Linking the Negative Binomial and Logarithmic Series Distributions via their Associated Series

Relacionando las distribuciones binomial negativa y logarítmica vía sus series asociadas

Mauricio Sadinle[a]

Departamento de Estadística, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia

Abstract

The negative binomial distribution is associated to the series obtained by taking derivatives of the logarithmic series. Conversely, the logarithmic series distribution is associated to the series found by integrating the series associated to the negative binomial distribution. The parameter of the number of failures of the negative binomial distribution is the number of derivatives needed to obtain the negative binomial series from the logarithmic series. The reasoning in this article could be used as an alternative method to prove that the probability mass function of the negative binomial distribution sums to one. Finally, an interpretation of the logarithmic series distribution is given by using the presented reasoning.

Key words: Convergent series, Logarithmic series distribution, Negative binomial distribution, Power series distributions.

Resumen

La distribución binomial negativa está asociada a la serie obtenida de derivar la serie logarítmica. Recíprocamente, la distribución logarítmica está asociada a la serie obtenida de integrar la serie asociada a la distribución binomial negativa. El parámetro del número de fallas de la distribución binomial negativa es el número de derivadas necesarias para obtener la serie binomial negativa de la serie logarítmica. El razonamiento presentado puede emplearse como un método alternativo para probar que la función de masa de probabilidad de la distribución binomial negativa suma uno. Finalmente, se presenta una interpretación de la distribución logarítmica usando el razonamiento planteado.

Palabras clave: distribución binomial negativa, distribución de series de potencias, distribución logarítmica, series convergentes.

[a]Student. E-mail: msadinleg@unal.edu.co

1. Introduction

In statistical literature, many pages have been dedicated finding relations between probability distributions. For example, Casella & Berger (2002) present an interesting diagram that illustrates the relationships between several common distributions through transformations or limits. Recently, an extension of this diagram has been presented by Leemis & McQueston (2008). Their paper shows the relationships among 76 univariate probability distributions: 19 discrete and 57 continuous.

The relations between some common discrete distributions with infinite support are well known since the work of Quenouille (1949), who showed the links between the logarithmic series, Poisson, and negative binomial distributions. Also Anscombe (1950) carries out several investigations in order to establish relations between the negative binomial and logarithmic series distributions by finding their sampling properties. He also presents the method for parameter estimation and some tests for departure from these distributions.

Power series have been used to find properties of discrete distributions. For example, Samaniego (1992) uses the geometric series and related equalities to find the moments of the geometric distribution. In another way, Casella & Berger (2002, p. 74) use the method of taking derivatives of the series $\sum_{n=0}^{\infty}(1 - \pi)\pi^n$ to find the mean of the geometric distribution. Following these type of reasonings, this paper uses the idea first presented by Noack (1950) of associating a convergent power series to a probability mass function (pmf) of some infinite discrete random variable.

Though probability distributions model real situations, their mathematical expressions are sometimes connected. In this paper, integrals and derivatives of the series associated to the negative binomial and logarithmic series distributions are used to find the relationship between these distributions. In Section 3 the parameter of the number of failures of the negative binomial distribution is found to be the number of derivatives of the logarithmic series needed to find the negative binomial series. The work in Section 3 can be also used to prove that the pmf of the negative binomial distribution sums to one. In Section 4 the above parameter is found to be related to the number of integrals taken from the series associated to a negative binomial distribution with a major parameter.

The relationship between the logarithmic series distribution and the negative binomial is not as explicit as the relation between the geometric distribution and the negative binomial distribution. However, many authors have showed the existent relationship through different ways (Fisher et al. 1943, Anscombe 1950, Ord 1967). The method explained here provides a simple way to find this relationship, which is presented in Section 4.

At the end of this article it is shown a diagram inspired by the work of Leemis & McQueston (2008) that summarizes the ideas presented.

2. From Convergent Power Series to Probability Mass Functions

Suppose that you have a convergent power series, i.e.

$$\sum_{n=l}^{\infty} a_n x^n = L(x)$$

with a_n depending only on n, and x a real number in the convergence interval $(-r, r)$. Then, it is possible to establish an associated pmf of some infinitely discrete random variable as follows

$$P(n \mid x) = \begin{cases} \frac{a_n x^n}{L(x)}, & \text{if } n = l, l+1, \ldots; \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

provided that $a_n x^n / L(x) \geq 0$ for all n. See the papers of Noack (1950) and Khatri (1959) for a general discussion about this distribution and its properties with $l = 0$, and the paper of Patil (1962) in which he introduces the generalized power series distribution when n varies within an arbitrary non-null subset of non-negative integers.

Example 1. It is well known that the series $\sum_{n=0}^{\infty} x^n / n!$ converges to e^x, for all x. If $x > 0$, then $e^{-x} x^n / n! > 0$ for $n = 0, 1, 2, \ldots$, and the pmf of the Poisson distribution is obtained.

In this paper we work with convergent series for $x \in (0, 1)$ which denotes a non-trivial probability of success in a Bernoulli trial. To follow a classical notation let us replace x by π.

For the developments presented in the following sections, take into account that if a power series converges for all $\pi \in (0, 1)$, then, also for all $\pi \in (0, 1)$,

$$\frac{d}{d\pi} \sum_{n=0}^{\infty} a_n \pi^n = \sum_{n=0}^{\infty} a_n \frac{\partial}{\partial \pi} \pi^n = \sum_{n=0}^{\infty} a_n n \pi^{n-1}$$

and

$$\int \sum_{n=0}^{\infty} a_n \pi^n d\pi = \sum_{n=0}^{\infty} a_n \int \pi^n d\pi = \sum_{n=0}^{\infty} \frac{a_n}{n+1} \pi^{n+1} + c$$

This is obtained from general theorems about power series. See, e.g., Apostol (1988, p. 529) and Casella & Berger (2002, p. 74) for results about convergent series.

3. From the Logarithmic Series to the Negative Binomial Distribution

The logarithmic series pmf is given by

$$P(n \mid \pi) = \frac{\pi^n}{-n \log(1 - \pi)}, \qquad n = 1, 2, \ldots$$

with $\pi \in (0, 1)$. Its associated power series is

$$\pi + \frac{\pi^2}{2} + \frac{\pi^3}{3} + \frac{\pi^4}{4} + \cdots + \frac{\pi^n}{n} + \cdots = -\log(1 - \pi) \tag{2}$$

Taking derivatives of this series with respect to π, the geometric series

$$1 + \pi + \pi^2 + \pi^3 + \cdots + \pi^n + \cdots = \frac{1}{1 - \pi} \tag{3}$$

is found.

Multiplying both sides by $(1 - \pi)$, it is found that

$$(1 - \pi) + (1 - \pi)\pi + (1 - \pi)\pi^2 + \cdots + (1 - \pi)\pi^n + \cdots = 1$$

where the generic term of this series, $(1 - \pi)\pi^n$, is the pmf of the well known geometric distribution, where $n = 0, 1, 2, \ldots$ represents the number of successes before the first failure in a sequence of independent Bernoulli trials with parameter π.

Now let us take derivatives at both sides of (3)

$$1 + 2\pi + 3\pi^2 + \cdots + n\pi^{n-1} + (n + 1)\pi^n + \cdots = \frac{1}{(1 - \pi)^2} \tag{4}$$

and following (1), i.e. in this case by multiplying by $(1 - \pi)^2$, one finds a series which sums to 1 with generic term

$$(n + 1)(1 - \pi)^2 \pi^n = \binom{n + 1}{1}(1 - \pi)^2 \pi^n = \binom{2 + n - 1}{2 - 1}(1 - \pi)^2 \pi^n$$

which is the pmf of the negative binomial distribution with parameters 2 and π, for $n = 0, 1, 2, \ldots$. Again, let us take derivatives of (4)

$$2 + 3 \times 2\pi + 4 \times 3\pi^2 + \cdots + (n + 1)n\pi^{n-1} + (n + 2)(n + 1)\pi^n + \cdots = \frac{2}{(1 - \pi)^3}$$

and the associated pmf is found to be, again as in (1),

$$\frac{(n + 2)(n + 1)}{2}(1 - \pi)^3 \pi^n = \binom{n + 2}{2}(1 - \pi)^3 \pi^n = \binom{3 + n - 1}{3 - 1}(1 - \pi)^3 \pi^n$$

for $n = 0, 1, 2, \ldots$ which is the pmf of the negative binomial distribution with parameters 3 and π.

Following this idea, taking k derivatives at each side of (2), one finds the series from which the negative binomial distribution with parameters k and π can be obtained. Given the previous work, to demonstrate this result (by mathematical induction), we need to assume that the result is true for k and demonstrate that it is true for $k + 1$, as follows:

Suppose that by taking k derivatives of (2) we find the series

$$(k-1)! + \frac{k!}{1!}\pi + \frac{(k+1)!}{2!}\pi^2 + \cdots + \frac{(k+n-1)!}{n!}\pi^n + \cdots = \frac{(k-1)!}{(1-\pi)^k} \qquad (5)$$

which is associated to the negative binomial distribution with parameters k and π, following (1)

$$\frac{(k+n-1)!}{(k-1)!n!}(1-\pi)^k\pi^n = \binom{k+n-1}{k-1}(1-\pi)^k\pi^n$$

for $n = 0, 1, 2, \ldots$. Let us take derivatives of (5) with respect to π

$$k! + \frac{(k+1)!}{1!}\pi + \frac{(k+2)!}{2!}\pi^2 + \cdots + \frac{(k+n)!}{n!}\pi^n + \cdots = \frac{k!}{(1-\pi)^{k+1}}$$

Following (1) we obtain

$$\binom{k+n}{k}(1-\pi)^{k+1}\pi^n, \qquad \text{for } n = 0, 1, 2, \ldots$$

which is the pmf of the negative binomial distribution with parameters $k + 1$ and π, and the result is demonstrated.

As stated above, taking the generic term of the series and dividing it by the sum of the series, we find the associated pmf. If instead of this if we take all the series and divide it by its sum, we will find a proof that the associated pmf sums to one. This could be useful if we start, for instance, from the widely known convergence of the geometric series, following the steps of this section we find the series associated to the negative binomial distribution and dividing by its sum, we demonstrate that the pmf of this distribution sums to one.

Casella & Berger (2002, p. 95) mentions that the traditional proof of the fact that the negative binomial pmf sums to one utilizes an extension of the binomial theorem that includes negative exponents. This usual proof leads to the use of binomial coefficients with negative integers, which is not taught in regular calculus courses and which many people find difficult to handle. Thus, the above results can be used to present a more pedagogical way to prove that the pmf of the negative binomial distribution sums to one.

4. From the Negative Binomial to the Logarithmic Series Distribution

In the previous section we found that by taking derivatives of the series associated to the negative binomial distribution with parameters k and π, the series

associated to the negative binomial distribution with parameters $k + 1$ and π is obtained.

Conversely, taking the series associated to the negative binomial with parameters k and π and integrating it, it can be found the series associated to the negative binomial with parameters $k - 1$ and π, taking the constant of integration to be equal to $(k - 2)!$ in the side of the series and zero in the side of the sum. This means that if we are studying the probability of getting n successes before k failures, and if we integrate the associated series, we will obtain a series with an associated pmf for studying the probability of getting n successes before $k - 1$ failures.

The geometric pmf, which is equivalent to the negative binomial pmf with parameters 1 and π, has associated the geometric series, and from the above idea, it is natural to think of integrating it to find a way of modeling the probability of getting n successes, taking the parameter of the number of failures as zero. The constant of integration $(k - 2)! = (1 - 2)! = (-1)!$ is not defined, so for convenience it is taken as zero in this case, and the following series is obtained

$$\pi + \frac{\pi^2}{2} + \frac{\pi^3}{3} + \cdots + \frac{\pi^n}{n} + \cdots = -\log(1 - \pi)$$

This series has the associated pmf

$$\frac{\pi^n}{-n\log(1 - \pi)}, \qquad \text{for } n = 1, 2, \ldots$$

which is the logarithmic series pmf. The interpretation of the above distribution is not simple. In fact, this distribution is a limit case of the negative binomial distribution, and it is found by Fisher et al. (1943) through a different reasoning. It is also presented by Anscombe (1950) as a limiting process of the negative binomial distribution by considering a sample of N Bernoulli trials ($N \neq 0$), letting N tend to infinity and k to zero in terms of the Gamma function, as follows

$$\frac{\Gamma(k + n)}{n!\Gamma(k)}(1 - \pi)^k \pi^n, \qquad \text{for } n = 0, 1, 2, \ldots$$

Ord (1967), by graphical methods, found in a natural way that the logarithmic series distribution occurs as the limit of the negative binomial, when the parameter of the number of failures tends to zero.

5. Conclusions

The parameter of the number of failures of the negative binomial distribution is the number of derivatives needed to obtain the negative binomial series from the logarithmic series. Thus the negative binomial pmf with parameters k and π is found by

$$P_{NB}(n \mid k, \pi) = \begin{cases} \frac{d^k}{d\pi^k}\big((n+k)^{-1}\pi^{n+k}\big) \big/ \frac{d^k}{d\pi^k}\big(-\log(1-\pi)\big), & \text{if } n = 0, 1, 2, \dots; \\ 0, & \text{otherwise.} \end{cases}$$

$$\sum_{n=1}^{\infty} \frac{\pi^n}{n} = -\log(1-\pi) \quad \xleftarrow[\substack{a_n\pi^n/L(\pi)}]{\sum a_n\pi^n = L(\pi)}\xrightarrow{} \quad LogS(\pi)$$

$$\frac{d}{d\pi} \bullet \Big\downarrow \qquad \Big\uparrow \int \bullet \, d\pi$$

$$\sum_{n=0}^{\infty} \pi^n = \frac{1}{1-\pi} \quad \xleftarrow[\substack{a_n\pi^n/L(\pi)}]{\sum a_n\pi^n = L(\pi)}\xrightarrow{} \quad G(\pi)$$

$$\frac{d}{d\pi} \bullet \Big\downarrow \qquad \Big\uparrow \int \bullet \, d\pi$$

$$\sum_{n=0}^{\infty} (n+1)\pi^n = \frac{1}{(1-\pi)^2} \quad \xleftarrow[\substack{a_n\pi^n/L(\pi)}]{\sum a_n\pi^n = L(\pi)}\xrightarrow{} \quad NB(2, \pi)$$

$$\frac{d}{d\pi} \bullet \Big\downarrow \qquad \Big\uparrow \int \bullet \, d\pi$$

$$\sum_{n=0}^{\infty} (n+2)(n+1)\pi^n = \frac{2}{(1-\pi)^3} \quad \xleftarrow[\substack{a_n\pi^n/L(\pi)}]{\sum a_n\pi^n = L(\pi)}\xrightarrow{} \quad NB(3, \pi)$$

$$\frac{d}{d\pi} \bullet \vdots \quad \Big\uparrow$$

$$\Big\downarrow \quad \vdots \int \bullet \, d\pi$$

$$\sum_{n=0}^{\infty} \frac{(k+n-2)!}{n!}\pi^n = \frac{(k-2)!}{(1-\pi)^{k-1}} \quad \xleftarrow[\substack{a_n\pi^n/L(\pi)}]{\sum a_n\pi^n = L(\pi)}\xrightarrow{} \quad NB(k-1, \pi)$$

$$\frac{d}{d\pi} \bullet \Big\downarrow \qquad \Big\uparrow \int \bullet \, d\pi$$

$$\sum_{n=0}^{\infty} \frac{(k+n-1)!}{n!}\pi^n = \frac{(k-1)!}{(1-\pi)^k} \quad \xleftarrow[\substack{a_n\pi^n/L(\pi)}]{\sum a_n\pi^n = L(\pi)}\xrightarrow{} \quad NB(k, \pi)$$
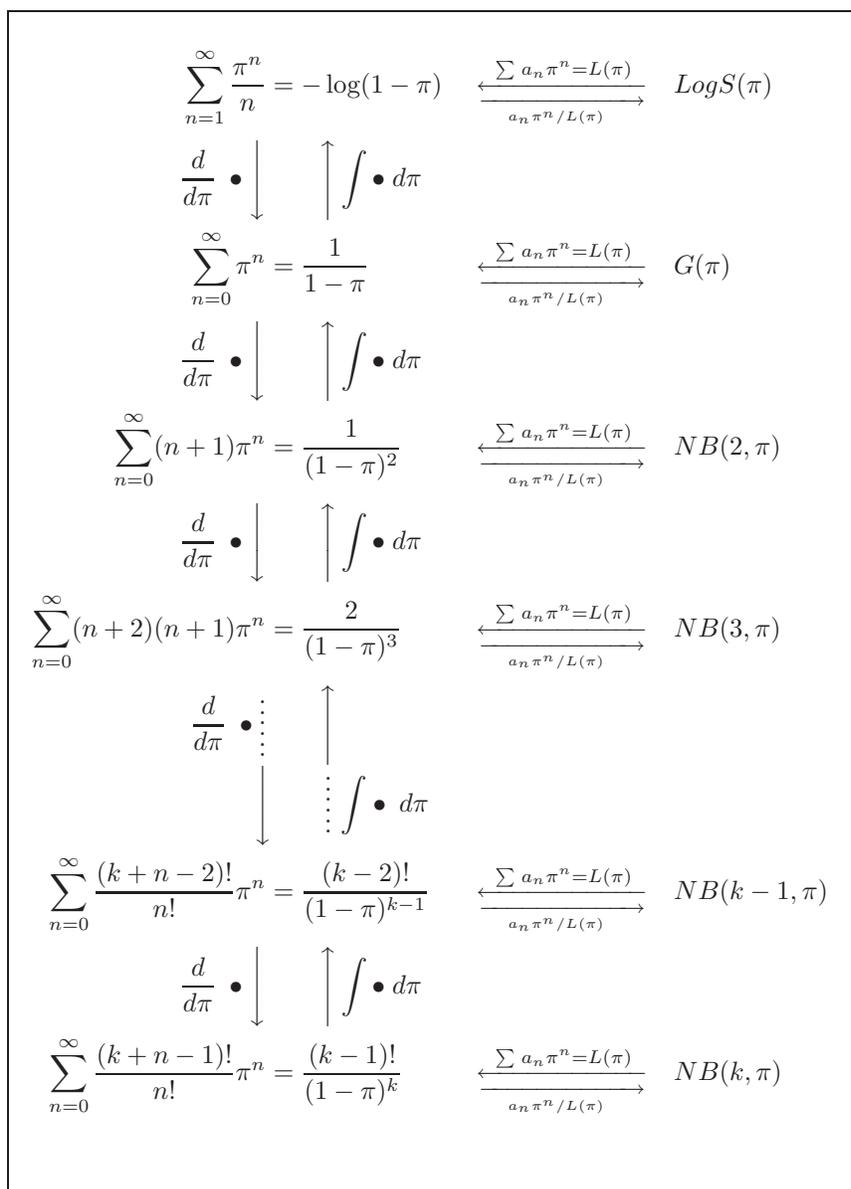
FIGURE 1: Relationships between logarithmic series distribution $LogS(\pi)$, geometric distribution $G(\pi)$ and negative binomial distribution $NB(n, \pi)$.

Conversely, this idea can be applied by replacing derivatives by i integrals with the analogous terms of the series associated to the negative binomial with parameters k and π. For $i = k - k'$ we can find the negative binomial with parameters k' and π, for $1 < k' < k$. For $i = k - 1$ we can find the geometric distribution, and finally, for $i = k$ we can find the logarithmic series distribution. Do not forget to take into account the appropriate integration constants.

The logarithmic series distribution is interpreted by using the presented method as a distribution for modeling the probability of getting n successes letting the number of failures of the negative binomial experiment tend to zero, as shown by Ord (1967).

Figure 1 shows the relationships between the different pmf of the above mentioned distributions. The link to go from one distribution to another one follows the next algorithm. First, choose a given pmf. Second, find the associated series to such pmf. Third, either differentiate or integrate that series as many times as needed to obtain the series associated to the pmf wanted. Fourth, divide the general term of the obtained series by its summation.

For instance, the first derivative of the geometric series, associated to the geometric distribution, produces the series associated to the negative binomial distribution with parameters 2 and π. While integrating once and taking the integration constant as zero, produces the logarithmic series, associated to the logarithmic series distribution. Similarly, if one would like to obtain the negative binomial pmf with parameters 3 and π from the geometric series, thus one should differentiate it twice and finally divide the general term of the series found by the summation.

Finally, if you do not follow the fourth step in the algorithm above, and instead of this you divide the series obtained in the third step by its sum, you will find a way to prove that the pmf, of the distribution associated to the series, sums to one.

Acknowledgments

References

Anscombe, F. J. (1950), 'Sampling Theory of the Negative Binomial and Logarithmic Series Distributions', *Biometrika* **37**(3/4), 358–382.

Apostol, T. M. (1988), *Calculus*, second edn, Reverté.

Casella, G. & Berger, R. L. (2002), *Statistical Inference*, second edn, Duxbury Thomson Learning, Pacific Grove, United States.

Fisher, R. A., Corbet, A. S. & Williams, C. B. (1943), 'The Relation between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population', *The Journal of Animal Ecology* **12**(1), 42–58.

Khatri, C. G. (1959), 'On Certain Properties of Power-Series Distributions', *Biometrika* **46**(3/4), 486–490.

Leemis, L. M. & McQueston, J. T. (2008), 'Univariate Distribution Relationships', *The American Statistician* **62**(1), 45–53.

Noack, A. (1950), 'A Class of Random Variables with Discrete Distributions', *The Annals of Mathematical Statistics* **21**(1), 127–132.

Ord, J. K. (1967), 'Graphical Methods for a Class of Discrete Distributions', *Journal of the Royal Statistical Society* **130**(2), 232–238.

Patil, G. P. (1962), 'On Homogeneity and Combined Estimation for the Generalized Power Series Distribution and Certain Applications', *Biometrics* **18**(3), 365–374.

Quenouille, M. H. (1949), 'A Relation between the Logarithmic, Poisson, and Negative Binomial Series', *Biometrics* **5**(2), 162–164.

Samaniego, F. J. (1992), 'Elementary Derivations of Geometric Moments', *The American Statistician* **46**(2), 108–109.