ガンマ・ダイバージェンスに基づく
ロバストかつスパースな一般化線形回帰

総合研究大学院大学　川島　孝行
統計数理研究所　藤澤　洋徳

1. はじめに　回帰モデルの同定は条件付密度関数 $f(y|x;\theta)$ の同定とも考えることができる．この問題は次の 回帰用の KL ダイバージェンス に基づいて考えることが可能である：$D_{KL}(g(y|x), f(y|x;\theta); g(x)) = \int D_{KL}(g(y|x), f(y|x;\theta)g(x)dx$．ただし $g(y|x)$ および $g(x)$ はデータを生成する分布である．$g(y|x)$ および $g(x)$ を，それぞれ経験密度関数 $\bar{g}(y|x)$ と $\bar{g}(x)$ で置き換えると，最尤推定に一致 する．ダイバージェンスに基づく回帰の枠組みでは，適切なパラメトリックモデルを選択することで，線形回帰・ロジスティック回帰・ポアソン回帰といった主要な回帰モデルの多くを一般化線形回帰の枠組みで捉えることができる．しかし，KL ダイバージェンスは，外れ値に弱い．そこで，我々は，ロバスト性をもたらすダイバージェンスとして知られている，ガンマ・ダイバージェンス (Fujisawa and Eguchi 2008) に基づく回帰を考える．さらに，これにスパース正則化を組み合わせることで，ロバストかつスパースな一般化線形回帰を達成する．ただし，単に適用すれば良い訳ではなく，それぞれの回帰モデルに応じて工夫を加えている．

2. ガンマ・ダイバージェンスに基づくロバストかつスパースな回帰　ガンマ・ダイバージェンスに基づく回帰は，相互エントロピー $d_\gamma(g(y|x), f(y|x;\theta); g(x)) = -E_{g(x,y)}[f(y|x;\theta)^\gamma / \left(\int f(y|x;\theta)^{1+\gamma} dy\right)^{\frac{\gamma}{1+\gamma}}]$．の経験推定を考え，スパース正則化を組み合わせることで，以下のように表される：

$$\arg\min_\theta -\log\left\{\frac{1}{n}\sum_{i=1}^n f(y_i|x_i;\theta)^\gamma \Big/ \left(\int f(y|x_i;\theta)^{1+\gamma} dy\right)^{\frac{\gamma}{1+\gamma}}\right\} + \lambda P(\theta). \tag{1}$$

$P(\theta)$ としては，L1 正則化 (Tibshirani 1996)，elastic net (Zou and Hastie 2005), SCAD (Fan and Li 2001) といった様々な正則化を考えることが可能である．以下に，具体的に，線形回帰・ロジスティック回帰・ポアソン回帰それぞれについての詳細を述べる．

2.1.　線形回帰　(1) の最適化は，通常の方法では，推定アルゴリズムの導出自体が困難となる．そこで，MM アルゴリズム (Hunter and Lange 2004) の考え方を用い，スパース正則化を加えた場合でも，目的関数を単調に減少させる推定アルゴリズム の導出を行った．数値実験および実データ（マイクロアレイデータ）の解析において，比較手法 (Alfons et al. 2013; Khan et al. 2007) に比べて非常に大きな改善を見せた (Kawashima and Fujisawa 2017)．この結果は昨年度の連合大会で報告した．

2.2. ロジスティック回帰　Kanamori and Fujisawa (2015) では，回帰問題において，外れ値の割合が説明変数 x に依存する場合は，位置・尺度分布に限定 したときのみ，スーパーロバスト性を示した．しかし，(1) に基づく回帰では，そのような仮定は必要なく，ロジスティック回帰の場合でも，スーパーロバスト性を得ることが可能である．線形回帰の場合と同じように，MM アルゴリズムの一例である，Bohning and Lindsay (1988) の方法を用いて，目的関数を単調に減少させる推定アルゴリズム の導出を行った．アルゴリズムの収束性に関しても，Miral (2013) の結果を用いることで，大域的収束性 が得られた．

2.3.　ポアソン回帰　ポアソン回帰の場合でも，Kanamori and Fujisawa (2015) の仮定は必要なく，外れ値の割合が説明変数 x に依存する場合でも，スーパーロバスト性を得ることが可能 である．しかしながら，ポアソン回帰の場合は，$\left(\int f(y|x_i;\theta)^{1+\gamma} dy\right)^{\frac{\gamma}{1+\gamma}}$ の項で，超幾何級数の計算が必要となる．そのため，(1) に基づく限り，通常の方法では効率的な推定アルゴリズムの導出が困難 となる．そこで，経験推定を経由することなく，直接，(1) の母集団レベルを最小化するオンライン最適化 の枠組みを用いて，この問題を克服した．特に，オンライン最適化の手法の中でも，非凸かつ滑らかでない目的関数を対象とした，Randomized Stochastic Projected Gradient (Ghadimi et al. 2016) を用いることで，大域的収束性を持つような，オンライン推定アルゴリズム を導出した．