

不均衡データに対する二値回帰モデルの挙動

慶應義塾大学・理工学部 清 智也

Tomonari Sei

Faculty of Science and Technology,
Keio University

概要

ロジスティック回帰モデルは、二値反応が不均衡な場合、ポアソン点過程で近似されることが知られている。本稿では、この現象が二値（二項）回帰モデルにおける多くのリンク関数に対して普遍的に成り立つことを示す。証明には極値理論の結果を用いる。ロジット、プロビット、および complementary log-log リンク関数に対しては、点過程の強度が指数型分布族になる。他のリンク関数の場合、変形指数型分布族が現れる。これらは Sei (2014) で示された結果である。本稿ではさらに、罰則付き最尤推定量の漸近許容性について議論する。

キーワード 二項回帰, 極値理論, 不均衡データ, ポアソン点過程, q -指数型分布族.

1 はじめに

$\{(X_i, Y_i)\}_{i=1}^m$ を, $\mathbb{R}^p \times \{0, 1\}$ 上の独立同一分布に従う m 個の確率変数とする。 X_i を条件づけた下での Y_i の条件付き分布は

$$P(Y_i = 1 \mid X_i, a, b) = G(a + b^T X_i), \quad a \in \mathbb{R}, \quad b \in \mathbb{R}^p, \quad (1)$$

で与えられると仮定する。ここで、 $G(\cdot)$ は 1 次元の累積分布関数であり、その逆関数 $G^{-1}(p) = \sup\{z : G(z) \leq p\}$ は一般化線形モデルにおけるリンク関数である。 X_i の周辺分布を $F(dX_i)$ と記す。累積分布関数 G として、実用上は

- ロジスティック分布 : $G(x) = e^x / (e^x + 1)$,
- 標準正規分布 : $G(x) = \int_{-\infty}^x (2\pi)^{-1/2} e^{-x^2/2} dx$,
- (負の方向の) ガンベル分布 : $G(x) = 1 - e^{-e^x}$

がよく用いられる。対応するリンク関数はロジット、プロビット、complementary log-log リンク関数である。これらの 3 つの例に対しては、式 (1) の対数尤度関数は凹となることが知られている (Wedderburn, 1976)。

本稿では、データが非常に偏っている場合を考える。言い換えれば、成功確率がほとんどゼロの場合であり、このようなデータを不均衡データという。そのようなケースが現れる例としては、不正検出、医療診断、政策分析などがある (Bolton and Hand, 2002; Chawla et al., 2004; Jin et al., 2005; King and Zeng, 2001)。

説明変数がない場合、ポアソンの少数の法則がよく知られている：もし

$$P(Y_i = 1) = \lambda/m + o(m^{-1}), \quad i = 1, \dots, m,$$

ならば, 正例の個数 $\sum_{i=1}^m Y_i$ は平均パラメータ λ のポアソン分布に分布収束する. この考察から, 不均衡データに対しては, (1) の真のパラメータ (a, b) は m に依存してもよいと考えるのが自然であろう. これを (a_m, b_m) と書く.

Owen (2007) は, $\sum_{i=1}^m Y_i$ を固定したまま m を増加させると, ロジスティック回帰モデルの最尤推定値が, ある指数型分布族の最尤推定量に収束することを示した. この結果は大まかには以下のように導かれる.

モデル (1) でロジスティック分布 $G(z) = e^z / (1 + e^z)$ を用いた場合を考える. 任意の α と β に対し, $a_m(\alpha) = -\log m + \alpha$, $b_m(\beta) = \beta$ とおく. すると, $m \rightarrow \infty$ の下で

$$P(Y_i = 1 \mid X_i, a_m(\alpha), b_m(\beta)) = \frac{e^{-\log m + \alpha + \beta^T X_i}}{1 + e^{-\log m + \alpha + \beta^T X_i}} = \frac{e^{\alpha + \beta^T X_i}}{m} + o(m^{-1}) \quad (2)$$

が成り立つ. ベイズの定理から, $Y_i = 1$ の下での X_i の条件付き密度 ($F(dx_i)$ に対する密度) は, 形式的に,

$$\frac{e^{\beta^T X_i}}{\int e^{\beta^T x} F(dx)} + o(1) \quad (3)$$

と計算される. これは十分統計量を X_i とする指数型分布族である (Owen, 2007).

注意 1. 正確に言えば, Owen (2007) はここで述べたものとは異なる設定を考慮しており, $Y_i = 0, 1$ の下での X_i の条件付き分布を任意の分布 F_0, F_1 とした上で, 最尤推定量の収束性を証明している. 我々の設定では, F_0 は漸近的に F に等しく, F_1 は F に対して密度 (3) を持たなければならない. したがって, この仮定が満たされなければ本稿の設定は誤特定の状況になる. この点については Sei (2014) で論じている.

Warton and Shepherd (2010) は, ロジスティック回帰モデルが, 尤度比の意味でポアソン点過程モデルに収束することを示した. このことは大まかには次のように確かめられる. 式 (2) より, \mathbb{R}^p の任意のコンパクト集合 A に対して, 確率 $P(Y_i = 1, X_i \in A)$ は近似的に $m^{-1} \int_A e^{\alpha + \beta^T x} F(dx)$ と表される. よって, ポアソンの少数の法則より, $X_i \in A$ かつ $Y_i = 1$ となるような観測値の個数は, 平均パラメータ $\int_A e^{\alpha + \beta^T x} F(dx)$ のポアソン分布に近似的に従うことが分かる. これは強度 (intensity) $e^{\alpha + \beta^T x} F(dx)$ のポアソン点過程である.

Baddeley et al. (2010) は, ポアソン点過程モデルをピクセルベースの二項回帰モデルで近似するときの性質について詳しく議論している. 彼らは, complementary log-log リンク関数がピクセルサイズの変更に対して整合的であることを示すとともに, split-pixel strategy という手法を提案し, 共変量が空間的に滑らかでない場合も近似がうまく機能することを示した.

Sei (2014) は, ロジスティック回帰モデル以外の二項回帰モデルの不均衡極限を考えた. ロジスティック回帰に対する結果から予想される通り, その極限はポアソン点過程になる. 注目すべき点は, 収束先の点過程の強度が, 一般に q -指数型分布族と呼ばれるクラスになることである. q -指数型分布族とは, 変形指数型分布族, あるいは α -分布族とも呼ばれ, 実数 q を使って特徴づけられる確率分布族であり, 統計物理や情報幾何において近年脚光を浴びている (Amari (1985); Amari and Nagaoka (2000); Amari and Ohara (2011); Naudts (2002, 2010); Tsallis (1988)). 例えば, 冒

頭に挙げたロジット, プロビット, complementary log-log リンクの場合にはいずれも $q = 1$ となる. 一方, G をコーシー分布としたときは $q = 2$, G を一様分布としたときは $q = 0$ となる. これらの結果は 2 節, 3 節でレビューされる.

また 4 節では, Sei (2014) で詳しく触れなかった話題として, 推定量の許容性について言及する.

2 二項回帰の不均衡極限

実数 q に対し, q -指数関数を

$$\exp_q(z) = \begin{cases} e^z, & \text{if } q = 1, \\ [1 + (1 - q)z]_+^{1/(1-q)}, & \text{if } q \neq 1, \end{cases} \quad (4)$$

により定義する. ここで, $[z]_+ = \max(z, 0)$, $[0]_+^{-1} = \infty$ と約束する. この変換はパラメータ $\lambda = 1 - q$ の Box-Cox 変換の逆変換に他ならない. 特に, $q < 1$ かつ $z \leq -1/(1 - q)$ のとき $\exp_q(z) = 0$ であり, $q > 1$ かつ $z \geq -1/(1 - q)$ のとき $\exp_q(z) = \infty$ となる. 関数 $\exp_q(z)$ は $q \geq 0$ のとき, またそのときに限り凸関数である.

さて, 二項回帰モデル (1) において, G に関する次の仮定を設けよう.

仮定 1. ある $q > 0$, $c_m \in \mathbb{R}$ および $d_m > 0$ が存在し, 各 $z \in \mathbb{R}$ に対して

$$G(c_m + d_m z) = \frac{1}{m} \exp_q(z) + o(m^{-1}), \quad m \rightarrow \infty \quad (5)$$

が成り立つ.

極値理論によれば, 式 (5) 以外の漸近形は存在しない (例えば de Haan and Ferreira (2006, Theorem 1.1.2 and 1.1.3)). 実数 q は G の左裾の構造を決定している. 例えばロジスティック分布は, 仮定 1 を満たし, $q = 1$, $c_m = -\log m$, $d_m = 1$ である. 他の例については 3 節で述べる.

仮定 1 の c_m, d_m を用いて, $(\alpha, \beta) \in \mathbb{R} \times \mathbb{R}^p$ に対して

$$a_m(\alpha) = c_m + d_m \alpha \quad \text{and} \quad b_m(\beta) = d_m \beta \quad (6)$$

と定義する. また, 真のパラメータ (回帰係数) が $(a_m(\alpha), b_m(\beta))$ のときの $\{(X_i, Y_i)\}_{i=1}^m$ の確率法則を $P_{m,\alpha,\beta}$ と記す.

さて, 式 (2) の類推が仮定 1 から得られる. 実際,

$$\begin{aligned} P_{m,\alpha,\beta}(Y_i = 1 \mid X_i) &= G(a_m(\alpha) + b_m(\beta)^T X_i) \\ &= G(c_m + d_m(\alpha + \beta^T X_i)) \\ &= \frac{1}{m} \exp_q(\alpha + \beta^T X_i) + o(m^{-1}) \end{aligned}$$

となる. よって, ロジスティック回帰のときと同様, 二項回帰モデルはポアソン点過程に収束することが期待される. これを以下示す.

主結果を述べる前に次の補題を用意する.

補題 1. $(\alpha, \beta) \in \mathbb{R} \times \mathbb{R}^p$ とする. また, A を \mathbb{R}^p のコンパクト集合とし, $\forall x \in A$ に対し $\exp_q(\alpha + \beta^T x)$ は有限であると仮定する. このとき次の式が成り立つ:

$$P_{m,\alpha,\beta}(Y_i = 1, X_i \in A) = \frac{\lambda(A)}{m} + o(m^{-1}). \quad (7)$$

ただし $\lambda(A) = \int_A \exp_q(\alpha + \beta^T x) F(dx)$ とする.

証明. $t := \alpha + \beta^T X_i$ の確率分布を $F^*(dt)$ とおく. また, $A^* = \{\alpha + \beta^T x \mid x \in A\}$ と定義する. 仮定より, A^* はコンパクトである. このとき,

$$\begin{aligned} P_{m,\alpha,\beta}(Y_i = 1, X_i \in A) &= \int_A G(a_m(\alpha) + b_m(\beta)^T x) F(dx) \\ &= \int_A G(c_m + d_m(\alpha + \beta^T x)) F(dx) \\ &= \int_{A^*} G(c_m + d_m t) F^*(dt) \end{aligned}$$

と書ける. 式 (7) を示すには,

$$\int_{A^*} G(c_m + d_m t) F^*(dt) = \frac{1}{m} \int_{A^*} \exp_q(t) F^*(dt) + o(m^{-1})$$

を言えば十分である. 仮定 1 より, 各 $t \in A^*$ に対して $mG(c_m + d_m t) = \exp_q(t) + o(1)$ となる. よって, $mG(c_m + d_m t)$ が $t \in A^*$ について一様に $\exp_q(t)$ に収束することを示せばよい. ところが, $mG(c_m + d_m t)$ は t について単調であり, かつ $\exp_q(t)$ は $t \in A^*$ について連続であるから, この一様収束性は一般論から導かれる (例えば Galambos (1987, Lemma 2.10.1)). \square

さて, データ $\{(X_i, Y_i)\}_{i=1}^m$ に対して,

$$N_m(A) = \#\{i \mid X_i \in A, Y_i = 1\}, \quad A \subset \mathbb{R}^p,$$

によって点過程 N_m を定義する. 上の式は, 集合 A に属するような X_i のうち, $Y_i = 1$ となるようなものの個数を表している.

定理 1. $P_{m,\alpha,\beta}$ の下で, 点過程 N_m は次の強度を持つポアソン点過程 N に法則収束する:

$$\lambda(dx) = \exp_q(\alpha + \beta^T x) F(dx). \quad (8)$$

正確には, 等式

$$\begin{aligned} \lim_{m \rightarrow \infty} P_{m,\alpha,\beta}(N_m(A_j) = \nu_j, 1 \leq j \leq J) \\ = P(N(A_j) = \nu_j, 1 \leq j \leq J) = \prod_{j=1}^J \frac{\lambda(A_j)^{\nu_j} e^{-\lambda(A_j)}}{\nu_j!} \end{aligned} \quad (9)$$

が, 任意の正整数 J , 非負整数 ν_j , 互いに排反なコンパクト集合 $A_j \subset \mathbb{R}^p$ (で $\exp_q(\alpha + \beta^T x) < \infty, x \in A_j$, を満たすもの) に対して成り立つ.

式 (9) は, Embrechts et al. (1997) にある点過程の法則収束の定義と整合的である.

定理 1 の証明. $\{A_j\}_{j=1}^J$ を互いに排反なコンパクト集合とする. $\{(X_i, Y_i)\}_{i=1}^m$ は独立同一分布に従うので, 確率変数 $\{N_m(A_j)\}_{j=1}^J$ の同時分布は

$$P_{m,\alpha,\beta}(N_m(A_j) = \nu_j, 1 \leq j \leq J) = \prod_{j=1}^J (p_{m,j})^{\nu_j} \left(1 - \sum_j p_{m,j}\right)^{m - \sum_j \nu_j}$$

という多項分布に従う. ただし,

$$p_{m,j} = P_{m,\alpha,\beta}(X_i \in A_j, Y_i = 1), \quad 1 \leq j \leq J,$$

とおいた. したがって補題 1 より, $(N_m(A_1), \dots, N_m(A_J))$ は独立なポアソン確率変数に法則収束し, その平均パラメータは $(\lambda(A_1), \dots, \lambda(A_J))$ である. \square

定理 1 より, 特にロジスティック回帰モデルは強度 $\exp(\alpha + \beta^T x)F(dx)$ のポアソン点過程モデルに収束する. これは Warton and Shepherd (2010) が示した事実と整合的である.

定義 1. 実数 $q \in \mathbb{R}$ に対し, 式 (8) を強度の q -指数型分布族と呼ぶ. 対応する点過程の確率法則を $P_{\alpha,\beta}^{(q)}$ と記す.

強度の q -指数型分布族は, 確率測度の q -指数型分布族に密接に関係している. 強度 (8) の全強度を

$$\Lambda_q(\alpha, \beta) = \int_{\mathbb{R}^p} \exp_q(\alpha + \beta^T x) F(dx) \quad (10)$$

と記すことにし, $\Lambda_q(\alpha, \beta) < \infty$ と仮定しよう. すると, $P_{\alpha,\beta}^{(q)}$ の尤度は

$$\frac{e^{-\Lambda_q(\alpha,\beta)}}{n!} \prod_{i=1}^n \exp_q(\alpha + \beta^T x_i) \quad (11)$$

と書ける. ここで, n の基準測度は非負整数上の計数測度であり, 各 i に対する x_i の基準測度は $F(dx_i)$ とする. 式 (11) において, 数 n は観測値の個数であり, その周辺分布は平均 $\Lambda_q(\alpha, \beta)$ のポアソン分布である. n を条件づけたとき, 各 x_i は独立に

$$\frac{\exp_q(\alpha + \beta^T x_i)}{\Lambda_q(\alpha, \beta)} \quad (12)$$

という密度をもつ確率分布に従う. 式 (12) は q -指数型分布族, 変形指数型分布族, あるいは α -分布族 ($\alpha = 2q - 1$) と呼ばれる. 密度 (12) は, 適切な θ と $\psi_q(\theta)$ を選んで $\exp_q(\theta^T x_i - \psi_q(\theta))$ という形に書くこともできる. (例えば Amari and Ohara (2011)). しかし, 本稿ではこの表現を用いない. その理由は, この表現を用いても, ポアソン点過程の尤度関数 (11) において $\Lambda_q(\alpha, \beta)$ が残ってしまうためである.

図 1 に q -指数型分布族への収束のイメージを示す.

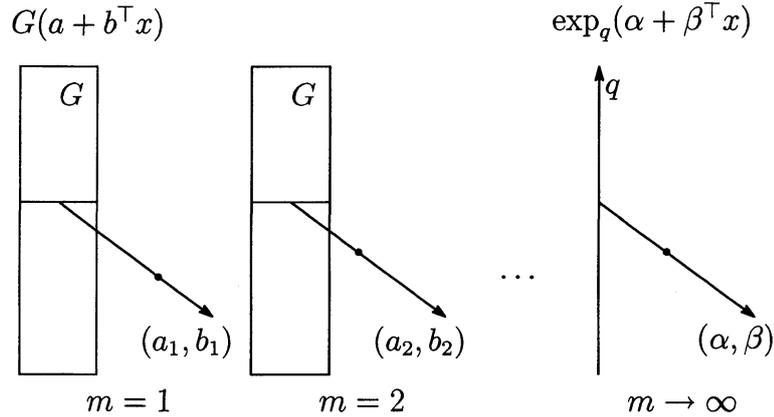


図 1: q -指数型分布族への収束のイメージを示す。モデル空間の確率分布は、 G と a, b の選び方によって定まる。ただし $F(dX_i)$ は固定して考える。一方、不均衡極限においては G がいずれかの実数 q に対応する（または極限を持たない）。

3 例

この節では、仮定 1 を満たす分布 G の例をいくつか与える。また、最尤推定の収束性に関する数値実験結果を与える。

なお、分布 G が仮定 1 を満たすとしても、列 (c_m, d_m) は一意に定まらない。一意的に定める方法は知られている（例えば Galambos (1987, Theorem 2.1.4–2.1.6)）が、以下では比較の見やすい形の (c_m, d_m) を用いることにする。

ロジスティック分布 $G(z) = e^z / (1 + e^z)$ とガンベル分布 $G(z) = 1 - \exp(-e^z)$ に対しては、いずれも

$$q = 1, \quad c_m = -\log m, \quad d_m = 1. \quad (13)$$

であることが容易にチェックできる。標準正規分布に対しては、

$$q = 1, \quad c_m = -(2 \log m)^{1/2} + \frac{\log(\log m) + \log(4\pi)}{2(2 \log m)^{1/2}}, \quad d_m = (2 \log m)^{-1/2}. \quad (14)$$

であることが知られている（Galambos (1987, Section 2.3.2)）。コーシー分布に対しては

$$q = 2, \quad c_m = -m/\pi, \quad d_m = m/\pi. \quad (15)$$

である。その他、 t -分布やパレート分布などの例については極値理論の本（Galambos (1987); Embrechts et al. (1997) など）を参照せよ。また、Ding et al. (2011) で提案された t -ロジスティック回帰については Sei (2014) で調べられている。

次に、数値実験結果について説明する。

表 1 と表 2 に、結果を示す。ここでは、標本サイズ m に対してデータを

$$(X_i, Y_i) = \begin{cases} (0.4 + 0.4(i-1)/(n-1), 1) & \text{if } i \in \{1, \dots, n\}, \\ ((i-n-1)/(m-n-1), 0) & \text{if } i \in \{n+1, \dots, m\} \end{cases} \quad (16)$$

とし、また $n = 10$ とした。二項回帰モデルに対しては、回帰係数の推定値 (\hat{a}, \hat{b}) を式 (6) で標準化したものを示している。また、ポアソン点過程に対しては、共変量の真の分布 $F(dx)$ は未知なので、 $\{X_i\}_{i=1}^m$ の経験分布でこれを代用した。この経験分布には $Y_i = 0$ のデータ（負例）も含まれることに注意する。表 1 より、プロビットリンクの収束は非常に遅く、理論的に収束するかどうかに興味を持たれる。他のリンク関数については、収束性は満足いくものと言えよう。

表 1: $q = 1$ のポアソン点過程モデルと二項回帰モデルの最尤推定値の比較。ロジット、プロビット、complementary log-log (cloglog) リンクが用いられている。標本は式 (16) とし、 n は 10 とする。標準化のための数列 (c_m, d_m) は式 (13), (14) を用いた。

m	Poisson process		logit		probit		cloglog	
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
10^2	1.6504	1.1737	1.6883	1.3067	2.0282	2.3030	1.6975	1.1883
10^3	1.6277	1.2246	1.6314	1.2373	1.9070	1.8777	1.6322	1.2260
10^4	1.6256	1.2294	1.6260	1.2307	1.8634	1.6725	1.6260	1.2295
10^5	1.6254	1.2299	1.6254	1.2300	1.8330	1.5642	1.6254	1.2299

表 2: $q = 1$ のポアソン点過程モデルと、コーシーリンク（コーシー分布の逆関数）による二項回帰モデルの最尤推定値の比較。標本は式 (16) とし、 n は 10 とする。標準化のための数列 (c_m, d_m) は式 (15) を用いた。

m	Poisson process		cauchit	
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
10^2	0.8662	0.0667	0.8632	0.0656
10^3	0.8626	0.0673	0.8623	0.0677
10^4	0.8622	0.0680	0.8622	0.0679
10^5	0.8621	0.0680	0.8622	0.0679

4 漸近許容性

本節では、Sei (2014) では詳しく触れなかった話題として、推定量の漸近許容性について議論する。ここでは仮定 1 において $q = 1$ の場合、すなわち定理 1 において収束先が指数型分布族の場合のみを考える。また、本稿で漸近許容性と言っているのは単に極限の統計モデルにおける許容性のことであり、LeCam-Hájek 流の漸近許容性とは異なることをあらかじめ断っておく。

推定値として，罰則付き最尤推定量

$$(\hat{a}_m, \hat{b}_m) = \operatorname{argmax} \sum_{i=1}^{m+2} \{Y_i \log G(a + b^T X_i) + (1 - Y_i) \log(1 - G(a + b^T X_i))\} \quad (17)$$

を考える。ここで， $(X_{m+1}, Y_{m+1}) := (\bar{X}_m, 0)$ と $(X_{m+2}, Y_{m+2}) := (\bar{X}_m, 1)$ は疑似データであり， \bar{X}_m は $\{X_i\}_{i=1}^m$ の標本平均である。共変量の母平均を $\mu = \int \xi F(d\xi)$ とおけば，これらの疑似データはそれぞれ $m \rightarrow \infty$ の下で $(\mu, 0), (\mu, 1)$ に収束する。

以下，正例の個数を $n = \#\{1 \leq i \leq m \mid Y_i = 1\}$ とおき（これは確率変数），正例に対する共変量全体を $\{x_i\}_{i=1}^n = \{X_i \mid 1 \leq i \leq m, Y_i = 1\}$ とおくことにする。またその標本平均を $\bar{x} = n^{-1} \sum_{i=1}^n x_i$ とおく。不均衡極限の尤度関数が式(11)で与えられることに注意すると，式(17)の推定量 (\hat{a}_m, \hat{b}_m) は，次の極限に法則収束すると考えられる：

$$(\hat{\alpha}, \hat{\beta}) = \operatorname{argmax} \left\{ (n+1)\alpha + \beta^T(n\bar{x} + \mu) - \int e^{\alpha + \beta^T \xi} F(d\xi) \right\}. \quad (18)$$

ここで，収束性は (\hat{a}_m, \hat{b}_m) を一旦式(6)によって標準化して考える。この推定量の収束は厳密には証明できていない。しかし，少なくとも目的関数が各点で法則収束することは示される。以下では最初から式(18)を考察の対象とする。

ここでさらに，次の式でパラメータを (α, β) から (Λ, β) に変換する：

$$\Lambda = \int e^{\alpha + \beta^T \xi} F(d\xi) \quad \leftrightarrow \quad \alpha = \log \Lambda - \log \int e^{\beta^T \xi} F(d\xi).$$

このようにすると，式(18)の目的関数が Λ と β で分離されるという利点がある：

$$((n+1) \log \Lambda - \Lambda) + \beta^T(n\bar{x} + \mu) - (n+1)\psi(\beta), \quad \psi(\beta) := \log \int e^{\beta^T \xi} F(d\xi).$$

特に， Λ の推定量は $\hat{\Lambda} = n+1$ であり， β の推定量 $\hat{\beta}$ は

$$\frac{n\bar{x} + \mu}{n+1} = \frac{\int \xi e^{\hat{\beta}^T \xi} F(d\xi)}{\int e^{\hat{\beta}^T \xi} F(d\xi)} = \nabla \psi(\hat{\beta}), \quad \nabla := (\partial / \partial \beta_i)_{i=1}^p, \quad (19)$$

を満たす。

以下，推定量 $(\hat{\Lambda}, \hat{\beta})$ は Kullback-Leibler (KL) 損失の下で許容的であることを説明する。まず，推定量 $(\hat{\Lambda}, \hat{\beta})$ の KL 損失は次のように表される：

$$\text{KL}((\Lambda, \beta), (\hat{\Lambda}, \hat{\beta})) = \left(\Lambda \log \frac{\Lambda}{\hat{\Lambda}} - \Lambda + \hat{\Lambda} \right) + \Lambda \left((\beta - \hat{\beta})^T \nabla \psi(\beta) - \psi(\beta) + \psi(\hat{\beta}) \right).$$

これは， n が Λ を平均とするポアソン分布に従うこと，また n を条件付けたとき $\{x_i\}_{i=1}^n$ が独立に指数型分布族 $\exp(\beta^T x_i - \psi(\beta))$ に従うことから分かる。

さて，ポアソン分布の平均パラメータ Λ に対する推定量 $\hat{\Lambda} = n+1$ は KL 損失の下で許容的であることが Ghosh and Yang (1988) によって示されている。より正確には，ある Λ のプロパー事前分布の列 $\{\pi_k(\Lambda)\}_{k=1}^\infty$ が存在して，この列に対応するべ

イズ推定量 Λ_k が積分リスクに関して $r(\pi_k, \hat{\Lambda}) - r(\pi_k, \hat{\Lambda}_k) \rightarrow 0$ ($k \rightarrow \infty$) を満たすことが示されている。この結果と Blyth の定理から許容性が従う。

我々の設定に戻って、 $(\hat{\Lambda}, \hat{\beta})$ の許容性を示すには、やはりベイズ推定量の列を考えればよい。 Λ と β の事前分布は独立とし、 Λ については上記の $\pi_k(\Lambda)$ を用いる。一方、 β に関しては次の事前分布を用いる：

$$\pi(\beta) \propto \exp(\beta^T \mu - \psi(\beta)).$$

ただし、 $\mu = \int \xi F(d\xi)$ は以前定義したものと同一である。まず、この事前分布はプロパーであることが (F のサポートのアフィン包が \mathbb{R}^p という仮定の下で) 示される。また、この事前分布に対するベイズ推定量は罰則付き最尤推定に一致することが、以下のように示される (注：一般に指数型分布族の自然パラメータについてはベイズ推定量と MAP 推定量が一致する)。 β に関係する部分の事後損失は

$$\int \pi(\beta|n, x) \left\{ (\beta - \hat{\beta})^T \nabla \psi(\beta) - \psi(\beta) + \psi(\hat{\beta}) \right\} d\beta, \quad x := \{x_i\}_{i=1}^n,$$

と書けるので、ベイズ推定量は

$$\int \pi(\beta|n, x) \nabla \psi(\beta) d\beta = -\nabla \psi(\hat{\beta})$$

の解である。しかし、いま事後分布は $\pi(\beta|n, x) \propto \exp(\beta^T (n\bar{x} + \mu) - (n+1)\psi(\beta))$ であるから、部分積分を用いると

$$-\frac{n\bar{x} + \mu}{n+1} = -\nabla \psi(\hat{\beta})$$

となる。これは式 (19) と一致する。

以上から、プロパーな事前分布の列 $\pi_k(\Lambda)\pi(\beta)$ を考えると、そのベイズ推定量は $(\hat{\Lambda}_k, \hat{\beta})$ となる。そして、 $(\hat{\Lambda}, \hat{\beta})$ と $(\hat{\Lambda}_k, \hat{\beta})$ の積分リスクの差は 0 に収束することが示される。よって、Blyth の定理から、 $(\hat{\Lambda}, \hat{\beta})$ は許容的であることが示された。

同様に、ベイズ予測についても、ポアソン分布に関する既存の結果 (Komaki, 2004) を用いれば、漸近許容的な推定量を構築できると考えられる。

参考文献

- Amari, S., 1985. Differential-geometrical methods in statistics. Berlin: Springer.
- Amari, S., Nagaoka, H., 2000. Methods of information geometry (Translations of Mathematical Monographs). Oxford University Press.
- Amari, S., Ohara, A., 2011. Geometry of q-exponential family of probability distributions. Entropy 13, 1170–1185.
- Baddeley, A., Berman, M., Fisher, N.I., Hardegen, A., Milne, R.K., Schuhmacher, D., Shah, R., Turner, R., 2010. Spatial logistic regression and change-of-support in Poisson point processes. Electron. J. Statist. 4, 1151–1201.

- Bolton, R.J., Hand, D.J., 2002. Statistical fraud detection: a review. *Statist. Sci.* 17, 235–249.
- Chawla, N.V., Japkowicz, N., Koltz, A., 2004. Editorial: special issue on learning from imbalanced data sets. *ACM SIGKDD Explorations Newsletter* 6, 1–6.
- Ding, N., Vishwanathan, S.V.N., Warmuth, M., Denchev, V., 2011. t -logistic regression. *J. Mach. Learn. Res.* 12, 1–55.
- Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. *Modelling extremal events*. Berlin: Springer.
- Galambos, J., 1987. *The asymptotic theory of extreme order statistics*. Malarbar: Robert E. Krieger Publishing Company.
- Ghosh, M., Yang, M.C., 1988. Simultaneous estimation of Poisson means under entropy loss. *Ann. Statist.* 16, 278–291.
- de Haan, L., Ferreira, A., 2006. *Extreme value theory, an introduction*. New York: Springer.
- Jin, Y., Rejesus, R.M., Little, B.B., 2005. Binary choice models for rare events data: a crop insurance fraud application. *Applied Economics* 37, 841–848.
- King, G., Zeng, L., 2001. Logistic regression in rare events data. *Political Analysis* 9, 137–163.
- Komaki, F., 2004. Prediction of independent Poisson observables. *Ann. Statist.* 32, 1744–1769.
- Naudts, J., 2002. Deformed exponentials and logarithms in generalized thermostatics. *Physica A* 316, 323–334.
- Naudts, J., 2010. The q -exponential family in statistical physics. *J. Phys.: Conf. Ser.* 201, 012003.
- Owen, A.B., 2007. Infinitely imbalanced logistic regression. *J. Mach. Learn. Res.* 8, 761–773.
- Sei, T., 2014. Infinitely imbalanced binomial regression and deformed exponential families. *J. Statist. Plan. Infer.* in Press.
- Tsallis, C., 1988. Possible generalization of Boltzmann-Gibbs statistics. *J. Statist. Phys.* 52, 479–487.
- Warton, D.I., Shepherd, L.C., 2010. Poisson point process models solve the “pseudo-absence problem” for presence only data in ecology. *Ann. Applied Statist.* 4, 1383–1402.
- Wedderburn, R.W.M., 1976. On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika* 63, 27–32.