entropy

MDPI

Article

Linear Bayesian Estimation of Misrecorded Poisson Distribution

Huiqing Gao [1], Zhanshou Chen [1,2,*] and Fuxiao Li [3]

[1] School of Mathematics and Statistic, Qinghai Normal University, Xining 810008, China; huiqinggao0820@outlook.com
[2] The State Key Laboratory of Tibetan Intelligent Information Processing and Application, Xining 810008, China
[3] Faculty of Science, Xi'an University of Technology, Xi'an 710048, China; lifuxiao@xaut.edu.cn
* Correspondence: chenzhsh@qhnu.edu.cn

Abstract: Parameter estimation is an important component of statistical inference, and how to improve the accuracy of parameter estimation is a key issue in research. This paper proposes a linear Bayesian estimation for estimating parameters in a misrecorded Poisson distribution. The linear Bayesian estimation method not only adopts prior information but also avoids the cumbersome calculation of posterior expectations. On the premise of ensuring the accuracy and stability of computational results, we derived the explicit solution of the linear Bayesian estimation. Its superiority was verified through numerical simulations and illustrative examples.

Keywords: misrecorded Poisson distribution; linear Bayesian estimation; bootstrap

check for updates

Citation: Gao, H.; Chen, Z.; Li, F. Linear Bayesian Estimation of Misrecorded Poisson Distribution. *Entropy* **2024**, *26*, 62. https://doi.org/10.3390/e26010062

Academic Editors: Carlos Alberto De Bragança Pereira, Paulo Canas Rodrigues and Mark Andrew Gannon

Received: 21 November 2023
Revised: 2 January 2024
Accepted: 9 January 2024
Published: 11 January 2024

1. Introduction

Count data, which are typically characterized as data points that cannot be collected consecutively or are represented as 0 or natural numbers like 1, 2, 3, etc., are commonly encountered in various aspects of daily life. These data points are often effectively approximated by binomial or Poisson distributions. However, certain challenges may emerge during the actual counting process. When determining the number of defects per unit or item, there is a possibility that the recorder may inaccurately classify some units that actually contain one defect as perfect or defect-free while accurately recording other values. More specifically, some observations with a value of 1 are misclassified and reported as 0, while others, such as 2, 3, and 4, are correctly recorded. For this type of data containing erroneous records, direct modeling with a Poisson distribution is not appropriate. Instead, it can be approximated using a misrecorded Poisson distribution.

The misrecorded Poisson distribution, first proposed by Cohen [1], is often denoted as the Cohen–Poisson (CP) or MSR-Poisson distribution. Since then, many authors have studied this distribution and its various forms. For surveys, Dorris and Foote [2] provided a comprehensive survey and analysis of the impact of inspection errors on statistical quality control procedures. They discuss their effects on various aspects of statistical quality control procedures, including control charts, sample size, and inspection efficiency. Subsequently, the article introduces methods for estimating error probabilities and adjusting parameters in Poisson distributions, along with strategies for designing compensatory plans to mitigate the impact of inspection errors. All these efforts are grounded in Cohen's seminal work, which developed multiple estimators for error probability assessment and parameter recalibration in Poisson distributions, thereby establishing a significant theoretical framework. Gupta et al. [3] studied when the observed frequency of zeros is significantly higher or lower than predicted by the model; the adjusted random variable can be described as a mixture of two Poisson distributions. This distribution can be regarded as a mixture of two distributions, one of which is degenerate at zero. Zhang et al. [4] study the zero-and-one inflated Poisson (ZOIP) distribution, develop likelihood-based inference methods, and provide simulations and real data examples to illustrate the proposed methods. The key

properties of the ZOIP distribution are established, including five equivalent stochastic representations and other important distributional properties. Maximum likelihood estimates of parameters are obtained through Fisher scoring and the expectation–maximization algorithm, and bootstrap confidence intervals and hypothesis testing methods are provided for parameters in large samples. Liu [5] et al. introduce a new multivariate ZAP distribution, building on the traditional multivariate Poisson distribution. A key feature of this distribution is its ability to model count vectors that are zero-truncated, zero-deflated, or zero-inflated with a more flexible dependency structure. This means that correlation coefficients between components can be both positive and negative. They introduce an expectation–maximization (EM) algorithm for calculating the maximum likelihood estimates (MLEs) and posterior modes of parameters. Bagui and Mehra [6] gave a historical background for the Poisson distribution and also described some of its applications in the early days and more. Additionally, they introduced a new ratio method that demonstrates convergence to the normal distribution, which may attract attention. In the context of Poisson processes, considering the swift progress in current industrial and technological spheres, statistical models must advance to keep up with the times, adapting to handle data with higher frequencies and greater complexity.

The Hawkes process model, noted for its self-excitation and time-dependence, serves as an extension to the Poisson process. This extension offers more sophisticated and adaptable tools for analyzing and forecasting error patterns in ever-changing environments. For relevant research results, refer to Zhang et al. [7] and Wang and Zhang [8]. Lamprinakou et al. [9] proposed a new epidemiological model using the Hawkes process to study the spread of COVID-19. It focuses on estimating unobserved infection cases, offering insights into the disease's transmission dynamics.

Furthermore, numerous monographs and doctoral dissertations have investigated the misrecorded Poisson distribution and its various forms. Xu [10] studied the statistical properties of a Poisson generalized inverse Gaussian distribution, a negative binomial distribution, a Poisson inverse Gaussian distribution, and a Poisson inverse gamma distribution. The model they studied is similar to the model studied in this article, which is very helpful for the study of properties and parameter derivation methods in this article. Johnson [11] has effectively synthesized the findings from various studies on the misrecorded Poisson distribution and its parameter estimation, encompassing research conducted by Cohen as well as contributions from other researchers in the field. Djuraš [12] discussed the one-displaced misrecorded Poisson and size-biased misrecorded Poisson distributions, deriving their parameter estimators. Tuwei [13] derived the probability density function from the probability generating function and provided expressions for the mean, the variance, and the relationship between parameters based on this probability density function.

While the majority of previous studies have focused on classical estimation methods, Angers and Biswas [14] employ Bayesian methods for parameter estimation and prediction. Specifically, they contemplate a zero-inflated generalized Poisson model, which encompasses three parameters: the zero-inflation parameter, the dispersion parameter of the generalized Poisson distribution, and the mean parameter. For the purpose of Bayesian estimation, the article utilizes appropriate prior distributions and employs a Monte Carlo integration technique via importance sampling to obtain the posterior distributions. Subsequently, the expected values of the posterior distributions are computed, serving as the Bayesian estimates of the parameters. Rodrigues [15] studied the zero-inflated distributions from a Bayesian point of view using a data augmentation algorithm. Wang [16] considered a 0–1 expansion Poisson regression model with covariates and proposed a Bayesian estimation of the model parameters.

Compared to traditional Bayesian estimation methods, it is worth noting that the linear Bayesian method, initially proposed by Hartigan [17], has gained popularity as a simple Bayesian approach for parameter estimation. Rao [18] studied the linear Bayesian method from the perspective of linear optimization and argued that this method better considers the prior uncertainty and uses this "incomplete" information to construct linear Bayesian

parameter estimates. LaMotte [19] introduced a linear Bayesian estimator that has the smallest total mean square error among all linear estimators. Hesselager [20] demonstrated that if the average risk of an empirical linear Bayesian estimate converges to the risk of the corresponding linear Bayesian estimate, then it is asymptotically optimal in the usual sense. Goldstein [21] adeptly modified the linear Bayesian estimator, specifically designed for estimating the mean of a distribution whose form is unknown, by incorporating the use of an estimate derived from sample variance. Hoffmann [22] applied the linear Bayesian method to estimate an unknown parameter vector in the linear regression model with ellipsoidal parameter constraints. The conditions under which certain linear empirical Bayesian estimators are superior to the standard estimator for an arbitrary $k \geq 1$ are given by Samaniego and Vestrup [23]. Wei and Zhang [24] verified under the Predictive Pitman Closeness criterion and Posterior Pitman Closeness that the linear Bayesian method outperforms the generalized least squares method for linear models. More recently, Lin [25] employed Monte Carlo simulations and empirical computations to compare the outcomes of constrained least-squares estimation and constrained linear Bayesian estimation, as well as their distances from the Bayesian estimates. This approach serves to validate the superiority of constrained linear Bayesian estimation over constrained least-squares estimation. Tao [26] provided an empirical linear Bayesian approach, conducting an empirical analysis using a Bayesian model for multiple insurance contracts with Pareto distributions. This method does not rely on any prior distribution information. Liu [27] studied the parameter estimation problem of singular linear models with equality constraints and concluded that the higher the degree of singularity of the model, the more obvious the superiority of the best homogeneous linear Bayesian unbiased estimation over the least-squares estimation. Chen [28] introduces a linear Bayesian method for estimating parameters in extreme value distributions from Type II censored samples. This approach, which combines Bayesian and optimization techniques, offers a straightforward and practical solution, overcoming the complexity often associated with classical Bayesian estimation. The method's advantages, particularly for small or heavily censored samples, are validated through numerical experiments, demonstrating its effectiveness when compared to maximum likelihood and unbiased estimations. The models, parameter estimation methods, and criteria for assessing their effectiveness used in the aforementioned work have significantly contributed to the research presented in this paper. These models, encompassing a range of statistical, probabilistic, and computational approaches, form the fundamental basis of our analysis. They provide a structured and systematic framework that aids in the coherent understanding and interpretation of the complex data sets we are dealing with. The selection of a particular model is heavily dependent on the specific nature of the data under study and the overarching research question, with each model introducing its own set of assumptions and perspectives to the analysis.

Parameter estimation methods are a key component in refining and fine-tuning these models. These methods are instrumental in extracting significant insights from the data, as they focus on determining the values of the model parameters that most accurately represent the observed data. Depending on the model's nature and the data's characteristics, various techniques like maximum likelihood estimation, Bayesian inference, or least-squares fitting are utilized. The precision and reliability of these methods are paramount, as they have a direct bearing on the validity and credibility of our research findings. Therefore, the judicious selection and application of these methods are critical.

Given the notable advantages of linear Bayesian estimators, this paper specifically focuses on parameter estimation for the misrecorded Poisson distribution using linear functions derived from sample data. We employ the criterion of minimizing the mean square error to formulate linear Bayesian estimation expressions for the distribution's parameters. To our knowledge, no existing studies have explored linear Bayesian estimation specifically for parameters within the misrecorded Poisson distribution. Traditional methods often fall short in addressing the unique challenges posed by misrecorded data, such as biased estimates or increased error variance. Therefore, our method fills a significant gap in

the statistical methodology by providing a tailored solution for this specific distribution. Furthermore, we use the criterion of minimizing the mean square error to formulate linear Bayesian estimation expressions for the distribution's parameters. This criterion ensures that the estimations are not only accurate but also consistent by minimizing the average of the squares of the errors—the difference between the estimator and what is estimated. This approach optimizes the balance between bias and variance in the estimations, leading to more reliable results.

Our work stands out not only for its innovative approach to handling the misrecorded Poisson distribution but also for its potential wide-ranging applications in various fields. In areas where misrecorded count data are common, such as epidemiology, environmental studies, and quality control in manufacturing, the implications of this research are substantial. By introducing new insights and techniques for managing such distributions, our study promises to enhance the accuracy and reliability of data analysis in these fields, leading to more informed and dependable conclusions. This could mark a significant advancement in statistical methods, opening doors to more robust and sophisticated data analysis techniques in these critical areas of study.

The rest of this paper is structured as follows: Section 2 offers a detailed examination of the misrecorded Poisson distribution and delves into its maximum likelihood estimation. In Section 3, we introduce and elaborate on a linear Bayesian estimation method, highlighting its superior estimation results, primarily based on the principle of the mean square error matrix. Section 4 presents numerical simulations and example analyses. Lastly, in Section 5, we summarize the entire study, provide our conclusions, point out the shortcomings of the article, and suggest future research directions

2. Misrecorded Poisson Distribution

The probability mass function of the misrecorded Poisson distribution is

$$P(X=x)=\begin{cases} e^{-\lambda}(1+\lambda\phi) & x=0 \\ e^{-\lambda}\lambda(1-\phi) & x=1 \\ e^{-\lambda}\frac{\lambda^x}{x!} & x=2,3,\ldots \end{cases}, \tag{1}$$

where $\lambda>0$, $0<\phi<1$. ϕ is the probability of misclassification or the proportion of ones that are reported as zeros. The misrecorded Poisson distribution is reduced to a Poisson distribution when $\phi=0$. Also, all the observations falling in class one will be reported as zero when $\phi=1$; thus, it is a zero–one modified distribution. We can say that the misrecorded Poisson distribution is zero-inflated (there is an excess of zeros) and one-deflated (there are fewer ones than expected). If $\phi_0=\lambda\phi e^{-\lambda}>0$, then the Poisson's zero probability is increased (zero inflation), while the Poisson's one probability is decreased for $\phi_1=-\lambda\phi e^{-\lambda}<0$.

The corresponding probability generating function, the mean, the variance, and the second, third, and fourth moments are given as, respectively,

$$\begin{aligned} G_X(s)&=e^{-\lambda}(1+\lambda\phi)+\lambda e^{-\lambda}(1-\phi)s+\sum_{x=2}^{\infty}\frac{(\lambda s)^x e^{-\lambda}}{x!} \\ &=e^{-\lambda}(1+\lambda\phi)+\lambda s e^{-\lambda}(1-\phi)+\left\{e^{-\lambda(s-1)}-\lambda s e^{-\lambda}-e^{-\lambda}\right\} \\ &=e^{-\lambda}+\lambda\phi e^{-\lambda}+\lambda s e^{-\lambda}-\lambda\phi s e^{-\lambda}+e^{-\lambda(s-1)}-\lambda s e^{-\lambda}-e^{-\lambda} \\ &=\lambda\phi e^{-\lambda}-\lambda\phi s e^{-\lambda}+e^{-\lambda(s-1)}, \\ E(X)&=\mu=\lambda(1-\phi e^{-\lambda}), \\ Var(X)&=C_2(0)=\sigma_X^2=\lambda^2+(1+\phi e^{-\lambda})(1-\lambda(\phi e^{-\lambda})), \\ E(X^2)&=\mu_{(0)}=\lambda^2+\lambda(1-\phi e^{-\lambda}), \end{aligned}$$

$$E(X^3) = \mu_{(0,0)} = \lambda^3 + 3\lambda^2 + (1 - \phi e^{-\lambda}),$$
$$E(X^4) = \mu_{(0,0,0)} = \lambda^4 + 6\lambda^3 + 7\lambda^2 + (1 - \phi e^{-\lambda}).$$

assuming that the random variables $x_1, x_2, \ldots\ldots, x_N$ are samples consisting of N observations from the misrecorded Poisson distribution, with n_0 denoting the number of zero observations and n_1 denoting the number of one observations. The likelihood function is

$$L(x_1, x_2, \cdots x_N; \lambda, \phi) = [e^{-\lambda}(1 + \phi\lambda)]^{n_0}[(1 - \phi)\lambda e^{-\lambda}]^{n_1}\prod * e^{-\lambda}\lambda^{x_i}/x_i$$
$$= e^{-N\lambda}(1 + \phi\lambda)^{n_0}(1 - \phi)^{n_1}\lambda^{\sum_{i=1}^{N} x_i}\left[\prod * x_i!\right]^{-1}, \quad (2)$$

where $\prod *$ is the product of $x_1, x_2, \ldots\ldots, x_N$ that are neither zero nor one. Taking the logarithms of Equation (2) and letting its partial derivatives with respect to λ and ϕ, respectively, equal zero, we obtain the following equation:

$$\partial \log L/\partial\lambda = -N + n_0\phi/(1 + \phi\lambda) + \sum_{i=1}^{N} x_i/\lambda = 0.$$
$$\partial \log L/\partial\phi = n_0\lambda(1 + \phi\lambda) - n_1/(1 - \phi) = 0,$$

and then we have:

$$\lambda^2 - (\overline{x} - 1 + n_0/N)\lambda - (\overline{x} - n_1/N) = 0, \quad (3)$$
$$\phi = [n_0 - n_1/\lambda]/(n_0 + n_1), \quad (4)$$

where $\overline{x} = \sum_{i=1}^{N} x_i/N$ is the mean of the sample. Solving the above equation yields the maximum likelihood estimates $\hat{\lambda}_{MLE}$ and $\hat{\phi}_{MLE}$ for λ and ϕ:

$$\hat{\lambda}_{MLE} = \left[(\overline{x} - 1 + n_0/N) + \sqrt{(\overline{x} - 1 + n_0/N)^2 + 4(\overline{x} - n_1/N)}\right]/2, \quad (5)$$

$$\hat{\phi}_{MLE} = \left(n_0 - n_1/\hat{\lambda}_{MLE}\right)/(n_0 + n_1). \quad (6)$$

Translating the solution of this estimate into matrix form to express it, we have

$$\hat{\theta}_{MLE} = (\hat{\lambda}_{MLE}, \hat{\phi}_{MLE})' = \begin{pmatrix} 1/2 & 0 \\ 0 & \frac{1}{n_0+n_1} \end{pmatrix}\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \triangleq AT, \quad (7)$$

where

$$A = \begin{pmatrix} 1/2 & 0 \\ 0 & \frac{1}{n_0+n_1} \end{pmatrix}, T = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}. \quad (8)$$

3. Linear Bayesian Estimation

3.1. The Expressions of Linear Bayesian Estimation

Definition 1. *Assume that the prior distribution function, $\pi(\theta)$ of $\theta = (\lambda, \phi)'$, satisfies the family of prior distributions:*

$$\xi = \left\{\pi(\theta) : E\|\theta\|^2 < \infty\right\}. \quad (9)$$

Let $\hat{\theta}_{LB} = BT + b$ be a linear estimate of the parameter θ under the statistic T, in which B and b are 2×2- and 2×1-dimensional unknown matrices, respectively. Then, $\hat{\theta}_{LB}$ is called the linear Bayesian estimate of θ if $\hat{\theta}_{LB}$ satisfies the following conditions:

$$E_{(T,\theta)}(\hat{\theta}_{LB} - \theta) = 0, \quad (10)$$

$$R(\hat{\theta}_{LB}, \theta) = \min_{B,b} E_{(T,\theta)} L(\hat{\theta} - \theta), \tag{11}$$

where $E_{(T,\theta)}$ denotes the expectation of the joint distribution of T and θ.

The function

$$L(\hat{\theta}_{LB}, \theta) = (\hat{\theta}_{LB}, \theta)' D(\hat{\theta}_{LB}, \theta), \tag{12}$$

where D is a 2×2 positive definite matrix, is called the squared loss function. Minimizing the squared loss, we can obtain the expressions of B and b, which are given in the following theorem.

Theorem 1. *Suppose that the prior distribution, $\pi(\theta)$, of the parameter vector $\theta = (\lambda, \varphi)'$ in the misrecorded Poisson distribution (1) satisfies the family of prior distributions (9). Then, under the squared loss function, the linear Bayesian estimate, $\hat{\theta}_{LB}$, of θ is*

$$\hat{\theta}_{LB} = (A - AWM)T + AWMA^{-1}E\theta, \tag{13}$$

$$b = (I - B)E(\theta) = (I - B)\mu, \tag{14}$$

where

$$W = E[Cov(T, \theta)], \tag{15}$$

$$M = \left[W + A^{-1}Cov(\theta)(A^{-1})'\right]^{-1}. \tag{16}$$

Proof.

According to the condition in Equation (10), we have

$$0 = E[E(\hat{\theta}_{LB} - \theta|\theta)] = E[E(BT + b) - \theta],$$

so

$$BE_{(T,\theta)}(T) + b = E\theta,$$

$$b = E\theta - BA^{-1}E\theta. \tag{17}$$

According to the condition in Equation (11), we have

$$\begin{aligned}
E_{(T,\theta)} L(\hat{\theta}, \theta) &= E_{(Y,\theta)}\{[BT + b - \theta]' D[BT + b - \theta]\} \\
&= E_{(T,\theta)} L(\hat{\theta}, \theta) \\
&= E_{(T,\theta)}\Big\{[BT + E\theta - BA^{-1}E\theta - \theta]' D[BT + E\theta - BA^{-1}E\theta - \theta]\Big\}. \\
&E_{(T,\theta)}\Big\{tr(D[BT + E\theta - BA^{-1}E\theta - \theta][BT + E\theta - BA^{-1}E\theta - \theta]')\Big\} \\
&= E_{(T,\theta)}\Big\{tr(D[B(T - A^{-1}E\theta) - (\theta - E\theta)][B(T - A^{-1}E\theta) - (\theta - E\theta)]')\Big\} \\
&= tr(DE_{(T,\theta)}\Big\{[B(T - A^{-1}E\theta) - (\theta - E\theta)][B(T - A^{-1}E\theta) - (\theta - E\theta)]'\Big\}) \\
&= tr(DE_{(T,\theta)}\Big\{B(T - A^{-1}E\theta)(T - A^{-1}E\theta)'B' - B(T - A^{-1}E\theta)(\theta - E\theta)' \\
&\qquad -(\theta - E\theta)(T - A^{-1}E\theta)'B' + (\theta - E\theta)(\theta - E\theta)'\Big\}) \\
&= tr(BDE_{(T,\theta)}[(T - A^{-1}E\theta)(T - A^{-1}E\theta)']B') - tr(DBA^{-1}Cov(\theta)) \\
&\qquad -tr(DCov(\theta)(A^{-1})'B') + tr(DCov(\theta)).
\end{aligned} \tag{18}$$

In the above equation,

$$E_{(T,\theta)}[(T - A^{-1}E\theta)(T - A^{-1}E\theta)']$$

$$= E_{(T,\theta)}[TT' - T(A^{-1}E\theta)' - A^{-1}E\theta T' + A^{-1}E\theta(A^{-1}E\theta)']$$
$$= E_{(T,\theta)}TT' + A^{-1}E\theta(A^{-1}E\theta)' - E_{(T,\theta)}T(A^{-1}E\theta)' - (A^{-1}E\theta)E_{(T,\theta)}T']$$
$$= E_{(T,\theta)}TT' + A^{-1}E\theta(A^{-1}E\theta)' - A^{-1}E\theta(A^{-1}E\theta)' - A^{-1}E\theta(A^{-1}E\theta)'$$
$$= E_{(T,\theta)}TT' - A^{-1}E\theta(A^{-1}E\theta)'$$
$$= E[E(TT'\Big|\theta)] - A^{-1}E\theta(A^{-1}E\theta)'$$
$$= E\Big\{E[(T - A^{-1}\theta + A^{-1}\theta)(T - A^{-1}\theta + A^{-1}\theta)'\Big|\theta]\Big\} - A^{-1}E\theta(A^{-1}E\theta)'$$
$$= E\Big\{E[A^{-1}\theta(A^{-1}\theta)' + (T - A^{-1}\theta)(T - A^{-1}\theta)' + (T - A^{-1}\theta)(A^{-1}\theta)'$$
$$+(A^{-1}\theta)(T - A^{-1}\theta)'\Big|\theta]\Big\} - A^{-1}E\theta(A^{-1}E\theta)'$$
$$= E\Big\{A^{-1}\theta(A^{-1}\theta)' + E[(T - A^{-1}\theta)(T - A^{-1}\theta)'] + E\Big[(T - A^{-1}\theta)(A^{-1}\theta)'\Big]$$
$$= E[A^{-1}\theta(A^{-1}\theta)' + E((T - A^{-1}\theta)(T - A^{-1}\theta)')] - A^{-1}E\theta(A^{-1}E\theta)'$$
$$+E(A^{-1}\theta(A^{-1}\theta)')\Big\} - A^{-1}E\theta(A^{-1}E\theta)'$$
$$= E[A^{-1}\theta(A^{-1}\theta)'] + E(Cov(T\Big|\theta)) - A^{-1}E\theta(A^{-1}E\theta)'$$
$$= E(Cov(T\Big|\theta)) + A^{-1}[E(\theta\theta') - E(\theta)E(\theta)'](A^{-1})$$
$$= W + A^{-1}Cov(\theta)(A^{-1})'.$$

Therefore,

$$E_{(T,\theta)}L(\hat{\theta},\theta) = tr(DB(W + A^{-1}Cov(\theta)(A^{-1})')B') - tr(DBA^{-1}Cov(\theta))$$
$$-tr(DCov(\theta)(A^{-1})'B') + tr(DCov(\theta)).$$

By taking a partial derivative of the matrix B in the above equation, we have

$$\frac{\partial R(\hat{\theta},\theta)}{\partial B} = 2DB[W + A^{-1}Cov(\theta)(A^{-1})'] - 2DCov(\theta)(A^{-1}) = 0,$$
$$B = A - AW[W + A^{-1}Cov(\theta)(A^{-1})']^{-1} = A - AWM, \tag{19}$$

where

$$W = E[Cov(T,\theta)]$$
$$= E[(T - A^{-1}\theta)(T - A^{-1}\theta)'\Big|\theta]$$
$$= E\left\{E\left[\begin{pmatrix}t_1\\t_2\end{pmatrix} - \begin{pmatrix}2 & 0\\0 & \frac{1}{n_0+n_1}\end{pmatrix}\begin{pmatrix}\lambda\\\phi\end{pmatrix}\right]\left[\begin{pmatrix}t_1\\t_2\end{pmatrix} - \begin{pmatrix}2 & 0\\0 & \frac{1}{n_0+n_1}\end{pmatrix}\begin{pmatrix}\lambda\\\phi\end{pmatrix}\right]'|\theta\right\}$$
$$= \begin{bmatrix}E(t_1 - 2\lambda)^2 & 0\\0 & E(t_2 - (n_0+n_1)\lambda)^2\end{bmatrix}$$
$$= \begin{bmatrix}Var(t_1) & 0\\0 & Var(t_2)\end{bmatrix}. \tag{20}$$

Theorem 1 proved. □

An expression for linear Bayesian estimation, as detailed in the previously stated Theorem 1, has been constructed. This expression is derived while adhering to the constraints of unbiasedness and focused on the minimization of risk.

3.2. The Superiority of Linear Bayesian Estimation

In this subsection, we focus on examining the superiority of linear Bayesian estimation. To evaluate the effectiveness of this estimator, we utilize the mean square error (MSE), a well-regarded measure for assessing how closely an estimator approximates the actual parameter it seeks to estimate. Additionally, for a more careful comparison, we also consider the mean square error matrix (MSEM). To assess the superiority of $\hat{\theta}_{LB}$ according to the MSEM criterion, let us start by defining the MSEM.

Definition 2. *Let $\hat{\theta}$ be an estimator of the parameter vector θ. The MSE of $\hat{\theta}$ is defined as*

$$MSE(\theta) = E[(\hat{\theta} - \theta)'(\hat{\theta} - \theta)], \tag{21}$$

and the MSEM of $\hat{\theta}$ is defined as

$$MSEM(\theta) = E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)']. \tag{22}$$

Let the parameter vector θ be estimated in two different ways as $\hat{\theta}_1$ and $\hat{\theta}_2$. If

$$MSEM(\hat{\theta}_2) - MSEM(\hat{\theta}_1) \geq 0,$$

or

$$(MSEM(\hat{\theta}_2) - MSEM(\hat{\theta}_1) \geq 0), \tag{23}$$

then $\hat{\theta}_1$ is said to be better than $\hat{\theta}_2$ under the MSEM or MSE criterion, and it is clear that the MSEM criterion is stronger than the MSE criterion.

From the previous section, the maximum likelihood estimation, $\hat{\theta}_{MLE}$, can be expressed as the product of the statistic T and the matrix A, and the linear Bayesian estimation, $\hat{\theta}_{LB}$, is also constructed from the statistic T. Therefore, in this section, the MSEM criterion is applied to compare $\hat{\theta}_{LB}$ and $\hat{\theta}_{MLE}$.

Theorem 2. *The linear Bayesian estimation, $\hat{\theta}_{LB}$, defined in (7) is preferred to the maximum likelihood estimation, $\hat{\theta}_{MLE}$, defined in (13) under the MSEM criterion, i.e.,*

$$MSEM(\hat{\theta}_{LB}) \leq MSEM(\hat{\theta}_{MLE})$$

Proof.

$$\begin{aligned}
MSEM(\hat{\theta}_{LB}) &= E_{(T,\theta)}[(\hat{\theta}_{LB} - \theta)(\hat{\theta}_{LB} - \theta)'] \\
&= E[Cov(\hat{\theta}_{LB} - \theta|\theta)] + Cov[E(\hat{\theta}_{LB} - \theta|\theta)] \\
&= E[Cov(\hat{\theta}_{LB}|\theta)] + Cov[E(\hat{\theta}_{LB} - \theta|\theta)] \\
&= E[Cov(A - AWM)T + AWMA^{-1}E\theta\Big|\theta)] \\
&+ Cov[E(AT - AWM(T - A^{-1}E\theta\Big) - \theta\Big|\theta)] \\
&= (A - AWM)E(Cov(T|\theta)(A - AWM)') \\
&+ Cov[E(AT\Big|\theta) - E(AWM(T - A^{-1}E\theta\Big|\theta) - \theta)] \\
&= (A - AWM)W(A - AWM)' + Cov(-AWME(T - A^{-1}E\theta\Big|\theta)) \\
&= (A - AWM)W(A - AWM)' + Cov(AWMA^{-1}E[AT - E\theta\Big|\theta])
\end{aligned}$$

$$= A[(I - WM)W(I - WM) + WMA^{-1}Cov(\theta)(A^{-1})'MW]A'$$
$$= A[W - 2WMW + WM(W + A^{-1}Cov(\theta)(A^{-1})')MW]A'$$
$$= A[W - WMW]A'.$$

However,

$$MSEM(\hat{\theta}_{MLE}) = E_{(T,\theta)}[(\hat{\theta} - \theta)(\hat{\theta} - \theta)']$$
$$= E\{E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)'|\theta]\}$$
$$= E\{E[(AT - \theta)(AT - \theta)'|\theta]\}$$
$$= AWA'.$$

According to the defining equation of W, we can easily find that W is positive and definite, and because of the definition of the matrix M, both A^{-1} and $(A^{-1})'$ are positive definite matrices. Thus, Theorem 2 is proved. □

4. Numerical Simulation and Empirical Application

4.1. Numerical Simulation

In this section, we evaluate the performance superiority of classical and linear Bayesian estimation methods for the misrecorded Poisson distribution using numerical simulations. By choosing a range of different parameter values, we compute the mean square error (MSE) for both classical and linear Bayesian estimations relative to these values. A smaller error indicates a more accurate estimation. The procedures for these numerical simulations are detailed in the following steps.

1. Generate a dataset that accurately follows the characteristics of the misrecorded Poisson distribution, ensuring to use the predefined true parameter values as a foundational reference.

2. Proceed to calculate and derive parameter estimates, utilizing and comparing different established estimation methods for the analysis.

3. Calculate the mean and variance of the statistic T defined in Equation (20) using the bootstrap sampling method, and then use these values in the computation.

4. Repeat the above steps J times to obtain the mean of the estimate and give the MSE.

We set the sample size to N = 200 and 800 and set the parameters in the misrecorded Poisson distribution (1) to very between $\lambda = 0.7$ and $\phi = 0.7$; $\lambda = 0.7$ and $\phi = 0.9$; $\lambda = 0.9$ and $\phi = 0.7$; and $\lambda = 0.9$ and $\phi = 0.9$. All the simulations are performed via $J = 100$ repetitions. The results are listed in Tables 1 and 2. In these two tables, $\hat{\lambda}_{LB}$ and $\hat{\phi}_{LB}$ represent the linear Bayesian estimates of λ and ϕ, $\hat{\theta}_{MLEMSE}$ denotes the MSE for the maximum likelihood estimation, $\hat{\theta}_{LBMSE}$ denotes the MSE of the linear Bayesian estimation, and $\|\hat{\theta}_{MLEMSE} - \hat{\theta}_{LBMSE}\|_2$ is the Euclidean distance between $\hat{\theta}_{MLEMSE}$ and $\hat{\theta}_{LBMSE}$.

Table 1. Estimation results.

	λ	ϕ	$\hat{\lambda}_{MLE}$	$\hat{\phi}_{MLE}$	$\hat{\lambda}_{LB}$	$\hat{\phi}_{LB}$
N = 200	0.7	0.7	0.704053	0.694272	0.703608	0.694273
		0.9	0.712559	0.893031	0.711961	0.893031
	0.9	0.7	0.89567	0.692927	0.895587	0.692927
		0.9	0.897774	0.89222	0.897602	0.89222
N = 800	0.7	0.7	0.702591	0.699152	0.702557	0.699152
		0.9	0.69286	0.89726	0.692883	0.89726
	0.3	0.8	0.294824	0.795203	0.294818	0.795203
	0.8	0.3	0.794149	0.294536	0.794162	0.294536

Table 2. The MSE and distance from the true value.

	λ	ϕ	$\hat{\theta}_{MLEMSE}$	$\hat{\theta}_{LBMSE}$	$\|\hat{\theta}_{MLEMSE}-\hat{\theta}_{LBMSE}\|_2$
N = 200	0.7	0.7	0.007665	0.007364	0.001971974
		0.9	0.00576	0.005474	0.002007057
	0.9	0.7	0.005215	0.005035	0.001461555
		0.9	0.006006	0.005752	0.001724022
N = 800	0.7	0.7	0.00149	0.001474	0.000222702
		0.9	0.001428	0.001411	0.000225155
	0.3	0.8	0.001506	0.001492	0.000243135
	0.8	0.3	0.003181	0.003162	0.000233711

From Table 1, we can see that all the estimates are close to the true values regardless of the values of λ and ϕ. Importantly, our analysis reveals a key trend: as the sample size increases, all the estimates progressively converge towards their true values. This pattern is evident for both the maximum likelihood estimation and the linear Bayesian estimation, underscoring their consistency as estimators. The reliability of these approaches is evidenced in Table 2, which presents a noteworthy distinction in the mean square error (MSE) values between them. Specifically, the MSE for the linear Bayesian estimation is lower than that for maximum likelihood estimation. Remarkably, the Euclidean distance between $\hat{\theta}_{MLEMSE}$ and $\hat{\theta}_{LBMSE}$ diminishes as the sample size increases, suggesting that the advantage of linear Bayesian estimation becomes more pronounced in smaller sample sizes.

4.2. Empirical Application

In this section, we illustrate our proposed method using two sets of real data.

4.2.1. Empirical Application 1

The first dataset for our research originates from Ladislaus Bortkiewicz's acclaimed book '*Das Gesetz der Kleinen Zahlen*', published in 1898 [29]. This particular data set gained historical significance when Cohen introduced the misrecorded Poisson distribution model for the first time. It comprises a collection of statistical records detailing the annual fatalities among Prussian soldiers due to horse-related incidents from 1875 to 1884. The data encompass detailed records from 14 distinct regiments and one guard unit, providing a diverse and comprehensive dataset for the analysis. The nature of these incidents inherently suggests the suitability of a Poisson distribution for modeling the frequency and patterns of these rare events. In an effort to validate and corroborate the findings discussed in this paper, we undertook some modifications to the original dataset. Specifically, we adjusted 20 entries that were originally noted as having a value of 1, revising them to 0. This adjustment is carefully documented in Table 3, allowing for a more accurate analysis in line with our study objectives.

Table 3. Soldiers who were kicked to death by horses between 1875 and 1984.

Number per Deaths of Army Corps per Year	Number of Observations	
	Original Data	Altered Data
0	109	129
1	65	45
2	22	22
3	3	3
4	1	1
5	0	0

We now fit a misrecorded Poisson distribution to this modified dataset using both maximum likelihood estimation and linear Bayesian estimation. From Table 3, we have $N = 200$, $n_0 = 129$, $n_1 = 45$, $\overline{x} = 102/200 = 0.51$, $n_0/N = 0.645$, and $n_1/N = 0.225$. Substituting

these values into the maximum likelihood estimation, as shown in Equations (5) and (6), we have

$$\hat{\lambda}_{MLE} = \left[(\bar{x} - 1 + n_0/N) + \sqrt{(\bar{x} - 1 + n_0/N)^2 + 4(\bar{x} - n_1/N)}\right]/2$$

$$= \left[(0.51 - 1 + 0.645) + \sqrt{(0.51 - 1 + 0.645)^2 + 4 \times (0.51 - 0.225)}\right]/2$$

$$= 0.61695.$$

$$\hat{\phi}_{MLE} = \left(n_0 - n_1/\hat{\lambda}\right)/(n_0 + n_1)$$

$$= \left((129 - 45)/\hat{\lambda}\right)/(129 + 45)$$

$$= 0.322187.$$

From the unchanged raw data [29], we have that $\lambda = 0.610$ and $\varphi = 20/65 = 0.308$, which is the proportion of ones that were misclassified in the process of altering the original data for this illustration. The specific proof process can be found in Appendix A. Utilizing linear Bayesian estimation (13), we obtain $\hat{\lambda}_{LB} = 0.6167831$ and $\hat{\varphi}_{LB} = 0.3221758$. Subsequently, we can compute the MSE by comparing the true values of the two parameters with the combined estimations from above: $MSE(\hat{\theta}_{MLE}) = 0.0001247861$ and $MSE(\hat{\theta}_{LB}) = 0.0001235136$, so $MSE(\hat{\theta}_{LB}) \leq MSE(\hat{\theta}_{MLE})$. This indicates that the linear Bayesian estimation has better performance than the maximum likelihood estimation.

4.2.2. Empirical Application 2

The second dataset for our research, originating from Yang [30], is presented in Table 4. The frequent occurrence of accidents at intersections in the area increases the probability of traffic accidents due to the vehicle type, traffic volume, and linear congestion. People can analyze data to discover the changes and distribution patterns of traffic volume over time and space. These findings help develop effective traffic control measures and reduce accident rates under different traffic conditions.

Table 4. Traffic volume at intersection C.

The Number of Passes for Small-Sized Cars	Number of Observations	
	Original Data	Altered Data
0	15	18
1	5	2
2	9	9
3	6	6
4	7	7
5	1	1
6	2	2

The investigation was conducted at the T-shaped intersection of West Ring Road outside the University City and Xingguang Road under the University City in Guangzhou Province. The survey focuses on the traffic volume of motor vehicles at this intersection, which was collected through manual monitoring.

According to Table 4, we can calculate from the original data that $\lambda = 1.911$ and $\phi = 3/5 = 0.6$. Substituting these values into the maximum likelihood estimation, as shown in Equations (5) and (6), we have $\hat{\lambda}_{MLE} = 2.101127$ and $\hat{\phi}_{MLE} = 0.8524065$. Utilizing linear Bayesian estimation (13), we obtain $\hat{\lambda}_{LB} = 1.817069$ and $\hat{\phi}_{LB} = 0.852164$. Therefore, similarly to the previous example, we can obtain $MSE(\hat{\theta}_{MLE}) = 0.7400561$, $MSE(\hat{\theta}_{LB})$ =0.4422734, and $MSE(\hat{\theta}_{LB}) \leq MSE(\hat{\theta}_{MLE})$. This still indicates that the linear Bayesian estimation is superior to the maximum likelihood estimation.

5. Conclusions

This research study primarily focuses on an in-depth examination of the unique characteristics and distinct properties of linear Bayesian estimation, specifically applied to the misrecorded Poisson distribution. A key aspect of this study is providing robust empirical evidence that demonstrates the enhanced performance of linear Bayesian estimation when compared to maximum likelihood estimation, particularly within the framework of this distribution. Moreover, to reinforce these theoretical insights, the study employs a comprehensive approach, encompassing both rigorous validation through a series of detailed numerical simulations and the use of illustrative examples to further elucidate these findings. This multifaceted approach not only solidifies the theoretical underpinnings but also showcases the practical applicability of linear Bayesian estimation in real-world scenarios.

The method employed in this paper may not capture all dynamic characteristics when dealing with complex real-world situations. Due to the development of big data and advanced statistical methods and the need for more complex models in recent years, we consider extending the traditional framework of the Poisson distribution with the Hawkes process, thereby providing more effective tools and methodologies for solving practical problems.

Author Contributions: Writing—original draft, H.G.; Writing—review & editing, Z.C. and F.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of China (No. 12161072) and the Natural Science Foundation of Qinghai Province (No. 2019-ZJ-920).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

1. Proof process in Empirical Application 1

According to Definition 1, $\theta = (\lambda, \phi)'$, we define $MES(\hat{\theta}_{MLE}) = \frac{1}{2}((\hat{\lambda}_{MLE} - \lambda)^2 + (\hat{\phi}_{MLE} - \phi)^2)$ and $MES(\hat{\theta}_{LB}) = \frac{1}{2}((\hat{\lambda}_{LB} - \lambda)^2 + (\hat{\phi}_{\mathrm{LB}} - \phi)^2)$ to represent the mean square errors of the population for both parameters in the two estimation methods. The true values in this example are $\lambda = 0.610$ and $\phi = 0.308$, where $\lambda = \frac{\sum(x_i \times f_i)}{N}$. f_i represents the observed frequency of each event occurrence and ϕ is the probability of misclassification or the proportion of ones that are reported as zeros. The estimated values calculated based on the maximum likelihood estimation are $\hat{\lambda}_{MLE}$ =0.61695 and $\hat{\phi}_{MLE} = 0.322187$, and those obtained from the linear Bayesian estimation are $\hat{\lambda}_{LB} = 0.6167831$ and $\hat{\phi}_{LB} = 0.3221758$. Hence, we have two parameters with the combined estimations $MSE(\hat{\theta}_{MLE}) = 0.0001247861$ and $MSE(\hat{\theta}_{LB}) = 0.0001235136$. Therefore, it follows that $MSE(\hat{\theta}_{LB}) \leq MSE(\hat{\theta}_{MLE})$. Proof completed.

References

1. Cohen, J.R.; Clifford, A. Estimating the parameters of a modified Poisson distribution. *J. Am. Stat. Assoc.* **1960**, *55*, 139–143. [CrossRef]
2. Dorris, A.L.; Foote, B.L. Inspection errors and statistical quality control: A survey. *AIIE Trans.* **1978**, *10*, 184–192. [CrossRef]
3. Gupta, P.L.; Gupta, R.C.; Tripathi, R.C. Analysis of zero-adjusted count data. *Comput. Stat. Data Anal.* **1996**, *23*, 207–218. [CrossRef]
4. Zhang, C.; Tian, G.L.; Ng, K.W. Properties of the zero-and-one inflated Poisson distribution and likelihood-based inference methods. *Stat. Its Interface* **2016**, *9*, 11–32. [CrossRef]
5. Liu, Y.; Tian, G.L.; Tang, M.L. A new multivariate zero-adjusted Poisson model with applications to biomedicine. *Biom. J.* **2019**, *61*, 1340–1370. [CrossRef] [PubMed]
6. Bagui, S.C.; Mehra, K.L. The Poisson Distribution and Its Convergence to the Normal Distribution. *Int. J. Stat. Sci.* **2020**, *36*, 37–56.
7. Zhang, Q.; Lipani, A.; Kirnap, O.; Yilmaz, E. Self-Attentive Hawkes Process. In Proceedings of the 37th International Conference on Machine Learning, Virtual, 13–18 July 2020; Volume 119, pp. 11183–11193.

8. Wang, L.W.; Zhang, L. *Hawkes Processes for Understanding Heterogeneity in Information Propagation on Twitter*; School of Science, Beijing University of Posts and Telecommunications: Beijing, China, 2022.
9. Lamprinakou, S.; Gandy, A.; McCoy, E. Using a latent Hawkes process for epidemiological modelling. *PLoS ONE* **2023**, *18*, e0281370. [CrossRef] [PubMed]
10. Xu, J. Study on a Class of Mixed Poisson and Its Zero-Inflated Distribution. Master's Thesis, Chang'an University, Chang'an, China, 2022.
11. Johnson, N.L. *Univariate Discrete Distributions*; John Wiley and Sons: Hoboken, NJ, USA, 2004.
12. Djuraš, G. *Generalized Poisson Models for Word Length Frequencies in Texts of Slavic Languages*; Graz University of Technology Press: Graz, Austria, 2012.
13. Tuwei, K.E. *Power Series Distribution Sand Zero-Inflated Models*; University of Nairobi Press: Nairobi, Kenya, 2014.
14. Angers, J.F.; Biswas, A. A Bayesian analysis of zero-inflated generalized Poisson model. *Comput. Stat. Data Anal.* **2003**, *42*, 37–46. [CrossRef]
15. Rodrigues, J. Bayesian Analysis of Zero-Inflated Distributions. *Commun. Stat.—Theory Methods* **2003**, *32*, 281–289. [CrossRef]
16. Wang, Y.Q. Bayesian Estimation Based on Zero-and-One-Inflated Poisson Regression Model. Master's Thesis, Central China Normal University, Wuhan, China, 2022.
17. Hartigan, J.A. Linear Bayesian Methods. *J. R. Stat. Soc. Ser. B Methodol.* **1969**, *31*, 446–454. [CrossRef]
18. Rao, C.R. *Linear Statistical Inference and Its Applications*, 2nd ed.; John Wiley and Sons: Hoboken, NJ, USA, 1973.
19. LaMotte, L.R. Bayes linear estimators. *Technometrics* **1978**, *20*, 281–290. [CrossRef]
20. Hesselager, O. Rates of risk convergence of empirical linear Bayes estimators. *Scand. Actuar. J.* **1992**, *1*, 88–94. [CrossRef]
21. Goldstein, M. General Variance Modifications for Linear Bayes Estimators. *J. Am. Stat. Assoc.* **1983**, *78*, 616–618. [CrossRef]
22. Hoffmann, K. A subclass of Bayes linear estimators that are minimax. *Acta Appl. Math.* **1996**, *43*, 87–95. [CrossRef]
23. Samaniego, J.F.; Vestrup, V.E. On improving standard estimators via linear empirical Bayes methods. *Stat. Probab. Lett.* **1999**, *44*, 309–318. [CrossRef]
24. Wei, L.S.; Zhang, W.P. The superiorities of Bayes Linear Minimum Risk Estimation in Linear Model. *Commun. Stat. Theory Methods* **2007**, *36*, 917–926. [CrossRef]
25. Lin, P.P. Linear Bayesian Estimation under Constraint Conditions. Master's Thesis, Beijing Jiaotong University, Beijing, China, 2018.
26. Tao, R.F. Linear Bayesian Estimation of Parameters in Pareto Distribution. Master's Thesis, Jiangxi Normal University, Beijing, China, 2019.
27. Liu, X.H. Linear Bayes Estimators in Singular Linear Model. Master's Thesis, Beijing Jiaotong University, Beijing, China, 2022.
28. Chen, T. Linear Bayes Estimator of the Extreme Value Distribution Based on Type II Censored Samples. Master's Thesis, Beijing Jiaotong University, Beijing, China, 2021.
29. von Bortkiewicz, L. *Das Gesetz der Kleinen Zahlen*; Teubner Press: Leipzig, Germany, 1898.
30. Yang, C.Q. Investigation and Analysis of Traffic Flow in Urban Road Intersection. Master's Thesis, Guangzhou University, Guangzhou, China, 2018.