo, contudo, não são transferíveis para o português, motivo pelo qual se optou por manter as palavras originais. Para o restante do livro, foram usadas traduções pertinentes aos respectivos contextos. (N. do T.)

nada. O desenvolvimento econômico começou a se dar muito mais rapidamente do que a taxa de crescimento da população, como continuou a fazer até os dias de hoje, apesar das ocasionais crises financeiras globais.²⁷

FIGURA I.2: PIB PER CAPITA GLOBAL, 1000-2010



Ocorre que a explosão de informações produzida pelo tipo móvel nos fez muito bem. Foram necessários 330 anos — e milhões de mortos em campos de batalha em toda a Europa — para que essas vantagens se estabelecessem.

O paradoxo da produtividade

Enfrentamos apuros sempre que o aumento da quantidade de informações ultrapassa nossa capacidade de processá-las. Os últimos quarenta anos da história humana sugerem que ainda podemos levar um bom tempo para traduzir as informações em conhecimento útil e que, se não tivermos cuidado, podemos até dar um passo para trás nesse meio-tempo.

O termo “era da informação” não é exatamente novo. Seu uso começou a ser disseminado no final da década de 1970. A expressão afim, “era do

computador”, foi usada ainda antes, a partir de 1970.²⁸ Nessa época, o uso de computadores se popularizava aos poucos, em laboratórios e no meio acadêmico, mesmo que o computador pessoal ainda não fosse comum. Dessa vez, não foram necessários três séculos para que o crescimento da tecnologia da informação produzisse benefícios tangíveis para a sociedade humana. Mas levou de quinze a vinte anos.

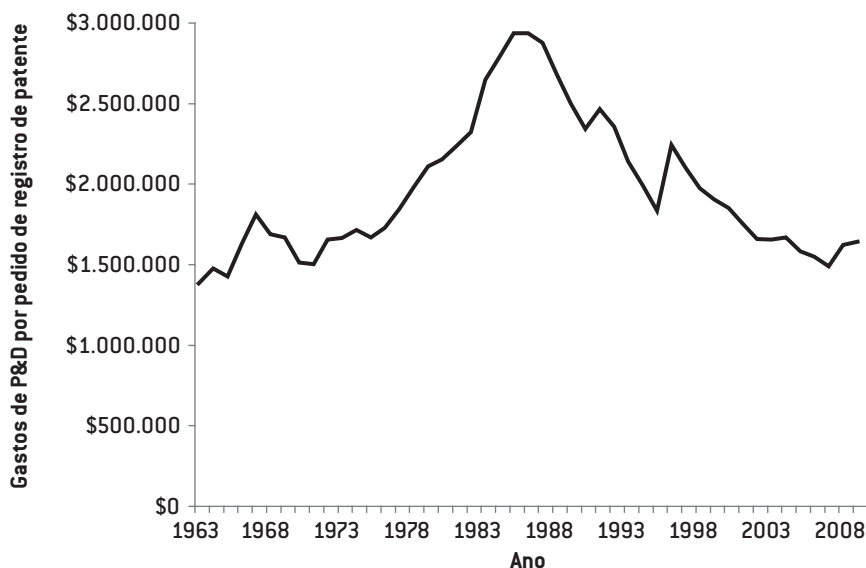
A década de 1970 foi o ponto alto da “grande quantidade de teoria aplicada à quantidade mínima de dados”, disse-me Paul Krugman. Tínhamos começado a usar computadores para produzir modelos do mundo, mas demoramos algum tempo para reconhecer que eles ainda eram muito rudimentares e hipotéticos e que a precisão que os computadores podiam gerar não substituía a precisão das nossas previsões. Em áreas que vão da economia à epidemiologia, foi uma época em que se realizaram previsões audaciosas, que fracassaram com igual frequência. Em 1971, por exemplo, alegou-se que seríamos capazes de prever terremotos dentro de uma década,²⁹ um problema que hoje, passados quarenta anos, ainda não estamos perto de resolver.

Ao contrário, o *boom* dos computadores das décadas de 1970 e 1980 gerou um *declínio* temporário na produtividade econômica e científica. Economistas chamaram o fenômeno de paradoxo da produtividade. “Podemos ver a era do computador em toda parte, menos nas estatísticas sobre produtividade”, escreveu o economista Robert Solow, em 1987.³⁰ Os Estados Unidos passaram por quatro recessões entre 1969 e 1982.³¹ A segunda metade da década de 1980 foi um período em que a economia do país esteve mais forte, mas não podemos dizer o mesmo sobre os outros países do mundo.

É mais difícil medir o progresso científico do que o progresso econômico.³² No entanto, uma de suas características é o número de patentes registradas, especialmente em relação aos investimentos em pesquisa e desenvolvimento. Se é mais barato produzir uma nova invenção, significa que estamos usando bem nossas informações e transformando-as em conhecimentos úteis. Se é mais caro, significa que estamos vendo sinais nos ruídos e desperdiçando nosso tempo com pistas falsas.

Na década de 1960, os Estados Unidos gastavam quase 1,5 milhão de dólares (ajustados segundo a inflação)³³ por pedido de registro de patente³⁴ feito por um inventor americano. No entanto, esse número *aumentou* no início da era da informação, em vez de cair, alcançando seu pico, cerca de 3 milhões de dólares, em 1986.³⁵

FIGURA I.3: GASTOS DE PESQUISA E DESENVOLVIMENTO POR PEDIDO DE REGISTRO DE PATENTE



Na década de 1990, à medida que adotamos pontos de vista mais realistas a respeito do que essa nova tecnologia poderia possibilitar, nossa produtividade em pesquisa começou a melhorar novamente. Andávamos por menos ruas sem saída; os computadores aos poucos melhoravam nossa vida cotidiana e ajudavam a economia do país. As histórias de previsões, em geral, estão relacionadas a progresso a longo prazo e regressão a curto prazo. Muitas coisas que parecem previsíveis a longo prazo frustram nossos melhores planos nesse meio-tempo.

A promessa e as armadilhas do “Big Data”

A expressão da moda hoje é “Big Data” (bancos de dados de tamanho bem maior do que os que em geral conhecemos). A IBM calcula que estamos gerando diariamente 2,5 quintilhões de bytes de dados, mais de 90% dos quais criados nos últimos dois anos.³⁶

Esse crescimento exponencial de informações costuma ser visto como uma panaceia, como aconteceu aos computadores na década de 1970. Chris Anderson, editor da revista *Wired*, escreveu, em 2008, que o mero volume de dados eliminaria a necessidade de teoria e até de método científico.³⁷

Este é um livro pró-ciência e pró-tecnologia e, em minha opinião, muito otimista. Entretanto, argumenta que essas visões estão totalmente equivocadas. Os números, em si, nada dizem. Nós falamos por eles. Nós os imbuímos de significado. Como César, podemos interpretá-los de maneiras que sirvam aos nossos interesses e desvinculadas de sua realidade objetiva.

Previsões baseadas em dados podem se concretizar — e podem falhar. É quando negamos nosso papel no processo que se elevam as chances de fracasso. Antes de exigir mais de nossos dados, precisamos exigir mais de nós mesmos.

Essa atitude pode parecer surpreendente para quem está familiarizado com minha história. Sou conhecido por usar dados e estatísticas para fazer previsões bem-sucedidas. Em 2003, entediado durante um trabalho de consultoria, projetei um sistema chamado Pecota, que tentava prever as estatísticas dos jogadores de beisebol da liga profissional americana. O projeto continha uma série de inovações — suas previsões eram probabilísticas, por exemplo, delineando uma gama de resultados possíveis para cada jogador —, e, quando comparamos os resultados, descobrimos que seu desempenho era superior ao de sistemas concorrentes. Em 2008, fundei o site FiveThirtyEight, com o objetivo de prever os resultados da eleição que se aproximava. As previsões do FiveThirtyEight prognosticaram corretamente o vencedor da corrida presidencial em 49 dos cinquenta estados americanos, bem como o vencedor de todas as 35 eleições para o Senado dos Estados Unidos.

Após a eleição, diversos editores me procuraram, querendo aproveitar o sucesso de livros como *Moneyball* e *Freakonomics*, que contavam a história de nerds que conquistaram o mundo. Este livro foi concebido nesta linha: como uma investigação das previsões baseadas em dados para áreas que vão do beisebol à segurança nacional, passando por finanças.

Entretanto, ao conversar com mais de cem especialistas de mais de uma dúzia de áreas ao longo de quatro anos, ler centenas de artigos de periódicos e livros e viajar pelo mundo, de Las Vegas a Copenhague, percebi, em minha investigação, que a previsão na era do Big Data não andava bem das pernas. Tive sorte em alguns aspectos: primeiro, por ter alcançado o sucesso apesar dos muitos erros que descreverei e, segundo, por ter escolhido bem minhas batalhas.

O beisebol, por exemplo, é um caso excepcional. Por acaso, é uma exceção especialmente rica e reveladora, e o livro examina por que isso acontece — por que, uma década após *Moneyball*, viciados em estatística e olheiros trabalham em harmonia.

O livro apresenta alguns outros exemplos promissores. A previsão do tempo, que também envolve um misto de julgamento humano e capacidade de processamento, é um deles. Os meteorologistas têm uma péssima reputação, mas conseguiram realizar progressos notáveis, sendo capazes de prever, com três vezes mais precisão do que conseguiam fazer há um quarto de século, o ponto em que um furacão atinge terra firme. Enquanto isso, conheci os jogadores de pôquer e os apostadores em esportes que superavam Las Vegas e os programadores que desenvolveram o Deep Blue, da IBM, e derrubaram um campeão mundial de xadrez.

Contudo, esses casos de progresso nos modelos de previsão devem ser ponderados com relação a uma série de falhas.

Se há uma coisa que define os americanos — uma coisa que os torna excepcionais — é a crença na ideia de Cássio de que são capazes de controlar o próprio destino. Os Estados Unidos foram fundados durante a aurora da Revolução Industrial e por religiosos rebeldes que haviam constatado que o livre fluxo de ideias ajudara a difundir não apenas suas crenças religiosas, mas também dados da ciência e do comércio. A maioria dos pontos fortes e fracos da nação — a criatividade e a determinação, a arrogância e a impaciência — deriva da crença inabalável na ideia de escolher o próprio curso.

No entanto, o novo milênio teve um começo terrível para os americanos. Não se previram os ataques de 11 de setembro. O problema não era a carência de informações. Exatamente como aconteceu nos ataques a Pearl Harbor, seis décadas antes, todos os sinais estavam lá. Ocorre que as peças não foram juntadas. Carentes de uma teoria adequada sobre os possíveis comportamentos dos terroristas, os americanos estavam cegos aos dados, e, para todos, os ataques foram “desconhecidos desconhecidos”.

Houve também os fracassos de previsão generalizados que acompanharam a recente crise financeira global. A confiança ingênua em modelos e a incapacidade de reconhecer sua fragilidade diante da escolha de pressupostos geraram resultados desastrosos. Enquanto isso, de maneira mais rotineira, descobri que somos incapazes de prever recessões com mais do que alguns meses de antecedência, e não é por falta de tentativa. Apesar de avanços consideráveis no controle da inflação, nossos formuladores de políticas econômicas muitas vezes voam às cegas.

Os modelos de previsão usados na eleição presidencial de 2000 e publicados pelos cientistas políticos antes das votações previram a vitória esma-

gadora de Al Gore por onze pontos.³⁸ Quem ganhou as eleições foi George W. Bush. Em vez de constituírem resultados anômalos, fracassos de previsão como esses são relativamente comuns na política. Um estudo de longo prazo realizado por Philip E. Tetlock, da Universidade da Pensilvânia, revelou que resultados considerados *impossíveis* por cientistas políticos aconteciam em aproximadamente 15% dos casos. (Os cientistas políticos, porém, são provavelmente melhores do que os especialistas de televisão.)

Recentemente houve, como na década de 1970, uma retomada nas tentativas de prever terremotos, a maioria delas com técnicas altamente matemáticas e baseadas em dados. Entretanto, essas previsões prognosticaram tremores que nunca aconteceram e não conseguiram nos preparar para os que se concretizaram. O reator nuclear de Fukushima foi projetado para suportar terremotos de magnitude 8,6 em parte porque alguns sismólogos concluíram que era impossível ocorrerem tremores ainda mais intensos. Foi então que, em março de 2011, houve o terrível terremoto de 9,1.

As previsões estão falhando em uma série de disciplinas, muitas vezes com um grande prejuízo para a sociedade. Consideremos as pesquisas biomédicas. Em 2005, um médico e pesquisador que cresceu em Atenas, John P. Ioannidis, publicou um polêmico artigo intitulado “Why Most Published Research Findings Are False” [Por que a maioria das descobertas científicas publicadas é falsa].³⁹ O artigo estudou descobertas positivas documentadas em periódicos revisados por profissionais da área: descrições de previsões bem-sucedidas para hipóteses médicas testadas em experimentos de laboratório. O estudo concluiu que a maior parte desses achados poderia falhar quando aplicada no mundo real. Recentemente, a Bayer confirmou a hipótese de Ioannidis. Tentando realizar os experimentos, o laboratório não conseguiu reproduzir cerca de *dois terços* das descobertas positivas apresentadas em periódicos médicos.⁴⁰

O “Big Data” vai gerar progresso — um dia. Quando isso vai acontecer, e se, nesse ínterim, retrocederemos, dependerá de nós.

Por que o futuro assusta?

Biologicamente, não somos muito diferentes dos nossos antepassados. Entretanto, alguns pontos fortes da Idade da Pedra tornaram-se pontos fracos na era da informação.

Os seres humanos não dispõem de muitas defesas naturais. Não somos tão rápidos, nem tão fortes. Não temos garras ou presas, nem carapaça. Não cuspimos veneno. Não sabemos nos camuflar. E não podemos voar. Ao contrário, é nossa sagacidade o que nos permite sobreviver. Nossa mente é rápida. Fomos programados para detectar padrões e reagir a oportunidades e ameaças sem muita hesitação.

“Os seres humanos têm mais essa necessidade de encontrar padrões do que outros animais”, disse-me Tomaso Poggio, neurocientista do MIT que estuda como nosso cérebro processa as informações. “Reconhecer objetos em situações difíceis significa generalizar (...) Um bebê recém-nascido é capaz de reconhecer o padrão básico de um rosto (...) Essa capacidade foi adquirida por meio da evolução, não pelo indivíduo.”

O problema, afirma Poggio, é que esses instintos evolutivos às vezes nos levam a detectar padrões quando, na verdade, eles não existem. “As pessoas vêm fazendo isso o tempo todo”, explicou Poggio. “Encontrando padrões em ruídos aleatórios.”

O cérebro humano é notável — é capaz de armazenar por volta de três terabytes de informação.⁴¹ No entanto, isso corresponde a apenas um milionésimo dos dados que, segundo a IBM, o mundo produz *diariamente*. Por isso, temos de ser extremamente seletivos a respeito daquilo que escolhemos lembrar.

Alvin Toffler, no livro *Choque do futuro*, de 1970, previu algumas consequências do que chamou de “sobrecarga de informação”. Em sua opinião, nosso mecanismo de defesa seria simplificar o mundo de maneiras que confirmassem nossas tendenciosidades, mesmo que o mundo em si estivesse se tornando mais diverso e mais complexo.⁴²

Nossos instintos biológicos nem sempre estão bem adaptados ao mundo moderno inundado de informações. A menos que trabalhemos *ativamente* para nos conscientizarmos dessas tendenciosidades, o retorno proveniente de informação suplementar pode ser mínimo — ou ter efeito diminuidor.

A sobrecarga de informação após o nascimento do tipo móvel gerou maior sectarismo. Hoje, essas diferentes ideias religiosas poderiam ser atestadas com mais informação, mais convicção, mais “prova” — e menos tolerância a opiniões divergentes. O mesmo fenômeno parece ocorrer atualmente. O sectarismo político começou a aumentar com muito mais rapidez nos Estados Unidos, desde mais ou menos a época em que Toffler escreveu *Choque do futuro*, e pode estar se acelerando com o advento da internet.⁴³

Essas crenças sectárias podem perturbar a equação que diz que quanto maior a quantidade de informações, mais perto estaremos da verdade. Um estudo recente publicado na revista *Nature* constatou que quanto *mais* informados estavam os principais políticos a respeito do aquecimento global, *menos* concordavam entre si.⁴⁴

Enquanto isso, se a quantidade de informações está aumentando em 2,5 quintilhões de bytes por dia, o mesmo não pode ser dito sobre a quantidade de informações *úteis*. A maior parte é apenas ruído, que está aumentando de forma mais rápida do que o sinal. Há inúmeras hipóteses a testar e muitos conjuntos de dados a garimpar, mas uma quantidade relativamente constante de verdades objetivas.

O tipo móvel mudou a maneira como cometemos erros. Enganos rotineiros de transcrição já não eram tão comuns. Mas, quando havia um erro, ele era reproduzido muitas vezes, como no caso da “Wicked Bible”.

Sistemas complexos como a World Wide Web têm essa propriedade. Podem não falhar com tanta frequência quanto mecanismos mais simples, mas, quando falham, falham feio. O capitalismo e a internet, ambos extremamente eficientes na propagação de informações, criam igual potencial de disseminação para ideias ruins e boas. As primeiras podem produzir efeitos desproporcionais. Antes da crise financeira, o sistema estava alavancado de tal forma que uma única suposição descuidada dos modelos das agências de classificação de risco de crédito desempenhou um grande papel quando pôs abaixo todo o sistema financeiro global.

A regulamentação é uma maneira de solucionar esses problemas. Desconfio, porém, que se trata de uma desculpa para evitarmos procurar respostas dentro de nós mesmos. Precisamos parar e admitir: temos um problema para prever. Adoramos prever as coisas — no entanto, não somos muito bons nisso.

A solução para prever

Se a previsão é o problema central deste livro, é também sua solução.

A previsão é indispensável às nossas vidas. Cada vez que escolhemos uma rota para o trabalho, que decidimos se vamos sair de novo com aquela pessoa que conhecemos ou que reservamos dinheiro para épocas de vacas magras,

estamos fazendo uma previsão sobre o futuro e sobre como nossos planos afetarão as chances de um resultado favorável.

Nem todos esses problemas do dia a dia exigem um esforço de pensamento extenuante; só podemos alocar determinada quantidade de tempo a cada decisão. Entretanto, fazemos previsões muitas vezes ao dia, ainda que não nos demos conta disso.

Por esse motivo, este livro vê a previsão como um empreendimento compartilhado, e não como uma função executada por um grupo seletivo de especialistas ou profissionais. É divertido zombar dos profissionais quando suas previsões falham. No entanto, é preciso ter cuidado com nossa alegria diante da desgraça alheia. Dizer que nossas previsões não são piores do que as feitas pelos profissionais não é nenhum elogio.

Todavia, a previsão desempenha um papel particularmente importante na ciência. Alguns de vocês podem se sentir pouco à vontade com uma premissa que venho insinuando e que gostaria de explicitar: *nunca* seremos capazes de fazer previsões perfeitamente objetivas. Elas serão *sempre* manchadas pelo ponto de vista subjetivo.

Contudo, este livro é enfaticamente contra a ideia niilista de que não existe verdade objetiva. Afirma, ao contrário, que a crença nela — e o compromisso de buscá-la — é o primeiro pré-requisito para realizar previsões melhores. O próximo compromisso do especialista em previsões é constatar que percebe a verdade de forma imperfeita.

A previsão é importante porque conecta a realidade subjetiva à objetiva. Karl Popper, filósofo da ciência, reconheceu esse aspecto.⁴⁵ Para Popper, uma hipótese só era científica se fosse falsificável — no sentido de que pudesse ser testada no mundo real por meio de uma previsão.

O que nos faz hesitar é que as poucas ideias que foram testadas não se saíram tão bem e que muitas das nossas ideias não foram ou não podem ser testadas. É muito mais fácil testar a projeção da taxa de desemprego do que uma alegação sobre a eficácia dos pacotes de estímulo à economia. Em ciência política, podemos testar modelos usados para prever o resultado das eleições, mas uma teoria sobre a influência de mudanças em entidades públicas pode levar décadas para ser testada.

Não vou tão longe quanto Popper, que afirma que tais teorias são, portanto, não científicas ou não têm valor. Entretanto, o fato de que as poucas teorias que *podemos* testar apresentam resultados tão insatisfatórios sugere

que muitas das ideias que ainda *não testamos* também estejam erradas. Sem dúvida, convivemos com muitas ilusões que sequer percebemos.

Mas há um caminho a seguir. Não é uma solução baseada em ideias políticas pouco maduras — considerando-se, em particular, que passei a ver o sistema político americano como uma grande parte do problema. Ao contrário, a solução exige uma mudança de atitude.

Essa conduta é representada por algo que se chama teorema de Bayes, que apresento no Capítulo 8. Trata-se, nominalmente, de uma fórmula matemática, mas, na verdade, é muito mais do que isso. O teorema deduz que precisamos pensar diferente sobre nossas ideias e sobre como testá-las. Precisamos ficar mais à vontade com a incerteza e a probabilidade. Precisamos pensar com mais cuidado sobre os pressupostos e as crenças que levamos à análise de um problema.

Este livro divide-se em duas partes. Os sete primeiros capítulos diagnosticam o problema da previsão, enquanto os seis capítulos finais exploram e aplicam a solução de Bayes.

Cada capítulo gira em torno de determinado assunto e descreve-o com certo nível de profundidade. É inegável que se trata de um livro detalhado — em parte porque muitas vezes aí está o problema e também porque, na minha opinião, certa imersão em um tópico gera muito mais insights do que um resumo.

Em geral, os temas que escolhi são aqueles sobre os quais existem mais informações divulgadas ao público. Há menos exemplos de previsões que se baseiam em informações privadas (por exemplo, como uma empresa usa registros de seus clientes para projetar a demanda por um novo produto). Minha preferência é por tópicos que permitam que você verifique sozinho os resultados, em vez de apenas aceitar o que eu digo.

Um breve mapa para usar o livro

Este livro inclui exemplos das ciências naturais e sociais, de esportes e de jogos. Parte-se de casos de previsão relativamente simples, em que sucessos e fracassos são mais fáceis de ser demarcados, e chega-se às análises que exigem um pouco mais de sutileza.

Os Capítulos 1 a 3 consideram as falhas de previsões que cercam a recente crise financeira, os sucessos no beisebol e a seara das previsões políticas — em que algumas abordagens funcionaram bem, outras não. O objetivo é fazer você refletir sobre algumas das questões mais fundamentais subjacentes ao problema da previsão. Como podemos aplicar julgamentos aos dados sem sucumbir às nossas tendenciosidades? Em que casos a concorrência de mercado melhora as previsões e em que casos pode piorá-las? Como conciliar a necessidade de usar o passado como guia para reconhecer que o futuro pode ser diferente?

Os Capítulos 4 a 7 concentram-se em sistemas *dinâmicos*: o comportamento da atmosfera terrestre, que define o clima; o movimento das placas tectônicas, que pode causar terremotos; as complexas interações humanas responsáveis pelo comportamento da economia americana; e a disseminação de doenças infecciosas. Esses sistemas estão sendo estudados por alguns de nossos melhores cientistas, mas eles dificultam as previsões, que nem sempre foram bem-sucedidas nessas áreas.

Os Capítulos 8 a 10 voltam-se para soluções — primeiro, apresentando um apostador em esportes que aplica o teorema de Bayes com mais habilidade do que muitos economistas e cientistas e, depois, considerando dois outros jogos, o xadrez e o pôquer. Esportes e jogos, por seguirem regras bem definidas, são ótimos laboratórios nos quais testamos nossa capacidade de previsão. Ajudam-nos a entender melhor a aleatoriedade e a incerteza e proporcionam insights sobre como transformar informações em conhecimento.

O teorema de Bayes, porém, também pode ser aplicado a problemas mais existenciais. Os Capítulos 11 a 13 discorrem sobre três casos: aquecimento global, terrorismo e bolhas nos mercados financeiros. São questões difíceis para previsores e para a sociedade. Mas, se a sociedade americana aceitar o desafio, pode tornar os Estados Unidos, sua economia e o planeta um pouco mais seguros.

O mundo percorreu um longo caminho desde a invenção do tipo móvel. A informação deixou de ser um bem escasso; hoje, temos uma quantidade tão grande de informação que não sabemos como usá-la. Todavia, a quantidade de informações úteis é relativamente reduzida. Nós as percebemos de forma seletiva e subjetiva, sem dar atenção às distorções resultantes. Pensamos querer informação quando, na realidade, queremos conhecimento.

O sinal é a verdade. O ruído é o que nos distrai, afastando-nos da verdade. Este é um livro sobre o sinal e o ruído.