



MÉTODO DE FALESCHINI:

aplicação à lei de Bradford, elite (Price) e rotulação de *outliers*

Luis Fernando Maia Lima

 <https://orcid.org/0000-0002-4208-4914>.

✉ luis.fernando@unir.br.

🏢 Universidade Federal de Rondônia (UNIR) |

ROR  <https://ror.org/02842cb31> | Porto Velho, Brasil.

João Marcelo Brazão Protazio

 <https://orcid.org/0000-0002-7976-0576>.

✉ mprotazio@gmail.com.

🏢 Universidade Federal do Pará (UFPA) |

ROR  <https://ror.org/03q9sr818> | Belém, Brasil.

Maria Cristina Piumbato Innocentini Hayashi

 <https://orcid.org/0000-0003-1250-3767>.

✉ piumbato@gmail.com.

🏢 Universidade Federal de São Carlos (UFSCar) |

ROR  <https://ror.org/00qdc6m37> | São Carlos, Brasil.

Djuli Machado de Lucca

 <https://orcid.org/0000-0003-4505-0688>.

✉ djuli@unir.br.

🏢 Universidade Federal de Rondônia (UNIR) |

ROR  <https://ror.org/02842cb31> | Porto Velho, Brasil.

Eixo temático: Técnicas, Ferramentas e Infraestrutura

Modalidade: Resumo expandido

DOI: 10.22477/ix.ebbc.268

Resumo: O objetivo é apresentar o método de Faleschini para aplicação na lei de Bradford, elite (Price) e rotulação de *outliers*. O método de Faleschini é uma equação do quarto grau que considera a média, desvio padrão, assimetria e curtose, com uso de pseudo-parâmetros baseados em percentis e quartis. Aplica-se o novo método aos dados reais da área de Lubrificação apresentada por Bradford. O método de Faleschini retorna cinco zonas, a adaptação aos dados demonstra a existência de quatro zonas, resultados distintos da literatura da área. Conclui indicando a viabilidade do método de Faleschini e sugere novas contribuições.

Palavras-Chave: Faleschini. Bradford. Price. Rotulação. *Outliers*.

1 INTRODUÇÃO

A Ciência da Informação utiliza métodos quantitativos na análise de seus dados, como por exemplo, a lei de Bradford (Egghe, 1990), elite de Price e sua conexão com a rotulação (Silva, 2019) de *outliers* (Lima et al. 2017). Os dois principais modelos da lei de Bradford são a formulação verbal (divisão equitativa de artigos em zonas crescentes de número de periódicos), conforme Leimkuhler (1967) e a fórmula de Egghe (1990).

Algumas pesquisas parecem dividir - sem citar explicitamente algum critério - periódicos com mesma quantidade de artigos em zonas distintas, a fim de garantir um número constante do denominado multiplicador de Bradford. Em Egghe (1990), por exemplo, num conjunto de 25 periódicos com dois artigos cada, o autor postulou que 17 periódicos ficariam na 2ª zona, e os outros oito periódicos restantes ficariam na 3ª e última zona. Isso pode indicar uma tentativa de “adequação” dos dados reais ao modelo da lei (ou multiplicador) de Bradford, em virtude de não terem sido esclarecidos os critérios utilizados por Egghe (1990) para classificar em zonas distintas os periódicos cujos resultados se apresentaram idênticos.

Por sua vez, a conexão da lei de elite de Price (raiz quadrada do total de autores) com outliers de uma distribuição (dados) de frequência aparece em Lima et al. (2017), onde também se comenta o problema de ocorrência de adequação dos dados reais ao modelo teórico de Price.

Lado outro, as leis bibliométricas de Bradford, Zipf, Lotka e elite de Price podem ser visualizadas como distribuições (dados) de frequências discretas, e assim, podendo haver a existência de média aritmética (“ μ ”), desvio padrão (“ σ ”), coeficientes momentos de assimetria (“A”) e curtose (“B”), e também passível de uso de modelos matemáticos ou estatísticos.

Já Faleschini (1948) apresentou um modelo matemático genérico para dados quantitativos univariados, a fim de estudar o efeito da introdução de um novo valor ou variável de interesse (“x”) justamente nos parâmetros estatísticos da média, desvio padrão, assimetria e curtose. Posteriormente, Fiori e Zenga (2005) aperfeiçoaram a ideia de Faleschini (1948) para distribuições (dados) quantitativos univariados contínuos com o uso da variável normalizada “Z” e que permitem o cálculo da variável de interesse (“x”).

Por exemplo, ao se estudar a influência no desvio padrão de um novo valor (“x”) a um conjunto de dados, o trabalho de Faleschini (1948) gera uma equação do segundo grau, com duas raízes, e que portanto, dividem o conjunto de dados originais em três zonas (em que se evidencia a associação direta com a lei de Bradford e demais leis bibliométricas). Assim, reitera-se que há nítidas conexões do trabalho pioneiro (para distribuições discretas) de Faleschini (1948) e na evolução (para distribuições contínuas) de Fiori e Zenga (2005) para uso nas leis bibliométricas de Bradford, Zipf, Lotka e elite de Price, via uso dos quatro parâmetros estatísticos de uma distribuição de frequência: a média, desvio padrão, coeficientes momentos de assimetria e curtose.

Estes quatro parâmetros estatísticos citados, quando estudados por Faleschini (1948) e Fiori e Zenga (2005) em distribuições de frequência, geram uma equação do quarto grau. No mesmo sentido, não tem se observado na literatura pesquisas que pudessem estabelecer uma conexão dos trabalhos de

Faleschini (1948) e Fiori e Zenga (2005) à Ciência da Informação e outros campos do conhecimento, indicando que esta é uma investigação original e inédita.

O objetivo deste trabalho é apresentar o método de Faleschini para aplicações na Ciência da Informação - no caso, para a lei de Bradford - e apontar suas conexões com a elite de Price e rotulação de outliers, além de comparar os resultados com o modelo verbal de Bradford consoante Leimkuhler (1967) e a proposta de Egghe (1990).

2 O MÉTODO DE FALESCHINI

A equação de Faleschini (1948) retrabalhada por Fiori e Zenga (2005) é:

$$Z^4 - 2*B*Z^2 - 4*A*Z + B = 0 \quad (1)$$

$$Z = (x - \mu) / (\sigma) \quad (2)$$

A equação (1) pode ser resolvida em softwares específicos ou planilha Excel e gera um máximo de quatro raízes; se todas reais, geram quatro valores para “x” e, portanto definem, em tese, um máximo de cinco zonas para a lei de Bradford, em contraposição ao modelo verbal ou também ao modelo de Egghe (1990) que não fixam o número de zonas, deixando ser de livre escolha de quem realiza os cálculos.

Sejam então em ordem crescente de valores, quatro raízes reais da equação (1): Z_1 ; Z_2 ; Z_3 e Z_4 . A equação (2) realiza a conversão para a variável de interesse (“x”), também em ordem crescente de valores, respectivamente: x_1 ; x_2 ; x_3 e x_4 . Definem-se então as cinco zonas de Bradford:

Se $x < x_1$:= zona de raríssima produção de artigos. Pode ser compreendido como rotulação de outliers inferiores (Silva, 2019). No caso da lei de Lotka, autores com raríssimas contribuições à ciência.

Se $x_1 \leq x < x_2$:= zona de baixa produção de artigos. Se uso da lei de Lotka, autores com reduzidas contribuições à área.

Se $x_2 \leq x < x_3$:= zona de moderada produção de artigos. Para a lei de Lotka, autores com moderada contribuições ao campo do conhecimento.

Se $x_3 < x \leq x_4$:= zona de alta produção de artigos ou autores com alta contribuição à ciência (Lei de Lotka).

Se $x > x_4$:= zona de elevada produção de artigos ou núcleo de Bradford. Pode ser compreendido como rotulação de outliers superiores (Silva, 2019) e no caso da lei do elitismo de Price (Lima et al., 2017), ser caracterizada como a elite, ou seja autores com elevada contribuição ao conhecimento conforme lei de Lotka.

Para o uso das equações (1) e (2) em dados reais, as seguintes “pseudo” medidas, com base nos percentis e quartis amostrais, são sugeridos, e devem ser interpretados como contribuição inicial à área, passível de novos aperfeiçoamentos e contribuições:

$$\mu = (P85 + P15) / 2 \quad (3)$$

$$\sigma = (P85 - P15) / 2 \quad (4)$$

$$A = [(P90 - \mu) / (Q2 - P15)] + [(P10 - \mu) / (P85 - Q2)] \quad (5)$$

$$B = [(P90 - \mu) / (Q2 - P15)]^2 + [(P10 - \mu) / (P85 - Q2)]^2 \quad (6)$$

Onde:

P85:= octogésimo quinto percentil.

P15:= décimo quinto percentil.

P90:= nonagésimo percentil.

P10:= décimo percentil.

Q2 := segundo quartil.

Triola (2012) comenta que há diversas fórmulas para o cálculo dos percentis e quartis, sem consenso entre os especialistas. Para fins de uso das equações (3) a (6), será proposto ou adotado uma solução ou correção de continuidade para o cálculo dos percentis e quartil. Esta solução ou correção de continuidade adotada aqui encontra sua analogia em Triola (2012, p. 232-234), que a aplicou para uso da distribuição normal (contínua) como aproximação da distribuição binomial (discreta), a qual remeteremos o leitor para mais detalhes.

A Tabela 1, que apresenta dados originais da área de Lubrificação de Bradford, conforme Egghe (1990), será usada como modelo para ilustração tanto do cálculo dos percentis e quartis necessários, como para uso das equações (1) a (6).

Para os dados de Lubrificação da Tabela 1, a proposta de solução continuidade é a seguinte: para exatamente um artigo, "Fracum" é igual a zero. Para 1,50 artigos, "Fracum" é igual a 0,62195. A interpolação linear fornece: P10 = 1,080; P15 = 1,121 e Q2 = 1,402.

Agora, para 2,50 artigos, "Fracum" é igual a 0,77439; Para 3,50 artigos, "Fracum" é igual a 0,85366. A interpolação linear calcula: P85 = 3,454. Finalmente, para 4,50 artigos, "Fracum" é igual a 0,86585. Para 5,50 artigos, "Fracum" é igual a 0,90854. A interpolação linear gera: P85 = 5,300.

De posse de P10; P15; Q2; P85 e P90 do parágrafo anterior, as equações (3) a (6) geram: $\mu = 2,29$; $\sigma = 1,17$; $A = 10,12$ e $B = 115,00$. Os elevados valores de "A" e "B" reforçam a tese que trata-se de contribuição inicial à área, havendo necessidade de mais estudos e novas visões.

Tabela 1 - Dados originais de Bradford compilado de Egghe (1990).

Dados de Lubrificação				
Artigos	Periódicos	Facum	Fracum	Continuidade (Artigos)
1	102	102	0,62195	1,50
2	25	127	0,77439	2,50
3	13	140	0,85366	3,50
4	2	142	0,86585	4,50
5	7	149	0,90854	5,50

Dados de Lubrificação				
Artigos	Periódicos	Facum	Fracum	Continuidade (Artigos)
6	1	150	0,91463	6,50
7	3	153	0,93293	7,50
8	3	156	0,95122	8,50
9	1	157	0,95732	9,50
10	2	159	0,96951	11,50
13	2	161	0,98171	14,00
15	1	162	0,98780	16,50
18	1	163	0,99390	20,00
22	1	164	1,00000	22,00
Total	102	164	1	-

Fonte: Dados da pesquisa, compilados de Egghe (1990).

Legenda: “Facum” significa frequência acumulada de periódicos. “Fracum” significa frequência relativa acumulada de periódicos; “Continuidade (Artigos)” refere-se a solução de continuidade adotada para o cálculo dos percentis e quartis.

Inserindo-se “A” e “B” na equação (1); retornam as seguintes raízes, em ordem crescente: $Z_1 = -15,24$; $Z_2 = -0,63$; $Z_3 = 0,80$; e $Z_4 = 15,06$. Utilizando então “ μ ” e “ σ ”; as seguintes variáveis de interesse (em ordem crescente) retornam: $x_1 = -15,49$; $x_2 = 1,56$; $x_3 = 3,22$ e $x_4 = 19,86$. Portanto, para os dados de Lubrificação:

Se $x < (x_1 = -15,49)$; zona de raríssima produção de artigos. O menor valor dos periódicos é “1”; portanto, não há zona de raríssima produção de artigos. No caso de rotulação de outliers, não há presença de outliers inferiores.

Se $(x_1 = -15,49) \leq x < (x_2 = 1,56)$; Zona de baixa produção de artigos. Nesse caso, há 102 periódicos que publicaram somente um artigo cada, portanto 102 artigos nesta zona.

Se $(x_2 = 1,56) \leq x \leq (x_3 = 3,22)$; Zona de moderada produção de artigos. No caso, há 25 periódicos que publicaram dois artigos e 13 periódicos que publicaram três artigos, ou em suma, há 38 periódicos que publicaram entre dois a três artigos, resultando em um total de 89 artigos.

Se $(x_3 = 3,22) < x \leq (x_4 = 19,86)$; Zona de alta produção de artigos. Há 23 periódicos nesta zona, que publicaram entre quatro e 18 artigos sobre o tema.

Se $x > (x_4 = 19,86)$. Zona de elevada produção de artigos ou núcleo de Bradford. Há somente um periódico que publicou 22 artigos. Dada a sensibilidade da equação (1) ao parâmetro “B”; reitera-se novamente a necessidade de contribuições científicas a fim de modelar as equações (3) a (6). Caso se visualize a distribuição de frequência como distribuição de artigos por autores, a rotulação de outliers superiores e sua conexão com a elite de Price, significaria a presença de um autor como elite que apresentou 22 contribuições ao campo do conhecimento.

3 DISCUSSÃO

A Tabela 2 compara os modelos de Bradford (Leimkuhler, 1967), Egghe (1990) e o método de Faleschini em termos de periódicos por zona.

Tabela 2 - Comparação dos modelos de Bradford, Egghe e Faleschini

Zona	Bradford	Egghe	Faleschini
1ª (núcleo)	8 periódicos (total de 110 artigos, variando entre 9 a 22 artigos)	10 periódicos (total de 126 artigos, variando entre 8* a 22 artigos)	1 periódico (este único periódico apresentou 22 artigos)
2ª	29 periódicos (total de 133 artigos, variando entre 3 a 8 artigos)	35 periódicos (total de 133 artigos, variando entre 2** a 8* artigos)	23 periódicos (total de 182 artigos, variando entre 4 a 18 artigos)
3ª	127 periódicos (total de 152 artigos, variando entre 1 a 2 artigos)	119 periódicos (total de 136 artigos, variando entre 1 a 2** artigos)	38 periódicos (total de 89 artigos, variando entre 2 a 3 artigos)
4ª	Não aplicável	Não aplicável	102 periódicos (total de 102 artigos, cada periódico com apenas 1 artigo cada)

Fonte: Dados da pesquisa, compilados de Leimkuhler (1967) e Egghe (1990).

Notas:

* Egghe (1990) dividiu três periódicos com oito artigos cada da seguinte maneira (mas sem citar qual critério periódicos de mesma quantidade publicada ficam em zonas distintas): dois periódicos na 1ª zona (ou núcleo) e um periódico na 2ª zona.

** Egghe (1990) dividiu 25 periódicos com dois artigos cada da seguinte maneira (mas novamente sem citar qual critério periódicos de mesma quantidade publicada ficam em zonas distintas): 17 periódicos na 2ª zona e oito periódicos na 3ª zona.

Leimkuhler (1967) reportou que, para os dados de Lubrificação de Bradford, a formulação verbal gerou os seguintes resultados: 1ª zona (ou núcleo) contendo oito periódicos que publicaram 110 artigos. Observando as Tabelas 1 e 2, verifica-se que estes oito periódicos publicaram entre nove a 22 artigos. Para a 2ª zona, Leimkuhler (1967) apontou 29 periódicos com 133 artigos; e a verificação das Tabelas 1 e 2 revela que são periódicos que publicaram entre 3 a 8 artigos. Para a 3ª e última zona, Leimkuhler (1967) mostrou 127 periódicos com 152 artigos, e novamente, as Tabelas 1 e 2 indicam que trata-se de periódicos com 1 a 2 artigos publicados sobre o tema.

Com base em Leimkuhler (1967) e analisando as Tabelas 1 e 2, vale a pena reforçar aqui que Bradford preocupou-se em garantir a separação em zonas distintas de periódicos com diferentes números de artigos.

Já a proposta de Egghe (1990) apresenta as três zonas da seguinte maneira: 1ª zona (ou núcleo) com dez periódicos que publicaram 126 artigos; 2ª zona com 35 periódicos que produziram 133 artigos; a 3ª zona com 119 periódicos com 136 artigos. Egghe (1990) alegou que seus resultados são melhores

que os originais de Bradford, citado por Leimkuhler (1967).

Todavia, uma análise detalhada das Tabelas 1 e 2 revela que Egghe (1990) dividiu em zonas distintas periódicos com a mesma quantidade publicada de artigos, conforme detalhado na legenda da Tabela 2, gerando, em tese, o seguinte questionamento: qual o critério Egghe (1990) utilizou para separar em zonas diferentes periódicos cujos resultados se apresentaram idênticos? Neste sentido, advogamos que trata-se de inconsistência matemática, que pode ser melhorada pelo método de Faleschini proposto neste trabalho.

O método de Faleschini apresentou cinco zonas, contudo, a adequação aos dados reais apontou a existência de somente quatro zonas: a 1ª zona (ou núcleo, periódicos com elevada produção científica) com somente um periódico que publicou 22 artigos. A 2ª zona (periódicos com alta produção de artigos), com 23 periódicos que publicaram entre quatro a 18 artigos; a 3ª zona (moderada produção) com 25 periódicos que publicaram dois a três artigos; e a 4ª e última zona (baixa produção) com os 102 periódicos que publicaram somente um artigo. Não houve presença de raríssima produção, e há a adequação dos dados reais às zonas de Bradford.

Contudo, a presença de um único periódico como núcleo e os elevados valores de assimetria ("A") e curtose ("B") associados à distribuição de frequência dos dados de Lubrificação de Bradford (Leimkuhler, 1967) de fato sugerem a possibilidade de novas pesquisas e contribuições ao método de Faleschini.

4 CONSIDERAÇÕES FINAIS

Este trabalho apresentou e demonstrou a viabilidade do método de Faleschini para uso na Ciência da Informação, aplicando-o para a lei de Bradford e apontando conexões com a lei de elite de Price e rotulação de *outliers*.

O método de Faleschini oferece uma discussão científica que a comunidade acadêmica deve enfrentar ou apresentar soluções. Emergem, nesse sentido, questionamentos como: irá se aceitar a solução de continuidade para o cálculo dos percentis e quartis proposto? Quais novas fórmulas podem ser propostas para as equações (3) a (6)? Deve-se levar em conta o tamanho amostral? A equação (1) sempre irá apresentar quatro raízes reais? Pode-se aplicar o método de Faleschini para o cálculo do fator do impacto ou ser um substituto para o índice H de Hirsh, se esses forem vistos como distribuição de frequência?

Em que pese os questionamentos do parágrafo anterior, a rigor, o método de Faleschini apresenta vantagens conceituais em relação a Egghe (1990), ao levar justamente em conta a forma da distribuição dos dados, caracterizada por " μ ", " σ ", "A" e "B".

Trata-se de contribuição inicial à área, necessitando de pesquisas, ajustes e novas contribuições por parte da comunidade científica e acadêmica. Recomenda-se estudos na área da Ciência da Informação para verificar a aplicabilidade do método de Faleschini para o cálculo do fator de impacto e para o índice H de Hirsh, se estes forem encarados como distribuição de frequência. Dada a generalidade do método de Faleschini para distribuição de frequência de dados univariados, outros usos e utilizações



em qualquer ramo da Ciência podem ser vislumbradas.

REFERÊNCIAS

EGGHE, L. Applications of the Theory of Bradford's Law to the Calculation of Leimkuhler's Law and to the Completion of Bibliographies. **Journal of the American Society for Information Science**, [S. l.], v. 41, n. 7, p. 469-492, Oct. 1990. DOI: [https://doi.org/10.1002/\(SICI\)1097-4571\(199010\)41:7<469::AID-ASI1>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1097-4571(199010)41:7<469::AID-ASI1>3.0.CO;2-P). Disponível em: [https://asistdl.onlinelibrary.wiley.com/doi/10.1002/\(SICI\)-1097-4571\(199010\)41:7%3C469::AID-ASI1%3E3.0.CO;2-P](https://asistdl.onlinelibrary.wiley.com/doi/10.1002/(SICI)-1097-4571(199010)41:7%3C469::AID-ASI1%3E3.0.CO;2-P). Acesso em: 30 abr. 2024.

FALESCHINI, Luigi. Su alcune proprietà dei momenti impegnati nello studio della variabilità, assimetria e curtosi. **Statistica**, [S. l.], v. 8, n. 4, p. 503-513, out./dez. 1948.

FIORI, Anna Maria; ZENGA, Michele. The meaning of kurtosis, the influence function and an early intuition by L. Faleschini. **Statistica**, Milão, v. 65, n. 2, p. 135-144, 2005. DOI: <https://doi.org/10.6092/issn.1973-2201/82>. Disponível em: <https://rivista-statistica.unibo.it/article/view/82>. Acesso em: 30 abr. 2024.

LEIMKUHLER, Ferdinand F. The Bradford Distribution, **Journal of Documentation**, v. 23, n. 3, p.197-207, Mar. 1967. DOI: <https://doi.org/10.1108/eb026430>. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/eb026430/full/html>. Acesso em: 30 abr. 2024.

LIMA, Luís Fernando Maia et al. Proposta de um critério auxiliar para a determinação da elite científica. In: BORGES, Maria Manuel; CASADO, Elias Sanz (coord.). **A ciência aberta: o contributo da Ciência da Informação: atas do VIII Encontro Ibérico EDICIC**. Coimbra: Universidade de Coimbra, 2017. p. 301-310. Disponível em: <https://purl.org/sci/atas/edicic2017>. Acesso em: 9 abr. 2024.

SILVA, Kelly Cristina Ramos da. **Regras robustas para rotular outliers em dados de caudas leves e caudas pesadas**. 2019. 106 p. Tese (Doutorado em Ciências da Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2019. DOI: <https://doi.org/10.11606/T.55.2019.tde-29042019-145141>. Disponível em: <https://teses.usp.br/teses/disponiveis/55/55134/tde-29042019-145141/pt-br.php>. Acesso em: 9 abr. 2024.

TRIOLA, Mario F. **Introdução à Estatística**. 10. ed. Rio de Janeiro: LTC, 2012.