ON THE NON-SINGULARITY ISSUE FOR THE POISSON-GAMMA MODEL

Gintautas Jakimauskas
Vilnius University Institute of Mathematics and Informatics

Introduction

Let us consider the problem of the estimation of small probabilities in large populations (e.g., the estimation of probability of some disease, death, suicides, etc.). The number of corresponding events depends on the size of the population and the probability of the single event. It is assumed that the number of events in populations have a Poisson distribution with certain parameters.

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, Caldor, 1987), (Meza, 2003)) that Bayesian estimates of the unknown probabilities have a substantially smaller mean square error as compared with the mean square error of simple relative risk estimates.

Let us have two models of a distribution of unknown probabilities: the probabilities have a gamma distribution with the shape parameter $n > 0$ and scale parameter $a > 0$ (Poisson-gamma model), or logits of the probabilities have a Gaussian distribution with mean m and variance $s^2 > 0$ (Poisson-Gaussian model). In the case of the Poisson-Gaussian model it is known (see (Sakalauskas, 1995)) that if a certain non-singularity condition does not hold then the empirical Bayes estimates of unknown probabilities are equal to mean relative risk estimates, and corresponding distribution of logits of the probabilities have singular distribution with zero variance. In such a case, in practice it means that the distribution variance given by iterative procedures for finding distribution parameters converges to zero and we do not obtain parameters with finite values of both mean and variance.

In case of the Poisson-gamma model (see (Jakimauskas, 2012)) we have a similar non-singularity issue. In practice it means that shape and scale parameters given by iterative procedures for finding distribution parameters converge to infinity and we do not obtain finite values of shape and scale parameters.

Non-singularity conditions for both models depend only on population sizes and the number of observed events. We will consider the Poisson-gamma model for some sets of data and we will show the behaviour of iterative procedures in various situations. We will focus on the behaviour of a partial derivative by n of maximum likelihood function, which is essential for the non-singularity condition for the Poisson-gamma model.

Mathematical models

Let us have K populations $A_1, A_2, \ldots, A_K$, consisting of N_j individuals, resp., and some event (e.g., death or some disease), which can occur in these populations. We observe the number of events $\{Y_j\} = Y_j$, $j = 1, 2, \ldots, K$.

We assume that a number of events are caused by unknown probabilities $\{l_j\} = l_j$, $j = 1, 2, \ldots, K$, which are equal for each individual from the same population. Then the number of events $\{Y_j\}$ are a sample of independent random variables (r.v.'s) $\{\mathbf{Y}_j\} = \mathbf{Y}_j$, $j = 1, 2, \ldots, K$, with binomial distribution (resp., with parameters (l_j, N_j), $j = 1, 2, \ldots, K$). Clearly,

$$\mathbf{E}\mathbf{Y}_j = l_j N_j, \quad j = 1, 2, \ldots, K. \tag{1}$$

An assumption is often made (see, e.g., (Tsutakawa *et al.*, 1985), (Clayton, Caldor, 1987)) that r.v.'s $\{\mathbf{Y}_j\}$ have a Poisson distribution with parameters $l_j N_j$, $j = 1, 2, \ldots, K$,

$$\mathbf{P}\{\mathbf{Y}_j = m\} = h(m, \lambda_j N_j),\ m = 0, 1, \ldots;\ j = 1, 2, \ldots, K,$$

where $h(m,z) = e^{-z}\dfrac{z^m}{m!}$, $m = 0, 1, \ldots$, $z > 0$,

Under such an assumption we also have (1).

We will consider the mathematical model assuming that unknown probabilities $\{l_j\}$ are independent identically distributed (i.i.d.) r.v.'s with a distribution function F from the certain class F. Our problem is to get estimates of unknown probabilities $\{\hat{\lambda}_j\}$ from the observed number of events $\{Y_j\}$, assuming that $F \in$ F.

Regardless of the distribution of $\{l_j\}$ we can use a mean relative risk estimate $\overline{\lambda}^{MRR} = \dfrac{\sum_{j=1}^{K} Y_j}{\sum_{j=1}^{K} N_j}$, so we assume that $\{\overline{\lambda}_j^{MRR}\} \equiv \overline{\lambda}^{MRR}$, $j = 1, 2, \ldots, K$. Also we can use relative risk estimates $\{\overline{\lambda}_j^{RR}\} = \overline{\lambda}_j^{RR}$, $j = 1, 2, \ldots, K$, where

$$\overline{\lambda}_j^{RR} = \frac{Y_j}{N_j},\ j = 1, 2, \ldots, K. \tag{2}$$

Poisson-gamma model. Let us make an assumption that $\{l_j\}$ are i.i.d. gamma r.v.'s with a shape parameter $n > 0$ and scale parameter $a > 0$, i.e. d.f. F has a distribution density

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

Then $\mathbf{E}l_j = n / a$, and $\mathbf{D}l_j = n / a^2$. Moreover,

$$\mathbf{E}(\lambda_j \mid \mathbf{Y}_j = Y_j) = \frac{Y_i + \nu}{N_j + \alpha},\ j = 1, 2, \ldots, K. \tag{3}$$

Empirical Bayes estimate $\{\hat{\lambda}_j\}$, which is a certain compromise between the mean relative risk estimate $\{\overline{\lambda}_j^{MRR}\}$ and relative risk estimate $\{\overline{\lambda}_j^{R}\}$, is obtained by (3) using parameter estimates $(\hat{\nu}, \hat{\alpha})$.

Poisson-Gaussian model. We will consider the Bayes estimate $\{\tilde{\lambda}_j\}$, which is obtained under assumption that unknown probabilities are i.i.d. r.v.'s such that their logits

$$\alpha_j = \ln \frac{\lambda_j}{1-\lambda_j},\ j = 1, 2,\dots, K,$$

are i.i.d. Gaussian r.v.'s with mean *m* and variance s^2. In this case the conditional expectation of $\{l_j\}$ has the following form (see (Sakalauskas, 1995), (Gurevičius *et al.,* 2009)):

$$\mathbf{E}(\lambda_j \mid \mathbf{Y}_j = Y_j) = \frac{\int_{-\infty}^{\infty} \frac{1}{1+e^{-x}} h(Y_j, \frac{N_j}{1+e^{-x}}) \varphi(x; \mu, \sigma^2) dx}{D_j(\mu,\sigma^2)}, \quad (4)$$

$$D_j(\mu,\sigma^2) = \int_{-\infty}^{\infty} h(Y_j, \frac{N_j}{1+e^{-x}}) \varphi(x; \mu, \sigma^2) dx.$$

Empirical Bayes estimate $\{\tilde{\lambda}_j\}$, is obtained by (4) using parameter estimates $(\tilde{\mu}, \tilde{\sigma}^2)$. Note, that the formula (4) can be calculated using Hermite polynomials (see (Abramovich, Stegun, (1968)).

Considering the Poisson-gamma model, the corresponding maximum likelihood function has the following form:

$$L(\nu,\alpha) = \sum_{j=1}^{K} \Bigg(\ln \frac{\Gamma(Y_j+\nu)}{\Gamma(\nu)} + \nu \ln(\alpha) - $$
$$- (Y_j+\nu)\ln(N_j+\alpha) + Y_j \ln N_j \Bigg) \quad (5)$$

Considering the Poisson-Gaussian model, the corresponding maximum likelihood function has the following form:

$$L_{PG}(\mu,\sigma^2) = \sum_{j=1}^{K} \left(\ln D_j(\mu,\sigma^2) \right) \quad (6)$$

Maximum likelihood estimates are obtained by maximizing (5), resp., (6) and replacing corresponding parameter values in (3) or (4). In practice, approximate estimates $\{\hat{\lambda}_j\}$ and $\{\tilde{\lambda}_j\}$ are obtained using numerical methods (usually iterative procedures) for finding approximate parameter values $(\hat{\nu}, \hat{\alpha})$, resp., $(\tilde{\mu}, \tilde{\sigma}^2)$. Non-singularity conditions for considered models imply that iterative procedures will give finite and non-zero values for estimated parameters. Denote

$$P = \overline{\lambda}^{MRR} = \frac{\sum_{j=1}^{K} Y_j}{\sum_{j=1}^{K} N_j},$$

$Q = \sum_{j=1}^{K} Y_j - \sum_{j=1}^{K} (Y_j^2 - (N_j \cdot P)^2)$. For Poisson-gamma model the non-singularity condition has the following form: $Q < 0$.

Denote

$$Q_{PG} = \sum_{j=1}^{K} Y_j - \sum_{j=1}^{K} (Y_j - N_j \cdot P)^2.$$

Similarly, for the Poisson-Gaussian model the non-singularity condition has the following form (see (Sakalauskas, 1995)): $Q_{PG} < 0$.

Simulation results

An iterative procedure for obtaining maximum likelihood estimates for the Poisson-gamma model is given in (Clayton, Caldor, 1987). Denote

$$\hat{\theta}_j = \frac{Y_j + \nu}{N_j + \alpha},\ j = 1,2,\dots,K, \quad (8)$$

and two equations:

$$\frac{\nu}{\alpha} = \frac{1}{K} \sum_{j=1}^{K} \hat{\theta}_j, \quad (9)$$

$$\frac{\nu}{\alpha^2} = \frac{1}{K-1} \sum_{j=1}^{K} (1 + \frac{\alpha}{N_j})(\hat{\theta}_j - \frac{\nu}{\alpha})^2. \quad (10)$$

From (9) and (10) we obtain *a* as root of the quadratic equation and then we obtain *n* using (10). The iterative procedure starts from $\{\hat{\theta}_j\}_0 = \{\lambda_j^{RR}\}$, then from (9) and (10) we obtain $(n\ , a\ )_0$, from (8) we obtain $\{\hat{\theta}_j\}_1$, etc.

Partial derivatives of a maximum likelihood function for the Poisson-gamma model (5) have the following form:

$$\frac{\partial L(\nu,\alpha)}{\partial \alpha} = K\frac{\nu}{\alpha} - \sum_{j=1}^{K} \frac{Y_j+\nu}{N_j+\alpha} = K\left(\frac{\nu}{\alpha} - \frac{1}{K}\sum_{j=1}^{K} \theta_j \right), \quad (11)$$

$$\frac{\partial L(\nu,\alpha)}{\partial \nu} = \sum_{j=1}^{K} \sum_{s=0}^{Y_j-1} \frac{1}{\nu+s} + K \ln\alpha - \sum_{j=1}^{K} \ln(N_j+\alpha). \quad (12)$$

The non-singularity condition of the Poisson-gamma model is based on the following formula:

$$\lim_{\substack{\nu,\alpha\to\infty \\ \nu/\alpha = P\cdot(1+o(1/\nu))}} 2\nu^2 \frac{\partial L(\nu,\alpha)}{\partial \nu} = Q.$$

Because for sufficiently small values of *n* and *a,* the partial derivative (12) is strictly positive, if the non-singularity condition (7) holds, i.e., if *Q* is strictly negative, then there exist finite values of *n* and *a*, for which the partial derivatives (11) and (12) are equal to zero.

We will focus on the behaviour of the partial derivative (12) for various real or simulated data in cases when the non-singularity condition (7) holds, or when this condition does not hold. In the case when the non-singularity condition (7) holds, it is denoted by $(\hat{\nu}, \hat{\alpha})$ parameters found by the iterative procedure (9)–(10). In the case when the condition (7) does not hold, parameters are defined using empirical mean and variance of relative risk estimates (2). Denote functions

$$g(x) = \frac{2\nu^2}{|Q|} \cdot \frac{\partial L(\nu,\alpha)}{\partial \nu}\Bigg|_{\substack{\nu=\hat{\nu}x, \\ \alpha=\nu/P}},$$

$$h(x) = \frac{\partial L(\nu,\alpha)}{\partial \nu}\Bigg|_{\substack{\nu=\hat{\nu}x, \\ \alpha=\nu/P}}.$$

Let us consider a certain artificial set of data (dataset 1), which illustrates the behaviour of a partial derivative (12). Set number of populations $K = 60$, population sizes $N_j = 50000$, $j = 1, 2,\dots, K$, and number of events $Y_j = (350, 150, 310, 190, 400, 100, 250, 250,\dots,$

250). For dataset 1 we have $Q = Q_{PG} = -57200$. For comparison we will use dataset 2, with the same number of populations, but with population sizes $N_j = 5000$, $j = 1, 2,\ldots, K$, and the number of events $Y_j = (35, 15, 31, 19, 40, 10, 25, 25,\ldots, 25)$. For dataset 1 we have $Q = Q_{PG} = 778$. The simulation results show that the iterative algorithm (8)-(10) for dataset 1 gives values for $(\hat{\nu}, \hat{\alpha})$ after just few iterations. The same algorithm for dataset 2 does not converge and stops after reaching values about 10^{19}. Fig. 1–4 illustrates the behaviour of partial derivatives of corresponding maximum likelihood functions.

The given examples show that the non-singularity condition depends not only on relative risks, but it significantly depends on population sizes.

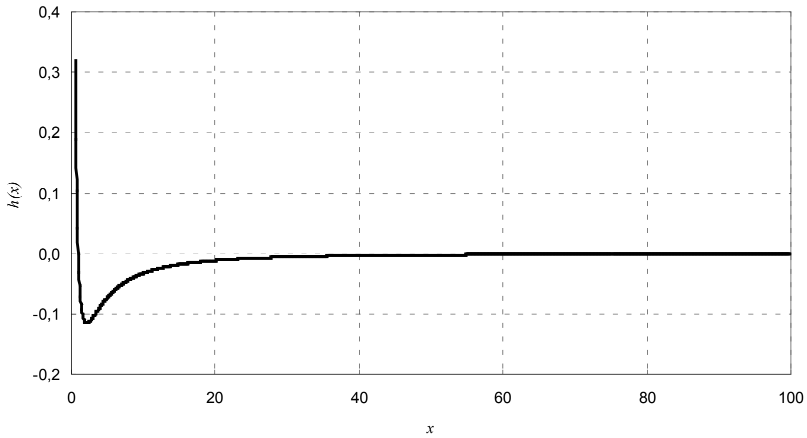

Fig.1. Function $h(x)$ for dataset 1.

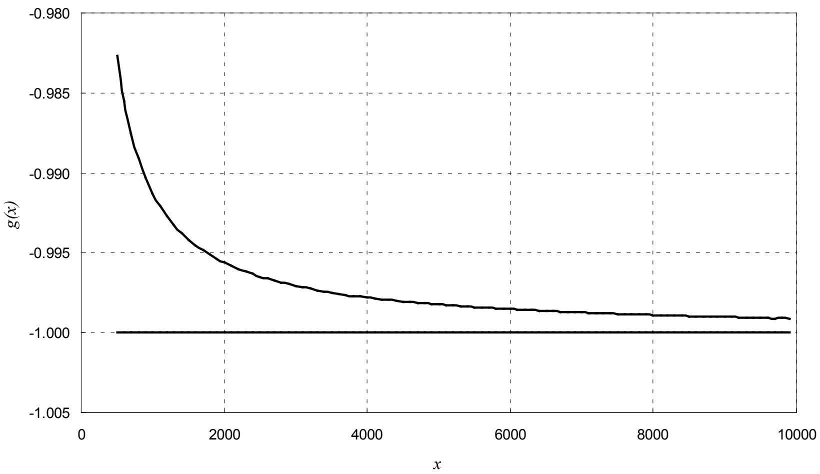

Fig.2. Function $g(x)$ for dataset 1.

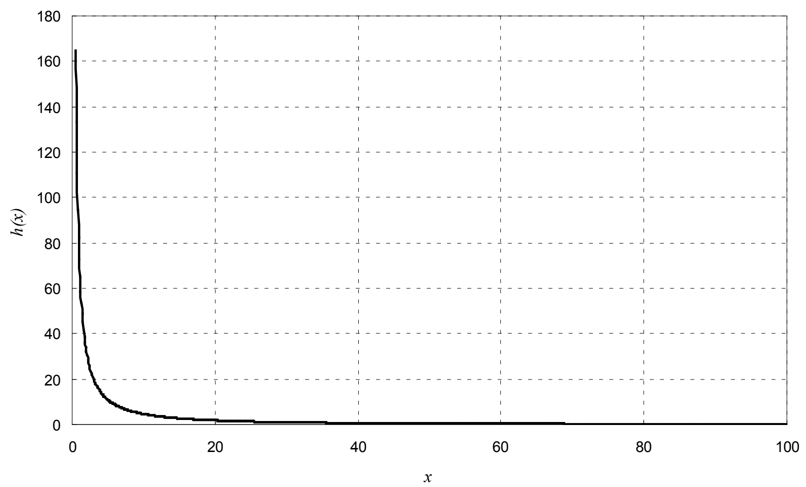

Fig.3. Function $h(x)$ for dataset 2.

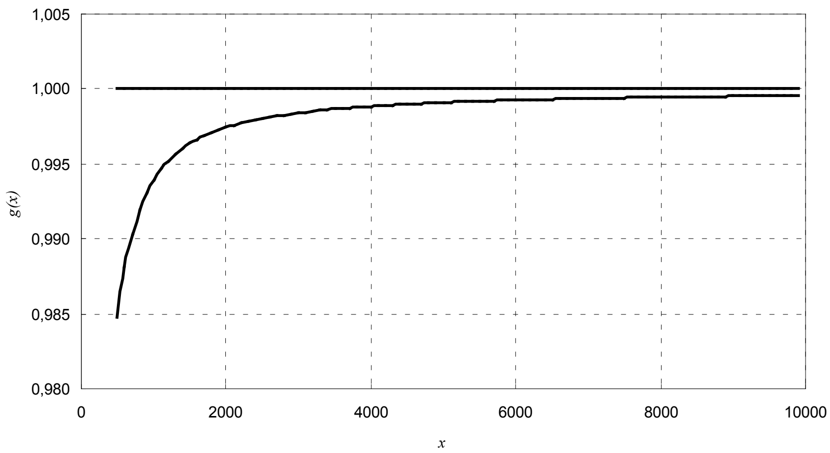

Fig.4. Function $g(x)$ for dataset 2.

References

1. Abramovich M., Stegun I. A., 1968, *Handbook of Mathematical Functions*. New York: Dover.
2. Clayton D., Kaldor J., 1987, Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*. Vol. 43. No. 3. P. 671–681.
3. Gurevičius R., Jakimauskas G., Sakalauskas L., 2009, Empirical Bayesian estimation of small mortality rates. *5th international Vilnius conference [and] EURO-mini conference „Knowledge-based technologies and OR methodologies for decisions of sustainable development (KORSD-2009)*. P. 290–295. Vilnius: Technika.
4. Jakimauskas G., 2012, Gamma and logit models in empirical Bayesian estimation of probabilities of rare events, *STOPROG 2012: Stochastic programming for implementation and advanced applications: proceedings of international workshop, July 3-6, 2012, Lithuania*. P. 43–48. Vilnius: Technika.
5. Meza J. L., 2003, Empirical Bayes estimation smoothing of relative risks in disease mapping. *Journal of Statistical Planning and Inference*. Vol. 112. P. 43–62.
6. Sakalauskas L., 1995, On Bayes analysis of small rates in medicine. *Proc. of the Internat. Conf. "Computer Data Analysis and Modeling", 1995, September 14-19, Minsk*. Vol. 1. P. 127–130. Minsk: Publishing centre BSU.
7. Tsutakava R. K., Shoop G. L., Marienfield C. J., 1985, Empirical Bayes estimation of cancer mortality rates, *Statistics in Medicine*. Vol. 4. No. 2. P. 201–212.

APIE NESINGULIARUMO SĄLYGĄ TAIKANT PUASONO-GAMA MODELĮ

Gintautas Jakimauskas

Santrauka

Straipsnyje nagrinėjama mažų tikimybių didelėse populiacijose (pvz., tam tikros ligos, mirčių, savižudybių ir t. t. tikimybių) vertinimo problema, darant prielaidą, kad įvykių skaičius turi Puasono skirstinį su tam tikrais parametrais, ir taikant nežinomų tikimybių gama skirstinio modelį.

Aptariami du nežinomų tikimybių pasiskirstymo modeliai: kai tikimybės pasiskirsčiusios pagal gama skirstinį (Puasono-gama modelis) ir tikimybių logitai pasiskirstę pagal Gauso modelį (Puasono-Gauso modelis). Kalbant apie pirmąjį modelį, žinoma, kad jei nėra išpildyta tam tikra nesinguliarumo sąlyga, nežinomų tikimybių empiriniai Bajeso įverčiai sutampa su vidutinės santykinės rizikos įverčiais, o atitinkamas logitų pasiskirstymas turi išsigimusį skirstinį su nuline dispersija. Puasono-gama modelyje situacija panaši: gama skirstinio parametrai, gaunami iteraciniais procesais, konverguoja į begalybę ir negaunamos baigtinių gama skirstinio parametrų reikšmės.

Nesinguliarumo sąlyga priklauso tik nuo populiacijų dydžių ir stebėtų įvykių skaičiaus. Nagrinėjamas Puasono-gama modelio taikymas atskiriems duomenų rinkiniams ir apibūdinamas iteracinių procesų veikimas. Aptariamas didžiausio tikėtinumo funkcijos dalinių išvestinių veikimas yra esminis Puasono-gama modelio nesinguliarumo sąlygai.

Prasminiai žodžiai: empirinis Bajeso vertinimas, Puasono-gama modelis.

ON THE NON-SINGULARITY ISSUE FOR THE POISSON-GAMMA MODEL

Gintautas Jakimauskas

Summary

The problem of estimation of small probabilities in large populations (e.g., the estimation of probability of some disease, death, suicides, etc.) is considered. The number of corresponding events depends on the size of the population and the probability of the single event. It is assumed that the number of events in populations has a Poisson distribution with certain parameters.

Let us have two models of distribution of unknown probabilities: the probabilities have a gamma distribution (Poisson-gamma model), or logits of the probabilities have a Gaussian distribution (Poisson-Gaussian model). In the case of the Poisson-Gaussian model it is known that if a certain non-singularity condition does not hold then empirical Bayes estimates of unknown probabilities are equal to mean relative risk estimates, and corresponding distribution of logits of the probabilities have singular distribution with zero variance. In the case of the Poisson-gamma model we have a similar non-singularity issue. In practice it means that the shape and scale parameters given by iterative procedures for finding distribution parameters converge to infinity and we do not obtain finite values of shape and scale parameters.

The non-singularity condition depends only on population sizes and the number of observed events. We will consider the Poisson-gamma model for some sets of data and we will show the behaviour of iterative procedures. We will focus on the behaviour of a partial derivative of the maximum likelihood function, which is essential for the non-singularity condition for the Poisson-gamma model.

Keywords: empirical Bayesian estimation, Poisson-gamma model.

Įteikta 2014-01-15