Revisiting Morrison's Series Approximation for Estimating the Parameters of the NBD

Peter S. Fader
www.petefader.com

Bruce G. S. Hardie
www.brucehardie.com†

January 2025

1 Introduction

When dealing with any real-world count data associated with human behaviour, we should assume that it is overdispersed (Hanley and Bhatnagar 2022). The "go to" distribution for characterising such data is the negative binomial distribution (NBD), which has pmf

$$P(X = x \mid r, \alpha) = \frac{\Gamma(r + x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha + 1}\right)^r \left(\frac{1}{\alpha + 1}\right)^x \tag{1}$$

and mean

$$E(X \mid r, \alpha) = r/\alpha \,. \tag{2}$$

Given the task of fitting the NBD to a dataset, most analysts would estimate the parameters using the method of maximum likelihood. However, this was not always the case. Greenwood and Yule (1920), who first proposed the gamma mixture of Poisson distributions as a chance mechanism that generates the NBD, used the method of moments. While it was known that method of moments estimators are not efficient (Jeffreys 1939) and maximum likelihood estimators proposed (e.g., Haldane 1941), maximum likelihood methods were rarely used because of their complexity given the computational tools at statisticians' disposal in that era (Anscombe 1949).

Anscombe (1949, 1950) proposed an alternative estimation method in which the sample mean and proportion of zeros are equated with their theoretical counterparts and we solve for the parameters. This is known as the method of mean and zeros.

Let p_0 be the observed proportion of zeros in the dataset and $\bar{x}$ the sample mean. This gives us two equations,

$$p_0 = \left(\frac{\alpha}{\alpha+1}\right)^r \tag{3}$$

$$\bar{x} = r/\alpha \,, \tag{4}$$

with two unknowns (r and α).

It follows from (4) that $\alpha = r/\bar{x}$. Substituting this in (3) gives us

$$p_0 = \left(\frac{r}{r+\bar{x}}\right)^r . \tag{5}$$

Finding the root of this equation gives us $\hat{r}$. It follows that $\hat{\alpha} = \hat{r}/\bar{x}$. Alternatively, it follows from (4) that $r = \alpha\bar{x}$. Substituting this in (3) gives us

$$p_0 = \left(\frac{\alpha}{\alpha+1}\right)^{\alpha\bar{x}} . \tag{6}$$

Finding the root of this equation gives us $\hat{\alpha}$, and $\hat{r} = \bar{x}\hat{\alpha}$. A solution to (5) and (6) exists if $-\bar{x}/\ln(p_0) > 1$; see Appendix A.

Anscombe (1949, 1950) showed that the method of mean and zeros estimator has a very high large-sample efficiency when the distribution is reverse J-shaped (with more zeros than ones). Many datasets to which to NBD is fitted have this shape and therefore "mean and zeros" became a popular method for estimating the parameters of the NBD.

There is no explicit solution to (5) or (6). The root can be found using iterative methods; see Appendix B. Evans (1953) and Chatfield (1969) derived tables that make it possible to get an approximate solution to (5) by simply interpolation.

In a paper that has received very little attention, Morrison (1969) used a series approximation to derive an explicit formula that gives us an accurate estimate of the root of (5) or (6). We now present a detailed derivation of Morrison's result.

2 Morrison's Series Approximation

At the heart of Morrison's solution is the idea of series reversion:

> "If you have a power series for a function $f(y)$, then it is often possible to get a power series approximation to the solution for y

in the equation $f(y) = x$. This power series effectively gives the inverse function f^{-1} such that $f(f^{-1}(x)) = x$. The operation of finding the power series for an inverse function is sometimes known as *reversion* of power series."[1]

Given the power series

$$y = a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5 + \cdots ,$$

the reversed series is

$$x = A_1 y + A_2 y^2 + A_3 y^3 + A_4 y^4 + A_5 y^5 + \cdots ,$$

where the new coefficients $(A_1, A_2, \ldots)$ can be expressed in terms of the original coefficients $(a_1, a_2, \ldots)$. Abramowitz and Stegun (1972, equation 3.6.25) presents expressions for $A_1, \ldots, A_7$. Methods for deriving the coefficients of higher-order terms can be found in a number of old mathematical references. These days, it is easier to delegate the task to software.

Letting $u = \bar{x}/(r + \bar{x})$, which implies $r = \bar{x}(1-u)/u$, (5) becomes

$$p_0 = (1-u)^{\bar{x}\frac{1-u}{u}} .$$

Taking the log of both sizes and rearranging terms gives us

$$\frac{\ln(p_0)}{\bar{x}} = \Big(\frac{1-u}{u}\Big) \ln(1-u) . \tag{7}$$

(We arrive at the same expression for (6) by letting $u = 1/(\alpha + 1)$, which implies $\alpha = (1-u)/u$.)

Noting that a Maclaurin series expansion for the natural logarithm is

$$\ln(1-x) = -\sum_{n=1}^{\infty} \frac{x^n}{n} = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \cdots , \qquad |x| < 1 ,$$

we can rewrite (7) as

$$\frac{\ln(p_0)}{\bar{x}} = \Big(\frac{1-u}{u}\Big)\Big(-u - \frac{u^2}{2} - \frac{u^3}{3} - \frac{u^4}{4} - \cdots \Big)$$

which is equivalent to

$$\frac{\ln(p_0)}{\bar{x}} + 1 = \frac{1}{2}u + \frac{1}{6}u^2 + \frac{1}{12}u^3 + \cdots + \frac{1}{i(i+1)}u^i + \cdots .$$

Reverting this series gives us a power series approximation to the solution for u.

[1] `https://reference.wolfram.com/language/tutorial/SeriesLimitsAndResidues.html#31107` (Accessed 2025-01-03)

Morrison reported the first six terms of the reverted series. Using Mathematica's `InverseSeries` command, we compute (and report in the table below) the coefficients for the first 15 terms.[2] We do not add extra terms because of numerical precision issues in basic numerical computing environments. For example, base-R starts changing the digits in a long integer after the fifteenth or sixteenth digit.[3] If you enter a number with more than 15 digits in an Excel worksheet cell, any digits past the fifteenth digit are changed to zero (Microsoft 2024). The denominator of A_{16} contains 16 digits.

i	a_i	A_i
1	$\frac{1}{2}$	2
2	$\frac{1}{6}$	$-\frac{4}{3}$
3	$\frac{1}{12}$	$\frac{4}{9}$
4	$\frac{1}{20}$	$-\frac{16}{135}$
5	$\frac{1}{30}$	$\frac{8}{405}$
6	$\frac{1}{42}$	$-\frac{16}{2835}$
7	$\frac{1}{56}$	$-\frac{32}{42525}$
8	$\frac{1}{72}$	$-\frac{128}{127575}$
9	$\frac{1}{90}$	$-\frac{32}{45927}$
10	$\frac{1}{110}$	$-\frac{103616}{189448875}$
11	$\frac{1}{132}$	$-\frac{1726784}{3978426375}$
12	$\frac{1}{156}$	$-\frac{54631168}{155158628625}$
13	$\frac{1}{182}$	$-\frac{19316224}{66496555125}$
14	$\frac{1}{210}$	$-\frac{13582336}{55857106305}$
15	$\frac{1}{240}$	$-\frac{159899648}{775793143125}$

This results in the following procedure for estimating r and α.

[2]Note that Morrison's values for $A_2, \ldots, A_6$ do not exactly match those reported here. We assume that this is due to rounding error in his calculations.

[3]This can be overcome by using packages such as `gmp` and `Rmpfr`. However, if we are worried about such accuracy, we probably should not be using this approach to estimating the NBD's parameters.

Given p_0 and $\bar{x}$:

$$v \leftarrow \frac{\ln(p_0)}{\bar{x}} + 1$$

$$u \leftarrow \sum_{i=1}^{15} A_i v^i$$

$$\hat{\alpha} \leftarrow \frac{1-u}{u}$$

$$\hat{r} \leftarrow \bar{x}\hat{\alpha}$$

To illustrate this, suppose the NBD with $r = 0.161$ and $\alpha = 0.129$ is the true data-generating process.[4] With reference to Figure 1, this implies the proportion of zeros is 0.705 and the mean is 1.248 (cells B4:B5, computed by evaluating (1) and (2)). We compute v in cell B7. Given the values of A_i and v^i in cells H3:I17, we compute u (cell B8) using the formula `=SUMPRODUCT(H3:H17,I3:I17)`. The resulting estimates of r and α are very close to the true values — see the absolute errors reported in cells B13:B14. Traditional root-finding methods recover the original parameters — see Appendix B.

	A	B	C	D	E	F	G	H	I
1	r	0.161			A_i				
2	alpha	0.129		i	num.	denom.		A_i	v^i
3				1	2	1		2.00000	0.72016
4	p_0	0.705		2	-4	3		-1.33333	0.51864
5	mean	1.248		3	4	9		0.44444	0.37350
6				4	-16	135		-0.11852	0.26898
7	v	0.72016		5	8	405		0.01975	0.19371
8	u	0.88574		6	-16	2835		-0.00564	0.13950
9	alpha (est)	0.12900		7	-32	42525		-0.00075	0.10047
10	r (est)	0.16100		8	-128	127575		-0.00100	0.07235
11				9	-32	45927		-0.00070	0.05211
12	Absolute error			10	-103616	189448875		-0.00055	0.03752
13	r	3.99E-06		11	-1726784	3978426375		-0.00043	0.02702
14	alpha	3.19E-06		12	-54631168	155158628625		-0.00035	0.01946
15	p_0	3.61E-06		13	-19316224	66496555125		-0.00029	0.01402
16	mean	0.00E+00		14	-13582336	55857106305		-0.00024	0.01009
17				15	-159899648	775793143125		-0.00021	0.00727

Figure 1: Implementing Morrison's series approximation

Microsoft Excel calculates formulas and stores the results with 15 significant digits of precision. If we set p_0 to 0.705 and $\bar{x}$ to 1.248 (i.e., round the true values to three d.p.), we get $\hat{r} = 0.161239$ and $\hat{\alpha} = 0.129198$. These are very close to the values we get using traditional root-finding methods ($\hat{r} = 0.161243$ and $\hat{\alpha} = 0.129201$) — see Appendix B.

[4] These parameter values are the maximum likelihood estimates (rounded to three d.p.) associated with fitting the NBD to the champagne purchasing data reported in Gourieroux and Visser (1997).

Morrison (1969) presents three examples to check the accuracy of his six-term series expansion.

- In example 1, $p_0 = 0.1$ and $\bar{x} = 4$. He obtains $\hat{r} = 2.26014$ and reports that $p_0 - \widehat{p_0} = -9.35 \times 10^{-7}$. Our fifteen-term series expansion gives us $\hat{r} = 2.26016$ and the resulting error in the estimate of p_0 is 1.15×10^{-10}.
- In example 2, $p_0 = 0.2$ and $\bar{x} = 4$, which corresponds to $r = 1$. He obtains $\hat{r} = 0.999701$ and reports that $p_0 - \widehat{p_0} = -4.85 \times 10^{-5}$. We obtain $\hat{r} = 1.00000$ with an errror in the estimate of p_0 of 9.89×10^{-8}.
- In example 3, $p_0 = 0.5$ and $\bar{x} = 3.75$, which corresponds to $r = 0.25$. Morrison's series expansion yields $\hat{r} = 0.249865$. We obtain $\hat{r} = 0.249898$.

As we would expect, adding the extra terms to the series expansion yields more precise estimates.

3 Relevance in the 21st Century

Sprott (1983, p. 457) comments that "[i]n the present age of computer sophistication, computational difficulty is no longer a justification for seeking alternative and inefficient estimation procedures in place of maximum likelihood estimation, particularly when there are only two or three parameters to be estimated." What was true in the early 1980s is even more true in the present day. It is a trivial exercise to fit the NBD to a dataset using maximum likelihood estimation; we can do this in a Excel spreadsheet (assuming an optimization add-in such as Solver is installed). So why would we care about the method of mean and zeros?

If we have the raw count data or a frequency distribution, we should not care; we would use maximum likelihood estimation. However, if the only data at hand are the sample mean and proportion of zeros, it is the only way to go.

Within marketing, two common brand performance measures are penetration and purchases per buyer (PPB). Penetration is the proportion of the sample that made at least one purchase in the time period of interest, and PPB is the average number of times the product was purchased by those that purchased it at least once in that same time period. We note that

$$\text{penetration} = 1 - p_0 \tag{8}$$

and

$$\text{PPB} = \frac{\bar{x}}{1 - p_0} \,. \tag{9}$$

Given these two numbers for any brand and the assumption that the distribution of the number of transactions can be characterized by the NBD, we can estimate the model parameters using the method of mean and zeros and then compute related brand performance measures without needing access to the raw transaction data. Morrison's series approximation means these can be computed immediately.

References

Abramowitz, Milton and Irene A. Stegun (eds.) (1972), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York, NY: Dover Publications.

Anscombe, F. J. (1949), "The Statistical Analysis of Insect Counts Based on the Negative Binomial Distribution," *Biometrics*, **5** (June), 165–173.

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

Banerjee, A. K. and G. K. Bhattacharyya (1976), "A Purchase Incidence Model with Inverse Gaussian Interpurchase Times," *Journal of the American Statistical Association*, **71** (December), 823–829.

Chatfield, C. (1969), "On Estimating the Parameters of the Logarithmic Series and Negative Binomial Distributions," *Biometrika*, **56** (August), 411–414.

Evans, D. A. (1953), "Experimental Evidence Concerning Contagious Distributions in Ecology," *Biometrika*, **40** (June), 186–211.

Gourieroux, C. and M. Visser (1997), "A Count Data Model with Unobserved Heterogeneity," *Journal of Econometrics*, **79** (August), 247–268.

Greenwood, Major and G. Udny Yule (1920), "An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents," *Journal of the Royal Statistical Society*, **83** (March), 255–279.

Haldane, J. B. S. (1941), "The Fitting of Binomial Distributions," *Annals of Eugenics*, **11** (January) 179–181.

Hanley, James A. and Sahir Bhatnagar (2022), "The "Poisson" Distribution: History, Reenactments, Adaptations," *The American Statistician*, **76** (November), 363-371.

Jeffreys, Harold (1939), *Theory of Probability*, Oxford: Clarendon Press.

Microsoft (2024), "Last digits are changed to zeros when you type long numbers in cells of Excel." https://learn.microsoft.com/en-us/office/troubleshoot/excel/last-digits-changed-to-zeros (accessed 2025-01-03).

Morrison, Donald G. (1969), "A Series Approximation for Negative Binomial Parameter Estimation," *Journal of Marketing Research*, **6** (August), 355–356.

Sprott, David A. (1983), "Estimating the Parameters of a Convolution by Maximum Likelihood," *Journal of the American Statistical Association*, **78** (June), 457–460.

Appendix A

It is often noted that a solution to (5) or (6) only exists if $-\bar{x}/\ln(p_0) > 1$. However, the logic of this is not widely documented. The following explanation is presented in Banerjee and Bhattacharyya (1976).

The function given in (5) is equivalent to

$$\ln(p_0) = r\,\ln\left(\frac{1}{1+\bar{x}/r}\right),$$

which is equivalent to

$$-\frac{\bar{x}}{\ln(p_0)} = \left(\frac{\bar{x}}{r}\right)\frac{1}{\ln(1+\bar{x}/r)}\,. \tag{A1}$$

Noting that for $u > 0$, $1/(1+u) < 1$, it follows that for every $y > 0$,

$$\int_0^y \frac{1}{1+u}\,du < \int_0^y 1\,du\,,$$

which implies $\ln(1+y) < y$ or $y/\ln(1+y) > 1$. Therefore the right-hand side of (A1) must be greater than 1, which means $-\bar{x}/\ln(p_0)$ must be greater than 1.

In a similar manner, it is easy to show that (6) is equivalent to

$$-\frac{\bar{x}}{\ln(p_0)} = \left(\frac{1}{\alpha}\right)\frac{1}{\ln(1+1/\alpha)}\,. \tag{A2}$$

Using the same logic as for (A1), the right-hand side of (A2) must be greater than 1, which means $-\bar{x}/\ln(p_0)$ must be greater than 1.

Note that the Poisson distribution is a limiting case of the NBD as r (and α) $\to \infty$ for a given mean $E(X)$. If X is distributed Poisson with mean λ, $E(X) = \lambda$ and $P(X = 0) = e^{-\lambda}$, i.e., $-E(X)/\ln\big(P(X = 0)\big) = 1$.

Appendix B

Given the task of finding the root of (5) or (6) using iterative methods, many analysts would make use of Newton's method (which sees us computing the first derivative of the function) or the secant method. However, the fixed-point iteration method is much easier to apply to this task.

In order to find the root of $f(x) = 0$, the fixed-point iteration method first sees us rewriting the equation as $x = g(x)$. Then, given an initial guess of x_0, we compute the sequence

$$x_{n+1} = g(x_n),\ n = 0, 1, 2, \ldots$$

hoping it converges, thereby giving us the root.

To illustrate this, consider (5). Taking the log of both sides and re-arranging terms, we have

$$r = \ln(p_0) \Big/ \ln\left\{ \left(\frac{r}{r+\bar{x}} \right) \right\}. \tag{B1}$$

Suppose the NBD is the true data-generating process, with $r = 0.161$ and $\alpha = 0.129$; these are the parameters estimates used in our numerical example in Section 2. With reference to Figure B1, this implies the true proportion of zeros is 0.705 and the mean is 1.248 (cells B4:B5). Setting r_0 to 1, we enter `=LN(B$4)/LN(D2/(D2+B$5))` in cell D3 and copy the formula down the column. To check for convergence, we enter `=D3=D2` in cell F3 and copy the formula down the column. We see that r_n converges reasonably quickly. (Is this equal to the true value of r? We enter `=D42=0.161` in cell B42 and see that the answer is "yes".)

	A	B	C	D	E	F
1	r	0.161	n	r		r_n = r_n-1?
2	alpha	0.129	0	1.0000		
3			1	0.4311		FALSE
4	p_0	0.705	2	0.2569		FALSE
5	mean	1.248	3	0.1975		FALSE
6			4	0.1755		FALSE
7	r (est)	0.1610	5	0.1668		FALSE
8	alpha (est)	0.1290	6	0.1634		FALSE
9			7	0.1620		FALSE
10	Absolute error		8	0.1614		FALSE
11	r	5.55E-17	9	0.1612		FALSE
12	alpha	5.55E-17	10	0.1611		FALSE
13	p_0	1.11E-16	11	0.1610		FALSE
14	mean	0.00E+00	12	0.1610		FALSE
15			13	0.1610		FALSE
41			39	0.1610		FALSE
42		TRUE	40	0.1610		TRUE

Figure B1: Implementing the fixed-point iteration method

As the values for p_0 and $\bar{x}$ reported in cells B4:B5 are stored with 15 significant digits of precision, it is not surprising that we recover the exact value of r and therefore α. Figure B2 reports the solution when we set p_0 to 0.705 and $\bar{x}$ to 1.248 (i.e., round the true values to three d.p.). We see that r_n converges reasonably quickly. Unsurprisingly, we do not recover the original value of r.

We can, of course, find the root using off-the-shelf software. We can do so using Excel's Solver add-in in the following manner. With reference to Figure B3a, we enter the values of p_0 and $\bar{x}$ in cells B1:B2. We enter a starting value for r of 1 in cell B3 and compute the implied value of α by entering `=B4/B2` in cell B4. Next, we enter the formula for $p_0 - P(X = 0 \,|\, r, \alpha)$ in cell B7: `=B1-(B5/(B5+1))^B4`. We use Solver to find the value of r that sets this cell to a value of 0 — see Figure B4 for the associated Solver settings. The solution is given in Figure B3b.

	A	B	C	D	E	F
1	r	0.161	n	r		r_n = r_n-1?
2	alpha	0.129	0	1.0000		
3			1	0.4315		FALSE
4	p_0	0.705	2	0.2572		FALSE
5	mean	1.248	3	0.1979		FALSE
6			4	0.1758		FALSE
7	r (est)	0.161243	5	0.1671		FALSE
8	alpha (est)	0.129201	6	0.1636		FALSE
9			7	0.1622		FALSE
10	Absolute error		8	0.1616		FALSE
11	r	2.43E-04	9	0.1614		FALSE
12	alpha	2.01E-04	10	0.1613		FALSE
13	p_0	1.11E-16	11	0.1613		FALSE
14	mean	0.00E+00	12	0.1613		FALSE
15			13	0.1612		FALSE
41			39	0.1612		FALSE
42			40	0.1612		TRUE

Figure B2: Fixed-point iteration solution with rounded p_0 and $\bar{x}$

	A	B
1	p_0	0.705
2	mean	1.248
3		
4	r (est)	1.000000
5	alpha (est)	0.801282
6		
7	p_0 - P(X=0)	0.26016

(a)

	A	B
1	p_0	0.705
2	mean	1.248
3		
4	r (est)	0.161243
5	alpha (est)	0.129201
6		
7	p_0 - P(X=0)	6.57E-08

(b)

Figure B3: Root finding using Solver

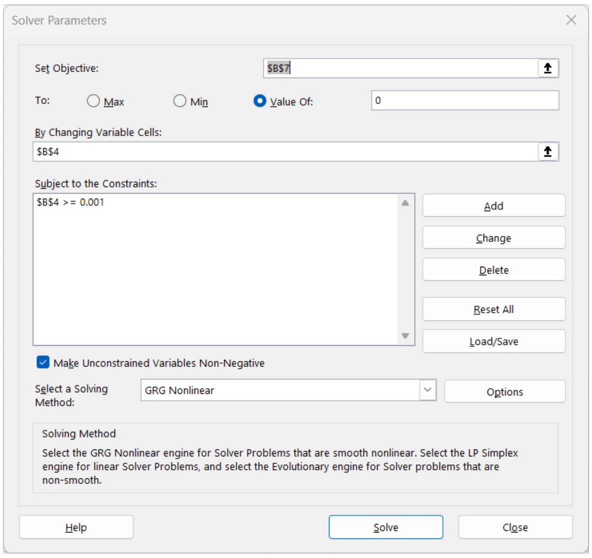

Figure B4: Solver settings

We note two things. First, our estimates of r and α are the same (at least to six d.p.) as those obtained using the fixed-point iteration method (as reported in Figure B2). Second, the value of $p_0 - P(X = 0)$ is not exactly zero.[a] This is due to the default convergence criteria used by Solver.

We could also formulate this as an optimization problem, setting our objective function as $(p_0 - P(X = 0))^2$ and instructing Solver to find the value of r that minimizes this function.[b]

[a]This means our estimates of r and α are not exactly the same as those obtained using the fixed-point iteration method. However, being the same to six d.p. is good enough in this context.

[b]It is important to check that the objective-function value is 0 at the minimum. Otherwise the solution is not a root.