数理解析研究所講究録
1506巻 2006年 167-176

Estimating a common slope of multiple strata in the Tweedie distribution using a conjugate prior

統計数理研究所　大西 俊郎 (Toshio Ohnishi)
The Institute of Statistical Mathematics

Abstract

The logarithmic link regression model with multiple strata based on the Tweedie distribution is investigated. Assuming a conjugate prior density only on the intercept parameters, we derive the optimum estimating function of the common slope parameter and discuss the conjugate analysis as to the intercept parameter vector. An interesting relation between the optimum estimating function and the optimum estimator of the intercept is observed.

Key Words: common slope, conjugate analysis, estimating function, generalized linear model, logarithmic link, Pythagorean relationship, standardized posterior mode, Tweedie distribution

1. Introduction

Let us consider an exponential family whose variance function is proportional to a power of the mean parameter. The exponential family is called the Tweedie distribution in Jørgensen (1997). For details see Chapter 4 of his monograph. Its density function is expressed as

$$p(x;\,\mu,\,\tau_0) = \exp\left[\tau_0\{c(\mu)x - M(\mu)\}\right] a(x;\,\tau_0) \quad (x \geq 0,\,\mu \in \mathbb{R}^+), \tag{1.1}$$

and the variance function is of the form

$$\frac{1}{\tau_0 c'(\mu)} = \frac{\mu^\xi}{\tau_0}, \tag{1.2}$$

where μ is the mean parameter, $\tau_0 c(\mu)$ is the canonical parameter as a function of μ, $\tau_0 M(\mu)$ is the cumulant as a function of μ, and $a(x;\,\tau_0)$ is the supporting measure. Here $\tau_0 > 0$ and $\xi \in [1,\,2]$ are assumed to be known.

When $\xi = 1$ with $\tau_0 = 1$, the density (1.1) is the probability function of the Poisson distribution $\mathrm{Po}(\mu)$ with mean μ. In the case of $\xi = 2$ the density (1.1) is the probability density function of the gamma distribution $\mathrm{Ga}(\mu,\,\tau_0)$ with density

$$\frac{\tau_0^\lambda}{\Gamma(\tau_0)} \frac{x^{\tau_0 - 1}}{\mu^{\tau_0}} \exp\left(-\tau_0 \frac{x}{\mu}\right).$$

When $1 < \xi < 2$, the density (1.1) is that of the compound Poisson distribution which $\sum_{i=1}^N Y_i$ follows where $N, Y_1, Y_2, \ldots$ are mutually independent, and N and Y_i are distributed according to $\mathrm{Po}(m)$ and $\mathrm{Ga}(\theta,\,\lambda)$, respectively, with

$$m = \tau_0 \frac{\mu^{2-\xi}}{2-\xi},\quad \theta = (2-\xi)\mu^{\xi-1},\quad \lambda = \frac{2-\xi}{\xi-1}.$$

It should be noted that the Tweedie distribution with $1 \leq \xi < 2$ has a positive probability in zero:

$$\Pr(x = 0) = \exp\left(-\tau_0 \frac{\mu^{2-\xi}}{2-\xi}\right).$$

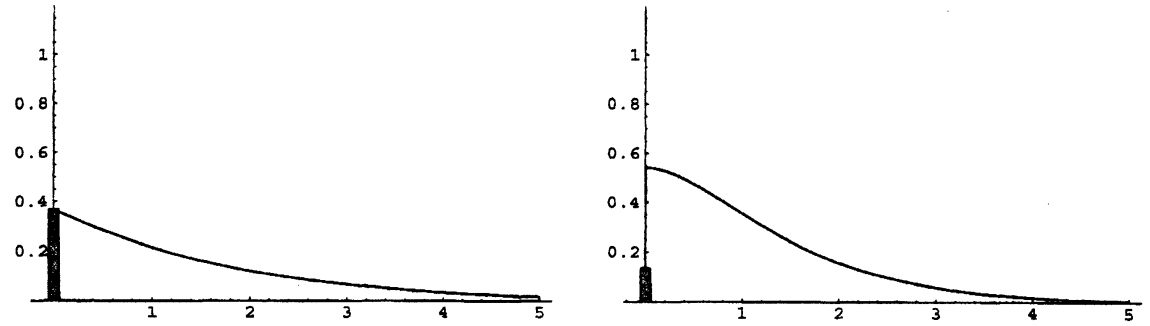
The Tweedie distribution with $1 < \xi < 2$ has a probability density function for $x > 0$ which is given by

$$p(x;\, \mu,\, \tau_0) = \exp\left\{-\tau_0 \left(\frac{\mu^{\xi-1}}{\xi-1}x + \frac{\mu^{2-\xi}}{2-\xi}\right)\right\} a(x;\, \tau_0),$$

where

$$a(x;\, \tau_0) = \frac{1}{x}\sum_{n=1}^{\infty} \frac{\tau_0^n \left\{\frac{1}{2-\xi}\left(\frac{\tau_0 x}{\xi-1}\right)^{(2-\xi)/(\xi-1)}\right\}^n}{\Gamma\left(\frac{2-\xi}{\xi-1}n\right) n!}.$$

Figure 1 draws the graphs of the density function when $\xi = 3/2$. Each bar at the origin indicates the probability $\Pr(x = 0)$.



(a) $(\mu,\, \tau_0) = \left(1,\, \frac{1}{2}\right)$ (b) $(\mu,\, \tau_0) = (1,\, 1)$

Figure 1: The graphs of the Tweedie density function for $\xi = 3/2$.

The Tweedie distribution is applied to analyze the dataset including the following three examples: (1) Compound Poisson distributions have been used for modeling the loss paid in the actuarial science; (2) Shono (2006) applied the Tweedie distribution to the CPUE (catch per unit effort) standardization of by-catch data including many observations with zero-catch; (3) Dr. Peter Dunn states that April total rainfall in Melbourne can be modeled with a Tweedie distribution with $\xi = 1.58$ in his website:

`http://www.sci.usq.edu.au/staff/dunn/Datasets/tech-glms.html#Tweedie`

We investigate a logarithmic link regression model $p(\boldsymbol{x};\, \boldsymbol{\alpha},\, \beta) = \prod_{k=1}^{K} p(\boldsymbol{x}_k;\, \alpha_k,\, \beta)$, which is based on the Tweedie distribution (1.1), where

$$p(\boldsymbol{x}_k;\, \alpha_k,\, \beta) = \prod_{i=1}^{n_k} p(x_{ki};\, e^{\alpha_k + \beta z_{ki}},\, \tau_0),$$

In the above $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K)$ with $\boldsymbol{x}_k = (x_{k1}, \ldots, x_{kn_k})$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ is the incidental intercept parameter vector, β is the common slope parameter of interest, and z_{ki} is the covariate. Although an extension to the vector slope parameter is straightforward, we focus on the scalar case for simplicity.

The aim of this paper is to propose an estimation procedure of β through a unified discussion. First, the conjugate analysis as to $\boldsymbol{\alpha}$ is discussed. Secondly, the optimum estimating function of β is derived. Finally, an interesting relation between the optimum estimator of $\boldsymbol{\alpha}$ and the optimum estimating function of β is observed.

2. Hybrid Bayesian approach

This section introduces an approach to the estimation of $(\boldsymbol{\alpha}, \beta)$. This is a result of the joint work with Professor Takemi Yanagimoto. Only on $\boldsymbol{\alpha}$ we assume a prior density of the form $\pi(\boldsymbol{\alpha}; \alpha_0, \delta) = \prod \pi(\alpha_k; \alpha_0, \delta n_k)$ where

$$\pi(\alpha_k; \alpha_0, \delta n_k) \propto \exp\{-\delta n_k d(\alpha_0, \alpha_k)\} b(\alpha_k).$$

Here $d(s, t)$ is appropriately chosen function satisfying the condition $d(s, t) \geq 0$ and $d(s, t) = 0$ iff $s = t$, $b(\alpha_k)$ is a non-negative function, and the hyper-parameters α_0 and δ are assumed to be known. Note that no prior density is assumed on β. Thus this may be called a hybrid Bayesian approach.

We propose an estimation procedure which consists of the following two steps:

Step 1. Estimate β by the solution $\hat{\beta}$ to the estimating function

$$\mathrm{E}_{\mathrm{post}}\left[l_\beta(\boldsymbol{x}; \boldsymbol{\alpha}, \beta) \right] = 0, \tag{2.1}$$

where $\mathrm{E}_{\mathrm{post}}[\,\cdot\,]$ stands for the posterior expectation, and $l_\beta(\boldsymbol{x}; \boldsymbol{\alpha}, \beta)$ is the score function of β.

Step 2. Estimate α_k by $\hat{\alpha}_k(\hat{\beta})$ where

$$\hat{\alpha}_k(\beta) = \underset{\alpha_k}{\mathrm{Argmax}}\left[p(\boldsymbol{x}_k; \alpha_k, \beta) \exp\{-\delta n_k d(\alpha_0, \alpha_k)\} \right] \tag{2.2}$$

for $k = 1, \ldots, K$.

Another expression of the estimating function (2.1) in Step 1 is given in the following proposition.

Proposition 2.1. *It holds that*

$$\frac{\partial}{\partial \beta} \log p_{\mathrm{marg}}(\boldsymbol{x}; \beta) = \mathrm{E}_{\mathrm{post}}\left[l_\beta(\boldsymbol{x}; \boldsymbol{\alpha}, \beta) \right],$$

where $p_{\mathrm{marg}}(\boldsymbol{x}; \beta)$ is the marginal density.

Proof. Differentiating $p_{\mathrm{marg}}(\boldsymbol{x}; \beta) = \int p(\boldsymbol{x}; \boldsymbol{\alpha}, \beta)\, \pi(\boldsymbol{\alpha}; \alpha_0, \delta)\, d\boldsymbol{\alpha}$, we have

$$\frac{\partial}{\partial \beta} p_{\mathrm{marg}}(\boldsymbol{x}; \beta) = \int l_\beta(\boldsymbol{x}; \boldsymbol{\alpha}, \beta)\, p(\boldsymbol{x}; \boldsymbol{\alpha}, \beta)\, \pi(\boldsymbol{\alpha}; \alpha_0, \delta)\, d\boldsymbol{\alpha}.$$

Since the posterior density is $p(\boldsymbol{x};\, \boldsymbol{\alpha},\, \beta)\,\pi(\boldsymbol{\alpha};\, \alpha_0,\, \delta)/p_{\mathrm{marg}}(\boldsymbol{x};\, \beta)$, the required result is obtained. □

An optimality of the estimating function (2.1) is shown in the following theorem. Let $\mathrm{E}_p[\,\cdot\,]$ and $\mathrm{E}_\pi[\,\cdot\,]$ denote the expectations with respect to the sampling and the prior densities, respectively. The criterion function in the following theorem is an extended version of the one in Godambe and Kale (1991) to the Bayesian framework.

Theorem 2.2. *The posterior mean* $\mathrm{E}_{\mathrm{post}}\,[\,l_\beta(\boldsymbol{x};\, \boldsymbol{\alpha},\, \beta)\,]$ *of the score function of* β *is optimum with respect to the criterion function*

$$\mathcal{M}[\,g(\boldsymbol{x};\, \beta)\,] = \frac{\mathrm{E}_\pi\,[\,\mathrm{E}_p[\,g^2(\boldsymbol{x};\, \beta)\,]\,]}{\{\mathrm{E}_\pi\,[\,\mathrm{E}_p[\,g_\beta(\boldsymbol{x};\, \beta)\,]\,]\}^2} \tag{2.3}$$

among the estimating functions unbiased in the sense that

$$\mathrm{E}_\pi\,[\,\mathrm{E}_p[\,g(\boldsymbol{x};\, \beta)\,]\,] = 0. \tag{2.4}$$

Proof. The criterion function (2.3) and the unbiasedness condition (2.4) can be rewritten as

$$\mathcal{M}[\,g(\boldsymbol{x};\, \beta)\,] = \frac{\mathrm{E}_{\mathrm{marg}}[\,g^2(\boldsymbol{x};\, \beta)\,]}{\left\{\mathrm{E}_{\mathrm{marg}}\left[\,\frac{\partial}{\partial\beta} g(\boldsymbol{x};\, \beta)\,\right]\right\}^2}, \tag{2.5}$$

and

$$\mathrm{E}_{\mathrm{marg}}[\,g(\boldsymbol{x};\, \beta)\,] = 0, \tag{2.6}$$

respectively, where $\mathrm{E}_{\mathrm{marg}}[\,\cdot\,]$ denotes the marginal expectation. Set $g_{\mathrm{B}}(\boldsymbol{x};\, \beta) = \mathrm{E}_{\mathrm{post}}\,[\,l_\beta(\boldsymbol{x};\, \boldsymbol{\alpha},\, \beta)\,]$. It follows from Proposition 2.1 that $g_{\mathrm{B}}(\boldsymbol{x};\, \beta) = (\partial/\partial\beta)\log p_{\mathrm{marg}}(\boldsymbol{x};\, \beta)$. Differentiation of both sides of (2.6) gives

$$\mathrm{E}_{\mathrm{marg}}\left[\,g(\boldsymbol{x};\, \beta)\, g_{\mathrm{B}}(\boldsymbol{x};\, \beta) + \frac{\partial}{\partial\beta} g(\boldsymbol{x};\, \beta)\,\right] = 0. \tag{2.7}$$

It follows from Schwarz's inequality that

$$\{\mathrm{E}_{\mathrm{marg}}[\,g(\boldsymbol{x};\, \beta)\, g_{\mathrm{B}}(\boldsymbol{x};\, \beta)\,]\}^2 \le \mathrm{E}_{\mathrm{marg}}[\,g^2(\boldsymbol{x};\, \beta)\,]\,\mathrm{E}_{\mathrm{marg}}[\,g_{\mathrm{B}}^2(\boldsymbol{x};\, \beta)\,].$$

This, together with (2.5) and (2.7), yields that

$$\mathcal{M}[\,g(\boldsymbol{x};\, \beta)\,] \ge \frac{1}{\mathrm{E}_{\mathrm{marg}}[\,g_{\mathrm{B}}^2(\boldsymbol{x};\, \beta)\,]}. \tag{2.8}$$

Setting $g(\boldsymbol{x};\, \beta) = g_{\mathrm{B}}(\boldsymbol{x};\, \beta)$ in (2.7), we have

$$\mathrm{E}_{\mathrm{marg}}[\,g_{\mathrm{B}}^2(\boldsymbol{x};\, \beta)\,] = -\mathrm{E}_{\mathrm{marg}}\left[\,\frac{\partial}{\partial\beta} g_{\mathrm{B}}(\boldsymbol{x};\, \beta)\,\right]. \tag{2.9}$$

The inequality $\mathcal{M}[g(x;\beta)] \geq \mathcal{M}[g_B(x;\beta)]$ follows from (2.5), (2.8) and (2.9). □

Theorem 2.2 is valid in more general setting. Consider a sampling density $p(x;\theta,\psi)$ with a parameter θ of interest and an incidental parameter ψ. Assume a prior density only on ψ. Then, the posterior mean of the score function for θ is optimum. Thus a Bayesian analysis which assumes a prior density on the incidental parameter works well. Recall that the optimum estimating function is obtained as a combination of the Bayesian and the likelihood approaches.

Next, we derive an optimality of the estimator (2.2) of α_k for a known β in Step 2, which is associated with the optimality of the posterior mode.

Proposition 2.3. *Set $\gamma_k = B(\alpha_k)$ with $B(\cdot)$ being the primitive function of $b(\cdot)$. Then, $\hat{\gamma}_k(\beta) = B(\hat{\alpha}_k(\beta))$ is the posterior mode. Therefore the minimizer $\hat{\alpha}_k(\beta)$ in (2.2) is equivalent to the posterior mode.*

Proof. Note that

$$\hat{\alpha}_k(\beta) = \underset{\alpha_k}{\text{Argmax}} \left\{ p(x_k;\alpha_k,\beta) \frac{\pi(\alpha_k;\alpha_0,\delta n_k)}{b(\alpha_k)} \right\}.$$

The prior density on γ_k which is proportional to $\pi(\alpha_k;\alpha_0,\delta n_k)/b(\alpha_k)$ is equivalent to the prior density $\pi(\alpha_k;\alpha_0,\delta n_k)$ on α_k. Thus, the above equality shows that $\hat{\gamma}_k(\beta)$ is the posterior mode when β is known. □

The estimator $\hat{\alpha}_k(\beta)$ was called the standardized posterior mode in Yanagimoto and Ohnishi (2005). Since the Jacobian factor $b(\alpha_k)$ is discarded, the estimation procedure has invariance with respect to the parameter transformation.

3. Conjugate analysis as to the intercept

We assume on α_k the prior density

$$\pi(\alpha_k - \alpha_0;\delta n_k) = \frac{1}{K(\delta n_k)} \exp\left[-\delta n_k \{u(e^{\alpha_k-\alpha_0};2-\xi) - u(e^{\alpha_k-\alpha_0};1-\xi)\}\right], \tag{3.1}$$

which is in the location family, where

$$u(x;\kappa) = \begin{cases} \log x & \text{for } \kappa = 0, \\ \dfrac{x^\kappa - 1}{\kappa} & \text{otherwise.} \end{cases}$$

In the expression (3.1) of the prior density α_0 and $\delta > 0$ are hyper-parameters which are assumed to be known, and $K(\delta n_k)$ is the normalizing constant. When $\xi = 1$ or $\xi = 2$, the density (3.1) is a transformed gamma density.

The Tweedie density (1.1) has the following representation

$$p(x;\mu,\tau_0) = \exp\left[-\tau_0\{u(\mu;2-\xi) - xu(\mu;1-\xi)\}\right] \tilde{a}(x;\tau_0),$$

where $\tilde{a}(x;\tau_0) = \exp\left[\tau_0\{c(1)x - M(1)\}\right] a(x;\tau_0)$. This is shown as follows. Solving the differential equation $c'(\mu) = \tau_0\mu^{-\xi}$ in (1.2), we have $c(\mu) = u(\mu;1-\xi) + c(1)$. Noting that $M'(\mu) = \mu c'(\mu) = \tau_0\mu^{1-\xi}$, we get $M(\mu) = u(\mu;2-\xi) + M(1)$.

Properties of the function $u(x;\,\kappa)$ are shown in the following lemma, which play an essential role in the proof of conjugacy. The proof is a straightforward calculation, and is omitted.

Lemma 3.1. (i) *It holds for any x, y and κ that*

$$u(xy;\,\kappa) = y^{\kappa}u(x;\,\kappa) + u(y;\,\kappa).$$

(ii) *Suppose that κ and ν are positive. Then, it holds for any $a > 0$ and $b > 0$ that*

$$\begin{aligned} &au(x;\,\kappa) - bu(x;\,-\nu) \\ &= \delta^*\{u(x/x_0;\,\kappa) - u(x/x_0;\,-\nu) - u(1/x_0;\,\kappa) + u(1/x_0;\,-\nu)\}, \end{aligned}$$

where

$$x_0 = (b/a)^{\frac{1}{\kappa+\nu}} \quad \text{and} \quad \delta^* = a^{\frac{\nu}{\kappa+\nu}}b^{\frac{\kappa}{\kappa+\nu}}.$$

Using this lemma, we get conjugacy of the assume prior density (3.1).

Proposition 3.1. *The posterior density corresponding to the prior density $\pi(\alpha_k - \alpha_0;\,\delta n_k)$ under the sampling density $p(\boldsymbol{x}_k;\,\alpha_k,\,\beta)$ is expressed as $\pi\big(\alpha_k - \hat{\alpha}_k(\beta);\,\delta_k^* n_k\big)$, where*

$$\hat{\alpha}_k(\beta) = \log \frac{\tau_0 \sum_{i=1}^{n_k} x_{ki}e^{(1-\xi)\beta z_{ki}} + \delta n_k e^{-(1-\xi)\alpha_0}}{\tau_0 \sum_{i=1}^{n_k} e^{(2-\xi)\beta z_{ki}} + \delta n_k e^{-(2-\xi)\alpha_0}},$$

$$\delta_k^* = \frac{1}{n_k}\left\{\tau_0 \sum_{i=1}^{n_k} x_{ki}e^{(1-\xi)\beta z_{ki}} + \delta n_k e^{-(1-\xi)\alpha_0}\right\}^{2-\xi}\left\{\tau_0 \sum_{i=1}^{n_k} e^{(2-\xi)\beta z_{ki}} + \delta n_k e^{-(2-\xi)\alpha_0}\right\}^{\xi-1}.$$

Therefore, the prior density $\pi(\alpha_k - \alpha_0;\,\delta n_k)$ is conjugate.

The family of distributions to which the conjugate prior density (3.1) belongs is derived through the following requisition:

> A density $p(x-\mu)$ in a location family should have a conjugate prior density of the form $\pi(\mu - m;\,\delta) \propto \{p(m-\mu)\}^{\delta}$.

This requisition also yields the normal and the von Mises distributions.

Now we discuss the conjugate analysis as to α_k for a known β, adopting the loss function

$$L\big(\check{\alpha}_k(\beta) - \alpha_k\big) = u\big(\exp\{\alpha_k - \check{\alpha}_k(\beta)\};\,2-\xi\big) - u\big(\exp\{\alpha_k - \check{\alpha}_k(\beta)\};\,1-\xi\big).$$

This is a Kullback-Leibler loss function, which follows from Lemma 3.2(ii). Note that the prior density (3.1) is proportional to $\exp\{-\delta n_k L(\alpha_0 - \alpha_k)\}$.

Properties of the assumed prior density (3.1) is given in the following lemma.

Lemma 3.2. *Set $\rho(\tau) = 1 - (2-\xi)(\xi-1)K'(\tau)/K(\tau)$.*
(i) *It holds that*

$$\begin{aligned} \mathrm{E}\left[\, e^{(2-\xi)\alpha_k} \,\middle|\, \pi(\alpha_k - \alpha_0,\,\delta n_k)\,\right] &= \rho(\delta n_k)e^{(2-\xi)\alpha_0}, \\ \mathrm{E}\left[\, e^{(1-\xi)\alpha_k} \,\middle|\, \pi(\alpha_k - \alpha_0,\,\delta n_k)\,\right] &= \rho(\delta n_k)e^{(1-\xi)\alpha_0}. \end{aligned}$$

(ii) *The Kullback-Leibler separator from* $\pi(\alpha_k - \alpha_{01};\, \delta n_k)$ *to* $\pi(\alpha_k - \alpha_{02};\, \delta n_k)$ *is expressed as*

$$\mathrm{KL}\big(\pi(\alpha_k - \alpha_{01};\, \delta n_k),\, \pi(\alpha_k - \alpha_{02};\, \delta n_k)\big) = \delta n_k \rho(\delta n_k) L(\alpha_{02} - \alpha_{01}).$$

The conjugate analysis of α_k for a known β is summarized in the following proposition.

Proposition 3.2. *A modified Pythagorean relationship*

$$\mathrm{E}_{\mathrm{post}}\Big[\, L\big(\check{\alpha}_k(\beta) - \alpha_k(\beta)\big) - L\big(\hat{\alpha}_k(\beta) - \alpha_k(\beta)\big) - \rho(\delta_k^* n_k) L\big(\check{\alpha}_k(\beta) - \hat{\alpha}_k(\beta)\big)\,\Big] = 0$$

holds for any estimator $\check{\alpha}_k(\beta)$. *Therefore, the estimator* $\hat{\alpha}_k(\beta)$ *is optimum under the loss function* $L\big(\check{\alpha}_k(\beta) - \alpha_k\big)$.

Proof. Since $\pi(\alpha_k - \alpha_0,\, \delta n_k) \propto \exp\{-\delta n_k L(\alpha_0 - \alpha_k)\}$, we have

$$\begin{aligned}
&\mathrm{KL}\Big(\, \pi\big(\alpha_k - \hat{\alpha}_k(\beta);\, \delta_k^* n_k\big),\, \pi\big(\alpha_k - \check{\alpha}_k(\beta);\, \delta_k^* n_k\big)\,\Big) \\
&= \delta_k^* n_k \mathrm{E}\Big[\, L\big(\check{\alpha}_k(\beta) - \alpha_k\big) - L\big(\hat{\alpha}_k(\beta) - \alpha_k\big) \;\Big|\; \pi\big(\alpha_k - \hat{\alpha}_k(\beta);\, \delta_k^* n_k\big)\,\Big].
\end{aligned}$$

Recall Proposition 3.1 and apply Lemma 3.2(ii) to the left-hand side of this equality. □

4. Optimum estimating function of the slope

The score function of β is expressed as $l_\beta(\boldsymbol{x};\, \boldsymbol{\alpha},\, \beta) = \sum l_{k\beta}(\boldsymbol{x}_k;\, \alpha_k,\, \beta)$ where

$$l_{k\beta}(\boldsymbol{x}_k;\, \alpha_k,\, \beta) = -\tau_0 \left\{ e^{(2-\xi)\alpha_k} \sum_{i=1}^{n_k} z_{ki} e^{(2-\xi)\beta z_{ki}} - e^{(1-\xi)\alpha_k} \sum_{i=1}^{n_k} x_{ki} z_{ki} e^{(1-\xi)\beta z_{ki}} \right\}. \tag{4.1}$$

This is shown by noting that

$$\log p(\boldsymbol{x}_k;\, \alpha_k,\, \beta) = -\tau_0 \sum_{i=1}^{n_k} \Big\{ u\Big(e^{\alpha_k + \beta z_{ki}};\, 2 - \xi\Big) - x_{ki} u\Big(e^{\alpha_k + \beta z_{ki}};\, 1 - \xi\Big) \Big\} + C,$$

where C is the term constant in β.

The optimum estimating function of β is expressed in terms of the optimum estimator $\hat{\alpha}_k(\beta)$ which is derived through the conjugate analysis in the previous section. This interesting relation is given in the following proposition.

Proposition 4.1. *It holds for any* $k \in \{1, \ldots, K\}$ *that*

$$\mathrm{E}_{\mathrm{post}}\big[l_{k\beta}(\boldsymbol{x}_k;\, \alpha_k,\, \beta)\big] = \rho(\delta_k^* n_k)\, l_{k\beta}\big(\boldsymbol{x}_k;\, \hat{\alpha}_k(\beta),\, \beta\big).$$

Therefore, the optimum estimating function is

$$\mathrm{E}_{\mathrm{post}}\big[l_\beta(\boldsymbol{x};\, \boldsymbol{\alpha},\, \beta)\big] = \sum_{k=1}^{K} \rho(\delta_k^* n_k)\, l_{k\beta}\big(\boldsymbol{x}_k;\, \hat{\alpha}_k(\beta),\, \beta\big).$$

Proof. Proposition 3.1 and Lemma 3.2 yield that

$$\mathrm{E}_{\mathrm{post}}\left[e^{(2-\xi)\alpha_k}\right] = \rho(\delta_k^* n_k) e^{(2-\xi)\hat{\alpha}_k(\beta)},$$
$$\mathrm{E}_{\mathrm{post}}\left[e^{(1-\xi)\alpha_k}\right] = \rho(\delta_k^* n_k) e^{(1-\xi)\hat{\alpha}_k(\beta)}.$$

This, together with (4.1), complete the proof. □

5. Numerical example

Let us consider a Poisson logarithmic link regression model where

$$x_{ki} \sim \mathrm{Po}(\mu_{ki}) \quad \text{with} \quad \log \mu_{ki} = \alpha_k + \beta z_{ki}.$$

This is a special case of the Tweedie logarithmic link regression model where $\xi = 1$ and $\tau_0 = 1$. For simplicity we deal with the following binary covariate case where $n_k = 2m_k$ and the covariates are given by

$$z_{ki} = \begin{cases} 1 & (1 \le i \le m_k), \\ 0 & (m_k + 1 \le i \le 2m_k). \end{cases}$$

It follows from Propositions 3.1 and 4.1 that the estimator is given by $(\hat{\boldsymbol{\alpha}}(\hat{\beta}), \hat{\beta})$ where

$$\hat{\alpha}_k(\beta) = \log\left\{\left(\sum_{i=1}^{n_k} x_{ki} + \delta n_k\right) \Big/ \left(\sum_{i=1}^{n_k} e^{\beta z_{ki}} + \delta n_k e^{-\alpha_0}\right)\right\},$$
$$\hat{\beta} = \log\left\{(1 + 2\delta e^{-\alpha_0}) \sum_{k=1}^{K}\sum_{i=1}^{m_k} x_{ki} \Big/ \sum_{k=1}^{K}\left(\sum_{i=m_k+1}^{2m_k} x_{ki} + 2\delta m_k\right)\right\}.$$

Note that the estimator $(\hat{\boldsymbol{\alpha}}(\hat{\beta}), \hat{\beta})$ approaches to the maximum likelihood estimator (MLE) when $\delta \to +0$.

We run a simulation in the case where $K = 20$, $n_1 = \cdots = n_{20} = 10$, and the true value of the parameters are

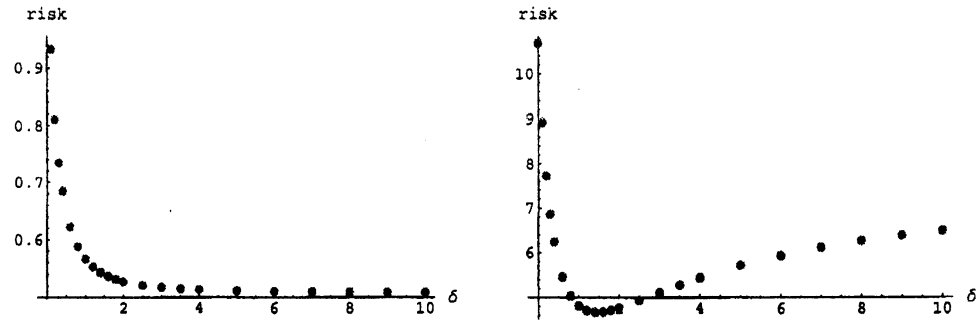
$$\boldsymbol{\alpha} = (1.10353, 1.00058, 0.700968, 0.963605, 0.784342, 0.960677, 1.18908,$$
$$0.923425, 0.581723, 0.910598, 1.35572, 0.872309, 1.03158, 0.948453,$$
$$1.33564, 0.473799, 1.11999, 0.931665, 1.01157, 1.44858),$$

and $\beta = 0.2$. The following two Kullback-Leibler loss functions are adopted:

$$L(\check{\beta}, \beta; \boldsymbol{\alpha}) = \sum_{k=1}^{K}\sum_{i=1}^{n_k} \mathrm{KL}(e^{\alpha_k + \check{\beta} z_{ki}}, e^{\alpha_k + \beta z_{ki}}),$$
$$L((\check{\boldsymbol{\alpha}}, \check{\beta}), (\boldsymbol{\alpha}, \beta)) = \sum_{k=1}^{K}\sum_{i=1}^{n_k} \mathrm{KL}(e^{\check{\alpha}_k + \check{\beta} z_{ki}}, e^{\alpha_k + \beta z_{ki}}).$$

The former is for estimation of β and the latter for that of $(\boldsymbol{\alpha}, \beta)$.

The graphs of the estimated risks are drawn by 5,000 iterations for selected values of the hyperparameter δ in Figure 2. Here the other hyperparameter α_0 is set as zero. Figure 2 indicates a superior performance of the proposed estimators over the MLEs.



(a) Estimation of β (b) Estimation of (α, β)

Figure 2: Estimated risks of the proposed estimation procedure for selected values of δ.

Let us examine the behavior of the proposed estimation procedure of β. It follows from Propositions 3.1 and 4.1 that

$$g_{\mathrm{B}}(\boldsymbol{x};\,\beta) \stackrel{\mathrm{def}}{=} \sum_{k=1}^{K}\sum_{i=1}^{m_k} x_{ki} - \frac{e^{\beta}}{1+e^{\beta}+2\delta e^{-\alpha_0}} \sum_{k=1}^{K}\left(\sum_{i=1}^{2m_k} x_{ki} + 2\delta m_k\right).$$

Taking the limit $\delta \to +0$, we set

$$g_{\mathrm{M}}(\boldsymbol{x};\,\beta) \stackrel{\mathrm{def}}{=} \sum_{k=1}^{K}\sum_{i=1}^{m_k} x_{ki} - \frac{e^{\beta}}{1+e^{\beta}} \sum_{k=1}^{K}\sum_{i=1}^{2m_k} x_{ki}.$$

The performance of the two estimating functions $g_{\mathrm{B}}(\boldsymbol{x};\,\beta)$ and $g_{\mathrm{M}}(\boldsymbol{x};\,\beta)$ is evaluated from Frequentists' point of view in the following proposition, which is consistent with the Figure 2 (a).

Proposition 5.1. *Assume that* $e^{\alpha_0}\sum_{k=1}^{K} m_k = \sum_{k=1}^{K} m_k e^{\alpha_k}$. *Then, it holds that*

$$\frac{\mathcal{M}_0[\,g_{\mathrm{B}}(\boldsymbol{x};\,\beta)\,]}{\mathcal{M}_0[\,g_{\mathrm{M}}(\boldsymbol{x};\,\beta)\,]} = \frac{1}{1+e^{\beta}}\left\{1 + \frac{e^{\beta}}{(1+2\delta e^{-\alpha_0})^2}\right\},$$

where

$$\mathcal{M}_0[\,g(\boldsymbol{x};\,\beta)\,] = \frac{\mathrm{E}_p[\,g^2(\boldsymbol{x};\,\beta)\,]}{\left\{\mathrm{E}_p\left[\,\frac{\partial}{\partial\beta} g(\boldsymbol{x};\,\beta)\,\right]\right\}^2}. \tag{5.1}$$

Therefore, the estimating function $g_{\mathrm{B}}(\boldsymbol{x};\,\beta)$ *with* $\delta > 0$ *is superior to the estimating function* $g_{\mathrm{M}}(\boldsymbol{x};\,\beta)$ *with respect to the criterion function* (5.1).

References

Godambe, V. P. and Kale, B. K. (1991). Estimating functions: An overview. *Estimating Functions* (Ed. V. P. Godambe), 3–20, Clarendon Press, Oxford.

Jørgensen, B. (1997). *The theory of dispersion models.* Chapman and Hall, London.

Shono, H. (2006). Applications of Tweedie distribution to the CPUE standardization. *Workshop on Prediction for marine resources 2006, Institute of Statistical Mathematics*

Yanagimoto, T. & Ohnishi, T. (2005). Standardized posterior mode for the flexible use of a conjugate prior, *J. Statist. Plann. Inference* **131**, 253–269.