Aplicação da distribuição Poisson zero truncada á produtividade de autores

Rubén Urbizagástegui Alvarado

Bibliotecário. University of California, Riverside.
Riverside, CA 97521-5900. USA.
e-mail: rubur@ucracl.ucr.edu

Descreve a natureza da distribuição zero truncada de Poisson conforme desenvolvida por Plackett. Apresenta as equações específicas quando a freqüência de zero observações não esta presente na amostra coletada. Visto que é comum encontrar-se dados deste tipo no campo da bibliometria, descreve-se passo a passo, a forma de aplicação do modelo aos dados estudados por Gupta & Kumar sobre a produção dos autores no campo da genética populacional e publicados no periódico Evolution.

Palavras-chave: Lei de Lotka; Distribuição Poisson zero truncado; Produtividade de autores; Bibliometria; Genética; Informetria; Cienciometria

Recebido em: 20.11.2003 Aceito em: 15.03.2004

Introdução

Vem sendo usada uma série de modelos para se estudar os fenômenos sociais no campo da bibliometria. Os três mais conhecidos e freqüentemente citados na literatura como leis, são: a Lei de Bradford, sobre a produtividade de artigos por periódicos especializados; a Lei de Zipf, sobre a freqüência de ocorrências de palavras num texto; e a Lei de Lotka, sobre a produtividade científica dos autores. A produtividade científica reflete-se freqüentemente na literatura na forma de artigos, patentes registradas, informes apresentados etc. Grande número de acadêmicos tem usado essas variáveis para medir a produtividade científica de indivíduos, organizações e companhias, em diferentes situações, assuntos e períodos. Um dos modelos usados para essa avaliação, mas ainda pouco explorado, tem sido o da distribuição de Poisson.

Tal distribuição proporciona um bom modelo para dados que representam o número de ocorrências de um evento específico numa determinada unidade de tempo ou espaço. Foi desenvolvida por Simeón Denis Poisson, 1781-1840, que a aplicou ao campo de eletrostática e magnetismo. Este modelo tem sido empregado para descrever a função de probabilidades de eventos como:

- a demanda de produtos
- a demanda de serviços
- o número de ligações numa central telefônica
- o número de chegadas de ônibus num terminal de passageiros
- o número de chegadas de aviões num aeroporto
- o número de acidentes, por trabalhador, numa fábrica
- o número de pessoas, por residência, que sofrem de enfermidades contagiosas.

Estes exemplos têm duas coisas em comum: uma, as ocorrências podem ser descritas em termos de *variáveis discretas* (tomam valores integrais de 0, 1, 2, 3, …., etc.), isto é, não se pode falar em valores fracionados como dois e meio, 3.75, etc., senão de valores inteiros; e outra, apresentarem uma taxa que caracterize o sucesso do evento produzido (a taxa é o número de ocorrências por intervalo de tempo ou de espaço). Por exemplo, não se pode dizer que um autor produziu um artigo e meio num período de 20 anos.

Visto que o modelo de Lotka não prediz a produção anual de artigos científicos, e que não leva em consideração os autores mais prolíficos, Mantel (1966) sugeriu o uso da distribuição de Poisson para estimar a probabilidade de ocorrência de um evento durante um período determinado. Aplicou o modelo ao número de contribuições por autor, listados numa bibliografia publicada pela Oficina Naval de Pesquisas em Engenharia Humana, do governo americano, de 1959 a 1960, com 332 autores produzindo 2255 artigos. O modelo descreveu com sucesso a distribuição de freqüências de autores e publicações, no período.

Igualmente, Huber (1998a) estudou uma amostra de 773 inventores que trabalharam para uma companhia durante os anos de 1960 a 1995, os quais tinham produzido seis patentes ou mais, num período de cinco anos ou mais. Esses 773 inventores foram responsáveis por 5254 patentes concedidas nesses anos. Esta produção representava 58% do total de patentes registradas pela companhia no período estudado. O autor observou que a maioria dos inventores mostravam um padrão de produtividade randômico que se ajustava muito bem à distribuição de Poisson. Nesse mesmo ano, Huber (1998b) estudou a taxa de crescimento e decréscimo indicando aprendizagem, envelhecimento, controle e ajuste aos objetivos da companhia. Desta vez, o estudo esteve limitado a uma amostra de 162 inventores que tinham produzido 1975 patentes. Usando o teste do qui-quadrado, a distribuição das patentes foi testada ao 0.10 nível de significância. Novamente, os dados se ajustaram muito bem ao modelo de Poisson.

Também Gupta e Kumar (1998) estudaram 11 grupos de dados procedentes do núcleo de periódicos representativos no campo da genética cobrindo o período de 1971 a 1980. Os autores observaram que a distribuição zero truncada de Poisson ajustou-se somente a cinco dos 11 grupos estudados. Gupta; Kumar e Rousseau (1998) usando uma bibliografia exaustiva sobre genética teórica, publicada por J. Felsentein em 1981, cobrindo a literatura sobre o assunto desde 1870 até 1980, identificaram um núcleo representativo de 11 revistas. A seguir, contaram o número de autores de cada artigo publicado em cada uma dessas revistas e estudaram a aplicabilidade da distribuição de Lotka, a distribuição geométrica e a distribuição truncada de Poisson. Os autores verificaram que a distribuição de Lotka não foi apropriada para o estudo da distribuição dos autores. No entanto, observaram que a distribuição geométrica e a distribuição zero truncada de Poisson mostraram-se como as mais adequadas aos dados estudados.

Deve-se realçar que a lei de Lotka tem sido testada em muitas recopilações de dados, mas o ajuste nem sempre tem sido apropriado. Este ponto não tem sido o suficientemente destacado na bibliografia publicada, ainda que as revisões do estado-da-arte de Vlachy (1980) e Potter (1981) tenham ajudado a reorientar o interesse dos pesquisadores para avaliar se outros tipos de distribuições poderiam proporcionar um melhor ajuste dos dados observados. Isso levou à introdução, nas pesquisas bibliométricas, de modelos estatísticos usados nas ciências naturais, tais como a distribuição hiperbólica, a distribuição logarítmica normal, a distribuição de Yule, a distribuição binomial, a série geométrica, a série logarítmica, a distribuição de Weinbull e, finalmente, a distribuição inversa generalizada de Gauss-Poisson.

É importante assinalar que esses modelos estatísticos não têm sido explorados na bibliometria latino-americana, embora estejam disponíveis na literatura publicada e sejam acessíveis. Talvez isso ocorra em razão da falta de familiaridade com os modelos estatísticos ou desconhecimento dos instrumentos matemáticos para testar os dados. É sabido que desentranhar complicados modelos estatísticos não é o passatempo favorito de muitos pesquisadores. Por essa razão, o objetivo deste trabalho é proporcionar um guia para analisar e testar a lei de Lotka sobre a produtividade científica dos autores, usando a distribuição de Poisson. Considera-se que este guia é necessário para apoiar a compreensão, adoção, aplicação e difusão deste modelo.

Natureza da distribuição de Poisson

Na distribuição de Poisson a variável discreta *x* pode tomar valores de *x* = 0, 1, 2, ….. n . Então, a probabilidade de obter-se exatamente *x* ocorrências é:

$$\int(x) = \frac{\mu^{\chi} e^{-\mu}}{x!} \qquad para\, x = 0, 1, 2, ... \qquad (1)$$

onde:

μ = *média aritmética de ocorrências por intervalo de tempo.*

e = 2.71828 (a base dos logaritmos naturais)

x = valores 0, 1, 2, …. da variável discreta x

Nesta equação **x!** representa o fatorial do valor de *x* = 0, 1, 2, 3, ….. Este fatorial é calculado da seguinte maneira:

Para x = 0! = 1

Para x = 1! = 1 × 1 = 1

Para x = 2! = 2 × 1 = 2

Para x = 3! = 3 × 2 × 1 = 6

Para x = 4! = 4 × 3 × 2 × 1 = 24

Para x = 5! = 5 × 4 × 3 × 2 × 1 = 120

Para x = 6! = 6 × 5 × 4 × 3 × 2 × 1 = 720

Para x = 7! = 7 × 6 × 5 × 4 × 3 × 2 × 1 = 5040

Para x = 8! = 8 × 7 × 6 × 5 × 4 × 3 × 2 × 1 = 40320

Para x = 9! = 9 × 8 × 7 × 6 × 5 × 4 × 3 × 2 × 1 = 362880

Para x =10! = 10 × 9 × 8 × 7 × 6 × 5 × 4 × 3 × 2 × 1= 3628800

e assim sucessivamente.

Como se pode observar na equação (1) acima, a distribuição de Poisson tem um único parâmetro desconhecido, simbolizado pela letra minúscula grega μ (mu). Ao se conhecer o valor de m pode-se calcular a probabilidade de toda a distribuição de freqüências. O parâmetro μ pode ser interpretado como a média aritmética de ocorrências por intervalo de tempo ou espaço, o qual caracteriza o processo que gera a distribuição de Poisson.

Não obstante, na maioria das observações bibliométricas, as contribuições dos autores que não produzem um único artigo ou trabalho escrito, ou seja, quando *x* = 0, não são observados nem relatados nos trabalhos publicados.

Normalmente, ao se coletar os dados da produtividade científica não é possível observar a produtividade dos autores, diga-se, os improdutivos. Isto é, existem situações em que algumas freqüências não são verificadas, como no caso dos autores com zero colaborações num determinado período. Nessas situações, tem-se uma distribuição de Poisson zero truncada, que é definida por,

$$P_r = \frac{\lambda^r e^{-\lambda}}{r!(1-e^{-\lambda})} \qquad (r=1, 2, 3...) \qquad (2)$$

onde:

λ = *média aritmética de ocorrências por intervalo de tempo.*
e = 2.71828 (a base dos logaritmos naturais)
r = valores 0, 1, 2, ... da variável discreta r

Para resolver esta equação e calcular os valores esperados ou teóricos, Plackett (1953) desenvolveu um método bastante eficiente para estimar λ (Lambda).

Aplicação da distribuição de Poisson zero truncada

A expectativa do parâmetro de Poisson quando os zero valores das freqüências da distribuição não são observadas nas amostras coletadas, como é o caso da maioria das distribuições bibliométricas, é muito complexa. A solução adotada pelo método da máxima probabilidade é realmente problemática já que se faz necessário resolver complicadas equações não-lineares. As propostas de solução, para este caso, procedem de diferentes campos mas, basicamente, da biologia. Vários aspectos da expectativa dos parâmetros que envolvem a distribuição de Poisson zero truncada tem sido considerados por Tippett (1932), Bliss (1948), David e Johnson (1952), Moore (1952), Rider (1953), Finney (1955), Hartley (1958), Irwin (1959) e Cohen (1960), entre outros.

Um exemplo de aplicação da distribuição Poisson zero truncado, segue o método sugerido por Plackett (1953), e utiliza os dados proporcionados por Gupta e Kumar (1998). Estes autores estudaram 11 revistas básicas do campo da genética populacional e aplicaram as distribuições do quadrado inverso, poder inverso generalizado, a distribuição binomial negativa, a distribuição geométrica, e a distribuição truncada de Poisson. Neste artigo, será aplicada somente a distribuição zero truncada de Poisson, usando o método sugerido por Plackett (1953). O processo é explicado passo a passo, porém utilizando somente os dados correspondentes ao periódico *Evolution*. Os dados apresentados pelos pesquisadores e referentes à produtividade dos autores são mostrados na TAB. 1:

TABELA 1 - Número de contribuições por autor no periódico E*volution*

Número de contribuições r	Número de autores Nr
1	56
2	7
3	3
6	1
TOTAL	67

Se se considerar esses dados como uma amostra de 67 autores, tem-se uma distribuição zero truncada, na qual a freqüência dos autores que não produziram um único artigo durante o período estudado estão ausentes, ou seja, os autores não foram contabilizados, e são desconhecidos. Esta é a forma mais comum em que os dados se apresentam nas distribuições bibliométricas da produtividade dos autores, em qualquer campo do conhecimento. Então, o estimado que se busca é:

$$\lambda^* = \frac{\sum_{r=2}^{\infty} r\, N_r}{N} \qquad (3)$$

1 Calcular λ^* (Lambda) ou a media da distribuição de freqüências

Para calcular λ^*, que representa a media da distribuição de freqüências, é necessário primeiro organizar os dados, como na TAB. 2:

TABELA 2 - Cálculo da media da distribuição das freqüências de λ^*

Nº de cotribuições r	Nº de autores N_r	Total de artigos produzidos rN_r
1	56	56
2	7	14
3	3	9
6	1	6
Total	67	

Se se considerar que um total de 85 artigos foram produzidos por 67 autores, quantos autores se estimariam que, no período estudado, produziriam somente 1 artigo? E quantos 2, 3, 4, … *n* artigos cada um? O modelo de Poisson deve responder da forma mais exata possível a estas perguntas.

a) A equação (3) indica um somatório da multiplicação rNr, que vai de r = 2 até o último valor do total de artigos, então, tomando os valores apresentados na TAB. 2, tem-se que: 14 + 9 + 6 = 29 ou, alternativamente, 85 – 56 = 29.

b) Solucionando a equação (3) tem-se:

$$\lambda^* = \frac{\sum_{r=2}^{\infty} r\, N_r}{N} = \frac{29}{67} = 0.43284$$

2 Calcular os valores esperados

Com o valor de $\lambda^* = 0.43284$ já conhecido, calcular os valores esperados usando a equação (2)

Para r = 1 (o número dos autores que produziram 1 artigo)

$$P_1 = \frac{0.43284^1 \times 2.718^{-0.43284}}{1!(1 - 2.718^{-043284})}$$

$$P_1 = \frac{0.43284 \times 0.64869}{1!(1 - 0.35131)} = \frac{0.28077898}{0.35131} = 0.799243236 \times 67 = 53.6$$

Para r =2 (o número dos autores que produziram 2 artigos)

$$P_2 = \frac{0.43284^2 \times 0.64869}{2!(0.35131)} = \frac{0.187350466 \times 0.64869}{2 \times 0.35131} = \frac{0.121532374}{0.70262}$$

$$P_2 = 0.172970273 \times 67 = 11.6$$

Para r = 3 (o número dos autores que produziram 3 artigos)

$$P_3 = \frac{0.43284^3 \times 0.64869}{3!(0.35131)} = \frac{0.081092776 \times 0.64869}{6 \times 0.35131} = \frac{0.052604073}{2.10786}$$

$P_3 = 0.024956151 \times 67 = 1.7$ e assim sucessivamente, para os valores de r = 4, 5 e 6, ou seja, o número dos autores que produziram 4, 5 e 6 artigos.

3 Comparação dos valores observados e esperados

Com os valores esperados ou teóricos calculados, elaborar uma tabela especial de tal maneira que facilite a comparação entre os valores observados e os valores esperados. Esta comparação é mostrada na TAB. 3,

TABELA 3 - Comparação dos valores observados e esperados

No. de contribuições	Freqüências observadas	Freqüências esperadas
r	Nr	rNr
1	56	53.6
2	7	11.6
3	3	1.7
4	0	1.7
5	0	0.02
6	1	0.001
Total	67	67.121

Nesta tabela, observa-se quão próximos ou afastados estão ambos os valores verificados e calculados. Quanto mais próximos estiverem os valores esperados dos valores encontrados, maior será a possibilidade de que os dados se ajustem ao teste qui-quadrado. E, inversamente, quanto mais afastados estiverem os valores esperados dos valores observados maior será a possibilidade de que os dados não se ajustem ao teste do qui-quadrado. Observe também se os totais são os mesmos, quase os mesmos ou divergentes para ambos valores verificados e calculados. Neste caso, as divergências entre ambas observações são mínimas.

4 Estabelecer as hipóteses e selecionar um nível de significância (recomenda-se $\alpha = .05$)

O que se irá testar é se a distribuição de Poisson zero truncada obtida experimentalmente é homogênea ou não. Isto é, a probabilidade de que um elemento incluído na amostra é o mesmo (igualmente provável) para todos os elementos nessa mesma situação. Portanto, serão estabelecidas as hipóteses da seguinte forma:

H_o = a distribuição representa as contagens de r = 1, 2, 3 .....

Ha ≠ a distribuição não representa as contagens de r = 1, 2, 3 .....

5 Calcular o teste estatístico x^2 (qui-quadrado), usando a seguinte equação:

$$x^2 = \sum_{1}^{n} \frac{(f_o - f_t)^2}{f_t}$$

onde,

f_o = valores das freqüências observadas
f_t = valores das freqüências teóricas, esperadas ou calculadas

Para facilitar o cálculo do x^2 (qui-quadrado), deve-se elaborar uma tabela como a que se indica a seguir. Não obstante, é necessário chamar a atenção para o fato de que, quando se aplica o teste x^2 , nenhuma das freqüências observadas deve ser inferior a 5. Ao se encontrar freqüências menores do que 5, será necessário reunir as freqüências adjacentes em grupos não inferiores a 5, para que, dessa forma, o x^2 não perca validez nem consistência. Um exemplo de cálculo do x^2 com agrupamento de freqüências adjacentes inferiores a 5 são mostradas na TAB. 4 na qual as freqüências de 3 a 6 foram agrupadas para produzir um caso extremo e especial de 4 freqüências observadas.

TABELA 4 - Cálculo do x^2 com agrupamento das freqüências observadas inferiores a 5

x	*fof*	*ft*	$(fo - ft)$	$(fo - ft)^2$	$\frac{(fo - ft)^2}{ft}$
1	56	53.6	2.4	5.7600	0.1075
2	7	11.6	-4.6	0.2116	0.0182
3-6	4	1.921	2.079	4.3222	2.2500
				X^2	= 2.3757

6 Especificar a região de rejeição das hipóteses ao nível de significância de $\alpha = .05$

a) Determinar os graus de liberdade:

$$df = k – h - 1 = 3 – 1 – 1 = 1$$

onde,

df = são os graus de liberdade

k = é o número de pares de dados observados depois de agrupar os valores inferiores a 5. Como se mostra na TAB. 4, ao se agrupar as freqüências 3 a 6, o valor de k, neste caso, é igual a 3.

h = é o número de parâmetros usados na solução da equação (2). Neste caso é só um, o parâmetro λ.

1 = é o número de restrições usadas para os cálculos dos valores esperados. Neste caso é 1.

Quando se trata da distribuição de Poisson, o número de graus de liberdade sempre será igual ao número de casos menos 2, porque se emprega a média e a freqüência total para estimar as freqüências esperadas.

b) Usando o nível de significância de $\alpha = .05$ na Tabela do x^2 de qualquer texto estatístico (TAB. 1, em Anexo) encontrar a região de rejeição.

Se o $x^2 > x^2_{.05}$ rejeitar a hipótese nula (aceitar a Ha)

Se o $x^2 < x^2_{.05}$ aceitar a hipótese nula (rejeitar a Ha)

c) Usando a TAB. 1 dos valores críticos do $x^2_{.05}$ (Anexo 1) correspondentes a 1 grau de liberdade, encontra-se um valor igual a 3.84146. Assim, a região de rejeição da hipótese nula é a parte escurecida da FIG. 1 situada à direita. A região de aceitação é a parte à esquerda dessa figura.

Então,

FIGURA1 Gráfico das regiões de aceitação e rejeição das hipóteses.

7 Interpretação

Como o x^2 calculado (igual a 2.3757) é menor que o valor crítico do $x^2_{.05}$ (igual a 3.84146), aceita-se a hipótese nula, ao nível de significância de 0.05. Por isso, conclui-se que as proporções são homogêneas e procedem de uma distribuição do tipo Poisson. Em outras palavras, esta distribuição dos autores produtores de literatura sobre genética populacional se ajusta à distribuição da produtividade dos autores quando se usa o modelo Poisson zero truncado.

8 Dispersão dos valores observados e calculados

Conforme se mostra na FIG. 2, a aproximação ou afastamento entre os valores observados e calculados podem ser melhor visualizados e observados no traçado da dispersão dos dados de ambas as distribuições. Reomenda-se incluir esta figura na redação do informe final.

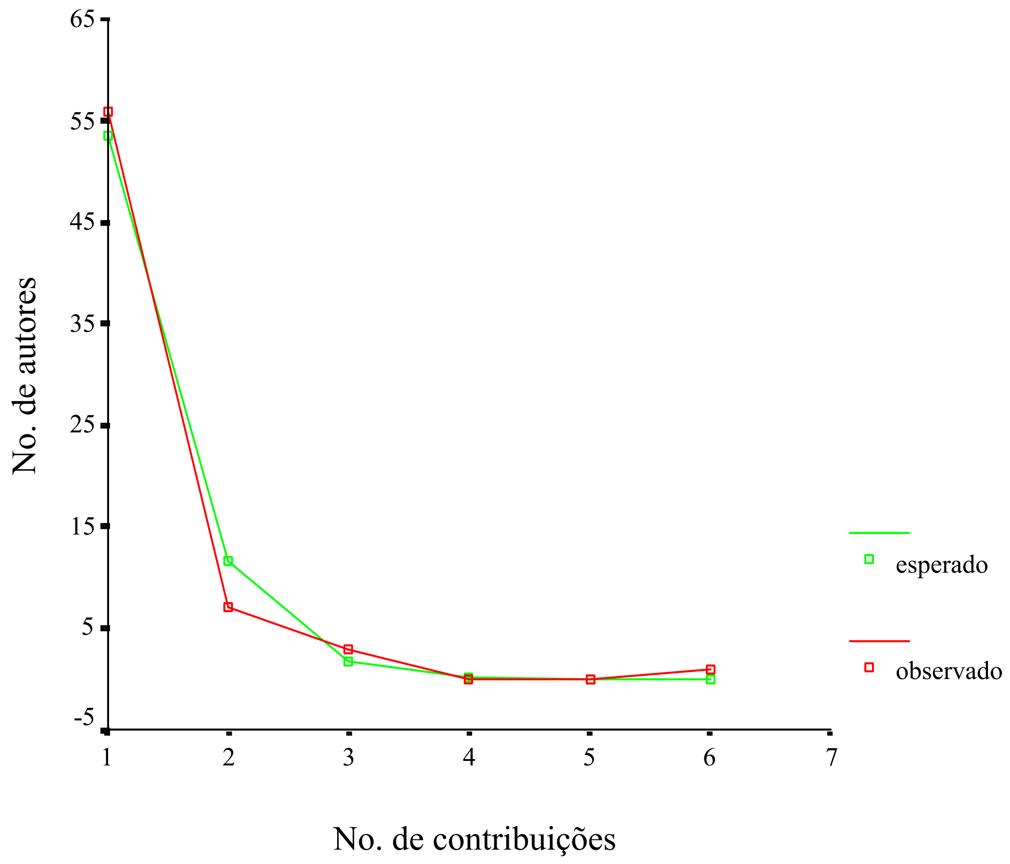


FIGURA 2- Gráfico de dispersão das frequências observadas e esperadas

Conclusão

Utilizando-se o método de estimação do parâmetro λ proposto por Plackett (1953), verificou-se que λ foi igual a 0.43284. Com este valor, estimaram-se adequadamente as freqüências esperadas da produtividade dos autores. O teste x^2 (qui-quadrado) foi usado para avaliar o ajuste dos dados observados e dos dados esperados. Ao 0.05 nível de significância, e com 1 grau de liberdade, observou-se que o valor crítico do x^2 foi igual a 3.84146, muito maior do que o valor do x^2 calculado de 2.3757. Portanto, aceita-se a hipótese nula e conclui-se que esta literatura ajusta-se muito bem à lei de Lotka, usando-se a distribuição de Poisson zero truncado.

A classe zero da distribuição truncada de Poisson

Como assinalado anteriormente, na maioria das observações bibliométricas, a distribuição das contribuições dos autores que não produzem um único artigo ou trabalho escrito (quando $x = 0$) não é observada nem relatada nos trabalhos publicados. Para situações deste tipo (quando se quer conhecer quantos autores não publicaram um único artigo), Dahiya e Gross (1973) propõem um método de expectativa bastante adequado.

Se n_o designa o número de freqüências da classe zero que não foram observadas na amostra, então o total da amostra completa é $N = n_o + n$. Se se permitir que $\hat{\lambda}$ e representem os esperados da máxima probabilidade de λ e N respectivamente, então esses valores são as soluções de:

$$\frac{\hat{\lambda}}{(1 - \partial^{-\hat{\lambda}}} = \hat{X}* \qquad (1.2)$$

e

$$\hat{N} = \frac{\sum_{x=1}^{R} x n_x}{\hat{\lambda}} = \frac{n}{\hat{p}} \qquad (1.3)$$

onde,

$$\widetilde{X}* = \begin{cases} \frac{N}{n}\widetilde{X} & \text{se} \quad n \geq 1 \\ 0 & \text{se} \quad n = 0, \end{cases} \qquad (1.4)$$

e,

$\widetilde{X} = \frac{\sum_{x=1}^{R} x n_x}{N}$ representa a média da amostra completa de N observações.

Para facilitar a solução da equação (1.2), Cohen (1960) proporciona uma tabela especial, que está incluída no Anexo 2. Uma vez que se tenha estimado λ, usando a equação (1.3), pode-se estimar facilmente o valor de $\hat{N}$. Em seguida, calcula-se n_o usando a equação (1.5) abaixo

$$\hat{n}_o = \hat{N} - n = \frac{n\hat{q}}{\hat{p}} \qquad (1.5)$$

onde,

$$\hat{q} = e^{-\hat{\lambda}} \quad \text{e} \quad \hat{p} = \hat{\lambda}$$

Cálculo da classe zero da distribuição truncada de Poisson

Será apresentado um exemplo do cálculo das freqüências não observadas da distribuição de Poisson, usando os dados de Gupta e Kumar (1998) em que os autores fornecem os seguintes dados, correspondentes a trabalhos publicados no periódico *Evolution*. Como se pode verificar na TAB. 5, os pesquisadores não oferecem o número dos autores que não produziram um único artigo, ou seja, aqueles que, teoricamente, tiveram zero artigos publicados, não estão presentes na amostra coletada.

TABELA 5 - Número de contribuições por autor no periódico *Evolution*

Nº de contribuições r	Nº de autores N_r	Total de artigos produzidos rN_r
1	56	56
2	7	14
3	3	9
6	1	6
Total	67	85

1. Calcular $\hat{\lambda}$

Para calcular o valor de $\hat{\lambda}$ será usado o método sugerido por Cohen (1960)

a) Calcular a média:

$$\tilde{x} = \frac{85}{67} = 1.2687$$

b) Estimar $\hat{\lambda}$ usando a TAB. 2 (Anexo 2) e arredondando diretamente a duas ou três decimais, ou por interpolação linear:

Na TAB. 2, a média $x = 1.2687$, arredondados a duas decimais $x = 1.27$, corresponde a $\hat{\lambda} = 0.4987$, arredondados a três decimais, $\hat{\lambda} = 0.499$.

2. Calcular $\hat{N}$ usando a equação (1.3)

$$\hat{N} = \frac{\sum_{x=1}^{k} x n_x}{\hat{\lambda}} = \frac{n}{\hat{p}} \qquad (1.3)$$

$$\hat{N} = \frac{n}{\hat{p}} = \frac{85}{0.499} = 170.34 \pm 170.0$$

3. Calcular n_o usando a equação (1.5)

$$\hat{n}_o = \hat{N} - n = \frac{n\hat{q}}{\hat{p}} \tag{1.5}$$

a) Calcular $\hat{q} = e^{-\hat{\lambda}}$

$$\hat{q} = e^{-\hat{\lambda}} = 2.718^{-0.499} = 0.6072$$

b) Resolver a equação (1.5)

$$\hat{n}_o = \frac{85 \times 0.6072}{0.499} = \frac{51.6094}{0.499} = 103.4256 \pm 103.0$$

Então, pode-se afirmar que existiam 103 autores que, trabalhando dentro do campo da genética populacional, no período estudado por Gupta e Kumar (1998), não submeteram à avaliação um só artigo ao periódico Evolution. Em outras palavras, no campo da genética populacional existiam 103 autores que, no período estudado por Gupta e Kumar (1998), não tiveram produtividade de artigos. Portanto, a população total de autores no campo da genética populacional, plausível de submeterem artigos ao periódico Evolution seria de

$$N = n_o + n$$

$$N = 103 + 67 = 170$$

Este valor de 170 autores é perfeitamente coincidente com o valor encontrado na solução da equação (1.3).

Zero truncated Poisson distribution applicated to authors productivity

Describes the nature of the zero truncated Poisson distribution as developed by Plackett with specific equations for cases when zero observations are not present in the data collected. In the bibliometric area it is very common to find data without zero frequency distributions, an application process of this distribution model is described step-by-step. To illustrate the application, process data collected and studied by Gupta and Kumar on populational genetics literature, and published in the journal Evolution, are replicated.

Key-words: *Lotka's law; Zero truncated Poisson distribution; Author's productivity; Bibliometrics; Genetics; Informetrics; Scientometrics*

Referências

BLISS, C. I. Estimation of the mean and its error from incomplete Poisson distributions.*Connecticut Agricultural Experiment Station Bulletin,* 513, p.1-12, Jan. 1948.

COHEN, A. Clifford. Estimating the parameter in a conditional Poisson distribution.*Biometrics,* v. 16, n. 2, p. 203-211, June 1960.

DAHIYA, Ram C.; GROSS, Alan J. Estimating the zero class from a truncated Poisson sample. *Journal of the American Statistical Association,* v. 68, n. 343, p. 731-733, Sept. 1973.

DAVID, F. N.; JOHNSON, N. L. The truncated Poisson. *Biometrics,* 8(4):275-285, Dec.1952.

FINNEY, D. J.; VARLEY, G. C. An example of the truncated Poisson distribution. *Biometrics,* v. 11, n. 3, p. 387-394, Sept. 1955.

GUPTA, B. M.; KUMAR. Suresh. Scientific productivity in theoretical population genetics: a case study in core journals. *Library Science with a Sland to Documentation and Information Studies,* v. 35, n. 2, p. 89-97, June 1998.

GUPTA, B. M.; KUMAR. Suresh; ROUSSEAU, R. Applicability of selectedprobability distributions to the number of authors per article in theoretical population genetics: *Scientometrics*, v. 42, n.3, p. 325-334, 1998.

HARTLEY, H. O. Maximum likelihood estimation from incomplete data. *Biometrics,*v. 14, p.174-194, 1958.

HUBER, John C. Invention and inventivity as a special kind of creativity, with implications for general creativity. *Journal of Creative Behavior,* v. 32, n.1, p.58-72,1998a.

HUBER, John C. Invention and inventivity is a random Poisson process: a potential guide to analysis of general creativity. *Creativity Research Journal,* v. 11, n. 3, p. 231-241, 1998b.

IRWIN, J. O. On the estimation of the mean of a Poisson distribution from a sample with the zero class missing. *Biometrics,* v. 15, n. 2, p. 324-326, June 1959.

KENDALL, M. G. Natural law in the social sciences. *Journal of the Royal Statistical Society.* Series A, v. 124:1-19, 1961.

MANTELL, Leroy H. On laws of special abilities and the poduction of scientific literature. *American Documentation,* v. 17, n.1, p. 8-16, Jan. 1966.

MOORE*, Biometrika,* v. 39, p. 247-251, 1952.

PLACKETT, R. L. The truncated Poisson distribution. *Biometrics,* v. 9, n. 4, p. 485-497, Dec. 1953.

POTTER, William Gray. Lotka's law revisited. *Library Trends,* v.30, n. 1, p. 21-39, Summer 1981.

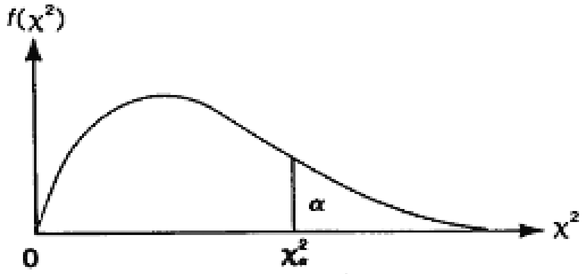
RIDER, Paul R. Truncated Poisson distributions. *Journal of the American Statistical Association,* v. 48, n. 264, p. 826-830, Dec. 1953.

TIPPETT, L. H. C. A modified method of counting particles. *Proceedings of the Royal Society,* Series A, v. 13, p. 434-446, Sept. 1932.

VLACHÝ, Jan. Evaluating the distribution of individual performance. *Scientia Yugoslavica,* v. 6, n. ¼, p. 267-275, 1980.

Anexo I

TABELA 1: Valores críticos de x^2



Graus de liberdade	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$
1	0,000039	0,000157	0,000982	0,003932	0,015791
2	0,010025	0,020101	0,050636	0,102587	0,210721
3	0,071722	0,114832	0,215795	0,351846	0,584374
4	0,206989	0,297110	0,484419	0,710723	1,063623
5	0,411742	0,554298	0,831212	1,145476	1,610308
6	0,675727	0,872090	1,237344	1,635383	2,204131
7	0,989256	1,239042	1,689869	2,167350	2,833107
8	1,344413	1,646497	2,179731	2,732637	3,489539
9	1,734933	2,087901	2,700389	3,325113	4,168159
10	2,155856	2,558212	3,246973	3,940299	4,865182
11	2,603222	3,053484	3,815748	4,574813	5,577785
12	3,073824	3,570569	4,403789	5,226029	6,303796
13	3,565035	4,106915	5,008751	5,891864	7,041505
14	4,074675	4,660425	5,628726	6,570631	7,789534
15	4,600916	5,229349	6,262138	7,260944	8,546756
16	5,142205	5,812212	6,907664	7,961646	9,312236
17	5,697217	6,407760	7,564186	8,671760	10,085186
18	6,264805	7,014911	8,230746	9,390455	10,864936
19	6,843971	7,632730	8,906516	10,117013	11,650910
20	7,433844	8,260398	9,590777	10,850811	12,442609
21	8,033653	8,897198	10,282900	11,591305	13,239598
22	8,642716	9,542492	10,982320	12,338015	14,041493
23	9,260425	10,195720	11,688550	13,090514	14,847956
24	9,886234	10,856360	12,401150	13,848425	15,658684
25	10,519650	11,523980	13,119720	14,611408	16,473408
26	11,160240	12,198150	13,843900	15,379157	17,291885
27	11,807590	12,878500	14,573380	16,151396	18,113896
28	12,461340	13,564710	15,307860	16,927875	18,939242
29	13,121150	14,256450	16,047070	17,708366	19,767744
30	13,786720	14,953460	16,790770	18,492661	20,599235
40	20,706540	22,164260	24,433040	26,509303	29,050523
50	27,990750	29,706680	32,357360	34,764252	37,688648
60	35,534490	37,484850	40,481750	43,187958	46,458888
70	43,275180	45,441720	48,757560	51,739278	55,328940
80	51,171930	53,540080	57,153170	60,391478	64,277844
90	59,196300	61,754080	65,646620	69,126030	73,291090
100	67,327560	70,064890	74,221930	77,929465	82,358136

Anexo II

1	2,705543	3,841459	5,023886	6,634897	7,879439
2	4,605170	5,991465	7,377759	9,210340	10,596635
3	6,251389	7,814728	9,348404	11,344867	12,838156
4	7,779440	9,487729	11,143287	13,276704	14,860259
5	9,236357	11,070498	12,832502	15,086272	16,749602
6	10,644641	12,591587	14,449375	16,811894	18,547584
7	12,017037	14,067140	16,012764	18,475307	20,277740
8	13,361566	15,507313	17,534546	20,090235	21,954955
9	14,683657	16,918978	19,022768	21,665994	23,589351
10	15,987179	18,307038	20,483177	23,209251	25,188180
11	17,275009	19,675138	21,920049	24,724970	26,756849
12	18,549348	21,026070	23,336664	26,216967	28,299519
13	19,811929	22,362032	24,735605	27,688250	29,819471
14	21,064144	23,684791	26,118948	29,141238	31,319350
15	22,307130	24,995790	27,488393	30,577914	32,801321
20	28,411981	31,410433	34,169607	37,566235	39,996846
25	34,381587	37,652484	40,646469	44,314105	46,927890
30	40,256024	43,772972	46,979242	50,892181	53,671962
40	51,805057	55,758479	59,341707	63,690740	66,765962
50	63,167121	67,504807	71,420195	76,153891	79,489978
60	74,397006	79,081944	83,297675	88,379419	91,951698
70	85,527043	90,531225	95,023184	100,425184	104,214899
80	96,578204	101,879474	106,628568	112,328793	116,321057
90	107,565009	113,145270	118,135893	124,116319	128,298944
100	118,498004	124,342113	129,561197	135,806723	140,169489

Anexo III

Tabela 2: $z = \lambda/(1 - e^{-\lambda})$

z	λ	z	λ	z	λ	z	λ	z	λ	z	λ
		1.155	0.2955	1.46	0.8115	2.30	1.9836	4.30	4.2379	7.6	7.5962
1.0005	0.0010	1.160	.3046	1.47	.8273	2.35	2.0464	4.35	4.2904	7.7	7.6965
1.0010	.0020	1.165	.3137	1.48	.8430	2.40	2.1086	4.40	4.3428	7.8	7.7968
1.0015	.0030	1.170	.3227	1.49	.8586	2.45	2.1703	4.45	4.3951	7.9	7.8971
1.0020	.0040	1.175	.3317	1.50	.8742	2.50	2.2316	4.50	4.4473	8.0	7.9973
1.0025	0.0050	1.180	0.3407	1.51	0.8897	2.55	2.2924	4.55	4.4994	8.1	8.0976
1.0030	.0060	1.185	.3497	1.52	.9052	2.60	2.3527	4.60	4.5515	8.2	8.1977
1.0035	.0070	1.190	.3586	1.53	.9207	2.65	2.4126	4.65	4.6034	8.3	8.2979
1.0040	.0080	1.195	.3675	1.54	.9361	2.70	2.4721	4.70	4.6553	8.4	8.3981
1.0045	.0090	1.200	.3764	1.55	.9514	2.75	2.5312	4.75	4.7071	8.5	8.4983
1.005	0.0100	1.205	0.3853	1.56	0.9667	2.80	2.5899	4.80	4.7588	8.6	8.5984
1.010	.0199	1.210	.3942	1.57	.9819	2.85	2.6483	4.85	4.8105	8.7	8.6986
1.015	.0299	1.215	.4030	1.58	.9970	2.90	2.7063	4.90	4.8622	8.8	8.7987
1.020	.0397	1.220	.4118	1.59	1.0121	2.95	2.7640	4.95	4.9137	8.9	8.8988
1.025	.0496	1.225	.4206	1.60	1.0272	3.00	2.8214	5.00	4.9652	9.0	8.9989
1.030	0.0594	1.230	0.4294	1.61	1.0422	3.05	2.8786	5.1	5.0679	9.1	9.0990
1.035	.0692	1.235	.4381	1.62	1.0571	3.10	2.9354	5.2	5.1704	9.2	9.1991
1.040	.0790	1.240	.4468	1.63	1.0720	3.15	2.9919	5.3	5.2728	9.3	9.2992
1.045	.0887	1.245	.4555	1.64	1.0869	3.20	3.0482	5.4	5.3750	9.4	9.3992
1.050	.0984	1.250	.4642	1.65	1.1017	3.25	3.1042	5.5	5.4770	9.5	9.4993
1.055	0.1081	1.26	0.4815	1.66	1.1165	3.30	3.1600	5.6	5.5789	9.6	9.5993
1.060	.1177	1.27	.4987	1.67	1.1312	3.35	3.2156	5.7	5.6806	9.7	9.6994
1.065	.1273	1.28	.5158	1.68	1.1458	3.40	3.2709	5.8	5.7821	9.8	9.7995
1.070	.1369	1.29	.5329	1.69	1.1604	3.45	3.3259	5.9	5.8836	9.9	9.8995
1.075	.1464	1.30	.5499	1.70	1.1750	3.50	3.3808	6.0	5.9849	10.0	9.9995
1.080	0.1559	1.31	0.5668	1.71	1.1895	3.55	3.4356	6.1	6.0871	10.1	10.0996
1.085	.1654	1.32	.5836	1.72	1.2040	3.60	3.4902	6.2	6.1873	10.2	10.1996
1.090	.1749	1.33	.6003	1.73	1.2185	3.65	3.5446	6.3	6.2883	10.3	10.2997
1.095	.1843	1.34	.6170	1.74	1.2329	3.70	3.5988	6.4	6.3893	10.4	10.3997
1.100	.1937	1.35	.6335	1.75	1.2472	3.75	3.6528	6.5	6.4902	10.5	10.4997
1.105	0.2031	1.36	0.6500	1.80	1.3184	3.80	3.7067	6.6	6.5910	10.6	10.5997
1.110	.2125	1.37	.6665	1.85	1.3885	3.85	3.7604	6.7	6.6917	10.7	10.6998
1.115	.2218	1.38	.6829	1.90	1.4578	3.90	3.8140	6.8	6.7924	10.8	10.7998
1.120	.2311	1.39	.6992	1.95	1.5261	3.95	3.8674	6.9	6.8930	10.9	10.8998
1.125	.2404	1.40	.7154	2.00	1.5936	4.00	3.9207	7.0	6.9936	11.0	10.9998
1.130	0.2496	1.41	0.7316	2.05	1.6603	4.05	3.9739	7.1	7.0942	11.2	11.1998
1.125	.2588	1.42	.7477	2.10	1.7263	4.10	4.0269	7.2	7.1946	11.3	11.2999
1.140	.2680	1.43	.7637	2.15	1.7916	4.15	4.0798	7.3	7.2951	11.5	11.4999
1.145	.2772	1.44	.7797	2.20	1.8562	4.20	4.1326	7.4	7.3955	12.0	11.9999
1.150	.2864	1.45	.7956	2.25	1.9202	4.25	4.1853	7.5	7.4958	12.5	12.5000