

GABRIEL MAIOLINI CAPEZ - 8992795

Resenha apresentado à Escola Politécnica da
Universidade de São Paulo sobre o livro
Fung, K. *Numbers Rule Your World: The Hidden
Influence of Probabilities and
Statistics on Everything You Do*, McGraw-Hill,
2010

Area de concentração: Probabilidade

Professor:

Prof. Dr. Luís Antônio Baccalá

Data: 10/05/2015

São Paulo

2015

Sumário

Identificação

Livro Estudado

Resumo

Resenha

Descrição do conteúdo (inclui críticas e conteúdo aprendido)

Tópicos para aprofundamento

Qualidade do texto

Conclusão

1 IDENTIFICAÇÃO

1.1 LIVRO ESTUDADO

O livro analisado nesta resenha trata-se do Fung, K. *Numbers Rule Your World: The Hidden Influence of Probabilities and Statistics on Everything You Do*, McGraw-Hill, 2010.

1.2 RESUMO

Este livro tenta demonstrar a influência de probabilidades sobre a vida do leitor, induzindo-o a refletir sobre ela enquanto transmite-lhe a “maneira estatística” de pensar através do estudo de dez casos cotidianos, que permitem a introdução e análise de quatro tópicos: média e variância, correlação e modelagem, diferenças entre grupos e funcionamento diferencial do item, e testes de hipótese - tanto a relação entre falsos negativos e positivos e seus custos assimétricos, quanto ao seu nível de significância -. Posteriormente, o autor aplica todos os conceitos a cada caso visto, ressaltando o pensamento estatístico.

2 RESENHA

2.1 Descrição do conteúdo

No primeiro capítulo, o livro aborda o conceito de média, inicialmente como uma ferramenta estatística que, no cotidiano, serve para facilitar a compreensão de certos eventos a partir da criação de artifícios que

na realidade não existem, mas que podem ser utilizados para traduzir uma idéia, como o “dia comum”, que pode ser entendido como um dia ensolarado, com temperatura média - neste caso, a média estatística -. Em seguida, esta construção é criticada por eliminar a diversidade e supersimplificar situações, esquecendo-se da variabilidade em torno desta média. Este conceito, de variabilidade, ou seja, o de quanto as coisas se desviam da média, é o ponto crucial deste capítulo e o tópico abordado em seus dois casos base, pois é ele a representação da inconsistência entre a repetição de certos fenômenos, ou experimentos, e o fator que ilustra a imprevisibilidade do resultado de um experimento, responsável por reduzir a confiabilidade sobre o resultado de sua repetição.

Para ilustrar a idéia da variabilidade, quatro casos são analisados, estando implícito o paradoxo de Braess - um ponto muito interessante que poderia ter sido explorado neste capítulo - que afirma que a adição de mais capacidade não necessariamente resulta em tempos menores de percurso:

O lado negativo:

Ao discutir as longas filas às quais os visitantes da Disney estão submetidos ao visitar as atrações de seus parques, tidas como fator principal de descontentamento entre eles, são causadas não pela falta de capacidade das atrações, mas pela distribuição desigual de clientes ao longo de delas, ou seja, pela variabilidade do padrão de distribuição de clientes em diferentes atrações e horários e, como solução, além de construir um sistema que minimiza o tempo médio, dentro do possível, o conceito de gerenciamento de percepção, ou seja, o de fazer com que o tempo de espera seja imperceptível ao cliente, é utilizado. Ao criar um sistema que fornece um horário de retorno com rápido acesso à atração, além de reduzir a variabilidade

por controlar o horário de retorno e assim dosar o fluxo, o cliente acredita que não teve um tempo de espera alto, apesar de na realidade, ele possivelmente ter esperado mais do que o tempo médio.

De forma análoga, o livro dispõe sobre a congestão de tráfego, sugerindo que a sua principal causa não é a falta de infraestrutura viária, mas a variabilidade do carregamento de trânsito em cada via e também não é o alto tempo médio de trânsito, mas a sua variabilidade, causada por acidentes e variações imprevisíveis de tempo em diversas rotas que causa ansiedade e frustração nos motoristas e, para reduzir o tempo médio e a variabilidade, foi desenvolvido com muito sucesso o controle de fluxo por meio do gerenciamento de acesso às vias, como observado por uma análise do tráfego anterior e posterior ao desligamento temporário do sistema, assumindo que qualquer variação neste intervalo era devido a isto.

Outro caso mencionado pelo autor é o de testes de doping que, devido a diferença biológica entre pessoas, não consegue estabelecer um limiar perfeito, capaz de detectar apenas atletas que consomem drogas.

O lado positivo:

No caso de seguros, mostra-se que eles arrecadam anualmente menos do que asseguram e, se não fosse pelos diferentes níveis de risco aos quais diferentes propriedades estão submetidas e aos diferentes momentos em que o seguro é acionado, seria impossível estabelecer tal sistema, pois ele seria facilmente sobrecarregado.

Já no segundo capítulo, é abordado o tópico de regressão a partir de dois modelos, ou seja, tentativas de descrever o desconhecido

utilizando somente o conhecido, para justificar a variabilidade e a noção de que um modelo não precisa reproduzir perfeitamente a situação, apenas aproximá-la suficientemente. Para tanto, o conceito crucial de correlação, ou seja, a medida da relação entre duas ou mais variáveis, ou tendência de movimento conjunto entre duas ou mais coisas, é utilizado.

No caso do modelo correlacional apresentado, é introduzida a idéia de que não é necessário conhecer as causas de um evento para poder decidir sobre ele e, que em muitos casos, é impossível determiná-las. No entanto, havendo informação suficiente para correlacionar diversos fatores precisamente, é possível obter grande sucesso, como observado em um dos casos base do capítulo, o da indústria de crédito, que deve ser capaz de modelar o risco de realizar um empréstimo que não será retornado, cuja causa é desconhecida, mas como possui-se informações suficientes sobre como este risco varia em função de algumas variáveis e, a partir delas, consegue-se estimar o risco. Analogamente, o autor informa que a partir de padrões de correlação, foi possível construir uma forma de minimizar o tempo de espera nas atrações

Já no caso do modelo causal, no qual as correlações não são suficientemente úteis e não implicam causalidade - um ponto que poderia ter sido mais desenvolvido pelo autor (correlação não implica causalidade) -, pois elas permitem apenas a obtenção de indicadores e, portanto, é necessário validar os resultados obtidos por meio de outras ferramentas, sendo uma delas o estudo de caso controle, abordada no livro. Esta estratégia visa obter uma referência, o grupo controle, composto por componentes comparáveis em todas as características, exceto a que deseja-se estudar e com uma distribuição de tipos de componentes o mais próxima possível do

experimento analisado. Aqui, como possível critério para auxiliar a decisão, foi introduzida a idéia dos critérios de Hill, tais como a presença de forte correlação, a precedência da causa em relação ao efeito, coerência com o conhecimento existente, etc.

Para demonstrar as idéias do modelo causal, o livro constrói dois exemplos, mas com ressalvas. O primeiro trata da epidemia de *E. coli* em setembro de 2006 nos Estados Unidos, que conseguiu ser identificada de forma rápida por meio das técnicas acima. Um questionário com 450 itens revelou o consumo de espinafre dentre 80% dos infectados, um consumo quase quatro vezes superior ao do grupo controle. Esta sugestiva correlação, associada ao fato que todos tinham sido contaminados pela mesma variação da bactéria e outros indicadores laboratoriais, associada a um pouco de sorte - o encontro de uma embalagem de espinafre na casa de um dos infectados - , permitiu que a fonte de contaminação fosse encontrada. No entanto, o autor alerta que neste caso, devido à baixa longevidade do espinafre e aos enormes prejuízos associados a cessação do consumo e venda do alimento foram enormes e, que em outros casos, o modelo determinou uma causa incorreta, novamente gerando prejuízos. Portanto, deve-se tomar cuidado ao utilizar o modelo causal, balanceando os riscos, possíveis prejuízos e a urgência da situação com a falta de informação disponível.

Ademais, em outro exemplo, o escritor alerta aos perigos de assumir incorretamente que correlação implica em causa. No caso de detectores de mentiras e sistemas detecção de terroristas, existem padrões de correlação, mas não é porque houve variação em uma medida que necessariamente o indivíduo está mentindo ou é um terrorista.

No terceiro capítulo, o livro discute dois pontos intrinsecamente relacionados e de vital importância no cotidiano: a falácia de analisar grupos dissimilares como um único e a técnica de funcionamento diferencial de itens (DIF) para tornar exames justos. O primeiro consiste na idéia de tratar não agregar grupos com características distintas em um único grupo, pois estas diferenças muitas vezes ocultam informações relevantes à análise sendo realizada. Então, sempre que possível, é desejável organizar grupos de acordo com seus atributos relevantes ao experimento, atentando para a possibilidade de que haja alguma hipótese implícita que torna os grupos não comparáveis. Assim, é possível minimizar as diferenças que interferem na análise e realizar comparações melhores e, em alguns casos, socialmente mais justas. Outra estratégia para evitar esta falácia, aparentemente mais eficaz, mas não justificada pelo autor, é o uso de randomização, ou seja, de atribuir aleatoriamente indivíduos ou componentes a serem testados a diferentes grupos, sempre que possível. Dessa forma, pode-se assumir que qualquer variação entre os dois grupos decorre do experimento e, não de suas diferenças.

O segundo conceito, DIF, é consequência do primeiro, mas aplicado a exames de seleção. Ele traduz a diferença de probabilidade entre pessoas de diferentes grupos, mas com mesma habilidade, de acertar um determinado teste. Idealmente, deseja-se que ele seja nulo quando os grupos foram particionados adequadamente, ou seja, quando estiver sendo comparado apenas examinandos com igual habilidade.

Para ilustrar esses conceitos, o livro apresenta três exemplos:

Quando se compara a segurança de vôo de todas as rotas das companhias americanas com as do terceiro mundo entre 1987 e 1996, observa-se que estas são responsáveis por 74% dos acidentes fatais com apenas 18% do total de vôos e, portanto, pode-se concluir erroneamente que é mais seguro voar com companhias americanas nos trajetos entre países do terceiro mundo e Estados Unidos. No entanto, quando se escolhe apenas rotas comparáveis, como as anteriores, observa-se que as companhias do terceiro mundo são tão seguras quanto as outras (62% de todos os vôos com 55% dos acidentes fatais).

Seguradoras contra desastres naturais perceberam após a onda de furacões de 2004, que levou muitas à falência, que era necessário aumentar as tarifas. No entanto, elas notaram que se não diferenciassem entre os clientes em áreas costeiras, de risco, e do interior, com baixíssimo risco, cobrando o mesmo preço altíssimo, estes seriam espantados e elas não conseguiram ser capazes de assegurar as propriedades costeiras, indo assim a falência no caso de alguma catástrofe. Portanto, concluíram que era necessário tratar os grupos de forma diferente.

Analogamente, no caso do SAT, ao observar que havia uma grande disparidade de pontuação entre brancos e negros, independentemente de suas habilidades, foi decidido eliminar questões que tinham uma diferença de acertos entre os dois grupos acima de um determinado limiar. No entanto, após vários estudos, observou-se que o método utilizado para descartar as questões era em si injusto e que a melhor forma de saber se não havia favorecimento de algum grupo era realizar a comparação de desempenho entre negros e brancos com igual habilidade. A partir deste momento, foi possível concluir que a disparidade anterior era

devida majoritariamente ao melhor acesso à educação de indivíduos brancos, que aumenta a nota média destes alunos.

No quarto capítulo, o conceito de erro estatístico em testes de hipótese, seus tipos, a relação entre eles e a assimetria do custo associada a cada um deles é estudado. Nele, o erro é definido como a detecção de um evento quando ele não ocorreu, tipo I, e a não detecção de um evento quando ele ocorreu, tipo II. Por meio de diagramas, aplicando implicitamente o teorema de Bayes, o autor demonstra que a medida em que reduzimos a probabilidade de um erro, a do outro aumenta. Portanto, é necessário encontrar um equilíbrio entre cada valor. Mas, devido ao custo assimétrico do erro de cada tipo em diferentes aplicações e a variabilidade existente entre diversas pessoas ou componentes, frequentemente se reduz um dos valores ao ponto de o sistema possuir um alto número de falsos negativos (tipo II) ou positivos (tipo I). No caso de um sistema de detecção de doping, um falso positivo pode destruir a carreira de um atleta e, devido a variabilidade biológica de cada um, é necessário configurar um teste que tenha uma quantidade muito pequena de erros tipo I. Consequentemente, o número de falsos negativos aumenta e vários atletas que realizam doping passam despercebido. Analogamente, em um sistema de detecção de terroristas, deseja-se minimizar o número de falsos negativos para um valor quase nulo, pois um único pode ser catastrófico. No entanto, o número de falsos positivos cresce absurdamente e, muitos inocentes podem ser acusados falsamente como terroristas, acusação que em vários países resulta em pena de morte. Portanto, muito cuidado é necessário para balancear o melhor possível os custos assimétricos de cada erro. Nota-se que neste caso, devido à raridade de terroristas

em relação a população, a razão de falsos positivos para falso negativo é muito grande.

Finalmente, no quinto capítulo, são introduzidas as idéias de que deve-se analisar anomalias no contexto global, duvidar de ocorrências raras, assumindo que elas são praticamente impossíveis, e de aplicar testes de hipótese e significância à diversos experimentos para validar os resultados obtidos e verificar se eles possuem significância estatística. Esta ocorre quando o valor-p, ou seja, o menor nível de significância que pode ser utilizado para rejeitar a hipótese nula para um certo conjunto de dados [OLOFSSON, Peter; *Probability, Statistics and Stochastic Processes*, Wiley, 2012.], é menor do que o nível de significância alpha utilizado, ou seja, a probabilidade de rejeitar a hipótese sendo que ela é verdadeira.

Para isto, foram demonstrados dois casos:

No primeiro, após a alegação de que funcionários de uma casa lotérica tinham cometido fraude, uma investigação foi lançada para verificar se havia ocorrido eventos similares. Durante a investigação, notou-se que muitos outros empregados obtiveram prêmios de loteria, muito mais do que seria esperado se a loteria fosse justa. Então, foi criada a hipótese nula de que os empregados ganharam de forma justa e o valor-p calculado. Como neste caso o valor foi de um em um quindecilhão, a hipótese não foi validada no nível de significância. Assim, concluiu-se que a investigação deveria ser aprofundada, pois havia mais de uma instância de fraude.

De forma análoga, no segundo, o autor cita a análise da hipótese de que duas companhias aéreas possuem a mesma probabilidade de

queda de um de seus aviões por Arnold Barret, concluindo que tal hipótese não era refutável.

2.2 Tópicos para futuro aprofundamento

Os tópicos que mais induziram interesse para um posterior aprofundamento foram o de modelos correlacionais para aprendizado de máquina, não abordado no livro, mas mencionado na seção de Notas, a teoria de filas, de extrema importância em telecomunicações e engenharia de tráfego, abordada superficialmente no primeiro capítulo, e o uso de testes de hipótese para verificar conformidade com modelos, como realizado no capítulo V, no caso da loteria, descrito acima.

2.3 Qualidade do texto e conclusão

O livro é bem escrito, adequadamente estruturado, e possui uma seleção lexical adequada a seu propósito, o de um livro coloquial que visa iluminar o papel que as probabilidades exercem sobre a vida do homem e transmitir a “maneira estatística” de pensar, ou seja, a de aplicar princípios da estatística - média e variância, correlação e modelagem, diferenças entre grupos e funcionamento diferencial do item, e testes de hipótese - na análise de diversas situações do cotidiano, além de atentar ao contexto em que as informações foram apresentadas. Ademais, a forma como cada conceito é inicialmente dividido e analisado em dois distintos casos base e depois reiterado na análise conjunta de todos os casos facilita a sua fixação e assim todo o conteúdo descrito nesta resenha foi adquirido.

No entanto, o modelo escolhido para analisar os casos alternadamente às vezes resulta em perda de contexto e dificulta a compreensão, o livro frequentemente reitera conclusões já extraídas ao invés de aprofundá-las e utiliza alguns termos que possuem significados distintos na literatura formal de forma igual.