Statistics and Probability Letters 114 (2016) 86–92

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Note on the singularity of the Poisson–gamma model

Gintautas Jakimauskas [a,*], Leonidas Sakalauskas [a,b]

[a] Vilnius University Institute of Mathematics and Informatics, Akademijos str. 4, Vilnius, LT-08663, Lithuania
[b] Vilnius Gediminas Technical University, Saulėtekio al. 11, Vilnius LT-10223, Lithuania

ARTICLE INFO

Article history:
Received 30 November 2015
Accepted 18 March 2016
Available online 25 March 2016

Keywords:
Empirical Bayesian estimation
Poisson–gamma model

ABSTRACT

The problem of singularity of the Poisson–gamma model is considered. As an addition to the known results, we present the property of the log-likelihood function which implies the singularity of this model. An example of the singularity issue is presented.

1. Introduction

Let us consider the problem of singularity in the Poisson–gamma model of count data applying the maximum likelihood method to obtain empirical Bayes estimates of unknown parameters. Reference to empirical Bayes methods for data analysis could be found in, e.g., Carlin and Louis (1996). A comprehensive list of methods for count data is presented, e.g., in Cameron and Trivedi (1998), Cameron and Trivedi (2005), Cameron and Trivedi (2009) and Hilbe (2011). General linear models are considered, e.g., in McCullagh and Nelder (1983).

The singularity, in this context, means that, in the process of estimating parameters of the selected model, variance of the mixing distribution (in this case, the gamma distribution) converges to zero. Sometimes this problem arises in probability estimation of rare events in large populations if event count estimates, based on the relative risk estimates, are too close to the observed event count values. In such a case the maximum likelihood estimates of unknown probabilities give zero variance of the mixing component. In practice, iterative optimization procedures, used for finding maximum likelihood estimates, converge to infinity values of distribution parameters, so we need to stop the procedures at some large values of the parameters and set the maximum likelihood estimates of unknown probabilities all equal to the mean relative risk estimate.

The solution of this problem (the existence of maximum of the likelihood function, considering symmetrical compound multinomial distribution, negative binomial distribution, and the Poisson mixture) is given in Levin and Reeds (1977).

It is worth mentioning, that the condition of non-singularity (i.e. condition of the existence of maximum of the likelihood function for finite parameter values) is not well known, at least in practical use. Mostly, rather a theoretical setting of the problem is used, so it is hard to find a connection with a practical setting of the problem. Recently, the existence of maximum likelihood estimators in Poisson–gamma hierarchical generalized linear models (HGLM) and negative binomial regression models has been studied in Gning and Pierre-Loti-Viaud (2013). An issue of singularity of the Poisson–Gaussian model was considered in detail in Sakalauskas (2010a,b), that implemented a non-singularity condition (which appears to

* Corresponding author.
E-mail addresses: gintautas.jakimauskas@mii.vu.lt (G. Jakimauskas), leonidas.sakalauskas@mii.vu.lt (L. Sakalauskas).

http://dx.doi.org/10.1016/j.spl.2016.03.007
0167-7152/

be the same as for the Poisson–gamma model) to the observed event counts, given the population sizes. In this note, we use the practical setting of the problem and present a property of the log-likelihood function that gives a clear explanation why the singularity of estimating parameters of the empirical Bayes Poisson–gamma model occurs. An example of the singularity issue is presented as well.

2. Main results

Let us have K populations consisting of $N_j, j = 1, \dots, K$, individuals, and some event (e.g., some disease) can occur in these populations. We observe the number of events $Y_j, j = 1, \dots, K$. We assume that for $j = 1, \dots, K$, each number of events Y_j is caused by an unknown probability λ_j, and these probabilities are equal for each individual from the same population. Moreover, assume that all the events in all the populations are independent. Under the selected mathematical model, $Y_j, j = 1, \dots, K$, are a sample of independent random variables $\mathbf{Y}_j, j = 1, \dots, K$, with a binomial distribution, respectively, with the number of experiments $N_j, j = 1, \dots, K$, and success probabilities $\lambda_j, j = 1, \dots, K$. Clearly, $\mathbf{E}(\mathbf{Y}_j) = \lambda_j N_j, j = 1, \dots, K$. Note that the corresponding variance is $\mathbf{E}(\mathbf{Y}_j - \mathbf{E}(\mathbf{Y}_j))^2 = \lambda_j(1 - \lambda_j)N_j, j = 1, \dots, K$.

Considering small probabilities $\lambda_j, j = 1, \dots, K$, an assumption is often made (see, e.g., Clayton and Kaldor, 1987; Tsutakava et al., 1985) that random variables $\mathbf{Y}_j, j = 1, \dots, K$, have a Poisson distribution with the parameters $\lambda_j N_j, j = 1, \dots, K$, respectively, i.e.,

$$\mathbf{P}\{\mathbf{Y}_j = m\} = h(m, \lambda_j N_j), \quad m = 0, 1, \dots, \ j = 1, \dots, K,$$

where

$$h(m, z) = e^{-z}\frac{z^m}{m!}, \quad m = 0, 1, \dots, z > 0.$$

Again, $\mathbf{E}(\mathbf{Y}_j) = \lambda_j N_j, j = 1, \dots, K$. Additionally, following the property of the Poisson distribution, $\mathbf{E}(\mathbf{Y}_j - \mathbf{E}(\mathbf{Y}_j))^2 = \mathbf{E}(\mathbf{Y}_j) = \lambda_j N_j, j = 1, \dots, K$. It is important to note that this model (Poisson model) should be used only for small probabilities, because, for this approximation, there is a non-zero probability for event count to exceed the population size. It is possible to use the Poisson model only if $\mathbf{P}\{\mathbf{Y_j} > N_j\} \approx 0, j = 1, \dots, K$, so that these small probabilities might be ignored.

The Poisson model distinguishes by the so-called equidispersion (equality of mean and variance) property, which follows from the property that the sum of the independent Poisson random variables has a Poisson distribution with the parameter equal to the sum of parameters of the components. The Poisson model is very simple (in fact, the benchmark model for count data is the Poisson distribution), but, in practice, very rarely real-world data have such a property. The Poisson model is a variant of the standard negative binomial model, termed NB2 (see Hilbe, 2011). The latter model may be regarded as more general and more representative of the majority of count models. It is worth mentioning, that the initial binomial model does not have an exact equidispersion property, so this model must be considered in a different way (this model is not in the scope of this article).

Mostly, due to the unobserved heterogeneity, variance of the real-world data exceeds the mean (overdispersion). Only in some special cases we observe the opposite feature, i.e., underdispersion. The latter cases require special mathematical models that will not be discussed here. In the case of overdispersed data, there is a simple method to add the additional variance to the model. We can assume that unknown probabilities are independent random variables with non-zero variance (as a mixing distribution). Variance of the compound distribution will be greater, and it may be adjusted by selecting the appropriate distribution parameters. Perhaps the most popular choice for the mixing distribution is a gamma distribution.

Empirical Bayes methods (Carlin and Louis, 1996) are widely used to represent the precision of the small-area parameters through their estimated posterior distributions, replacing the unknown variance component and other parameters by their maximum likelihood estimates. In the empirical Bayesian estimation the probabilities of events in populations are assumed random and have some certain distribution. It is well known (see, e.g., Clayton and Kaldor, 1987) that Bayesian estimates of unknown probabilities have a substantially smaller mean square error as compared with the mean square error of simple relative risk estimates.

We assume that unknown probabilities $\lambda_j, j = 1, \dots, K$, are independent identically distributed gamma random variables with the shape parameter $\nu > 0$ and the scale parameter $\alpha > 0$, i.e., the distribution function F has the distribution density

$$f(x) = f(x; \nu, \alpha) = \frac{\alpha^\nu x^{\nu-1}}{\Gamma(\nu)}e^{-\alpha x}, \quad 0 \le x < \infty.$$

Then $\mathbf{E}(\lambda_j) = \nu/\alpha$, and $\mathbf{E}(\lambda_j - \mathbf{E}(\lambda_j))^2 = \nu/\alpha^2, \ j = 1, \dots, K$. Given the observed number of events $Y_j, j = 1, \dots, K$, and population sizes $N_j, j = 1, \dots, K$, Bayes estimates for $\lambda_j, j = 1, \dots, K$, are (see, e.g., Clayton and Kaldor, 1987)

$$\mathbf{E}(\lambda_j \mid \mathbf{Y}_j = Y_j) = \frac{Y_j + \nu}{N_j + \alpha}, \quad j = 1, \dots, K. \tag{1}$$