

A produtividade dos autores sobre a Lei de Lotka

Ruben Urbizagastegui

Doutor em ciência da informação pela Universidade Federal de Minas Gerais.
E-mail: ruben@ucr.edu

Resumo

Estuda a produção estratificada dos autores produtores da literatura sobre a Lei de Lotka de 1922 a 2003 e analisa essa produtividade através dos modelos Poisson lognormal e Gauss Poisson inversa generalizada. Para tanto, faz uso dos três tipos de contagem da literatura produzida: contagem direta, contagem completa e contagem fracionada. Os dados da pesquisa são avaliados usando o teste qui-quadrado ao 0.05 nível de significância. Ambos os modelos ajustam-se muito bem à distribuição da literatura produzida, porém a distribuição Poisson Gauss inversa generalizada produz um chi-quadrado menor e prediz melhor o total de autores do que a distribuição Poisson Lognormal.

Palavras-chave

Lei de Lotka. Productividad de autores. Modelo Poisson lognormal. Modelo Gauss Poisson inverso generalizado. Cienciometria. Bibliometria. Informetria.

Authors' productivity on Lotka's Law

Abstract

The stratified literature produced from 1922 to 2003 on Lotka's law is analyzed by using the Poisson lognormal and the Poisson Gauss inverse generalized models. For this purpose, three types of how the literature produced is counted are used: straight, complete and fractioned ones. The data of the research are evaluated by using the chi-square test at 0.05 significance level. Both models fit very well into the distribution of literature produced, but the Poisson Gauss inverse generalized distribution produces a smaller chi-square and better anticipates the total number of authors than the Poisson Lognormal distribution.

Keywords

Lotka's law. Authors' productivity. Scientometrics. Bibliometrics. Informetrics. Poisson lognormal model. Gauss-Poisson inverse generalized model.

INTRODUÇÃO

A produção bibliográfica dos autores tem sido objeto de diversos estudos em numerosos campos do conhecimento; não obstante, o estudo aplicando a mesma metodologia a um campo quase homogêneo como a literatura sobre a lei de Lotka ainda não tem sido explorado. Na realidade, a produção bibliográfica como um indicador do trabalho científico vem sendo desenvolvida por numerosos pesquisadores (EGGHE, 2003; HUBER, 2002; BOOKSTEIN, 2001; TSAY; JOU & MA, 2000; WAGNER-DÖBLER & BERG 1999).

Este estudo sobre a produtividade dos autores é importante porque tal modelo tem se convertido no eixo central da pesquisa bibliométrica contemporânea. Para essa comunidade de pesquisadores em bibliometria, a geração de novos conhecimentos está diretamente relacionada ao desenvolvimento de novas pesquisas, sendo a regra que uma pesquisa só está terminada quando é publicada. Essa publicação pode ser concretizada em vários suportes físicos, tais como artigos de periódicos, capítulos de livros, trabalhos apresentados em congressos e similares canais de informação capazes de tornar públicos os resultados de uma pesquisa.

Segundo Ziman (1979), a função do cientista é produzir e publicar trabalhos originais, comunicá-los a seus pares e, assim, contribuir para o conhecimento público. Também Price (1975) afirma que qualquer descoberta do cientista, por maior ou menor que seja, só se converte em efetiva contribuição à ciência quando é julgada, publicada, incorporada ao estoque de conhecimentos e depois usada pelos pares. Segundo Castro (1985), fazer pesquisa não é o mesmo que fazer ciência, fazer ciência seria mais abrangente. No entanto, Lungarzo (1989) entende que o

conhecimento científico é submetido a uma série de testes, análises, controles que garantam pelo menos uma "chance" alta de obter informações verdadeiras e justificadas... O cientista recorre aos fatos reais para equipará-los a conhecimento. A elaboração desse conhecimento produz teorias. Por sua vez, essas teorias são submetidas novamente à realidade para se conferir sua validade.

Entende-se, então, que a atividade científica deve materializar-se em documentos escritos validados e legitimados pela comunidade científica. O conjunto

desses documentos, por sua vez, constitui importantes indicadores do estágio do desenvolvimento de uma área do saber. Todo cientista, especialmente o jovem, sofre pressões sociais, uma vez que anseia contribuir significativamente para o conhecimento científico. Isto porque tanto o mérito quanto a qualidade e originalidade de seus trabalhos fundamentam-se no julgamento de seus pares. A esse desejo de publicar tão rapidamente quanto possível junta-se o desejo de publicar tanto quanto possível. Embora a qualidade do trabalho precise ser, é claro, considerada juntamente com a quantidade, há uma convicção generalizada de que quanto mais um cientista puder publicar, melhor será para a sua reputação. Na opinião de Castro (1986),

a estratificação social dentro da ciência é razoavelmente rígida e estável. As regras para subir são conhecidas e baseadas em realizações tangíveis. Adquirem prestígio e poder aqueles que mais contribuem para o avanço da ciência. Mas não se pode esperar décadas para a confirmação dessa contribuição, como acontece com o Prêmio Nobel ... Desta forma, um reconhecimento mais imediato e pleno de gradações e nuances é dado pelas publicações. Tem maior cotação na bolsa de prestígio dos cientistas aqueles que publicam mais, aqueles que publicam em periódicos ou editoras mais prestigiadas e aqueles que são citados mais freqüentemente por seus pares.

De certo modo, na literatura dedicada a explorar a Lei de Lotka se verificará também um padrão estratificado da produção científica e essa estratificação poderá muito bem se ajustar ao modelo em estudo. Nesta pesquisa pretende-se verificar e evidenciar esse padrão do comportamento estratificado dos produtores de literatura sobre a Lei de Lotka e analisar se essa produtividade poderá ser adequadamente descrita pelos modelos Poisson lognormal e Gauss Poisson inverso generalizado.

REFERENCIAL TEÓRICO

Um dos problemas que emergem constantemente no mundo acadêmico com relação à prática científica é a produtividade dos seus participantes na forma de publicações, tanto nos seus aspectos quantitativos quanto nos qualitativos. Esta tendência tem dado lugar a reflexões críticas sobre o que agora é considerado como a obrigação de publicar, e a existência de um grupo de significativos contribuintes em qualquer campo do conhecimento. Por conseguinte, pode-se perguntar se a contribuição dos grandes produtores é de menor, igual, ou maior qualidade que a contribuição dos menores

produtores. Cattell (1910) afirmava que não se sabia se o progresso da ciência se devia ao grande número de fiéis trabalhadores ou à genialidade de uns poucos. Nisso parecia discordar de Lambroso (1891:120), sustentava “que o aparecimento de um único gênio era mais do que equivalente ao nascimento de centenas de mediócras”.

A característica distinta de um gênio, entre outras, seria sua grande produtividade de documentos científicos. Essa característica foi ilustrada pelas afirmações de Barron (1963: 139), no sentido de que uma “*volumosa produtividade é a regra e não a exceção entre os indivíduos que têm feito alguma notável contribuição à ciência*”. Essa é também a posição de Simonton (1984), ao afirmar que a característica de um gênio de outro tipo de profissional é sua imensa produtividade, e reafirma isto quando diz, citando Albert (1975), que

a mais comum falsa percepção de que um intelecto fenomenal contribui com um escasso seletivo número de trabalhos mestres, ou ainda um único magnus opum, é totalmente errado. Ao final de sua carreira, Darwin podia contabilizar 119 publicações, Einstein 248, e, em psicologia, Galton 227, Binet 277, James 307, Freud 330, e Maslow 165 (Albert *apud* Simonton, 1988: 60).

Outros cientistas advogam uma tese contrária, endossando o que vem sendo citado como a “Hipótese de Ortega”, que ressalta:

é necessário insistir no fato extraordinário porém inegável: a ciência experimental tem progredido graças ao trabalho de homens assombrosamente mediócras e ainda menos do que mediócras. Isto quer dizer que a ciência moderna, raiz e símbolo de nossa civilização atual, cede lugar até para o homem intelectualmente comum e lhe permite trabalhar com sucesso (ORTEGA y GASSET, 1932:110-111).

Duffrenoy (1938), observando a produtividade de autores no campo da biologia sugeria que os

autores [...] que publicam mais de cinco artigos constituem um grupo que não pertence à mesma série daqueles autores que contribuem com um, dois, três ou quatro artigos.

Williams (1944) rebatia estas afirmações sustentando que os escritores de cinco ou mais artigos poderiam estar numa categoria psicologicamente diferente, mas que ele estava tentado a dizer que apenas eram mais prolíficos ou mais verborreicos, mas que devemos lembrar que os

dados [de Dufrenoy] não levavam em consideração a extensão dos artigos, mas só a quantidade, e que era possível um longo artigo ser contabilizado por vários pequenos artigos.

Qual destas afirmações teria sentido numa sociedade ou numa comunidade científica na qual múltiplas variáveis condicionam a produtividade dos autores? Shockley (1957) sugeria que, se a produtividade era o produto multiplicativo de muitos fatores, todos normalmente distribuídos, a consequência seria uma distribuição lognormal altamente inclinada à direita e com uma alongada cauda superior; porém ele nunca especificou a natureza nem o número dos fatores teórica ou empiricamente envolvidos nessa produtividade.

Para encontrar essas respostas, um dos mecanismos seria analisar a contribuição feita pelos cientistas durante toda a sua vida, e para isso seria necessário ter acesso a uma amostra de dados sobre um período muito longo. Dessa forma seria possível achar respostas relativas aos muitos ou poucos colaboradores e aos fatores envolvidos nas suas produtividades. Na presente década, os pesquisadores se inclinam a sustentar que essas produtividades são o resultado de algum mecanismo do processo da vantagem acumulativa, mediante o qual o rico fica mais rico e o pobre fica ainda mais pobre. Na opinião de Simonton (1988:63),

este mecanismo é mais sociológico que psicológico, dependendo do que ele faz na seleção qualitativa do sistema de premiação científica. Porque o acesso a posições de pesquisa, espaços nos periódicos, bolsas de estudo, e honras e recompensas estão rigidamente restritas e aqueles que têm êxito primeiro obtêm um espaço que lhes permite ter êxito também depois, no entanto, aqueles que começam com o pé errado tendem a ter um padrão de fracassos que eventualmente os joga fora da competição pelos recursos. Sucesso gera mais sucesso; fracasso gera só fracasso.

Isso é também conhecido como o “Efeito Mateus”. É evidente que uma vez que um cientista atinge alta visibilidade profissional, sua oportunidade de publicar novos trabalhos será melhor do que daqueles colegas que ainda não estabeleceram essa visibilidade ou reputação, significando que “altas taxas de publicação se correlacionam fortemente com reconhecimento e distinções científicas” (VALCHY, 1970: 501).

Em virtude de diversos fatores sociais, os autores mais produtivos tendem a ser ainda mais produtivos no decorrer do tempo; contudo os autores menos produtivos

mostram uma tendência a declinar em produtividade. Nesse sentido, a distribuição da produtividade dos cientistas seria altamente elitista, com uma pequena proporção da totalidade da comunidade científica produzindo a maioria das contribuições, bem como

numerosos estudos já clássicos podem ser citados para demonstrar uma surpreendente correlação muito próxima entre os indicadores quantitativos da produtividade científica, por um lado, e o reconhecimento científico ou o alcance da eminência, como contribuinte de um campo científico, por outro lado (VLACHY, 1974:2).

Como já se sabe, a produtividade na forma de publicações, considerada como a parte com que diferentes pessoas contribuem para o progresso da ciência, foi estudada por Lotka (1926), que estabeleceu os fundamentos da *lei do quadrado inverso*, afirmando que o número de autores que fazem n contribuições num determinado campo científico é aproximadamente $1/n^2$ daqueles que fazem uma só contribuição, e que a proporção daqueles que fazem uma única contribuição é de mais ou menos 60 por cento.

Desde 1926, época em que Lotka estabeleceu esse modelo, muitos estudos têm sido realizados para pesquisar a produtividade dos autores em distintas disciplinas. Até dezembro de 2003, aproximadamente 390 trabalhos entre artigos, capítulos de livros, comunicações em congressos e literatura cinza tinham sido produzidos criticando, replicando e/o reformulando esta lei bibliométrica. Não obstante, apesar das numerosas pesquisas realizadas sobre este assunto, os resultados são contraditórios, conflitantes e parecem não proporcionar clara validade desta lei.

Desvios significativos da forma original observada por Lotka têm sido encontrados na prática e sugerido que a formulação do quadrado inverso seja teoricamente suspeita. Isto fez com que o modelo do quadrado inverso – ignorado e considerado apenas como uma curiosidade estatística por muito tempo – tenha se convertido em tópico de pesquisa intensiva. Esta intensidade parece ser consequência de suas implicações teóricas e de sua potencial importância para o gerenciamento da informação científica e técnica.

Portanto, formulações alternativas que se ajustem melhor aos dados observados, que não o modelo do quadrado inverso, estão em processo de pesquisa. Isto como resultado de estudos anteriores mostrando que tanto a abrangência quanto a aplicabilidade da Lei de Lotka ainda são muito limitadas:

primeiro, porque suas amostras não iam além dos dados originais analisados por Lotka e seus primeiros seguidores; segundo, porque alguns conceitos básicos envolvidos nesses estudos eram antecipados sem sequer ser profundamente analisados (VLACHY, 1976:46).

Também existem discordâncias em relação à forma do teste. Por exemplo, Nicholls (1989: 383) reclama que os resultados desses estudos não são comparáveis devido às diferenças substanciais na forma da medição, estimação dos parâmetros, formas dos testes, e ainda às interpretações do modelo. Também Oppenheim (1986) afirma que deve se enfatizar que a lei de Lotka tem sido testada em muitas coleções de dados, porém, o ajuste nem sempre tem sido bom.

Entretanto, foi sugerido que para a literatura da ciência da informação era necessário um expoente de $n = 3.5$, para proporcionar um bom ajuste aos dados empíricos (Voos, 1974). Já Schoor (1974: 33) propôs um expoente de $n = 4$, em vez de $n = 2$ no modelo do quadrado inverso de Lotka, afirmando que para

o caso da ciência da informação a produção acadêmica seguiria uma lei do poder quádruplo inverso, onde por cada cem contribuintes de um único artigo, aproximadamente seis contribuiriam com dois artigos, aproximadamente um contribuiria com três artigos, e quase nenhum autor produziria quatro artigos ou mais.

Worthen (1978) informou que quando se usava o modelo do quadrado inverso, a Lei de Lotka não se ajustava à literatura sobre produtos farmacêuticos. Também Radhakrishnan & Kernizan (1979) verificaram que o modelo do quadrado inverso de Lotka não se ajustava bem à literatura da ciência da computação. Não obstante, o modelo do cubo inverso descrevia melhor esta literatura.

Subramanyam (1979), analisando a literatura da ciência da computação de 1973 a 1975 (um período de três anos), constatou que esta literatura se ajustava bem ao modelo do quadrado inverso de Lotka e que o modelo do cubo inverso sugerido por Radhakrishnan & Kernizan (1979) não descrevia o comportamento dessa literatura. Essa falta de ajuste dos dados empíricos à distribuição teórica da produtividade dos autores levou Pao (1985, 1986) e Nicholls (1986, 1987, 1989) a propor o modelo do *poder inverso generalizado* como alternativa ao modelo do *quadrado inverso*.

Junto com essas críticas têm aparecido discordâncias relacionadas às formas de realizar a contagem na coleção dos dados relativos à autoria múltipla. Na literatura publicada, no momento se detecta a existência de três formas de contagem usadas: a *contagem direta*, quando somente os autores “sênior” ou principais (os autores nomeados em primeiro lugar) têm contada a sua contribuição e os autores secundários (colaboradores) são ignorados; a *contagem completa*, em que cada autor (principal e/o secundário) é computado com uma contribuição; e a *contagem ajustada*, na qual cada autor (principal e/o secundário) é contado com uma fração ou porção da contribuição total, isto é, se são identificados cinco autores de um só artigo, cada um seria computado com $1/5$ (uma quinta parte) do trabalho. Não obstante, alguns autores afirmam que a contagem direta e a contagem ajustada não produzem diferenças essenciais e que

os dois meios de contagem produzem a mesma coisa e que, por isto, não seria necessário considerar a contagem ajustada e que deveria se prestar maior atenção à contagem direta (Nath & Jackson, 1991: 207).

Outro ponto de conflito refere-se ao período de cobertura do estudo. Uma das condições de seu emprego se refere ao período da cobertura, que deve ser suficientemente longo, segundo Price (1963: 43), “para permitir que os que publiquem mais de um par de trabalhos, o façam”. Porém este limite temporal não tem sido fixado pelos estudos, ainda que “o habitual [seja] o modelo à literatura superior a uma década” (Terrada, 1973: 87). Também Potter (1981), em extensa revisão da literatura sobre a Lei de Lotka, afirma que

quando o período de cobertura é dez anos ou mais e a comunidade de autores é definida amplamente, a produtividade dos autores se aproxima à distribuição de freqüências observada por Lotka, conhecida como a Lei de Lotka.

Mas, outros estudiosos parecem não se importar muito com o período de cobertura e têm verificado que a produtividade dos autores se aproxima da distribuição de freqüências observada por Lotka, embora o período da cobertura seja de um ou mais anos. Por exemplo, os resultados de estudos anteriores mostram que tanto a abrangência quanto a aplicabilidade da Lei de Lotka ainda são muito limitadas, portanto,

uma das conclusões mais primárias sobre as relações de Lotka eram de que a mesma distribuição se mantinha para períodos tão longos como para um

centenário. Por outro lado, sempre havia necessidade intuitiva de que as fontes bibliográficas usadas para a contagem se estendessem por um número de anos suficientemente longo, como para permitir que os autores produzissem mais artigos ou outras contribuições quantificáveis (VLACHY, 1976:46).

Disto parecia discordar Bookstein (1977:208), ao afirmar que

uma das propriedades mais surpreendentes da Lei de Lotka é sua aparente falta de preocupação com o período considerado; assim, se analisamos as publicações num período de cinco anos ou num período maior, a distribuição se ajusta ao modelo e a lei permanece como verdadeira.

Outro ponto de conflito refere-se ao método empregado para a coleção dos dados. Por exemplo, Pao (1985) ressalta que “não existe um método uniforme para a coleta e organização dos dados” para testar a Lei de Lotka, assim como “muitos pesquisadores simplesmente deram 2 como o valor de n sem estimar o valor dos dados observados”. Esta preferência é atribuída ao fato de que os procedimentos de cálculo realizados dessa maneira são muito simples e fáceis de realizar. A mesma autora afirma também que “muitos pesquisadores evitaram fazer um teste apropriado do grau de ajuste dos dados observados” ante os dados esperados; como consequência, Vlachy (1974) teria encontrado sérias discrepâncias entre os dados empíricos e o ajuste da lei do quadrado inverso.

De certo modo, na literatura dedicada a explorar a Lei de Lotka se verificará também um padrão estratificado da produção científica e essa estratificação poderá muito bem se ajustar ao modelo de Lotka. Nesta pesquisa pretende-se verificar e evidenciar esse padrão do comportamento dos produtores de literatura sobre a Lei de Lotka e analisar se essa produtividade poderá ser adequadamente descrita pelos modelos Poisson lognormal e Gauss Poisson inverso generalizado.

METODOLOGIA

Este trabalho é empreendido como um intento de aplicar a Lei de Lotka à literatura de Lotka. Também se considera, aqui, a possibilidade de identificar os autores mais produtivos ou a “elite” desta subárea da bibliometria. Por essa razão, como unidades de análises foram tomados cada um dos autores e dos artigos, capítulos de livros e trabalhos apresentados em congressos, sobre a chamada “Lei de Lotka”, desde 1922 até 2003, isto é, um longo período de 82 anos.

Para identificar os autores contribuindo com artigos nesta área, foi feita uma busca usando a estratégia Dial Index em todas as bases de dados de DIALOG, com os termos Lotka?(5n)Law? Produzindo 50 bases de dados que pelo menos continham um artigo sobre o assunto pesquisado. Esta estratégia de busca produziu o total de 315 registros que, depois da depuração dos duplicados e falsas recuperações, foram acumulados no total de 275 referências bibliográficas. Essas referências foram depois trasladadas ao PROCITE 5.0 para a elaboração de uma base de dados específica sobre o assunto.

Posteriormente foi realizada criteriosa leitura de cada um dos artigos identificados na busca, dedicando especial atenção a cada citação efetuada no documento lido. Depois, cada citação referente à Lei de Lotka foi confrontada com a base de dados e nela incorporada, se não tivesse sido identificada na busca anterior no DIALOG. Também se realizaram buscas no Information Science Abstract (ISA), Library Literature (LL) e Library and Information Science Abstract (LISA). Com esta leitura minuciosa foi produzida uma bibliografia analítica da Lei de Lotka que lista o total de 390 referências bibliográficas contendo artigos de periódicos, monografias, capítulos de livros, comunicações em congressos, literatura gris, cartas dirigidas a editores de revistas especializadas em biblioteconomia e ciência da informação (URBIZAGASTEGUI, 2005). A bibliografia analítica dessas referências, produzidas entre 1922 e 2003, constitui o universo desta pesquisa.

O período coberto pela literatura recuperada é suficientemente longo para assegurar a publicação de artigos sobre este assunto nos periódicos do campo da ciência da informação e afins. Por isso, conforme a observação de Potter (1981), espera-se um bom ajuste desta literatura à Lei de Lotka. Para aferir os produtores de artigos, optamos pelo sistema de contagem direta, completa e ajustada. Isto significa, ainda, que os múltiplos autores de um só artigo serão contados e atribuídos como contribuintes à produção de cada artigo ou documento identificado no levantamento bibliográfico. Este procedimento também fornecerá mecanismos de comparação entre os parâmetros que compõem o modelo de Lotka.

Para a mensuração dos dados consideramos o modelo Poisson lognormal, bem como a distribuição Gauss-Poisson inversa generalizada. A natureza de ambos os modelos estatísticos são descritos mais adiante. Como nas distribuições bibliométricas geralmente a frequência de zero produção quase nunca é observada nos dados

coletados, para se obter as probabilidades da distribuição deve-se fazer uso da forma zero truncada.

A distribuição Poisson Lognormal

Segundo Steward (1994), geralmente a distribuição dos dados coletados para uma análise da produtividade dos autores é da forma J inverso zero truncada, com uma longa cauda de pequenos produtores. Isso faz com que o modelo Poisson-Lognormal seja um candidato ideal para testar este tipo de distribuições discretas. Ele descreve este modelo como uma distribuição composta, na qual a propensão subjacente δ dos cientistas para publicar um artigo segue uma distribuição lognormal. Dada a propensão subjacente δ específica de um cientista, sua probabilidade P_x de publicar x artigos segue um simples modelo Poisson:

$$P_x = \frac{\delta^x e^{-\delta}}{x!} \quad \text{para } x = 0, 1, 2, 3, \dots \quad (1)$$

Assim, a ordem dos valores observados de todos os autores que tenham o mesmo valor δ terão distribuição com uma média e uma variância de δ . Numa amostra de autores cujos logaritmos de δ estão normalmente distribuídos com uma média μ e um desvio padrão σ , as P_x da amostra total são proporcionadas por

$$P_x = \frac{1}{\sigma \sqrt{2\pi}} \frac{1}{x!} \int_0^\infty e^{-\delta} \delta^{x-1} \exp\left[-\frac{(\ln \delta - \mu)^2}{2\sigma^2}\right] d\delta \quad (2)$$

para $x = 0, 1, 2, 3, \dots$

Esta equação tem de ser estimada por métodos numéricos. Note-se que nela x pode ter valores de zero, sem criar nenhum problema para a estimação dos valores esperados. Se a distribuição é da forma zero truncada, Bulmer (1974) oferece as equações pertinentes.

A distribuição Gauss-Poisson inversa generalizada

A distribuição Gauss-Poisson inversa generalizada foi desenvolvida por Sichel (1982) como uma forma de compreender melhor o comportamento das compras de bens e serviços pelos consumidores num mercado de iguais probabilidades. Sichel (1982, 1985, 1986, 1992) tem experimentado sistematicamente este modelo, tanto que agora ele é conhecido como a “distribuição de Sichel”. A introdução deste modelo é considerada como um dos maiores eventos ocorridos nos estudos teóricos da ciência da informação contemporânea, embora, no início, muitos autores fossem céticos sobre seus

resultados e potencialidades (COCKS & BROOKES, 1986; BROOKES 1985, 1988). Não obstante, até hoje ainda são escassos os estudos aplicando esse modelo à produtividade dos autores. Tais estudos se reduzem aos trabalhos de Burrell & Fenton (1993), Urbizagastegui (2001, 2002, 2003), Urbizagastegui & Oliveira (2001), e Kuperman (2005).

A distribuição Gauss-Poisson inversa generalizada foi desenvolvida como um modelo de três parâmetros e como sendo aproximadamente disposta na forma de uma distribuição de Poisson. Então,

$$p(r | t\lambda) = \frac{e^{-t\lambda} (t\lambda)^r}{r!} \quad (3)$$

onde

$r = 0, 1, 2, 3, \dots, \infty$ $\lambda > 0$ e $t > 0$.

Para a distribuição dos λ 's (lambdas) toma-se a distribuição de Gauss inversa generalizada (Sichel, 1971), na forma de:

$$f(\lambda) = \frac{\left\{ \frac{2(1-\theta)^{\frac{1}{2}}}{\alpha\theta} \right\}^\gamma}{2K_\gamma \left\{ \alpha(1-\theta)^{\frac{1}{2}} \right\}} \lambda^{\gamma-1} e^{-\left(\frac{1}{\theta}\right)\lambda - \left(\frac{\alpha^2\theta}{4\lambda}\right)} \quad (4)$$

onde

$-\infty < \gamma < \infty$, $0 \leq \theta \leq 1$, $\alpha \geq 0$ e $K_\gamma(z)$ é a função modificada de Bessel de segunda ordem γ com argumento z .

A distribuição mista (4) é uma distribuição gama (Pearson tipo III), se $\gamma > 0$ e se deixa que $\alpha \rightarrow 0$. Entretanto, se $\gamma < 0$ e $\theta \rightarrow 0$, obtém-se a distribuição de Pearson tipo V. Por isso é que a distribuição (4) é mais rica que a distribuição gama original. Então, o número de unidades em determinado período t é dado por

$$\phi(r | t) = \int_0^\infty p(r | t\lambda) f(\lambda) d\lambda \quad (5)$$

A solução desta função integral foi dada por Sichel (1971) para o caso especial em que $t = 1$. Para a condição mais geral da equação (5) obtém-se:

$$\phi(r | t) = \frac{(1-\theta)_t^\gamma}{K_\gamma \left\{ \alpha_t (1-\theta)_t^{\frac{1}{2}} \right\}} \frac{\left(\frac{\alpha_t \theta_t}{2} \right)^r}{r!} K_{r+\gamma}(\alpha_t) \quad (6)$$

onde

$$\alpha_t = \alpha [1 + (t - 1)\theta]^{\frac{1}{\gamma}} \quad (7)$$

y

$$\theta_t = \frac{t\theta}{1 + (t - 1)\theta} \quad (8)$$

As variações particulares e as características de dependência em relação ao tempo e demais propriedades deste modelo são dadas por Sichel (1982) e não precisam de maiores considerações neste trabalho.

As probabilidades individuais da equação (6) são obtidas da equação (9) de recorrência geral, dada como:

$$\hat{\phi}(r) = \left(\frac{r + \gamma - 1}{r}\right) \theta_t \phi(r-1) + \frac{\alpha_t^2 \theta_t^2}{4r(r-1)} \quad (9)$$

Tudo o que inicialmente se necessita é simplesmente conhecer as duas primeiras probabilidades de $r = 0$ e de $r = 1$. Como a forma da distribuição da equação (6) se altera gradativamente conforme muda o parâmetro g , a equação (6) pode ser bem simplificada fazendo com que g tenha um valor de meio integral. Então se supõe que $\gamma = -\frac{1}{2}$ a priori. Daqui segue-se que a correspondente distribuição discreta da equação (6) é

$$\phi(r) = \left(\frac{2\alpha}{\pi}\right)^{\frac{1}{2}} \exp\{\alpha(1-\theta)^{\frac{1}{2}}\} \frac{\left(\frac{\alpha\theta}{2}\right)^r}{r!} K_{r-\frac{1}{2}}(\alpha) \quad (10)$$

Esta é a chamada distribuição de Gauss-Poisson inversa generalizada.

Como a distribuição representada pela equação (10) é da forma de um J fortemente inverso os parâmetros α e θ que a descrevem são eficientemente estimados da proporção da frequência dos zeros $\phi(0)$ e da \bar{r} meia da amostra. Este método é eficiente quando a proporção de zero observações de uma distribuição de frequências é relativamente alta. Existem três métodos para calcular os valores da distribuição Gauss-Poisson Inversa Generalizada, mas neste trabalho será usado o método da primeira proporção de zeros e a média da amostra para uma distribuição zero truncada.

Nos estudos das distribuições da produtividade autoral, o número de autores que não produzem um só item em determinado período às vezes não é observado. Então, a ocorrência dos “singletons”, isto é, o número de autores que produzem um só item nesse período é mais freqüente que os sugeridos pela distribuição logarítmica seriada ou a distribuição binomial negativa zero truncada.

Sichel (1971) tem demonstrado que o número das unidades compradas em determinado período t é dado pela equação (5). Se em casos como este se faz que na equação (6) $t = 1$, trunca-se a função da probabilidade, reverte-se o signo do parâmetro γ e deixa-se que α tenda a zero, obtém-se uma interessante distribuição de dois parâmetros que sempre têm a forma de um J invertido. Sua equação pode ser representada como

$$\phi(r) = \left[(1-\theta)^y - 1 \right]^{-1} \binom{y}{r} (-\theta)^r \quad (11)$$

onde

$$r = 1, 2, 3, \dots, \infty, \quad 0 < \gamma < 1, \quad y \quad 0 < \theta < 1.$$

Como a distribuição representada pela equação (11) com $0 < \gamma < 1$ é da forma de um J fortemente inverso, os parâmetros γ e θ que descrevem esta distribuição são eficientemente estimados da proporção da primeira freqüência $\phi(1)$ e da \bar{r} meia da amostra. A probabilidade dos primeiros singletons da equação (11) é dado por

$$\phi(1) = \frac{y\theta}{1 - (1-\theta)^y} \quad (12)$$

Depois, todas as outras probabilidades são calculadas da seguinte fórmula de recorrência geral:

$$\phi(r+1) = \left(\frac{r-y}{r+1}\right) \theta \phi(r) \quad (13)$$

Os dados foram codificados manualmente e analisados usando o pacote estatístico SSPS na versão 15.0 para Windows e Mathematica 5.0. O ajuste dos dados observados aos modelos Poisson-Lognormal e Gauss-Poisson inverso generalizado será avaliado usando o teste qui-quadrado ao 0,05 nível de significância.

RESULTADOS

No período estudado foram identificados 390 documentos produzidos por 376 autores. Uma classificação segundo os tipos de documentos pode ser observada na figura 1. Os produtores de literatura sobre este assunto têm preferência por apresentar suas contribuições na forma de artigos publicados nos periódicos do campo da ciência da informação; 86,7% das contribuições têm esta forma de publicação, deixando apenas um índice de 13% para os outros tipos. Nesse pequeno espaço, as contribuições apresentadas em anais de congressos (6,2%) ocupam um segundo estrato pouco significativo, assim como as publicações em capítulos de livros (3,1%), folhetos (1,8%) e cartas enviadas aos editores desses periódicos atingem apenas 2,3% das contribuições.

Esses resultados parecem naturais em termos da prática científica, pois afinal periódicos (eletrônicos ou não) continuam sendo o canal de comunicação formal por excelência para os cientistas e onde os artigos apresentam os resultados de pesquisa corrente na íntegra ou ainda resultados parciais. Já os livros ou capítulos representam conhecimento consolidado e cristalizado, mas que na produção sobre a Lei de Lotka mostram baixo índice (3,1%) de presença. Daí que pesquisas nesta subárea da bibliometria estejam em pleno desenvolvimento.

A figura 2 mostra as contribuições classificadas segundo os idiomas de publicação. Pode-se observar a hegemonia do inglês, com 310 (79,5%) das contribuições, deixando apenas a proporção de 20,5% para ser compartilhada com outros idiomas. Neste aspecto, as publicações em espanhol ocupam o segundo estrato, com 51 (13,1%), seguido do português, com 16 (4,1%), e assim sucessivamente o idioma alemão 4 (1%), chinês 4 (1%), turco, servo-croata e russo, com apenas 1 (0,3%) dos trabalhos produzidos.

Isto também não é nada surpreendente. Já é conhecida a hegemonia do idioma inglês na comunicação científica, tanto que, segundo Price (1971: 257), o inglês constitui pouco mais da metade da produção filosófica e científica do mundo, embora os cientistas não leiam muito e virtualmente nada em idiomas estrangeiros. Para ser mais preciso, "... olham a literatura estrangeira através de [seus] óculos obscuros que deixam ver somente uma décima parte daquilo que realmente existe" (Price, 1971:258).

A figura 3, a seguir, mostra a distribuição de trabalhos pelo número de autores, segundo a forma de contagem. Salienta-se a diferença entre os autores com uma só

FIGURA 1
Tipo de documentos produzidos

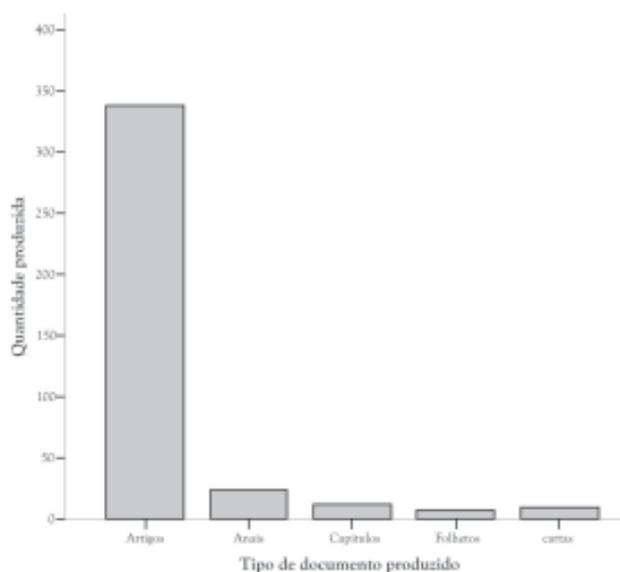
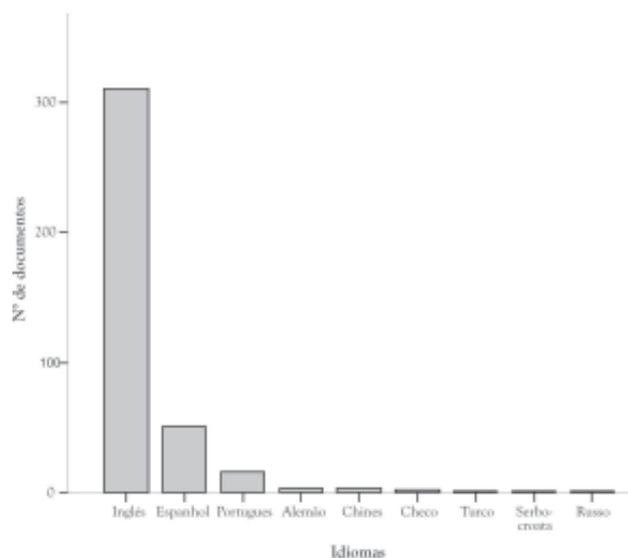


FIGURA 2
Documentos produzidos segundo os idiomas



produção e aqueles com dois, três ou mais. A queda do traçado é vertiginosa até os autores com duas contribuições, sendo esta maior naqueles com a contagem fracional, depois a contagem completa e, finalmente, menor na contagem direta.

Porém, entre aqueles que produziram mais de cinco trabalhos, atinge uma quase linearidade para as três formas de contagem. As caudas das três distribuições são alongadas, formando uma distribuição da típica forma J inverso. Esta distribuição é discreta porque pode assumir

somente números finitos ou contáveis, de valores integrais não-negativos. Essa forma de dispersão da literatura produzida pode ser bem explicada pela distribuição Poisson-Lognormal e Gauss-Poisson Inversa Generalizada.

Como mostra a tabela 1, quando se credita a produtividade de autores somente principais e se ignoram os autores colaboradores (*contagem direta*), no período estudado obteve-se o total de 229 autores que, em conjunto, produziram 381 trabalhos, já que nove documentos em forma de cartas dirigidas aos editores de publicações acadêmicas foram retirados do universo de análise. Desse total, 77,3% participaram na produção de um só artigo, mas em conjunto elaboraram somente 46,5% dos artigos, com média de 1,0 artigo por autor. No entanto, a produtividade média do total de autores é de 1,66 artigo por autor, com variância de 3,3 artigos, desvio padrão de 1,82 artigo e índice de dispersão igual a 1,97 artigo.

Com a forma de *contagem completa*, quando se credita a produtividade a todos os autores, independentemente de se eles participaram da produção do trabalho como principais ou colaboradores, segundo mostra a Tabela 2, foram observados 376 autores, sendo 76,86% participantes na produção de um só artigo; mas em conjunto o grupo de produtores únicos foi responsável somente por 40% dos trabalhos produzidos. Na faixa de autores com um só artigo, a média é de apenas 0,5 artigo por autor. Porém a produtividade média geral encontrada foi de apenas um artigo por autor, com variância de 4,84 artigos, desvio padrão de 2,2 índice de dispersão igual a 4,77 artigos.

Com a *contagem ajustada ou fracionada*, quando cada um dos autores principais e secundários são creditados com uma fração ou porção da contribuição total de artigos, identificou-se que 84% deles participaram na produção de 0,5 (médio) artigo, mas em conjunto elaboraram 45% dos trabalhos. Para este caso, os dados das contribuições tiveram que ser agrupados em intervalos de classe e estão mostrados na tabela 3, a seguir. A produtividade média geral continua sendo de apenas um artigo por autor, com variância de 2,0131 artigos, desvio padrão de 1,4188 artigos e índice de dispersão igual a 2,2134.

A percentagem de autores únicos (77% para a contagem direta e completa e 84% para a contagem fracionada) se assemelha à encontrada em outras áreas das ciências sociais e humanidades. Por exemplo, no campo da psicometria, Rivas & Peiró (1981) levantaram 80,33% de produtores únicos; nas humanidades, Murphy (1973) identificou a taxa de 76,4% de produtores únicos; Voos

FIGURA 3
Número de autores versus artigos produzidos

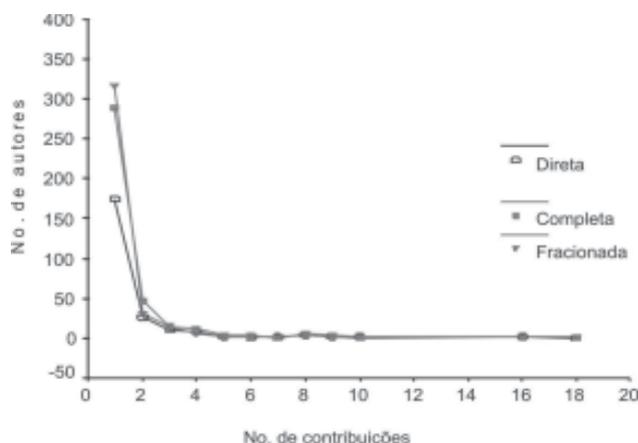


TABELA 1
Autores e produtividade de artigos segundo a contagem direta

Nº de contribuições por autor x	Nº de autores y	% de autores	Nº de artigos	% de artigos
1	177	0,7729	177	0,4646
2	23	0,1004	46	0,1207
3	10	0,0437	30	0,0787
4	7	0,0306	28	0,0735
5	2	0,0087	10	0,0262
6	1	0,0044	6	0,0157
7	1	0,0044	7	0,0184
8	4	0,0175	32	0,0840
9	1	0,0044	9	0,0236
10	2	0,0087	20	0,0525
16	1	0,0044	16	0,0420
Total	229	1,0000	381	1,0000

TABELA 2
Autores e produtividade de artigos segundo a contagem completa

Nº de contribuições por autor x	Nº de autores y	% de autores	Nº de artigos	% de artigos
1	289	0,7686	153	0,4016
2	45	0,1197	54	0,1417
3	15	0,0399	29	0,0761
4	10	0,0266	32	0,0840
5	3	0,0080	10	0,0262
6	3	0,0080	10	0,0262
8	5	0,0133	35	0,0919
9	3	0,0080	23	0,0604
10	1	0,0027	10	0,0262
16	1	0,0027	16	0,0420
18	1	0,0027	9	0,0236
Total	376	1,0000	381	1,0000

(1974) achou 87,6% na ciência da informação; McCreery & Pao (1984) encontraram 72% de contribuintes únicos no campo da etnomusicologia e Chung & Cox (1990); López-Muñoz & Rubio Valladolid (1995) observaram 71,21% no campo da psiquiatria. A elevada taxa de pequenos produtores parece ser característica das disciplinas em expansão e desenvolvimento na procura de sua institucionalização.

A tabela 4 mostra a demografia desta distribuição de freqüência, segundo a forma de contagem. Para a contagem direta, a variância é duas vezes maior do que sua média. Para a completa, a variância é quase 5 vezes maior do que a média. Para a contagem ajustada ou fracionada, é duas vezes maior do que a média. Esses valores indicam superdispersão no padrão do comportamento produtivo dos autores.

Essa apreciação é confirmada pelo índice de dispersão da distribuição, pois esse valor é maior do que a média, significando que os autores trabalham de forma isolada e sem maiores contatos com os pares, porém seguindo com atenção a produtividade da área. Em outras palavras, preferem trabalhar individualmente. O mesmo fenômeno já tinha sido observado por McCreery & Pao (1984) no campo da etnomusicologia. Estes autores afirmam que os cientistas sociais e humanistas em geral preferem trabalhar isoladamente e não em colaboração. Se considerarmos a ciência da informação e em especial a bibliometria como uma área afim das ciências sociais, talvez esse seja também o caso da produtividade de autores relacionados com a lei de Lotka.

O ajuste ao modelo Poisson lognormal

Aplicando-se o modelo Poisson lognormal encontrou-se que, para a contagem direta, $b = -6,3153279$ e $a = 2,5741725$. Com esses parâmetros já conhecidos, calcularam-se os valores esperados da distribuição, que ao 0,05 nível de significância e com 3 graus de liberdade produziram um qui-quadrado igual a 2,3225 menor do que o valor crítico de 7,81473, rejeitando-se, então, a hipótese nula de homogeneidade da distribuição. Portanto, concluiu-se que a distribuição dos produtores de artigos sobre a Lei de Lotka segundo a contagem direta ajusta-se, adequadamente, à distribuição Poisson lognormal. A figura 4 mostra a proximidade dos valores observados e esperados desta forma de contagem.

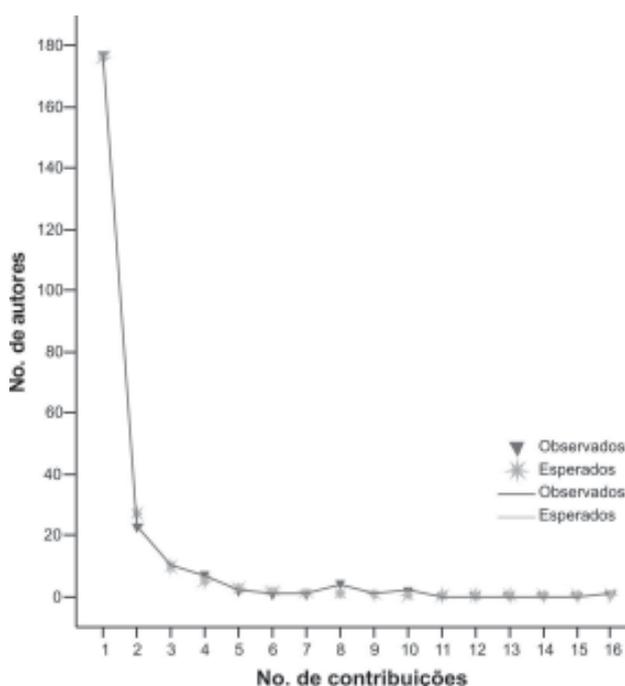
TABELA 3
Autores e produtividade de artigos segundo a contagem ajustada

Nº de contribuições por autor x	Nº de autores y	% de autores	Nº de artigos	% de artigos
0.5	316	0,8404	172	0,4514
1.5	31	0,0824	56	0,1470
2.5	13	0,0346	40	0,1050
3.5	5	0,0133	22	0,0577
4.5	1	0,0027	5	0,0131
5.5	2	0,0053	11	0,0289
6.5	1	0,0027	8	0,0210
7.5	3	0,0080	23	0,0604
8.5	2	0,0053	19	0,0499
12.5	1	0,0027	9	0,0236
13.5	1	0,0027	16	0,0420
Total	376	1,0000	381	1,0000

TABELA 4
Demografia segundo a forma de contagem

Contagem Parâmetros	Contagem direta	Contagem completa	fracionada
Média	1,6710	1,0133	0,9095
Variância	3,2966	4,8372	2,0131
Desvio padrão	1,8157	2,1994	1,4188
Índice de dispersão	1,9727	4,7737	2,2134
Grupo de elite	15,13	19,391	19,391

FIGURA 4
Valores observados e esperados da distribuição Poisson lognormal para a contagem direta

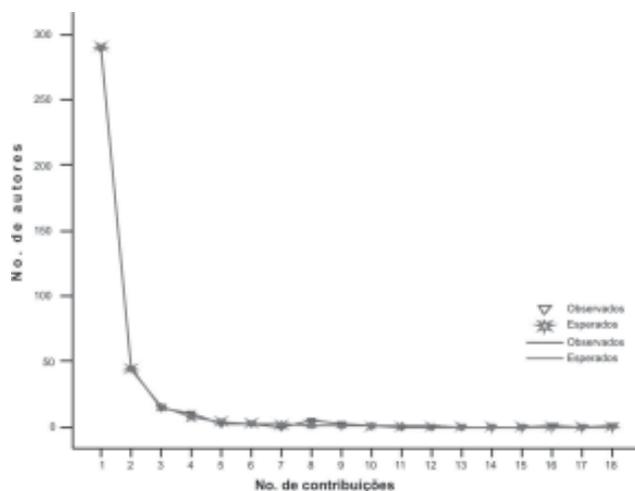


Para o método da contagem completa obteve-se $b = -7,5584044$ e $a = 2,7528968$. Com os parâmetros já conhecidos, estimaram-se os valores esperados da distribuição Poisson lognormal, que ao 0,05 nível de significância e com 4 graus de liberdade, produziram um qui-quadrado igual a 2,1405, menor do que o valor crítico de $p = 9,48773$, rejeitando a hipótese nula de homogeneidade da distribuição. Concluiu-se que a distribuição dos produtores de artigos sobre a Lei de Lotka ajusta-se, adequadamente, à distribuição Poisson lognormal. A figura 5 mostra a proximidade dos valores observados e esperados desta forma de contagem.

Para a contagem ajustada ou fracionada, obteve-se $b = -3,042388$ e $a = 2,193063$. Com os parâmetros já conhecidos, calcularam-se os valores esperados da distribuição Poisson lognormal, que ao 0,05 nível de significância e com 3 graus de liberdade produziram um qui-quadrado igual a 3,9951, menor do que o valor crítico de 7,81473, rejeitando, a hipótese nula de homogeneidade da distribuição. Inferiu-se que a distribuição dos produtores de artigos sobre a Lei de Lotka, quando se usa a contagem fracionada agrupada, ajusta-se de modo adequado à distribuição Poisson lognormal. A figura 6 mostra a proximidade dos valores observados e esperados desta forma de contagem.

A tabela 5 traz os dados estimados dos autores sobre a literatura da Lei de Lotka, bem como os parâmetros encontrados para o modelo da distribuição Poisson lognormal. Como o teste qui-quadrado é sensível a valores muito pequenos da distribuição, as freqüências menores que 5 foram acumuladas com as adjacentes para

FIGURA 5
Valores observados e esperados da distribuição Poisson lognormal pela contagem completa



produzir freqüências observadas iguais a 5. O modelo se ajusta muito bem à distribuição da produtividade dos autores e tem um desempenho satisfatório na predição dos valores esperados ou teóricos da distribuição.

FIGURA 6
Valores observados e esperados da distribuição Poisson lognormal para contagem fracionada

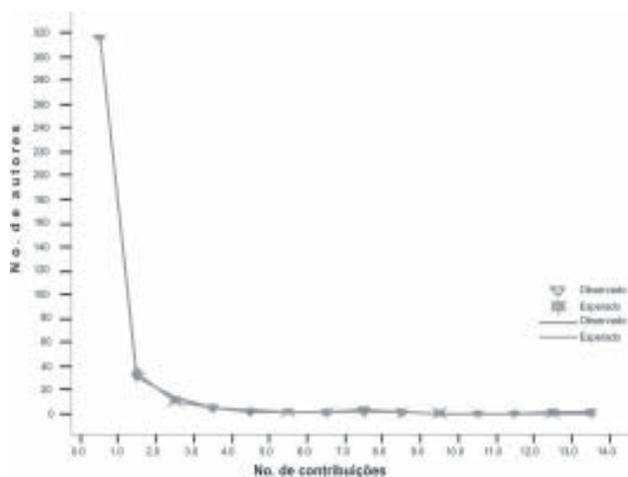


TABELA 5
Distribuição de freqüências da produtividade de autores segundo a distribuição Poisson Lognormal

Nº de contribuições r	Nº de autores observados (Direto)	Nº de autores observados (Completo)	Nº de autores estimados (Direto)	Nº de autores estimados (Completo)
1	177	289	175,72	290,68
2	23	45	27,05	44,51
3	10	15	9,68	15,65
4	7	10	4,78	7,62
5	2	3	2,79	4,40
6	1	3	1,81	2,82
7	1	0	1,25	1,93
8	4	5	0,91	1,40
9	1	3	0,69	1,05
10	2	1	0,54	0,81
11	0	0	0,43	0,64
12	0	0	0,35	0,52
13	0	0	0,29	0,43
14	0	0	0,24	0,36
15	0	0	0,21	0,30
16	1	1	0,18	0,26
17	—	0	—	0,22
18	—	1	—	0,19
Total	229	376	227,12	373,79

$\beta = -6,3153$
 $\alpha = 2,5742$
 $\chi^2 = 6,84$
 $P(X^2) = 9,48773$
 $df = 4$

$\beta = -7,5584$
 $\alpha = 2,7529$
 $\chi^2 = 3,69$
 $P(X^2) = 11,0705$
 $df = 5$

A tabela 6 exhibe os dados estimados dos autores sobre a literatura da Lei de Lotka, bem como os parâmetros encontrados para o modelo da distribuição Poisson Lognormal, segundo a contagem fracionada. Uma vez que o teste qui-quadrado é sensível a valores muito pequenos da distribuição, as freqüências menores que 5 foram acumuladas com as adjacentes para produzir freqüências observadas iguais a 5, conforme já visto. O modelo se ajusta corretamente à distribuição da produtividade dos autores e revela desempenho satisfatório na predição dos valores esperados ou teóricos da distribuição.

Ajuste ao modelo Gauss Poisson inverso generalizado

Há algum tempo vem-se insistindo que as distribuições bibliométricas são da forma J inverso, discreta, com uma longa cauda positivamente inclinada à direita e que suas variáveis randômicas r progridem em uma unidade. Então, Sichel (1971) propôs uma família de distribuições discretas que procedem da mistura da distribuição Poisson de parâmetro γ . Se este parâmetro γ é feito negativo, gera-se novo grupo de distribuições discretas. E se γ é conhecida *a priori*, os estimadores da probabilidade máxima para os parâmetros α e θ são simplificados.

Para facilitar esse cálculo, as equações foram desenvolvidas como uma distribuição discreta Gauss-Poisson inversa generalizada (GPIG), adequada para avaliar o ajuste das freqüências das publicações dos autores em qualquer campo do conhecimento. Também Sichel (1982, 1986) tem explorado vários métodos para calcular os parâmetros envolvidos nesta distribuição.

Seguindo suas propostas de observar a distribuição dos autores e artigos em um gráfico de dupla escala logarítmica, considerou-se que $\gamma = -1/2$ como valor *a priori*, portanto, o método da primeira proporção e a média seriam mais adequados para descrever a distribuição da produtividade de literatura sobre a Lei de Lotka.

Este método tem sido testado com sucesso na dispersão da produtividade de artigos em periódicos científicos (Sichel, 1985) e na literatura de física (Sichel, 1986). Na América Latina, só Urbizagastegui Alvarado (2003) tem explorado com êxito a aplicação deste modelo às cartas do arquivo privado de Getúlio Vargas e na antropologia brasileira (URBIZAGÁSTEGUI e OLIVEIRA, 2001).

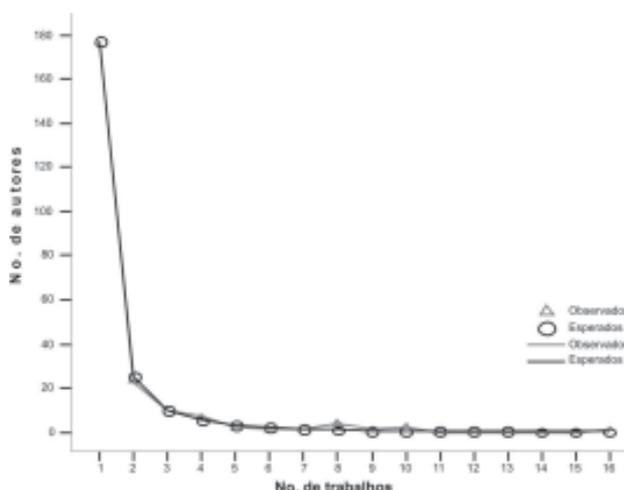
Aplicando-se tal método encontrou-se, para a contagem direta, $\alpha = 0,6933$ e $\theta = 0,917902$. Com esses parâmetros já conhecidos, calcularam-se os valores esperados da distribuição, que ao 0,05 nível de significância e com 3

TABELA 6
Distribuição de freqüência da produtividade de autores segundo a contagem fracionada

Nº de contribuições r	Nº de autores observados (agrupados)	Nº de autores estimados (PLN)
0.5	316	314,69
1.5	31	35,14
2.5	13	10,58
3.5	5	4,83
4.5	1	2,71
5.5	2	1,71
6.5	1	1,17
7.5	3	0,84
8.5	2	0,63
9.5	0	0,419
10.5	0	0,39
11.5	0	0,32
12.5	1	0,26
13.5	1	0,22
Total	376	373,98

$\beta = -3,0424$ $P(\chi^2) = 9,48773$
 $\alpha = 1,931$ $df = 4$
 $\chi^2 = 6,84$

FIGURA 7
Valores observados e esperados da distribuição Gauss-Poisson inversa generalizada zero truncada pela contagem direta

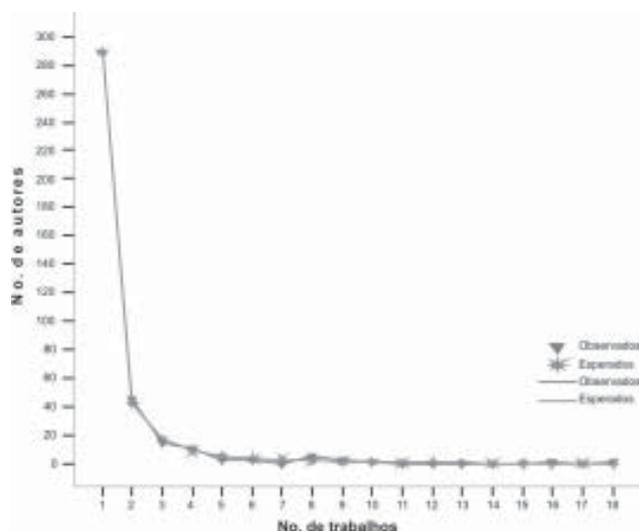


graus de liberdade, produziu um qui-quadrado igual a 0,8291 menor do que o valor crítico de $p = 7,81473$ rejeitando-se, então, a hipótese nula de homogeneidade da distribuição. Portanto, concluiu-se que a distribuição dos produtores de artigos sobre a Lei de Lotka, segundo a contagem direta, ajusta-se adequadamente à distribuição Gauss-Poisson inversa generalizada. A figura 7 mostra a proximidade dos valores observados e esperados desta forma de contagem.

Para o método da contagem completa encontrou-se que $a = 0,66723$ e $q = 0,89361$. Com os parâmetros já conhecidos, calcularam-se os valores esperados da distribuição Gauss-Poisson inversa generalizada zero truncada, que ao 0,05 nível de significância e com 4 graus de liberdade produziu um qui-quadrado igual a 1,3911 menor do que o valor crítico de 9,48773, rejeitando a hipótese nula de homogeneidade da distribuição. Deduziu-se que a distribuição dos produtores de artigos sobre a Lei de Lotka adequa-se à distribuição Gauss-Poisson inversa generalizada. A figura 8 mostra a proximidade dos valores observados e esperados desta forma de contagem.

Para a contagem ajustada ou fracionada estimou-se que $a = 0,194289$ e $q = 0,988847$. Como a tende a zero e seguindo Sichel (1982), estabeleceu-se que $g = 0,5$. Com esses parâmetros, estimaram-se os valores esperados da distribuição Gauss-Poisson inversa generalizada zero truncada, que ao 0,05 nível de significância e com 3 graus de liberdade produziu um qui-quadrado igual a 3,8669 menor do que o valor crítico de 7,81473, afastando a hipótese nula de homogeneidade da distribuição. Logo, a distribuição dos produtores de artigos sobre a Lei de Lotka, quando se considera a contagem fracionada, ajusta-se à distribuição Gauss-Poisson inversa generalizada. A figura 9 mostra a proximidade dos valores observados e esperados desta forma de contagem.

FIGURA 8
Valores observados e esperados da distribuição Gauss-Poisson inversa generalizada zero truncada pela contagem completa



A tabela 7 mostra os dados estimados dos autores produtores da literatura sobre a Lei de Lotka, bem como os parâmetros encontrados para o modelo da distribuição Gauss-Poisson inversa generalizada. Devido ao teste qui-quadrado mostrar-se sensível a valores muito pequenos da distribuição, as frequências menores que 5 foram acumuladas com as adjacentes para produzir frequências observadas iguais a 5. O modelo se ajusta muito bem à distribuição da produtividade dos autores e revela desempenho satisfatório na predição dos valores esperados ou teóricos da distribuição.

FIGURA 9
Valores observados e esperados da distribuição Gauss-Poisson inversa generalizada pela contagem ajustada

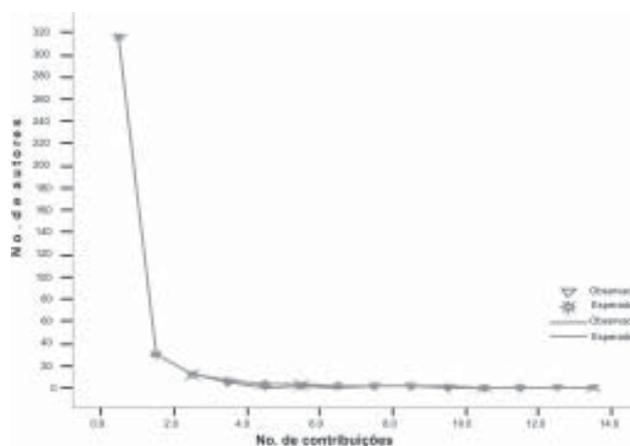


TABELA 7
Distribuição de frequência da produtividade de autores segundo a distribuição Gauss-Poisson inversa generalizada

Nº de contribuições r	Nº de autores observados (Direto)	Nº de autores observados (Completo)	Nº de autores esperados (Direto)	Nº de autores esperados (Completo)
1	177	289	177,015	288,992
2	23	45	24,917	42,968
3	10	15	9,962	17,058
4	7	10	5,273	8,890
5	2	3	3,201	5,295
6	1	3	2,109	3,417
7	1	0	1,468	2,714
8	4	5	1,062	1,920
9	1	3	0,791	1,398
10	2	1	0,603	1,041
11	0	0	0,469	0,789
12	0	0	0,369	0,607
13	0	0	0,295	0,473
14	0	0	0,238	0,372
15	0	0	0,194	0,296
16	1	1	0,159	0,237
17	—	0	—	0,191
18	—	1	—	0,155
Total	229	376	228,125	376,813

a = 0,6933
 $\theta = 0,9179$
 $\chi^2 = 0,8291$
 $P(X^2) = 7,81473$
 df = 3

a = 0,66723
 $\theta = 0,89361$
 $\chi^2 = 1,3911$
 $P(X^2) = 9,48773$
 df = 4

A tabela 8 exibe os dados estimados para os autores sobre a literatura da Lei de Lotka usando o método da *contagem ajustada ou fracionada*, bem como os parâmetros encontrados para o modelo da distribuição Gauss-Poisson inversa generalizada. Por ser o teste qui-quadrado sensível a valores muito pequenos da distribuição, as frequências menores que 5 foram acumuladas com as adjacentes para produzir frequências observadas iguais a 5. O modelo se ajusta muito bem à distribuição da produtividade dos autores, com desempenho satisfatório na predição dos valores esperados ou teóricos da distribuição.

CONCLUSÕES

Esta pesquisa demonstra que, no período estudado, a média geral da produtividade dos autores da literatura sobre a Lei de Lotka foi de 1,7, 1,0 e 0,9 trabalhos por autor, segundo a forma de contagem direta, completa e fracionada. Esta média é muito similar à média de 0,89 trabalhos por autor reportados por Rivas & Peiró (1981) no campo da psicometria, e está próxima à média de 1,83 trabalho por autor verificado por McCreery e Pao (1984) no campo da etnomusicologia.

O índice de dispersão variou entre 1,97 para a contagem direta, 1,93 para a contagem completa e 2,15 para a contagem fracionada, indicando que a dispersão desta literatura é muito pequena e que os autores trabalhavam de forma isolada, não havendo contato entre os pares, resultando numa produção em parceria reduzida. O resultado também é similar ao reportado por Rogge (1976) na antropologia americana, na qual observou que somente 20% dos artigos publicados no *The American Anthropologist* e 30% no *American Antiquity* eram elaborados em parceria.

A distribuição da produtividade dos autores da literatura de Lotka, segundo as três formas de contagem direta, completa e fracionada, ao 0,05 nível de significância e com 3 e 4 graus de liberdade, ajusta-se muito bem à distribuição Gauss-Poisson inversa generalizada zero truncada. Igualmente essas três formas de contagem da literatura produzida sobre a Lei de Lotka ajustam-se muito bem à distribuição Poisson-Lognormal, ao 0,05 nível de significância e com 4 e 5 graus de liberdade. Não obstante, a distribuição Poisson Gauss inversa generalizada produz um chi-quadrado menor e prediz melhor o total de autores do que a distribuição Poisson Lognormal.

TABELA 8
 Distribuição de frequência da produtividade de autores segundo a contagem ajustada

No. de contribuições r	No. de autores observados (agrupados)	No. de autores estimados (GPIG)
0.5	316	316,02
1.5	31	30,36
2.5	13	12,01
3.5	5	6,69
4.5	1	4,47
5.5	2	3,22
6.5	1	2,45
7.5	3	1,94
8.5	2	1,58
9.5	0	1,31
10.5	0	1,11
11.5	0	0,96
12.5	1	0,83
13.5	1	0,73
Total	376	383,78
a = 0,194 $\theta = 0,989$ $\chi^2 = 3,867$		$P(X^2) = 7,81473$ df = 3

Artigo submetido em 17/10/2008 e aceito em 22/12/2008.

REFERÊNCIAS

ALBERT, R. S. Toward a behavioral definition of genius. *American Psychologist*, n. 30, p. 140-151, 1975.

BARRON, Frank. The needs for order and for disorder as motives in creative activity. In: _____. *Scientific creativity*. New York: Wiley, 1963. p. 153-160.

BOOKSTEIN, Abraham. Implications of ambiguity for scientometric measurement. *Journal of the American Society for Information Science and Technology*, v. 52, n. 1, p. 74-79, Jan. 2001.

_____. Patterns of scientific productivity and social change: a discussion of Lotka's law and bibliometric symmetry. *Journal of the American Society for Information Science*, v. 28, n. 4, p. 206-210, July 1977.

BROOKES, Bertram C. Research note: the Inverse Gaussian Poisson Distribution (IGPD). *Journal of Information Science*, v. 11, p. 43-44, 1985.

_____. Comments on the scope of bibliometrics. In: INTERNATIONAL CONFERENCE ON BIBLIOMETRICS AND THEORETICAL ASPECTS OF INFORMATION RETRIEVAL, 1., 1987. *Proceedings...* Amsterdam: Leo Egghe and Ronald Rousseau; New York: Elsevier Science Publishers, 1988. p. 29-42.

BULMER, M. G. On fitting the Poisson lognormal distribution to species-abundance data. *Biometrics*, v. 30, n. 1, p. 101-110, Mar. 1974.

- BURRELL, Quentin L.; FENTON, Michael R. Yes, the GIGP really does work and it is workable! *Journal of the American Society for Information Science*, v. 44, n. 2, p. 61-69, Mar. 1993.
- CASTRO, Cláudio Moura. Há produção científica no Brasil?. *Ciência e Cultura*, São Paulo, v. 37, n. 7, p. 165-187, jul. 1985. Suplemento.
- _____. *Ciência e universidade*. Rio de Janeiro: Zahar, 1986.
- CATTELL, J. Mckeen. A further statistical study of American men of science. *Science*, v. 32, p. 633-648, Friday, Nov. 4, 1910.
- CHUNG, Kee H.; COX, Raymond A. K. Patterns of productivity in the finance literature: a study of bibliometric distributions. *The Journal of Finance*, v. 45, n. 1, p. 301-309, Mar. 1990.
- COCKS, T. M.; BROOKES, Bertram C. Sichel's unification of bibliometric frequency distributions. *Journal of Information Science*, n. 12, p. 45-51, 1986.
- DUFRENOY, Jean. The publishing behavior of biologists. *Quarterly Review of Biology*, n. 13, p. 207-210, 1938.
- EGGHE, Leo. Type/Token-Taken informetrics. *Journal of the American Society for Information Science and Technology*, v. 54, n. 7, p. 603-610, 2003.
- _____. Size-frequency and rank-frequency relations, power laws and exponentials: a unified approach. *Progress in Natural Science*, v. 13, n. 6, p. 478-480, June 2003.
- HUBER, John C. A new model that generates Lotka's law. *Journal of the American Society for Information Science and Technology*, v. 53, n. 3, p. 209-219, Feb. 2002.
- KUPERMAN, Victor. Productivity in the mailing lists: a bibliometric analysis. *Journal of the American Society for Information Science*, v. 57, n. 1, p. 51-59, Jan. 2006.
- LAMBROSO, Cesare. *The man genius*. London: Scott, 1891.
- LÓPEZ-MUÑOZ, Francisco; RUBIO VALLADOLID, G. La producción científica española em psiquiatria: estudo bibliométrico de las publicaciones de circulación internacional durante el periodo 1980-1993. *Anales de Psiquiatria*, v. 2, n. 2, p. 68-75, 1995.
- LUNGARZO, Carlos. *O que é ciência*. São Paulo, SP: Brasiliense, 1989.
- MCCREERY, Laurie S.; PAO, Miranda Lee. Bibliometric analysis of ethnomusicology. In: ASIS ANNUAL MEETING: challenges to an information society, 47., 1984, New York. *Proceedings...* White Plains, New York: Knowledge Industry Publications, Inc. for American Society for Information Science, 1984. p. 212-216.
- NATH, Ravinder; JACKSON, Wade M. Productivity of management information system researchers: does Lotka's law apply?. *Information Processing & Management*, v. 27, n. 2/3, p. 203-209, 1991.
- NICHOLLS, Paul Travis. Bibliometric modelling process and the empirical validity of Lotka's law. *Journal of the American Society for Information Science*, v. 40, n. 6, p. 279-385, 1989.
- _____. Empirical validation of Lotka's law. *Information Processing & Management*, v. 22, n. 5, p. 417-419, 1986.
- _____. Estimation of Zipf parameters. *Journal of the American Society for Information Science*, v. 38, n. 6, p. 443-445, Nov. 1987.
- ORTEGA Y GASSET, Jose. *The revolt of the masses*. New York: Norton, [1957, c1932].
- PAO, Miranda Lee. Lotka's test. *Collection Management*, v. 4, n. 1/2, p. 111-124, Spring/Summer. 1982.
- _____. Lotka's law: a testing procedure. *Information Processing & Management*, v. 21, n. 4, p. 305-320, 1985.
- _____. An empirical examination of Lotka's law. *Journal of the American Society for Information Science*, v. 37, n. 1, p. 26-33, Jan. 1986.
- POTTER, William Gray. Lotka's law revisited. *Library Trends*, v. 30, n.1, p. 21-39, Summer 1981.
- PRICE, John Derek de Solla. A calculus of science. *International Science and Technology*, v. 15, p. 37-43, Mar. 1963.
- _____. Galton revisited. In: LITTLE science, big science. New York: Columbia University Press, 1963. p. 33-61.
- _____. The expansion of scientific knowledge. *Annals of the New York Academy of Sciences*, v. 184, p. 257-259, June 1971.
- _____. Society's needs in scientific and technical information. *Annals of the New York Academy of Sciences*, v. 261, p. 126-136, 1975.
- RADHKRISHNAN, T.; KERNIZAN, R. Lotka's law and computer science literature. *Journal of the American Society for Information Science*, v. 30, n. 1, p. 51-54, Jan. 1979.
- RIVAS, Francisco; PEIRÓ, José María. Estado actual de la investigación psicométrica: una aproximación bibliométrica. In: CARPINTERO, Helio et al. *Psicología contemporánea: teoría y metodos cuantitativos para el estudio de su literatura científica*. Valencia: Alfapplus, 1981. p. 199-218.
- ROGGE, A. E. A look at academic anthropology: thorough a graph darkly. *American Anthropologist*, v. 78, n. 4, p. 829-843, Dec. 1976.
- SCHORR, Alan Edward. Lotka's law and library science. *RQ [Reference Quarterly]*, v. 14, n. 1, p. 32-33, Fall 1974.
- SHOCKLEY, William. On the statistics of individual variations of productivity in research laboratorios. *Proceedings of the Institute for Radio Engineers*, v. 45, n. 3, p. 279-290, Mar. 1957.
- SICHEL, H. S. On a family of discret distributions particularly suited to represent long-tailed frequency data. In: SYMPOSIUM ON MATHEMATICAL STATISTIC, 3., 1971. *Proceedings...* Pretoria, South África: C.S.I.R., 1971. p. 51-97.
- _____. Repeat-buying and the generalized inverse Gaussian-Poisson distribution. *Applied Statistics*, v. 31, n. 3, p. 193-204, 1982.
- _____. A bibliometric distributions which really works. *Journal of the American Society for Information Science*, v. 36, n. 5, p. 314-321, Sept. 1985.
- _____. The GIGP distribution model with applications to physics literature. *Czechoslovak. Journal of Physics*, v. B36, n. 1, p. 133-137, 1986.
- _____. Anatomy of the generalized inverse Gaussian-Poisson distribution with special applications to bibliometric studies. *Information Processing and Management*, v. 28, n. 1, p. 5-17, 1992.

- SIMONTON, Dean Keith. *Genius, creativity and leadership*. In: _____; OLIVEIRA, Marlene de. A produtividade dos autores na antropologia brasileira. *DataGramaZero*, v. 2, n. 6, p. 1-17, Dec. 2001. Massachussets: Harvard University Press, 1984. Cap. 5.
- _____. *Scientific genius: a psychology of science*. Cambridge, Mass.: Cambridge University Press, 1988.
- STEWART, John A. The Poisson-lognormal model for bibliometric/scientometric distributions. *Information Processing and Management*, v. 30, n. 2, p. 239-251, 1994.
- SUBRAMANYAM, Krishna. Lotka's Law and the literature of computer science. *IEEE Transactions on Professional Communication*, v. PC-22, n. 4, p. 181-191, Dec. 1979.
- TERRADA, Maria-Luz. La productividad de los autores médicos españoles (Ley de Lotka). In: LA LITERATURA médica española contemporánea: estudio estadístico y sociométrico. Valencia, España: Centro de Documentation e Informática Médica, 1973. p. 85-92.
- TSAY, Ming-Yueh; JOU, Shiow-Jen; MA, Sheau-Shin. A bibliometric study of semiconductor literature, 1978-1997. *Scientometrics*, v. 49, n. 3, p. 491-509, 2000.
- URBIZAGÁSTEGUI ALVARADO, Rubén. *Lotka's law on Lotka's literature: an exploration*, 2001. No prelo.
- _____. A lei de Lotka na bibliometria brasileira. *Ciência da Informação*, Brasília, v. 31, n. 2, p. 14-20, maio/ago. 2002.
- _____. La ley de Lotka: aplicación de la distribución gauss-poisson inversa generalizada a la productividad de autores. *Revista INFOBIB*, Lima, Perú, n. 2, p. 37-64, 2003.
- _____. *Lotka's law: an annotated bibliography*. [S.l.: s.n.], 2005.
- VLACHY, Jan. Physics journal in retrospect and comparisons. *Czechoslovak journal of physics*, v. B20, p. 501-526, 1970.
- _____. Distribution patterns in creative communities. In: WORLD CONGRESS OF SOCIOLOGY, 8., 1974. Toronto. [S.l.: s.n.], 1974.
- _____. Time factor in Lotka's law. *Probleme de Informare si Documentare*, v. 10, n. 2, p. 44- 87, mar./apr. 1976.
- VOOS, Henry. Lotka and Information Science. *Journal of the American Society for Information Science*, v. 25, n. 4, p. 270-272, July/Aug. 1974.
- WAGNER-DÖBLER, Roland; BERG, Jan. Physics 1800-1900: a quantitative outline. *Scientometrics*, v. 46, n. 2, p. 213-285, Oct. 1999.
- WILLIAMS, C. B. The number of publications written by biologists. *Annals of Eugenics*, n. 12, p. 143-146, 1944.
- WORTHEN, D. B. Short-lived technical literatures: a bibliometric analysis. *Methods of Information in Medicine*, v. 17, n. 3, p. 190-198, 1978.
- ZIMAN, John. *Conhecimento público*. Belo Horizonte: Itatiaia, 1979.