

Integrated Likelihood Inference in Poisson Distributions

Timothy Ruel

Department of Statistics and Data Science, Northwestern University

September 1, 2024

Abstract

The text of your abstract. 200 or fewer words.

Keywords: Directly standardized rate, Integrated likelihood ratio statistic, Maximum integrated likelihood estimator, Profile likelihood, Weighted sum, Zero score expectation parameter

1 Introduction

Consider a vector $\theta = (\theta_1, \dots, \theta_n)$ in which each component represents the mean of a distinct Poisson process. The purpose of this paper is to discuss the task of conducting likelihood-based inference for a real-valued parameter of interest $\psi = \tau(\theta)$, where $\theta \in \Theta \subset \mathbb{R}_+^m$ is unknown and $\tau : \Theta \rightarrow \Psi$ is a known twice continuously differentiable function. In particular, we will examine the utility of the integrated likelihood function as a tool for estimating ψ , using the performance of the more easily calculated profile likelihood as a benchmark.

We may obtain a sample of values from each Poisson process through repeated measurements of the number of events it generates over a fixed period of time. Suppose we have done so, and let X_{ij} represent the j th count from the i th sample, so that $X_{ij} \sim \text{Poisson}(\theta_i)$ for $i = 1, \dots, n$ and $j = 1, \dots, m_i$. The probability mass function (pmf) for a single observation $X_{ij} = x_{ij}$ is

$$p(x_{ij}; \theta_i) = \frac{e^{-\theta_i} \theta_i^{x_{ij}}}{x_{ij}!}, \quad x_{ij} = 0, 1, 2, \dots; \quad \theta_i > 0. \quad (1)$$

Denote the vector of counts from the i th process by $X_{i\bullet} = (X_{i1}, \dots, X_{im_i})$, its associated mean by $\bar{X}_{i\bullet} = \frac{1}{m_i} \sum_{j=1}^{m_i} X_{ij}$, and assume that all of the counts both within and between samples are measured independently. The likelihood function for an individual component θ_i based on the data $X_{i\bullet} = x_{i\bullet}$ is then equal to the product of the individual probabilities

of the observed counts. That is,

$$\begin{aligned}
L(\theta_i; x_{i\bullet}) &= \prod_{j=1}^{m_i} p(x_{ij}; \theta_i) \\
&= \prod_{j=1}^{m_i} \frac{e^{-\theta_i} \theta_i^{x_{ij}}}{x_{ij}!} \\
&= \left(\prod_{j=1}^{m_i} e^{-\theta_i} \right) \left(\prod_{j=1}^{m_i} \theta_i^{x_{ij}} \right) \left(\prod_{j=1}^{m_i} x_{ij}! \right)^{-1} \\
&= \left(e^{-\sum_{j=1}^{m_i} \theta_i} \right) \left(\theta_i^{\sum_{j=1}^{m_i} x_{ij}} \right) \left(\prod_{j=1}^{m_i} x_{ij}! \right)^{-1} \\
&= e^{-m_i \theta_i} \theta_i^{m_i \bar{x}_{i\bullet}} \left(\prod_{j=1}^{m_i} x_{ij}! \right)^{-1}.
\end{aligned} \tag{2}$$

We regard L as being a function of the parameter θ_i for fixed $x_{i\bullet}$. Since L is only useful to the extent that it informs our understanding of the value of θ_i , we are free to replace it with any other function differing from it by just a (nonzero) multiplicative term that is constant with respect to θ_i , provided that the result still satisfies the necessary regularity conditions, as this will not change any conclusions regarding θ_i that we draw from it. Hence, we may safely discard the term in parentheses on the final line of Equation 2 as it does not depend on θ_i and instead simply write

$$L(\theta_i; x_{i\bullet}) = e^{-m_i \theta_i} \theta_i^{m_i \bar{x}_{i\bullet}}. \tag{3}$$

The product of the likelihood functions for each component of θ then forms the basis of the likelihood function for θ itself:

$$\begin{aligned}
L(\theta; x_{1\bullet}, \dots, x_{n\bullet}) &= \prod_{i=1}^n L(\theta_i; x_{i\bullet}) \\
&= \prod_{i=1}^n e^{-m_i \theta_i} \theta_i^{m_i \bar{x}_{i\bullet}} \\
&= \left(\prod_{i=1}^n e^{-m_i \theta_i} \right) \left(\prod_{i=1}^n \theta_i^{m_i \bar{x}_{i\bullet}} \right) \\
&= e^{-\sum_{i=1}^n m_i \theta_i} \prod_{i=1}^n \theta_i^{m_i \bar{x}_{i\bullet}}.
\end{aligned} \tag{4}$$

It will be more convenient to work with the log-likelihood function, given by

$$\begin{aligned}
\ell(\theta) &= \log L(\theta) \\
&= \log \left(e^{-\sum_{i=1}^n m_i \theta_i} \prod_{i=1}^n \theta_i^{m_i \bar{x}_{i\bullet}} \right) \\
&= \log \left(e^{-\sum_{i=1}^n m_i \theta_i} \right) + \log \left(\prod_{i=1}^n \theta_i^{m_i \bar{x}_{i\bullet}} \right) \\
&= -\sum_{i=1}^n m_i \theta_i + \sum_{i=1}^n m_i \bar{x}_{i\bullet} \log \theta_i \\
&= \sum_{i=1}^n m_i (\bar{x}_{i\bullet} \log \theta_i - \theta_i).
\end{aligned} \tag{5}$$

When conducting research, it is common to encounter situations in which the question at hand would be better answered by knowing not the value of the full parameter of our statistical model but rather a real-valued function of it. That is, instead of estimating θ

scalar parameter ψ taking values in a set $\Psi \subset \mathbb{R}$ such that $\psi = g(\theta)$, where $g : \Theta \rightarrow \Psi$ has two continuous derivatives.

Note that while ℓ is a function of the n -dimensional θ , ψ is just a scalar. This decrease in dimension implies the existence of an $(n - 1)$ -dimensional nuisance parameter λ in the model that hinders or outright precludes inference regarding ψ and consequently must be eliminated from the log-likelihood function before proceeding. Furthermore, ψ need not explicitly be equal to one of the components of θ but instead may be defined as the output of any function g taking θ as input and satisfying the requirements mentioned above.

We refer to ψ and λ as being *implicit* parameters in such cases. In general, so it is rare in practice to encounter a situation in which a closed form expression for a nuisance parameter exists.

The standard procedure for eliminating λ from the log-likelihood function involves choosing some method with which to summarize $\ell(\theta)$ over its possible values while holding ψ fixed in place. This effectively reduces $\ell(\theta)$ to a simpler function depending on ψ alone,

having replaced each dimension of θ that depends on λ with a static summary of the values in its parameter space. We call this new function a pseudo-log-likelihood function for ψ and denote its generic form as $\ell(\psi)$. As we encounter specific types of pseudo-log-likelihoods, we will introduce more specialized notation as needed. Note that while it usually has properties resembling one, $\ell(\psi)$ is not itself considered a genuine log-likelihood function, and there will always be some degree of information contained within the data lost as a result of the nuisance parameter's elimination.

Perhaps the most straightforward method of summarization we can use to construct $\ell(\psi)$ is to maximize $\ell(\theta)$ over all possible values of θ for a fixed value of ψ . This yields what is known as the *profile* log-likelihood function, formally defined as

$$\ell_p(\psi) = \sup_{\theta \in \Theta: g(\theta) = \psi} \ell(\theta).$$

In the case where an explicit nuisance parameter exists, Equation 2 is equivalent to replacing λ with its conditional maximum likelihood estimate given ψ :

$$\ell_p(\psi) = \ell(\psi, \hat{\lambda}_\psi).$$

2 Integrated Likelihood Functions

3 Application to Poisson Models

We now turn our attention to the task of using the ZSE parameterization to construct an integrated likelihood that can be used to make inferences regarding a parameter of interest derived from the Poisson model described in the introduction. We will

4 Inference for the Weighted Sum of Poisson Means

Consider the weighted sum

$$Y = \sum_{i=1}^n w_i X_i,$$

where each w_i is a known constant greater than zero. Suppose we take for our parameter of interest the expected value of this weighted sum, so that

$$\psi \equiv E(Y) = \sum_{i=1}^n w_i \theta_i.$$

The maximum likelihood estimate (MLE) for θ_i is simply the sample mean $\hat{\theta}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} X_{ij}$.

Examples