Linear regression for Poisson count data: a new semi-analytical method with applications to COVID-19 events

Massimiliano Bonamente*

Department of Physics and Astronomy, University of Alabama in Huntsville, Huntsville, AL, United States

This study presents the application of a new semi-analytical method of linear regression for Poisson count data to COVID-19 events. The regression is based on the maximum-likelihood solution for the best-fit parameters presented in an earlier publication, and this study introduces a simple analytical solution for the covariance matrix that completes the problem of linear regression with Poisson data for one independent variable. The analytical nature of both parameter estimates and their covariance matrix is made possible by a convenient factorization of the linear model proposed by J. Scargle. The method makes use of the asymptotic properties of the Fisher information matrix, whose inverse provides the covariance matrix. The combination of simple analytical methods to obtain both the maximum-likelihood estimates of the parameters and their covariance matrix constitutes a new and convenient method for the linear regression of Poisson-distributed count data, which are of common occurrence across a variety of fields. A comparison between this maximum-likelihood linear regression method for Poisson data and two alternative methods often used for the regression of count data—the ordinary least−square regression and the χ^2 regression—is provided with the application of these methods to the analysis of recent COVID-19 count data. The study also discusses the relative advantages and disadvantages among these methods for the linear regression of Poisson count data.

KEYWORDS

probability, statistics, maximum-likelihood, ordinary least-squares, Poisson distribution, Cash statistic, parameter estimation

1. Introduction

Integer-count data based on the Poisson distribution are quite common in real-life applications, from Poisson's use of his name-sake distribution for the study of civil and criminal cases [1], Lord Rutherford's characterization of the rate of decay of radioactive isotopes [2] and Clarke's application to the number of bombs that fell in London during World War II [3] and to the distribution of low-count political events [4], such as the number of seats lost by the president's party in a mid-term election [5] or the study of mutations in bacteria [6]. The common occurrence of the Poisson distribution lies primarily in its association with the Poisson process, whereby this distribution describes the probability of occurrence of events—e.g., in a period of time or for a spatial region—from a process with a fixed rate of occurrence (e.g., [7]). Among the data analysis methods for Poisson-based count data, linear regression plays a fundamental role, given that the linear model is arguably the simplest model that enforces a correlation between the independent variables (or regressors) and the dependent measured variables.

The maximum-likelihood estimate of a two-parameter linear model for integer-valued count data can be accomplished with the use of the Poisson log-likelihood, also known as the *Cash* or *C* statistic [8]. Despite its common occurrence, it has been elusive to find a simple analytical expression for the maximum-likelihood estimates of the parameters for linear Poisson regression (see, e.g., Section 2.2 of [9]), even in the case of one independent variable. This restriction has resulted in a limited use of the maximum-likelihood method for the regression of count data. This is in contrast to the simple analytical solutions that are available for the ordinary least-square (OLS) regression for data with equal variances (i.e., homoskedastic data) or with different variances (i.e., heteroskedastic data), and for the regression with the χ^2 statistic, which is a maximum-likelihood method for heteroskedastic data of common use for normally-distributed data (e.g., [10]). Such simple analytical solutions, which also extend to the more general multiple linear regression with more than two adjustable parameters (e.g., see [11] for a review), are now a standard of practice for linear regression. Their use is in fact so widespread that it is not uncommon to see χ^2-based methods of regression even for Poisson count data for which the method is not justified, or rather it is known to lead to known biases that have been documented as early as [12] and more recently by [13] and [14].

Motivated by the need to provide a simple maximum-likelihood method for the linear regression of count data (sometimes also referred to as *cross-section* data) that explicitly accounts for the Poisson distribution of the measurements, in [15] we have reported the first semi-analytical solution for the parameters of linear regression with the Poisson-based *C* statistic. In this study, we continue the investigation of the same problem and report a new analytical solution for the covariance matrix of the two model parameters, based on the Fisher information matrix [16]. The combination of the new results presented in this study and those provided in [15] thus provide a complete treatment of the maximum-likelihood linear regression with Poisson-distributed count data. The availability of these new analytical methods for the linear regression of count data, therefore, represents a key step toward a wider use of an unbiased method of regression for such integer-count data, without the need to resort to either numerical solutions or the use of alternative and less accurate methods, such as those based on least-squares or the χ^2 statistics.

There are alternatives to the modeling of count data with the simple one-parameter Poisson distribution considered in this study. One of the key limitations of the Poisson regression is the inability to properly model *overdispersed* data (see, e.g., discussion in Chapter 3 of [9]). For such overdispersed data, other data models such as the negative binomial distribution can be used [17]. Alternatively, the standard Poisson maximum-likelihood regression can be modified to a *quasi*-MLE regression with empirical variance functions that reflect the overdispersion (e.g., [18]), that is also used in the generalized linear model (GLM) literature (e.g., [19]). Nonetheless, the simple linear regression of Poisson count data is an essential tool for the data analyst, and the development of simple analytical tools for its application is the subject of this study.

The article is structured as follows: Section 2 summarizes the model for the regression with count data, which is described in

full in [15]; Section 3 describes methods to obtain the covariance matrix, with Section 3.3 describing the properties of the matrix for the Poisson linear regression under consideration and Section 3.4 presents an alternative method to estimate the covariance matrix based on the error propagation method. Section 4 provides an application of the results to recent COVID-19 data with comparison to the least-squares and χ^2 methods. Finally, Section 5 contains a discussion and the conclusions.

2. Model for the regression of count data with the *C* statistic

The data under consideration are in the form of independent pairs of measurements (x_i, y_i), where y_i are Poisson-distributed variables, for $i = 1, \ldots, N$, and x_i are fixed values of the independent variable.

2.1. Parameterization of the linear model

A convenient parameterization of the linear model was proposed by [20], and it is in the form

$$f_S(x) = \lambda(1 + a(x - x_A)), \tag{1}$$

where x_A is a constant that usually coincides with the beginning of the range of the independent variable x, and λ and a are the two parameters. This parameterization is thus equivalent to a linear model with an overall slope λa, and intercept $\lambda(1 - a x_A)$, yet this factorization has algebraic andvantages when used in the Poisson log-likelihood. The parent mean of the Poisson distribution of the variable y_i is then given as

$$\mu_i = f_S(x_i) \, \Delta x_i, \tag{2}$$

where Δx_i is the width of the i-th bin of data, which is not required to be uniform among the measurements.

2.2. Maximum-likelihood regression

Upon minimization of the logarithm of the Poisson likelihood of the data with model, in [15] we have shown that the method of maximum likelihood yields two simple equations for the determination of the parameter estimates, which are summarized in the following. The first is a analytical relationship between the two parameters:

$$\lambda(a) = \frac{M}{R\left(1 + a\dfrac{R}{2}\right)}, \tag{3}$$

where M is the sum of all integer counts y_i, and R is the range of the independent variable, typically chosen to be $R = x_B - x_A$. This equations enables the determination of $\hat{\lambda}$ from $\hat{a}$. The second is an

equation that must be solved to obtain $\hat{a}$, and it is conveniently cast as the zero of a function $F(a)$:

$$F(a) = 1 + \frac{R}{2}\left(a - \frac{M}{g(a)}\right) = 0, \ \text{with} \ g(a) = \sum_{i=1}^{N} y_i \frac{(x_i - x_A)}{1 + a(x_i - x_A)}.$$
$$(4)$$

In general, solution of (4) to obtain the estimate $\hat{a}$ can be obtained numerically. The inherently analytical method leading to (3) and (4) and the need for a numerical solution of the latter constitute this new *semi-analytical* method for the linear regression of Poisson count data. The properties of the two functions $F(a)$ and $g(a)$ are described in detail in [15], and the key properties are reported in the following.

2.3. Identification of points of singularity and zeros of *g(a)*

It is immediate to identify the n points of singularity of $g(a)$ and its zeros, using the property that $g'(a) < 0$ between points of singularity. Since it is also true that $F'(a) < 0$ between the function's points of singularity, which are the zeros of $g(a)$, and the $n - 1$ zeros of $F(a)$ can be easily found because of the continuity of the function between singularities. In particular, the *acceptable* solution—defined as the pair of best-fit parameters $(\hat{a}, \hat{\lambda})$ that gives a non-negative model throughout its support, as required by the counting nature of the Poisson distribution—is either the first or the last zero according to the asymptotic value of the function $F(a)$ for $a \rightarrow \pm\infty$. The acceptable solution $\hat{a}$ can therefore be found with elementary numerical methods from the equation $F(a) = 0$.

2.4. Applicability of the model to data with gaps

Both Equations (3) and (4) apply to data with arbitrary bin size, i.e., the bin size Δx_i in (2) need not be uniform. Moreover, when there are gaps in the data, defined as intervals of the x variable without measurements, the two equations are modified with the introduction of a modified range R_m defined as

$$R_m = \frac{R^2 - 2S_G}{R - R_g}$$
$$(5)$$

where R_G is the sum of the ranges $R_{G,j}$ of all the gaps, and S_G is a constant defined by

$$S_G = \sum_{j=1}^{g} R_{G,j}(x_{G,j} - x_A),$$

where $x_{G,j}$ is the midpoint of the j-th gap. The equations are modified to[1]

$$\begin{cases} F(a) = 1 + \dfrac{aR_m}{2} - \dfrac{MR_m}{2g(a)} = 0 \\[4mm] \lambda(a) = \dfrac{M}{R\left(1 + a\dfrac{R}{2}\right) - (R_G + aS_G)}. \end{cases}$$
$$(6)$$

Such modifications for the presence of gaps in the data do not introduce any additional statistical complication to the analysis of either the maximum-likelihood estimates or the covariance matrix. Gaps in the data are quite common in data analysis, for example, when a range of wavelengths in the spectrum of an astronomical source is unavailable (e.g., because of detector inefficiencies, as in [21]) or when a time interval for the light-curve of a source is unobserved (e.g., [22]).

The reason for the parameterization of the linear model according to (1) is that the two resulting Equations (3) and (4) can be used to identify the maximum-likelihood parameter estimates more easily than for the standard parameterization $a + b\,x$, which leads to two coupled non-linear equations instead (as noted, e.g., in Chapter 16 of [11] or in [15]). This is the main reason for the [20] parameterization used in this study. Another advantage of the present method is that a solution of Equations (3) and (4) is always possible, whereas the same is not true for the simpler parameterization (as shown in [11]). Moreover, as will be shown in the next section, these equations lead to a very convenient analytical solution for the covariance matrix, which represents the main new result of this study.

A more general multiple linear regression that includes several independent variables x_j as predictors for a Poisson-distributed variable y can also be treated in a similar way as the present case of a simple linear regression. The results of the multiple linear regression method for Poisson data will be presented in a separate article.

3. Evaluation of the covariance matrix

3.1. General properties of the covariance matrix

The maximum-likelihood estimators of the linear model parameters $\hat{\theta} = (\hat{\lambda}, \hat{a})$ from Poisson count data are normally distributed but only in the asymptotic limit of a large number of measurements N (see, e.g., [9, 23]). In this asymptotic limit, the maximum-likelihood estimators are also unbiased and efficient, with their variance reaching the *Cramér-Rao* lower bound (see, e.g., [24–26]). Within this limit, it is thus possible to approximate the covariance or error matrix of the parameters, defined by

$$\varepsilon = \text{E}[(\hat{\theta} - \theta_0)(\hat{\theta} - \theta_0)^T],$$

---

1 This modification is justified in Lemma 6.7 of [15].

as the inverse of the *Fisher information matrix*,

$$\hat{\varepsilon} = -\left(E\left[\frac{\partial^2 \ln \mathscr{L}}{\partial \theta \, \partial \theta'}\right]\right)^{-1}_{\theta=\theta_0} = -\left(E\left[\sum_{i=1}^{N}\frac{\partial^2 \ln f(x_i)}{\partial \theta \, \partial \theta'}\right]\right)^{-1}_{\theta=\theta_0}. \tag{7}$$

In this application, $f(x_i)$ is the Poisson distribution for the measurement at a value x_i of the indepedendent variable, with a μ_i parameter given by (2), and the expectation in (7) is taken with respect to this distribution.[2] In practice, Equation (7) means that the maximum-likelihood estimates of the parameters of the linear model converge in distribution to a normal distribution with the covariance matrix given by (7) or

$$(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \hat{\varepsilon}).$$

Fisher was the first to associate the concept of *information* with the second derivative—or curvature—of the logarithm of the likelihood [16], and similar definitions of information have been derived from this idea (e.g., [27–29]). Given that the true parameter values θ_0 are unknown, the information matrix is evaluated at $\theta = \hat{\theta}$ in recognition of the fact that the maximum-likelihood estimates are unbiased and thus asymptotically converge to the true-yet-unknown values (e.g., [16], Chapter IX of [30], or Chapter 33 of [25]).

3.2. Evaluation of the covariance matrix for the Poisson linear regression with the Scargle parameterization

The negative of the logarithm of the likelihood, which is related to the C statistic by $C = -2\ln\mathscr{L} + \text{const}$, can be written as

$$-\ln\mathscr{L} = \lambda(R-R_G) + \lambda a S_1 - M\ln\lambda - \sum_{i=1}^{N} y_i \ln(1+a(x_i-x_A)), \tag{8}$$

where the constant S_1 is defined by

$$S_1 = \sum_{i=1}^{N}(x_i-x_A)\Delta x_i = \frac{R^2 - 2S_G}{2}, \tag{9}$$

and the second equation connects S_1 with the constants defined in Section 2, and it is an immediate consequence of the model linearity. The first derivatives are

$$\begin{cases} -\dfrac{\partial \ln\mathscr{L}}{\partial a} = \lambda S_1 - g(a) \\[2mm] -\dfrac{\partial \ln\mathscr{L}}{\partial \lambda} = (R-R_G) + a S_1 - \dfrac{M}{\lambda} \end{cases} \tag{10}$$

in which, when set to zero, they yield the usual maximum-likelihood estimates described in Section 2 and in more detail in

---

2 The notation $\partial^2 \ln\mathscr{L}/\partial\theta\,\partial\theta'$ indicates that diagonal elements are second derivatives with respect to the same parameter, and off-diagonal elements have cross-derivatives with respect to two different parameters. The symbol ^{T} indicates the transpose.

[15]. From these, it is immediate to evaluate the second-order derivatives and thus obtain

$$-\frac{\partial^2 \ln\mathscr{L}}{\partial\theta\,\partial\theta'} = \begin{bmatrix} \dfrac{M}{\lambda^2} & S_1 \\[3mm] S_1 & G(a) \end{bmatrix} \tag{11}$$

with a new function G defined as

$$G(a) = -\frac{\partial g(a)}{\partial a} = \sum_{i=1}^{N}\frac{y_i(x_i-x_A)^2}{(1+a(x_i-x_A))^2}. \tag{12}$$

The expectation of (11) is immediately calculated with the consideration that

$$E[y_i]\big|_{\hat{\theta}} = \mu_i\big|_{\hat{\theta}} = \hat{\lambda}(1+\hat{a}(x_i-x_A))\Delta x_i,$$

where Δx_i is the size of the i-th bin of data. Given that all terms in (11) are linear in the measurements y_i, the expectation is found as

$$-E\left[\frac{\partial^2 \ln\mathscr{L}}{\partial\theta\,\partial\theta'}\right]\bigg|_{\hat{\theta}} = \begin{bmatrix} \dfrac{(R-R_G)+\hat{a}S_1}{\hat{\lambda}} & S_1 \\[3mm] S_1 & \hat{\lambda}\,H(\hat{a}) \end{bmatrix}, \tag{13}$$

with

$$H(a) = \sum_{i=1}^{N}\frac{(x_i-x_A)^2}{1+a(x_i-x_A)}\Delta x_i \tag{14}$$

defined as a convenient analytical function to describe one of the terms of the information matrix. Notice how the bin size appears explicitly in this function, unlike in the case of the function $g(a)$ in (4), which is independent of the bin size.

With the simple analytical expression available for the second-order derivatives of the logarithm of the likelihood (13), it is finally possible to provide the asymptotic estimate of the covariance matrix as its inverse as

$$\hat{\varepsilon} = \begin{bmatrix} \hat{\sigma}_a^2 & \hat{\sigma}_{a\lambda}^2 \\[2mm] \hat{\sigma}_{a\lambda}^2 & \hat{\sigma}_\lambda^2 \end{bmatrix} = \frac{1}{\Delta}\begin{bmatrix} \hat{\lambda}\,H(\hat{a}) & -S_1 \\[3mm] -S_1 & \dfrac{(R-R_G)+\hat{a}S_1}{\hat{\lambda}} \end{bmatrix}, \tag{15}$$

with Δ as the determinant of the matrix (13),

$$\Delta = \big((R-R_G)+\hat{a}S_1\big)H(\hat{a}) - S_1^2.$$

This simple analytical form for the covariance matrix of the maximum-likelihood estimates for linear regression to Poisson data is the main result of this study. Its properties are analyzed in the following section.

3.3. Properties of the covariance matrix

Basic results on the sign of the terms in the covariance matrix (15) are provided in this section. First, it is necessary to recall the definition of *acceptability* of the estimates of the model parameters as values that ensure a non-negative parent Poisson mean, which

was briefly introduced in Section 2. Acceptable solutions for $\hat{a}$ and $\hat{\lambda}$ require that the quantities

$$\hat{\lambda}(1 + \hat{a}(x_i - x_A))$$

are all non-negative. This results in values of the two parameters to be found in two possible ranges, namely, (a) for $\hat{a} \leq -2/\Delta x_1$, where $\hat{\lambda} < 0$ or (b) for $\hat{a} > -1/(R - \Delta x_N/2)$, where $\hat{\lambda} > 0$ (see Lemma 5.2 of [15] for a proof). In the following, it is assumed that x_A is the beginning of the range of the independent variable x.

Lemma 3.1 (Sign of the determinant Δ). *The determinant Δ in (15) is a positive definite for all acceptable solutions of the regression coefficients.*

Proof: The determinant can be written as

$$\Delta = \sum_{i=1}^N \frac{(x_i - x_A)^2}{1 + \hat{a}(x_i - x_A)} \Delta x_i \cdot \sum_{i=1}^N (1 + \hat{a}(x_i - x_A))\Delta x_i$$
$$- \left(\sum_{i=1}^N (x_i - x_A)\Delta x_i \right)^2 .$$

This is in the form

$$\Delta = \sum_{i=1}^N \frac{r_i^2}{s_i} \cdot \sum_{i=1}^N s_i - \left(\sum_{i=1}^N r_i \right)^2$$

with $r_i = (x_i - x_A)\Delta x_i$ and $s_i = (1 + \hat{a}(x_i - x_A))\Delta x_i$. Notice that the terms s_i are either all positive or all negative, therefore making it such that one can use $|s_i|$ in their place in the previous equation,

$$\Delta = \sum_{i=1}^N \frac{r_i^2}{|s_i|} \cdot \sum_{i=1}^N |s_i| - \left(\sum_{i=1}^N r_i \right)^2 .$$

With the substitution $a_i = r_i/\sqrt{|s_i|}$ and $b_i = \sqrt{|s_i|}$, the determinant becomes

$$\Delta = \sum_{i=1}^N a_i^2 \cdot \sum_{i=1}^N b_i^2 - \left(\sum_{i=1}^N a_i b_i \right)^2 \geq 0,$$

according to the Cauchy-Schwartz inequality. □

It is also possible to show that the variances of $\hat{a}$ and $\hat{\lambda}$ have the proper sign, and that the covariance between them is negative.

Lemma 3.2 (Signs of variance and covariance terms). *The diagonal terms in the covariance matrix (15) are positive, and the off-diagonal terms are negative, for all acceptable solutions of the linear regression coefficients.*

Proof: The term S_1 is positive according to its definition (9), and therefore, the covariance is always negative.

According to (14), the variance of the estimate $\hat{a}$ is

$$\hat{\sigma}_a^2 = \frac{\hat{\lambda}H(\hat{a})}{\Delta} = \frac{\hat{\lambda}}{\Delta} \sum_{i=1}^N \frac{(x_i - x_A)^2}{1 + a(x_i - x_A)} \Delta x_i.$$

Given the discussion at the beginning of this section, $\hat{\lambda}$ and each of the terms $1 + \hat{a}(x_i - x_A)$ must have the same sign for acceptable

values of the regression coefficients. The variance is, therefore, always positive, as it should.

For the same reasons, the variance of $\hat{\lambda}$ is immediately shown to have the proper sign:

$$\hat{\sigma}_\lambda^2 = \frac{1}{\Delta\hat{\lambda}} \left((R - R_G) + \hat{a}S_1 \right) = \frac{1}{\Delta} \frac{\sum_{i=1}^N (1 + \hat{a}(x_i - x_A))\Delta x_i}{\hat{\lambda}} \geq 0.$$

The alternative approach to achieve a non-negative Poisson mean for all values of the model parameters would be to perform an exponential transform so that the expectations for the Poisson-distributed mesurements is, for example, of the type $\mu_i = \exp(a + b\,x_i)$ (see, e.g., Section 2.2 of [9]). In practice, this corresponds to a linear model for the logarithm of the means, $\log \mu_i = a + b\,x_i$. Although this transformation does indeed provide a Poisson mean μ_i that is always non-negative, the model is no longer *linear* between x and y, and therefore unsuitable for those cases where a linear relationship is required, e.g., because of the physical nature of the two variables. Moreover, even with this exponential transformation, the maximum-likelihood method of regression for Poisson data does not have an analytical solution with the usual parameterization of the linear model [9].

3.4. An alternative approximation for the covariance matrix

A different analytical approximation for the covariance matrix of the coefficients of linear regression can be given using the *error propagation* method, also known as the *delta* method (e.g., [9, 31]). The method consists of treating the parameters as a function of the independent measurements, $\theta_j = \theta_j(y_k)$, and approximating the variances and covariances with the first-order terms of their Taylor series:

$$\hat{\sigma}_{\theta_i \theta_j}^2 \simeq \sum_{k=1}^N \left(\frac{\partial \theta_i}{\partial y_k} \right)\Big|_{\hat{\theta}} \left(\frac{\partial \theta_j}{\partial y_k} \right)\Big|_{\hat{\theta}} \sigma_k^2, \qquad (16)$$

where σ_k^2 is an estimate of the variance of the measurement y_k. The regression model described in Section 2 does not lead to best-fit parameters $\hat{\theta}$ that are described as a function of the measurements in a closed form. Nonetheless, it is possible to take the derivatives of both (3) and (4), or more generally (6), with respect to y_j to obtain the derivatives $\partial \theta_i/\partial y_k$ as follows.

Start with

$$\lambda R \left(1 + a\frac{R}{2} \right) - \lambda(R_G + aS_G) = M,$$

which is the most general equation to obtain $\hat{a}$ in the presence of gaps in the data and of non-uniform binning. The equation can be brought in a simpler form as

$$\lambda(R - R_G) + a\lambda \left(\frac{R^2}{2} - S_G \right) = \sum_{i=1}^N y_i,$$

leading to

$$\frac{\partial \lambda}{\partial y_j}\left(R - R_G + a\left(\frac{R^2}{2} - S_G\right)\right) + \frac{\partial a}{\partial y_j}\lambda\left(\frac{R^2}{2} - S_G\right) = 1,$$

and therefore

$$\frac{\partial \lambda}{\partial y_j} = \frac{1 - \lambda\left(\frac{R^2}{2} - S_G\right)\frac{\partial a}{\partial y_j}}{R - R_G + a\left(\frac{R^2}{2} - S_G\right)}. \qquad (17)$$

Alternatively, this partial derivative can be expressed in terms of a alone as

$$\frac{\partial \lambda}{\partial y_j} = \frac{R - R_G + \left(\frac{R^2}{2} - S_G\right)\left(a - M\frac{\partial a}{\partial y_j}\right)}{\left(R - R_G + a\left(\frac{R^2}{2} - S_G\right)\right)^2}, \qquad (18)$$

where the correction for the presence of gaps in the data is in the constants R_G and S_G, both of which become null when the data spans a continous range between x_A and x_B, with $R = x_B - x_A$.

A similar procedure yields the partial derivatives of a from

$$\left(a + \frac{2}{R_m}\right)g(a) = M,$$

which is the general expression corresponding to the equation $F(a) = 0$ needed to find $\hat{a}$. The equation can be written in a more convenient way as

$$\sum_{i=1}^{N} \frac{y_i\left(2/R_m(x_i - x_A) - 1\right)}{1 + a(x_i - x_A)} =, 0$$

and from this, taking the derivative with respect to y_j,

$$\frac{\partial a}{\partial y_j} = \frac{1 - 2/R_m(x_j - x_A)}{1 + a(x_j - x_A)} \cdot \frac{1}{g_2(a) - 2/R_m G(a)}, \qquad (19)$$

with the aid of a new function defined by

$$g_2(a) = \sum_{i=1}^{N} \frac{y_i(x_i - x_A)}{(1 + a(x_i - x_A))^2}. \qquad (20)$$

The result is proven by noting that, in taking the total differential of the term in the sum for $i = j$, there is an extra term due to the presence of the random variable y_j.

The partial derivatives in (18) and (19) can now be used in (16) to provide an estimate of the covariance matrix from the error propagation method. The variance σ_k^2 of the measurement y_k can be estimated as either the measurement y_k itself or as the best-fit model $\hat{\mu}_k = \mathrm{E}[y_k/\hat{\theta}] = \hat{\lambda}(1 + \hat{a}(x_k - x_A))\Delta x_k$. In the application provided below in Section 4, the parent expectation $\hat{\mu}_k$ is used for the numerical result.

TABLE 1 Daily number of events (deaths) reported at the beginning of the pandemic.

Day	Number of events	Cumulative
0	0	0
1	1	1
2	2	3
3	3	6
4	4	10
5	2	12
6	0	12
7	3	15
8	4	19
9	3	22
10	4	26
11	5	31
12	6	37
13	6	43
14	7	50
15	10	60
16	8	68
17	23	91
18	26	117
19	45	162

Data are from [32]. Day 0 is 28 February 2020, the day prior to the first reported death, which occurred on 29 February 2020.

4. Application to COVID-19 data

This section provides an application of the methods described in this study for the maximum-likelihood linear regression of Poisson data with the [20] parameterization and a comparison with other methods of regression commonly used for count data.

The daily number of deaths caused by the COVID-19 pandemic in the United States provides an example of count data that can be analyzed with the linear regression model presented in this study. The data are obtained from the *New York Times* [32], and they are reported in Table 1.[3] These data follow the data model of Section 2, whereby the number of events are collected in a time interval with a uniform size of length $\Delta x = 1$ day, and the measurements in the column "number of events" are independent of one another.

This section presents two sets of linear regressions to these data with the C statistic to illustrate the novel method of analysis described in this study and in [15]. It also presents the comparison with two popular regression methods, the ordinary least squares regression and the fit based on the χ^2 statistic. The comparison among these methods is used to illustrate the advantages of the new method of linear regression for Poisson data and to discuss assumptions and limitations of the different methods.

---

3   The *New York Times* archive reports the cumulative number of deaths, Table 1 also reports the daily counts.
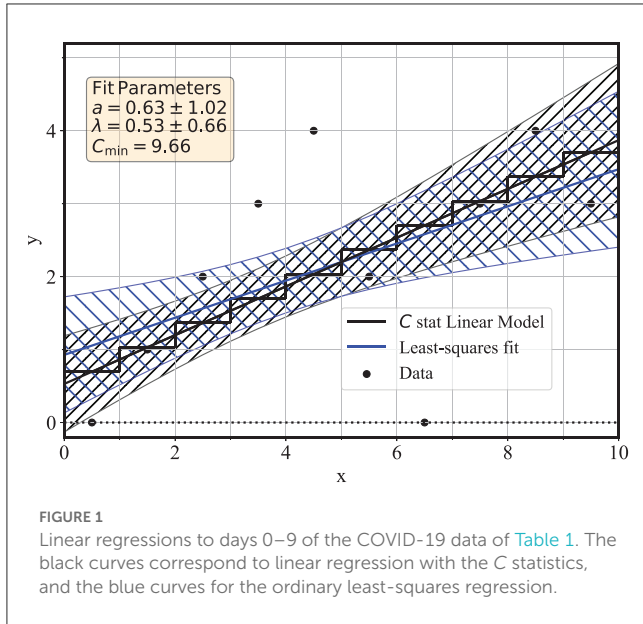
FIGURE 1
Linear regressions to days 0−9 of the COVID-19 data of Table 1. The black curves correspond to linear regression with the C statistics, and the blue curves for the ordinary least-squares regression.

4.1. Fit to days 0−9 with uniform binning

For this application, the data for days 0 (day prior to first case recorded) through 9 from Table 1 are fit to the linear model, with a total of $N = 10$ data points.

4.1.1. The maximum-likelihood parameter estimates

The best-fit model according to (1) is shown in Figure 1 as the black line, with the step-wise continuous black line indicating the best-fit model $\hat{\mu}_i$. The two lines intersect because the data are collected over bins with unit size ($\Delta x = 1$).

The method of maximum-likelihood regression (see Section 2) makes use of two functions $g(a)$ and $F(a)$ that are reported in Figure 2. In this application, there is a total of $M = 22$ counts for $N = 10$ bins and $n = 8$ unique non-zero bins. According to (4), this results in $n = 8$ points of singularity for $g(a)$ marked as green crosses in Figure 2, and $n - 1 = 7$ roots or zeros of $g(a)$ marked as blue dots between consecutive singularities, with $g'(a) < 0$ in each interval. The roots of $g(a)$ are also points of singularity for $F(a)$, whose zero(s) are the estimates of the parameter a. Of the $n - 2 = 6$ solutions, only the last one is *acceptable*, i.e., it is the only value $\hat{a}$ that gives a non-negative model throughout its support, as required for Poisson counts (see Section 5 of [15]). Having identified the last singularity of $F(a)$ as the last zero of $g(a)$, the acceptable solution $\hat{a}$ is marked as the rightmost green dot in Figure 2.

It is clear that it is not necessary to identify and calculate all the roots of $g(a)$ in order to determine $\hat{a}$ but just the last one between the last two singularities of $g(a)$. The computational burden is therefore very limited, even for data with a large number of events. Figure 2 report all points of singularities and the zeros for both $g(a)$ and $F(a)$ only for illustration purposes.
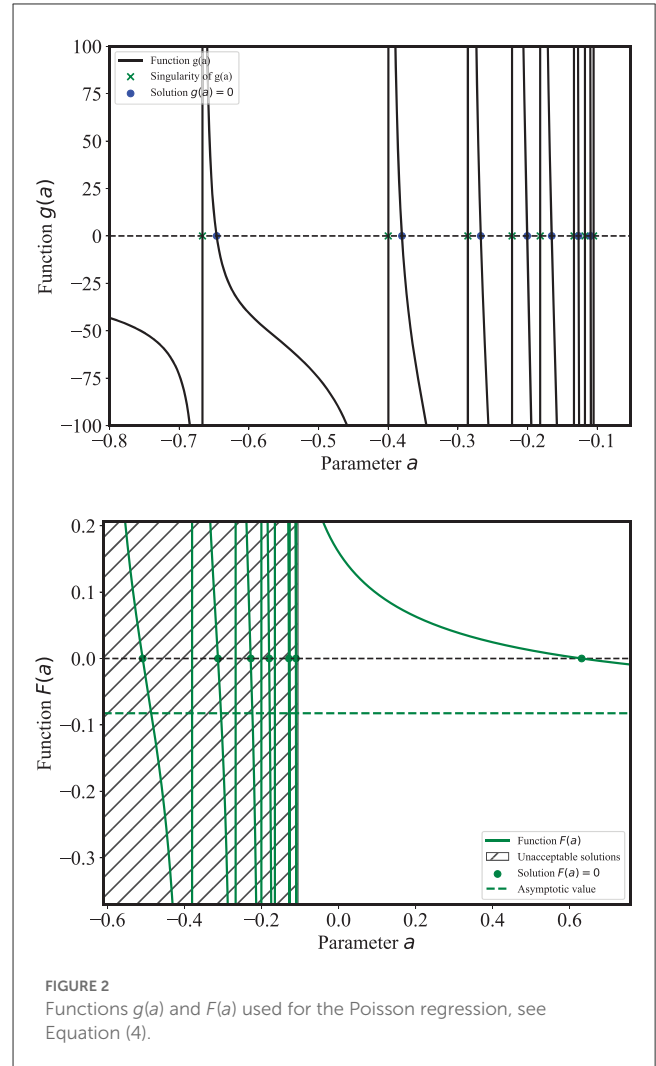

FIGURE 2
Functions $g(a)$ and $F(a)$ used for the Poisson regression, see Equation (4).

4.1.2. The covariance matrix

The error matrix evaluated according to (15) is

$$\hat{\varepsilon} = \begin{bmatrix} 0.346 & -0.665 \\ -0.665 & 1.044 \end{bmatrix}$$

with the estimated standard deviations of the parameters also reported in the figure for estimated parameters $\hat{a} = 0.63 \pm 1.02$ and $\hat{\lambda} = 0.53 \pm 0.66$.

4.1.3. The log-likelihood or C statistic

Moreover, the fit statistic is evaluated according to

$$C_{\min} = \sum_{i=1}^{N} (\hat{\mu}_i - y_i + y_i \ln (y_i/\hat{\mu}_i)), \qquad (21)$$

as reported in Equation 3 of [15]. Hypothesis testing with the C statistics requires the evaluation of critical values of its distribution, which is known only in the asymptotic limits of a large number of measurements N and for large parent means when $C_{\min}$ is approximately distributed as $\chi^2(N - m)$ (see, e.g., [11, 33]), with

$m = 2$ adjustable parameters. Although a detailed discussion of the method of hypothesis testing with the C statistics goes beyond the scope of this study, it is nonetheless useful to provide a brief description of its application for this example since hypothesis testing is an integral part of the overall regression. In the present example of small Poisson means, the approximation of $C_{\min}$ with a $\chi^2(N - m)$ distribution is not accurate, and in [14], we proposed an approximate method to estimate critical values that applies also to the low-mean case. The method consists of using the parent mean and variance of each term in (21) in place of the mean and variance of a $\chi^2(1)$ distribution (which are, respectively, 1 and 2) and use the central limit theorem to ensure an approximately normal distribution for the statistic. Following this method, the critical value of the $C_{\min}$ distribution for $N - m = 8$ degrees of freedom is 13.6 [compared to a corresponding critical value of 13.4 for the $\chi^2(8)$ distribution], and therefore, these data are consistent with the null hypothesis of a linear model at the 90% confidence level. Notice how both the parameter estimation and the goodness-of-fit statistic naturally account for measurements with zero counts, such as bins 0 and 6.

4.1.4. The confidence band

The gray-hatched band associated with the best-fit model highlights another strength of the availability of the error matrix for the C statistic regression, which makes it is possible to estimate the variance of any function of the model parameters. In particular, it is often useful to estimate the variance of the *expectation* of the $f_S(x)$ function for a fixed value of the independent variable x, i.e.,

$$\hat{f}_S(x) = \mathrm{E}[f_S(x)/x,\hat{\theta}] = \hat{\lambda}(1 + \hat{a}(1 - x)),$$

or, equivalently, for the expectation of the number of counts in a bin centered at x and with arbitrary bin size,

$$\hat{y}_x = \mathrm{E}[y/x,\hat{\theta}] = \hat{\lambda}(1 + \hat{a}(1 - x))\,\Delta x,$$

where x is any value of the regressor variable, not limited to the values x_i that correspond to the measurements. The variance of the variable $\hat{y}_x$ is represented by the variance of y values along a vertical line for a given x. The range $\hat{y}_x \pm \sigma_{y_x}$ is usually referred to as the *confidence band* of the best-fit model and is represented by the gray hatched region in Figure 1. The variance $\sigma_{y_x}^2$ is estimated by the error-propagation method, with

$$\hat{\sigma}_{y_x}^2 = \hat{\sigma}_\lambda^2 \left(\frac{\partial y_x}{\partial \lambda}\right)^2\bigg|_{\hat{\theta}} + \hat{\sigma}_a^2 \left(\frac{\partial y_x}{\partial a}\right)^2\bigg|_{\hat{\theta}} + 2\hat{\sigma}_{\lambda a}^2 \left(\frac{\partial y_x}{\partial \lambda}\right)\left(\frac{\partial y_x}{\partial a}\right)\bigg|_{\hat{\theta}}.$$

The standard error σ_{y_x} is therefore estimated immediately from the error matrix $\hat{\varepsilon}$ in (15) and from the analytical form for y_x. Similar conditderations can be extended to any function of the model parameters. Another useful application is to the overall slope of the model, which is measured as $\hat{a} \cdot \hat{\lambda} = 0.33 \pm 0.15$ in this example.

4.2. Comparison to OLS regression

For comparison, Figure 1 also reports the ordinary least-squares fit to the same data, with uniform weights for all the points,

with the linear model following the usual parameterization $f(x) = a + bx$. For completeness, a brief review of the OLS estimator is provided below.

4.2.1. OLS estimators

The OLS estimators are given by the usual formulas:

$$\hat{\theta}_{OLS} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} \sum_{i=1}^N x_i^2 \cdot \sum_{i=1}^N y_i - \sum_{i=1}^N x_i \cdot \sum_{i=1}^N x_i y_i \\ N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \cdot \sum_{i=1}^N y_i \end{bmatrix} \tag{22}$$

with

$$\Delta = N \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i\right)^2.$$

The OLS regression is known to provide an unbiased estimate of the parameters, regardless of the parent distribution of the measurements, due to the linearity of the model (see, e.g., [9, 23]). As a result, Figure 1 shows that the C statistics and the OLS linear regressions provide qualitatively similar best-fit values.

4.2.2. Estimate of the covariance matrix for OLS

Under the assumption of a normal distribution with equal variance for all the measurements, the OLS method is equivalent to the maximum-likelihood method, and the value for the common variance is required to estimate the covariance matrix (e.g., see Chapter 11 of [11]). For measurements that are not normally distributed, such as the data model presented in this study and for this specific data example, it is still possible to provide an estimate of the parameter variances for the OLS estimators, *if* the OLS estimator are interpreted as a function of the measurements y_i, without requiring homoskedasticity or normality. In this case, the linear OLS is no longer a maximum-likelihood estimator, but it retains the convenient property of unbiasedness.

Using the assumption that $y_i \sim \mathrm{Poisson}(\theta_0 + x_i\theta_1)$ and that the measurements are independent, it follows that $\mathrm{Var}(y_i) = \theta_0 + x_i\theta_1$, and therefore

$$\mathrm{Var}(\hat{\theta}_{OLS}) \simeq \frac{1}{\Delta^2} \begin{bmatrix} \sum_{i=1}^N \left(\sum_{i=1}^N x_i^2 - x_i \sum_{i=1}^N x_i\right)^2 \cdot (\theta_0 + x_i\theta_1) \\ \sum_{i=1}^N \left(Nx_i - \sum_{i=1}^N x_i\right)^2 \cdot (\theta_0 + x_i\theta_1) \end{bmatrix} \tag{23}$$

Naturally, since the true parameters are unknown, one can only use these variances by substituting $\hat{\theta}$ for θ in the right-hand side of (23). This simple result is due to the linearity of the model, and the independence of the measurements.

It is necessary to emphasize that the approximate variances derived in (23) are an *ad-hoc* estimate, and they do differ from the standard OLS error matrix[4]

$$\varepsilon_{OLS} = \frac{\sigma^2}{\Delta} \begin{bmatrix} \sum_{i=1}^N x_i^2 & -\sum_{i=1}^N x_i \\ \cdots & N \end{bmatrix} \tag{24}$$

---

4   See, e.g., Equation (11.19) of [11] or Example 19.6 of [34].

where σ^2 is the common (and unknown) variance. In the usual case of the OLS with equal variances, one would normally proceed with the estimation of σ^2 via the sum of square residuals, divided by $N - k$,

$$\hat{\sigma}^2 = \frac{1}{N-k} \sum_{i=1}^{N} (y_i - (\hat{a} + \hat{b}\, x_i))^2 \qquad (25)$$

where $k = 2$ corresponds to the two adjustable parameters of the model. But such estimator would be biases for the case of heteroskedastic data and, therefore, not meaningful for this Poisson regression. This is the reason for the development of the *ad-hoc* approximation (23).

Moreover, an estimate for the covariance of the linear OLS estimators can be obtained by the usual error propagation method, namely,

$$\hat{\sigma}_{ab}^2 = \sum_{j=1}^{N} \frac{\partial \hat{a}}{\partial y_j} \frac{\partial \hat{b}}{\partial y_j} \sigma_j^2,$$

where σ_j^2 is the variance of the j-th measurement, in this application assumed to be Poisson-distributed. Notice that this method can also be used to obtain the error matrix (24) (e.g., see Section 11.3 of [11]). The result is

$$\hat{\sigma}_{ab}^2 = \frac{1}{\Delta^2} \left(\left(N \sum_{i=1}^{N} x_i^2 + \left(\sum_{i=1}^{N} x_i \right)^2 \right) \sum_{j=1}^{N} x_j \sigma_j^2 \right.$$
$$\left. - \sum_{i=1}^{N} x_i \left(\sum_{i=1}^{N} x_i^2 \sum_{j=1}^{N} \sigma_j^2 + N \sum_{j=1}^{N} x_j^2 \sigma_j^2 \right) \right), \qquad (26)$$

which becomes the same as in (24) under the assumption of homoskedasticity. In (26), however, the variances to be used are the non-uniform Poisson-estimated $\sigma_j^2 = \hat{a} + x_j \hat{b}$. In summary, it is still possible to estimate a covariance matrix for the linear OLS in place of the maximum-likelihood and Poisson-based (15). These estimates, (23) and (26), are obtained by a *post facto* use of the Poisson distribution in the usual OLS regression estimators, which in general do not require the specification of a parent distribution. The data of Figure 1 yield linear OLS parameters of $a = 0.93 \pm 0.80$ and $b = 0.25 \pm 0.16$ for a covariance $\hat{\sigma}_{ab}^2 = -0.11$, using (23) and (26). The confidence band according to these errors is reported as the blue hatched area in Figure 1.

For the purpose of comparison, the parameter uncertainties for the standard OLS regression according to (24) and (25) are also calculated. First, from the OLS best-fit parameters, Equation (25) yields an estimated common data variance of $\hat{\sigma}^2 = 1.78$, and accordingly, the OLS error matrix yields errors that result in $a = 0.93 \pm 0.85$ and $b = 0.25 \pm 0.15$. In this application, the errors calculated for this "standard" OLS model is similar to those of the "modified" OLS model ($a = 0.93 \pm 0.80$, $b = 0.25 \pm 0.16$) and also to those with the "proper" Poisson regression of Section 4.1 ($\hat{a} = 0.63 \pm 1.02$ and $\hat{\lambda} = 0.53 \pm 0.66$, with overall slope $\hat{a} \cdot \hat{\lambda} = 0.33 \pm 0.15$). This chance agreement is in part the result of the fact that the Poisson data are consistent with the model, as determined by the analysis of the C statistics in Section 4.1.3. This agreement, however, is not guaranteed in every case, and the standard OLS

regression for heteroskedastic data is not recommended (see, e.g., the discussion in Section 3.1 of the textbook by [9]).

Finally, it is worth noticing that a weighted least-squares (WLS) linear regression, with weights equal to the variances of the measurements[5], is not recommended in this case. In fact, the WLS estimators depend on the unknown (and unequal) variances, and therefore, it is not possible to determine *a priori* a best-fit WLS model from which the Poisson variances for the error analysis is inferred, as was done in (23).

4.2.3. Goodness-of-fit analysis for OLS regression

Another difference between the maximum-likelihood C statistics regression and the OLS regression is with regard to goodness-of-fit testing. When the common variance is unavailable, as is generally the case for the OLS regression, the test statistics often used is

$$t = \frac{\hat{b}}{\hat{\sigma}_b^2} \sim t(N-2), \qquad (27)$$

where

$$\hat{\sigma}_b^2 = \frac{(1-r^2)}{N-2} \cdot \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \qquad (28)$$

is the sample variance of the slope b, and r is the linear correlation coefficient. Notice that this is a non-parametric estimate of the variance of the slope b, in that no assumptions on the distribution of the data are required. The linear correlation coefficient is defined by $r^2 = b\, b'$, where b is the usual slope of the linear regression of Y given X, and b' is the slope of the linear regression of X given Y (see, e.g., Chapter 14 of [11]). This variance differs from that in (23), where the parent distribution of the measurement is used. From the observations that $a = \bar{y} - b\,\bar{x}$ and that $(1 - r^2) \sum (y_i - \bar{y})^2 / (N-2)$ is the non-parametric unbiased estimator of the sample variance of y, (28) yields a non-parametric estimate of the variance of the parameter a of the regression as

$$\hat{\sigma}_a^2 = \hat{\sigma}_b^2 \left(\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2 + \bar{x}^2 \right), \qquad (29)$$

(see, e.g., Sections 11.5 and 14.3 of [11]).[6] The test statistics (27) assumes that the parent slope is null ($b = 0$), and it is distributed like a Student's t distribution with $N - 2$ degrees of freedom, and it is sometimes referred to as the *Wald statistics*, after [35]. For these data, the linear correlation coefficient is $r^2 = 0.273$ for a value of $t = 1.73$, corresponding to a p-value of 0.12. Therefore, in the case of the OLS regression, the t-statistic provides an alternate non-parametric means to test the null hypothesis that the data are uncorrelated or with $b = 0$.

---

5 The weighted least-squares regression is discussed, for example, in heteroskedastic 19.17 of [34].

6 These formulas are implemented in the `scipy.stats.linregress` software.
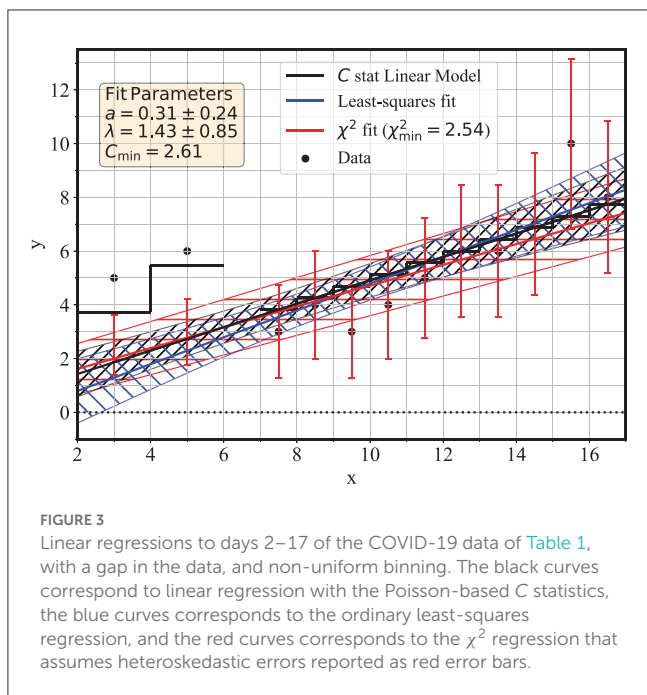
FIGURE 3
Linear regressions to days 2−17 of the COVID-19 data of Table 1, with a gap in the data, and non-uniform binning. The black curves correspond to linear regression with the Poisson-based C statistics, the blue curves corresponds to the ordinary least-squares regression, and the red curves corresponds to the χ^2 regression that assumes heteroskedastic errors reported as red error bars.

4.3. Fit to days 2−16 with a gap in the data, and non-uniform binning

In this example, the data for days 0 and 1 were ignored, the data for days 2–3 and 4–5 were combined into bins of size $\Delta x = 2$ days, the data for day 6 (where no events were reported) were ignored, and the datapoints for days 7–16 were kept with the original binning of $\Delta x = 1$ days, for a total of $N = 12$ measurements. These choices were simply made for illustration purposes. Specifically, it is generally not advisable to *ignore* a bin where no counts were recorded since the non-detection is actually positive information; a case with no counts-in-bin was provided in the previous example. Moreover the linear regression—and any non-linear regression as well—is sensitive to the choice of bin size. The choice of bin sizes must be dictated by an understanding of the data at hand and of the scientific goals of the regression.

4.3.1. Poisson regression
The results of the fit to these data are shown in Figure 3, for best-fit values of $\hat{a} = 0.31 \pm 0.24$, $\lambda = 1.43 \pm 0.85$, and an error matrix of

$$\hat{\varepsilon} = \begin{bmatrix} 0.716 & -0.202 \\ -0.202 & 0.060 \end{bmatrix}.$$

It is illustrative to compare this estimate to the one from the error propagation method described in Section 3.4, which yields an estimated error matrix of

$$\hat{\varepsilon}_{ep} = \begin{bmatrix} 0.687 & -0.200 \\ -0.200 & 0.061 \end{bmatrix},$$

where the variance of the measurements y_k were approximated with the estimated parent means $\hat{\mu}_k$. This covariance matrix is nearly

identical to the one based on the information matrix, and the small differences between $\hat{\varepsilon}$ and $\hat{\varepsilon}_{ep}$ provide an example of the agreement between the two estimates. Figure 3 also shows the confidence band for the Poisson regression as the gray hatched area.

4.3.2. Comparison with OLS and χ^2 regression
For comparison, a maximum-likelihood regression using the χ^2 statistic was also performed for this example, and the results are shown as the red line and confidence band. The ordinary least-squares regression is also shown in blue. For the method of regression that minimizes the χ^2 statistic, the data are assumed to be normally-distributed, with variances equal to the number of counts in each bin. Given the variable bin size, the continuous lines, therefore, represent a *density*, i.e., the number of counts per unit interval of the x variable, which in this case it is time in units of days. This is reflected in the relationship between the function (1) in units of counts per unit y, and the parent mean in (2) in units of pure counts, which are related by the bin size Δx_i. Accordingly, in the χ^2 fit with non-uniform binning, the fit statistic is

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - \mu_i}{\hat{\sigma}_i} \right)^2$$

where μ_i is given by (2) with a bin size $\Delta x_i = 2$ for $i = 1, 2$ and $\Delta x_i = 1$ for the other bins. Equivalently, the χ^2 regression can be performed directly to the function (1) (or to the linear model with the standard parameterization $a + bx$) by re-scaling both the measurements y_i and the errors $\hat{\sigma}_i$ to counts per unit y since this is the usual assumption in standard software that is used for maximum-likelihood regression for normal data.[7] Following this method, the first two data points have a count rate of one half of the values reported by the black dots, and the $\pm\hat{\sigma}$ error bars for the χ^2 regression are illustrated as red vertical lines; those error bars have no meaning for the C statistic or the OLS regressions.

The overall slope of the C statistic regression line is $a\lambda = 0.44 \pm 0.11$, while for the χ^2 regression it is 0.39 ± 0.13, and for the least-squares fit, it is 0.50 ± 0.08. The differences, albeit relatively small in this application, are the result of the different assumptions used in the modeling of the data, namely, the choice of Poisson vs. normal distribution for the number of counts in each bin.

5. Discussion and conclusion

This study has reviewed the [15] maximum-likelihood method of linear regression for Poisson count data and has presented an analytical method for the calculation of the covariance matrix for the parameters of a linear model. For this method of regression, it is found that the parameterization of the model of Equation 1, proposed by [20], is quite convenient due to its factorization of the two parameters, which results in convenient algebraic properties for the logarithm of the Poisson likelihood.

The method to obtain the covariance matrix is based on the Fisher information matrix, whose inverse approximates the

---

7 For example, `linregress` or `curve_fit` in the `scipy` libraries. Such rescaling of units preserves the statistical properties of the measurements.

covariance matrix in the asymptotic limit of a large number of measurement. The key result of this study is that it is possible to obtain a simple analytical expression for the covariance matrix (Equation 15) that can be used for the Poisson linear regression. This new result, together with the maximum-likelihood estimates of the parameters presented in [15], offers a simple and accurate method to perform linear regression on a variety of Poisson-distributed count data, including data with non-uniform binning and with the presence of gaps in the coverage of the independent variable. This method of regression for Poisson-distributed count data is, therefore, available for use in a variety of data applications, such as the one in the biomedical field presented in Section 4.

The study has also shown that an alternative method for the estimation of the covariance metrix, based on the error propagation or delta method, and also provides a simple analytical estimate that is in general good agreement with the method based on the information matrix. The main limitation in the use of the covariance matrix estimated from the information matrix (as well as that from the error propagation method) is that it applies only in the limit of a large number of measurements, where the maximum-likelihood estimates of the parameters are normally-distributed (e.g., [16, 25]). In this limit, the variance of the parameters and the covariance can be used for hypothesis testing assuming normal distribution of the parameters.

The maximum-likelihood linear regression to Poisson-distributed count data, presented in this study and in [15], have the advantages of being *unbiased* and *asymptotically efficient*, in the sense that, in the asymptotic limit of a large number of measurements, the estimators of the linear model parameters have the minimum variance bound according to the *Cramér-Rao* theorem ([24, 25], see also [34, 36]). Given that maximum-likelihood estimators are also known to be *consistent* in general (see, e.g., Chapter 18 of [34]), this method of regression has very desirable properties for the estimation of the linear model parameters. On the contrary, the ordinary least-squares regression does retain the convenient property of unbiasedness, even for the type of heteroskedastic variances that apply to these Poisson-distributed data. The OLS, however, is not guaranteed to be efficient, in the sense that the variances are in general larger than those of the maximum-likelihood method, and they can also be biased (e.g., [34, 37]). As a result, the OLS should be viewed as a less accurate method for the regression of Poisson-distributed count data.

Two other methods of linear regression also discussed in this study, the weighted least-squares and the χ^2 methods, suffer from more fundamental shortcomings when applied to integer-count Poisson data. In both cases, it is necessary to know *a priori* the unequal variances of the data, in order to proceed with the estimation, and this information is simply not available. In particular, the χ^2 method is often applied by making the assumption that the variances are equal to the square root of the counts, *de facto* assuming that the data are normally distributed and with variances equal to the *measured* counts. Although the Poisson distribution does converge to a normal distribution in the limit of a large number of counts (e.g., see Section 3 of [11]), the fact that the variance is approximated with the measured counts leads to a bias that remains even in the case of large-count data (e.g., [13, 14]). It is therefore not appropriate to use either of these two methods for the regression of Poisson-distributed count data.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Poisson SD. *Recherches sur la probabilit des jugements en mati re criminelle et en mati re civile*. Paris (1837).

2. Rutherford E, Chadwick J, Ellis C. *Radiations from Radioactive Substances*. Cambridge University Press (1920).

3. Clarke RD. An application of the Poisson distribution. *J Instit Act.* (1946) 72:481.

4. King G. Statistical models for political science event counts: bias in conventional procedures and evidence for the exponential Poisson regression model. *Am J Polit Sci.* (1988) 32:838–63.

5. Campbell JE. Explaining presidential losses in midterm congressional elections. *J Polit.* (1985) 47:1140–57.

6. Luria SE, Delbrück M. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics.* (1943) 28:491–511.

7. Ross SM. *Introduction to Probability Models*. San Diego, CA: Academic Press (2003).

8. Cash W. Parameter estimation in astronomy through application of the likelihood ratio. *Astrophys. J.* (1979) 228:939.

9. Cameron C, Trivedi PK. *Regression Analysis of Count Data.* 2nd ed. Cambridge University Press (2013).

10. Greenwood P, Nikulin M. *A Guide to Chi-Squared Testing.* Wiley (1996).

11. Bonamente M. *Statistics and Analysis of Scientific Data.* 3rd ed. Singapore: Springer (2022).

12. Lewis D, Burke CJ. The use and misuse of the chi-square test. *Psychol Bull.* (1949) 46:433–89.

13. Kelly BC. Some aspects of measurement error in linear regression of astronomical data. *Astrophys. J.* (2007) 665:1489–506.

14. Bonamente M. Distribution of the C statistic with applications to the sample mean of Poisson data. *J Appl Stat.* (2020) 47:2044–65. doi: 10.1080/02664763.2019.1704703

15. Bonamente M, Spence D. A semi–analytical solution to the maximum–likelihood fit of Poisson data to a linear model using the Cash statistic. *J Appl Stat.* (2022) 49:522–52. doi: 10.1080/02664763.2020.1820960

16. Fisher RA. On the mathematical foundations of theoretical statistics. *Philos Trans R Soc Lond Ser A.* (1922) 222:309–68.

17. Hilbe JM. *Modeling Count Data.* Cambridge University Press (2014).

18. Cameron AC, Trivedi PK. Econometric models based on count data. Comparisons and applications of some estimators and tests. *J Appl Econometr.* (1986) 1:29–53.

19. McCullagh P, Nelder JA. *Generalized Linear Models.* 2nd ed. Chapman & Hall/CRC (1989).

20. Scargle JD, Norris JP, Jackson B, Chiang J. Studies in astronomical time series analysis. VI. Bayesian block representations. *Astrophys. J.* (2013) 764:167. doi: 10.1088/0004-637X/764/2/167

21. Ahoranta J, Nevalainen J, Wijers N, Finoguenov A, Bonamente M, Tempel E, et al. Hot WHIM counterparts of FUV O VI absorbers: evidence in the line-of-sight towards quasar 3C 273. *Astronomy and Astrophysics.* (2020) 634:A106. doi: 10.1051/0004-6361/201935846

22. Levine AM, Bradt H, Cui W, Jernigan JG, Morgan EH, Remillard R, et al. First results from the all-sky monitor on the Rossi x-ray timing explorer. *The Astrophysical Journal Letters.* (1996) 469:L33.

23. Eadie WT, Drijiard D, James M, Roos M, Sadoulet B. *Statistical Methods in Experimental Physics.* North Holland (1971).

24. Rao CR. Information and accuracy attainable in the estimation of statistical parameters. *Bull Calcutta Math Soc.* (1945) 37:81.

25. Cramer H. *Mathematical Methods of Statistics.* Princeton, NJ: Princeton University Press (1946).

26. Amemiya T. *Advanced Econometrics.* Harvard University Press (1985).

27. Shannon CE, Weaver W. *The Mathematical Theory of Communication.* Urbana, IL: University of Illinois Press (1949).

28. Kullback S, Leibler RA. On information and sufficiency. *Ann Math Stat.* (1951) 22:79–86.

29. Akaike H. A new look at the statistical model identification. *IEEE Trans Automatic Control.* (1974) 19: 716–23.

30. Fisher RA. *Statistical Methods for Research Workers.* Edinburgh: Oliver and Boyd (1925).

31. Bevington PR, Robinson DK. *Data Reduction and Error Analysis for the Physical Sciences.* 3rd ed. McGraw Hill (2003).

32. New York Times. *GitHub* (2021). Available online at: https://github.com/nytimes/covid-19-data/blob/master/us.csv

33. Lampton M, Margon B, Bowyer S. Parameter estimation in X-ray astronomy. Astrophys. J.. (1976) 208:177–90.

34. Kendall M, Stuart A. *The Advanced Theory of Statistics. Vol.2: Inference and Relationship.* 3rd ed. New York, NY: Hafner (1973).

35. Wald A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans Am Math Soc.* (1943) 54: 426–82.

36. Rao CR. *Linear Statistical Inference and Its Applications.* 2nd ed. John Wiley and Sons (1973).

37. Swindel BF. On the bias of some least-squares estimators of variance in a general linear model. *Biometrika.* (1968) 55:313–6.