ベイジアンモデルへの経験ベイズ修正

林　　由　子

要旨

James-Stein 推定量は最尤推定量よりも，より良い推定値をもたらす縮小推定量として多くの関心を集めてきた。この James-Stein 推定量は経験ベイズ法を支持する論拠としても非常に有益である。しかしこれに対して，大規模データで取り扱われる問題においては，対象となる分布が多数存在し，十分に近いとは考えられない平均を持つ混合分布となる。それゆえ，James-Stein 推定量を適用することは適切ではない。こうした大規模データを対象とした問題に対して局所的な修正を試みた研究が近年見られる。一方，ベイジアンモデルを考えた場合，事前分布および超事前分布の設定は結果に大きく影響をもたらす。本稿では，ベイジアンモデルの事後的予測確率に基づいた階層ベイズモデルを対象として超事前分布を仮定することなく，複雑な形状を取る超事前分布に対する局所的経験ベイズ修正の適用を試みた。その結果，良好な修正結果が得られることが示された。

1　は　じ　め　に

Efron（2011）は，Tweedie's formula に経験ベイズ法を適用したバイアスの局所的修正方法を提案した[1)]。これは，事前分布を特定することなく修正を行うものであり，周辺分布の推定値を用いた経験ベイズ法に基づいた修正方法である。

この Efron（2011）による局所的経験ベイズ修正は，ベイズモデルに基づいているにも関わらず事前分布を特定する必要がないという点において優れた手法である。さらに Efron（2011）では，周辺分布の密度関数の推定段階において一般に用いられるカーネル密度関数や混合分布のように周辺分布に正規分布を仮定することなく，Lindsey's Method を用いて周辺分布の推定を行っている。この周辺分布の密度関数の推定方法については，その後いくつかの展開が見られる。たとえば Wager（2014）ではノンパラメトリックな推定方法を，Simon and Simon（2013）では，高次の近似に基づいた周辺分布の密度関数の推定方法の提案を行っている。また Tan et al.（2015）では Simon and Simon（2013）の近似方法にデータ間の相関を考慮した周辺分布の推定に対する考察が行われている。

一方，大規模データを対象とした研究においても多くのベイジアンモデルが用いられており，たとえば大規模データを分析する際に重要な分析手法である False Dicovery Rate（FDR）に対しても，ベイジアンモデルを対象とした Bayesian FDR が開発されている

1）この修正を Efron（2011）は Robbins（1956）が1947年の Tweedie の論文を引用していることから Tweedie's formula と呼んでいるが，Koenker and Gu（2016）では，このベイズ修正は1926年にすでに Dyson によってもたらされていることを指摘している。

（Müller et al.（2007）。また外れ値除去を考慮するために，t 分布を用いて大規模データを推定する際には，解析的に解くことが困難であるため，ベイジアンモデルを用いる必要がある（Gottardo et al.（2006）, Hayashi et al.（2016））。それゆえ，大規模データを対象とした階層構造を持つベイジアンモデルに対する局所的経験ベイズ修正の適用を行うことは大変重要である。

そこで本稿では，ベイジアンモデルから得られた事後的予測確率に基づいた局所的経験ベイズ修正の適用可能性についての検討を行う。本稿の構成は以下のとおりである。まず２章において経験ベイズ法について概説を行った後，Efron（2011）による経験ベイズ修正の紹介を行っていく。３章においてベイジアンモデルから得られた事後確率への経験ベイズ修正の適用可能性を提案し，４章でシミュレーションデータによる結果を示す。そして５章でまとめを行う。

2　経験ベイズ修正

以下ではまず経験ベイズ法についての概説を行っていく。ベイズの定理を応用し分布に適用した以下の関係がベイズ統計学において中心的役割を果たす。経験ベイズ法は，以下のような事前分布のパラメター θ が階層構造を持ったランダムパラメターであるモデルに対して，周辺分布の推定値を用いるものである。

Stein パラドックスは最尤推定量が最良の推定量とならないことを示したものとして知られている。この最尤推定量よりも James-Stein 推定量は，位置パラメターについての Gaussian 階層モデルを想定した場合，３以上の位置パラメターを推定する際に優れた結果をもたらす。さらにデータ $X_i \sim N(\mu_i, 1)$ に対して，位置パラメターの超パラメターに標本からの平均値を用いるという Lindley（1962）の提案により一般には以下の形式で用いられる。それゆえこの場合にはパラメターの数は４以上をあつかう。

$$\mu_k^{JS} = \bar{X} + \left\{1 - \frac{K-3}{\sum_{k=1}^{K}(X_k - \bar{X})^2}\right\}^{+}(X_k - \bar{X}) \qquad (1)$$

ここで $\bar{X}$ は，X_k の周辺分布の平均である。また“＋”はカッコ内が $max\{0, a\}$ となる値を取ることを意味する。

Efron and Morris（1973）は経験ベイズ法を用いた James-Stein 推定量の導出を行い，経験ベイズ法の重要性を指摘した。彼らの導出は以下のとおりである。それぞれの位置母数（μ_k）が超事前分布の平均（μ）を持つ正規分布にしたがっているならば（$\mu_k = \mu + \delta_k$, $\delta_k \sim N(0, WI)$），μ_k の事後分布は，ベイズの定理より以下となる。

$$\mu_k^{B} = X_k - \frac{1}{1+W}(X_k - \mu) \qquad (2)$$

この μ の不偏推定量は X_k の周辺分布の平均 $\bar{X}$ である。周辺分布の分散（$1+W$）の逆数は逆カイ二乗分布に従うことから期待値が $K-2$ となる。それゆえ不偏推定量を用いた経験ベイズ推定が James-Stein 推定量と同一であることが示される。つまり，Efron and Morris（1973）は，経験ベイズ推定によって最尤法で得られる推定量よりも総二乗誤差を

損失の基準とした場合に良い推定量が得られることから，経験ベイズ法の重要性を示したのである。

2.1 局所的経験ベイズ修正

James-Stein 推定量が位置母数の超パラメターに対して縮約されるのに対して，Efron (2011) は Robbins (1956) で展開された Tweedie の分布に基づく局所的ベイズ修正に対して，推定された周辺分布を用いた経験ベイズ法の適用を行った。ここでは Robbins (1956) のベイズ修正及び Efron (2011) で展開された経験ベイズ修正についての紹介を行っていく。

以下の尤度が Tweedie 分布に基づくモデルを考える。

$$\begin{cases} \eta \sim g(\eta) \\ y|\eta \sim f_\eta(y) = \exp[\eta y - \psi(\eta)] f_0(y) \end{cases} \tag{3}$$

上のモデルのもとで，η の事後分布は以下となる。

$$g(\eta|y) = f_\eta(y) g(\eta) / f(y) \tag{4}$$

但し，$f(y) = \int f_\eta(y) g(\eta) d\eta$ である。

この事後分布は，$\lambda(y) = \log(f(y)/f_0(y))$ とおくことで，以下のように整理できる。

$$g(\eta|y) = e^{\eta y - \lambda(y)} g(\eta) e^{-\psi(\eta)} \tag{5}$$

いま，指数分布族を Hogg and Tanis (2009) に従って以下のように定義する。

$$f(x|\theta) = exp[p(\theta)K(x) + S(x) + q(\theta)] \tag{6}$$

このとき $w = \sum_{i=1}^{n} K(x_i)$ について

$$f(w|\theta) = R(w) exp[p(\theta)w + nq(\theta)] \tag{7}$$

となるので，期待値と分散はそれぞれ

$$\begin{cases} E(w) = -n \dfrac{q'(\theta)}{p'(\theta)} \\ VAR(w) = n \dfrac{1}{p'(\theta)^3} \{p''(\theta) q'(\theta) - q''(\theta) p'(\theta)\} \end{cases} \tag{8}$$

となる。これを用いて事後分布において y を上式の θ とおくと，$p(y) = y$, $q(y) = \lambda(y)$ となるので，期待値および分散は以下のように整理される。

$$\begin{cases} E(\eta|y) = \lambda'(y) \\ VAR(\eta|y) = \lambda''(y) \end{cases} \tag{9}$$

$f(y)$ と $f_0(y)$ の対数値をそれぞれ $l(y) = \log(f(y))$, $l_0(y) = \log(f_0(y))$ とおくと η の事後平均および事後分散は以下となる。

$$\eta|y \sim (l'(y) - l_0'(y), l''(y) - l_0''(y)) \tag{10}$$

正規分布を仮定すると $(y \sim N(\mu, \sigma^2))$, $\mu = \sigma^2 \eta$ であるから，$\eta = \dfrac{\mu}{\sigma^2}$ となり，$l_0(y) = -\dfrac{y^2}{2\sigma^2} - \log(\sqrt{2\pi\sigma^2})$ であるから，以下が成立する。

$$\mu|y \sim (y+\sigma^2 l'(y), \sigma^2(1+\sigma^2 l''(y))) \tag{11}$$

それゆえ，位置パラメターについて以下のベイズ修正が導出される。

$$E(\mu|y)=y+\sigma^2 l'(y) \tag{12}$$

つまり，平均 μ の事後平均は，データの情報による不偏推定値にベイズ修正項を加えたものとなる。Efron (2011) では，周辺分布に基づく修正項を観測されたデータから推定する経験ベイズ法に基づいたバイアスの修正を提案している。

$$\hat{\mu}_i \equiv \hat{E}(\mu_i|y_i)=y_i+\sigma^2 \hat{l}'(y_i) \tag{13}$$

$l'(y_i)$ の推定値を得るために，Efron (2011) では，y を K 個のビンに分けた度数が独立なポアソン分布に従うとの仮定に基づく Lindsey's Method による，J 次のポアソン回帰から周辺分布の推定を行っている。Efron (2011) において指摘されているように，2次のスプラインを仮定すると，このベイズ修正は，James-Stein 推定量と同様に位置母数のハイパーパラメターに向かって縮約される。

以上のことから，y_i が平均 μ_i，分散1の正規分布に従うならば，μ_i の事前分布にかかわりなく，μ_i の経験ベイズ修正は以下で与えられる。

$$\hat{\mu}_i=y_i+\hat{l}'(y_i) \tag{14}$$

2.2 修正結果の評価

(13) 式で与えられた経験ベイズ推定値は，観測値の数が1の時には最尤法に，無限大となれば真の尤度が得られるため真のベイズ推定値となる。この推定値を用いることで生じる損失について Efron (2011) では，以下のように損失の評価を行っている。

y の周辺分布が既知であるとき，ある y_0 のもとでのベイズ修正値は，$\mu^*(y_0)=y_0+l'(y_0)$ であるのに対して，観測値に基づく経験ベイズ修正値は $\hat{\mu}(y_0)=y_0+\hat{l}'(y_0)$ となる。ここで $E\{(\mu-\mu^*(y_0))|y_0, Y\}=0$ を仮定すると，y_0 で条件付けられた regret（期待値の損失）は以下で得られる。

$$\begin{aligned} R(y_0) &= E[(\mu-\hat{\mu}(y_0))^2-(\mu-\mu^*(y_0))^2|y_0] \\ &= E[(\hat{l}'(y_0)-l'(y_0))^2|y_0] \end{aligned} \tag{15}$$

それゆえ y_0 で条件付けられた $\hat{l}'(y_0)$ の分散となる。

Efron (2011) は，ポアソン回帰による周辺分布の推定値 ($\hat{l}'(y_0)$) に基づいたこの期待損失の理論的近似値を以下のように示している。J 次のポアソン回帰による周辺分布の推定値にそれぞれのビンの中央にくる値 x_k による説明変数 $F(x)=(1, x, x^2, \ldots, x^J)$ を x で微分した $G(x)=(0, 1, 2x, \ldots, Jx^{(J-1)})$ を用いて，$y_0=x_k$ のもとで以下が得られる。

$$\hat{l}'(y_0)=G(x_k)\hat{\beta} \tag{16}$$

それゆえ regret は，最尤法によるポアソン回帰のパラメター β の推定値の分散共分散行列用いて以下から得られることが示されている。

$$\begin{aligned} R(y_0) &\approx var(\hat{l}'(y_0)) \\ &\doteq G(x_k)DG(x_k)^t \end{aligned} \tag{17}$$

ここで $D=\widehat{cov}(\hat{\beta})$ は最尤法による漸近的推定値をあらわす。

3　事後的スコアの分布の推定

本稿では，事後確率から事後的な統計スコアを求めることで超事前分布の仮定を置くことなく，Efron（2011）の経験ベイズ修正の適用を試みる。つまり超事前分布が未知にもかかわらず個々の位置母数の事後的な統計スコアの分布から位置母数についての修正を行うものである。それゆえ（14）式における y_i が個々の位置母数の事後分布から得られた事後的統計スコアに，$\hat{l}'(y)$ を事後的統計スコアの分布の推定値として修正を行う。

この個々の位置母数についての事後的統計スコアが指数分布族に属しない場合，Efron（2011）の経験ベイズ修正は適用できない。しかし大規模データを分析する際には個々のスコアは正規分布よりもむしろ t 値として得られる場合が多い。この t 分布は指数分布族に属しないため，Efron（2011）で提案された経験ベイズ修正は適用できない。このことは，t 分布の対数密度関数はその裾において凹関数とはならないため，観測値が十分に母数に近くない限り，経験ベイズ修正を適用することで偏りがさらに大きくなってしまうことを意味する。

この問題に対して Efron（2011）では，t 分布の分布関数より下側確率を求め，これを標準正規分布の逆関数を用いて標準正規分布に従う z 値を得ることで経験ベイズ修正の適用を行う提案がなされている。

$$z=\Phi^{-1}(F_\nu(t)) \qquad t\sim t_\nu(\kappa) \tag{18}$$

ここで Φ は，標準正規分布の分布関数を，$F_\nu(t)$ は自由度 ν，位置パラメター κ の t 分布の分布関数を表す。本稿においてもこの変換に基づいて，事後的統計量を標準正規分布に従うスコアへと変換することを試みる。

尤度が正規分布に従うベイジアンモデルにおいて，尺度母数が既知でない場合には無情報事前分布を用いたとしても事後分布は t 分布に従う。

データが t 分布に従うならば，標本平均を標本分散の平方根で標準化したものは z 値ではなく t 値となる。それゆえ，推定された分散の値を用いた標準化によって z 値を得ることはできない。この問題に対して Efron and Zhang（2011）では，頑強標準偏差を用いた標準化を行っている。本稿においても彼らの手法に従って，事後確率からこの頑強標準偏差の推定を行う。

まず，null グループに属するデータの平均は 0 であると仮定し，以下のような平均を表すパラメター β に対する事後的予測確率を考える（Ventrucci and Scott（2011））。

$$p_B\equiv Pr\{\beta>0|data\} \tag{19}$$

Null グループに属する時この事後的予測確率が一様分布に従う性質から，p_B はこれの下側確率とみなすことができる（Hayashi et al.（2016））。

p_B は，十分に大きな A 回の Markov chain Monte-Carlo（MCMC）イテレーションから以下のように求めることができる。

$$p_B\doteqdot\frac{1}{A}\sum_{rep=1}^{A}I\{\beta_{rep}>0|y\} \tag{20}$$

ここで β_{rep} は MCMC の replication を表す。

それゆえここで標準正規分布の逆関数を用いた以下の事後的スコアは標準正規分布に従う。

$$B = \Phi^{-1}(p_B) \tag{21}$$

しかし事後的予測確率を（20）式から推定する場合，イテレーションの回数が小さい場合には p_B は，0または1の値をとり，有限のスコアを得ることができない。さらに，十分に大きなイテレーション回数，たとえば $A=10{,}000$ を用いた場合の有限の最小スコアは，$p_B=0.0001$ つまり $B=-3.719$ となり，これ未満の値を得ることができない。そこで本稿では，以下の手続きから事後的スコアの推定を行っていく。

以下の（22）式で定義される $p_{B_i}=0.50$ となる推定値を用いる。つまり β_i の推定値は事後分布の中央値（β_i^*）とする。

$$p_{B_i} \equiv Pr\{\beta_i > \beta_i^* | data\} \tag{22}$$

これを正規分布の逆関数をもちいて変換を行うと，中央値（β_i^*）を平均にもつ正規分布に従う変数に変換を行ったものとなる。これにより，標準正規分布に従うとの仮定を満たす。それゆえ事後的スコアは，事後分布の中央値（β_i^*）を頑強標準偏差（S_R）で割った以下の式から推定を行う。

$$B_n = \frac{\beta_i^*}{S_R} \tag{23}$$

Efron and Zhang（2011）では，16パーセンタイルから84パーセンタイルの距離の2分の1の値を頑強標準偏差として用いている。全て null グループならば，p_B も p 値と同様に一様分布に従うことから，16パーセンタイルから84パーセンタイルの距離は標準正規分布の標準偏差2個分となる。そこで本稿においても β_i の事後的スコアの分布の16パーセンタイルから 84パーセンタイルの距離の2分の1の値を頑強標準偏差として用いる。

4 シミュレーション分析

本節ではシミュレーションデータに基づいて事後的予測確率値から事後的統計スコアを求め，Efron（2011）のベイズ修正の適用を行っていく。

Efron（2011）は，μ の事前分布に $e^{-\mu}$ を用いたシミュレーションに基づいて計算された（15）式と（17）式の regret の比較から $N=1000$ のデータセットを用いた場合，99パーセンタイルまでの近似がかなり良いことを示している[2)]。そこで本稿では，シミュレーションには各データセットの標本数を30とし，データセット数には1000を用いた（$N=1000$）。このうち100のデータセットについて標本数15の2グループに分け，non-null データセットを作成した。また簡単化のため2標本に等分散を仮定し，以下の線形モデルを用いて β_j についての事後確率を求める。

2）しかし99.9パーセンタイルといった極値に対しては経験ベイズによる近似があまりよくないことに注意を要すると述べられている。

$$y_{ji}=\alpha_j+\beta_j D_i+u_{ji} \qquad j=1, \ldots, 1000.\ i=1, \ldots, 30. \tag{24}$$

ここで誤差項は標準正規分布を用いた $u_{ji}\sim N(0, 1)$。また D_i は，null グループのデータは $D_i=0$ を non-null グループのデータは $D_i=1$ をとる状態変数である。

4.1 推定モデル

下層レベルの推定には尤度に正規分布を，位置パラメターと尺度パラメターの事前分布には，それぞれ Jefferey の事前分布を用いた。モデルの概要は以下のとおりである。

$$\begin{cases} f(y_{ji}|\mu_{ji}, \sigma_j)=N(\mu_{ji}, \sigma_j^2) \\ \alpha_j, \beta_j \overset{D}{\sim} Uniform \\ \log(\sigma_j) \overset{D}{\sim} Uniform \end{cases} \tag{25}$$

ここで，$\mu_{ji}=\alpha_j+\beta_j D_i$ である。推定には OpenBUGS（v.3.2.3）を用いた。イタレーション回数は，11,000回であり，最初の 1,000回分を取り除いている。

また事後分布の統計スコアの分布の推定には，Efron（2011）と同様に Lindsey's Method により，R(3.1.1) の一般化線形モデルのパッケージ（glm）を用いてヒストグラムの度数からポアソン回帰推定を行った。

4.2 推定結果

このモデルで全てのデータが null グループに属しているならば，（23）式の頑強標準偏差を用いて得られたものは（21）式で得られた統計スコアに近似される。図１は間隔0.05を用いたヒストグラムであり，図２はこれらのヒストグラムを用いて２次と５次のポワソン回帰分析によって得られた事後的統計スコアの分布の推定結果である。これらの結果より，頑強標準偏差を用いた結果は，逆関数を用いた結果に対してかなり良い近似となっていることが分かる。このことは，（23）式のスコアを用いることを支持するものである。

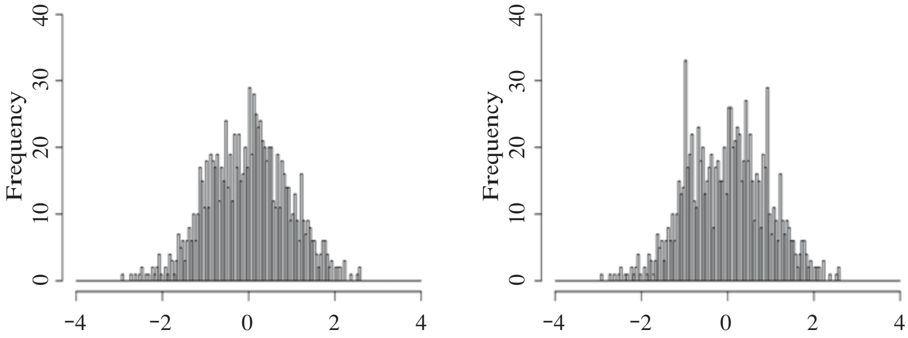
次に non-null グループの分布の違いによるモデルの動きを見るために，２種類の平均と３種類の分布についてのシナリオを用いて比較を行っていく。

図３はそれぞれのシミュレーションモデルから発生させたデータに間隔0.25を用いたヒストグラムとそれらのヒストグラムの度数を用いた５次のポアソン回帰から推定したスプラインである。

＜シミュレーションシナリオ＞

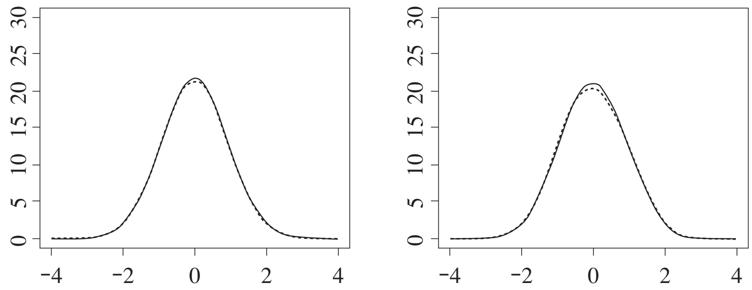
１．900の $\beta_j=0$，100の $\beta_j=1$
２．900の $\beta_j=0$，100の $\beta_j=3$
３．900の $\beta_j=0$，100の $\beta_j \overset{D}{\sim} N(1, 1)$
４．900の $\beta_j=0$，100の $\beta_j \overset{D}{\sim} N(3, 1)$
５．900の $\beta_j=0$，100の $\beta_j \overset{D}{\sim} Expon(1)$
６．900の $\beta_j=0$，100の $\beta_j \overset{D}{\sim} Expon(1/3)$

ここで指数分布は $Expon(\lambda)=\lambda exp^{-\lambda x}$ で定義されたものである。



(a) 頑強標準偏差によるスコア　　　(b) 逆関数によるスコア

図１：スコアのヒストグラム（全て null グループ）



(a) ２次のポアソン回帰　　　(b) ５次のポアソン回帰

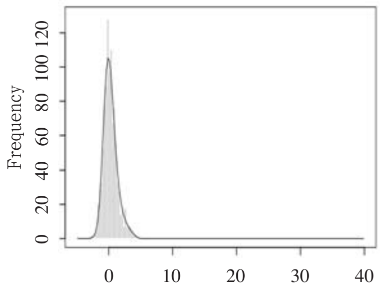
図２：ポアソン回帰による推定された事後的スコアの分布（全て null グループ）。直線は頑強標準偏差を用いた，点線は逆関数によって得られたスコアを用いた結果を表す

図４は，それぞれのベイズ修正項をプロットしたものである。仮に２次のスプラインであるならば，ベイズ修正項は直線に近づく。これに対して図４より，いずれもシナリオにおいても２つのモードに基づく修正がなされていることが分かる。つまり局所的修正が働いていることが見てとれる。また図５は経験ベイズ修正の前後を比較したものであり，特にシナリオ２の修正結果において顕著に修正されていることが分かる。

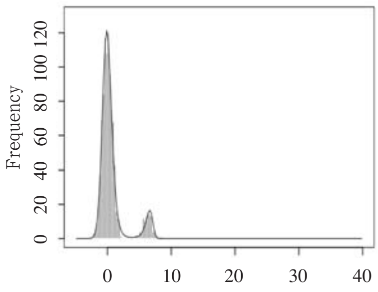
５　ま　と　め

大規模データで取り扱われる問題においては，対象となる分布が多数存在し，十分に近いとは考えられない平均を持つ混合分布となる。こうした大規模データを対象とした問題に対して局所的な修正を試みた研究が近年いくつか見られる。本稿では，Efron (2011) による局所的経験ベイズ修正のベイジアンモデルへの適用可能性について検討を行った。

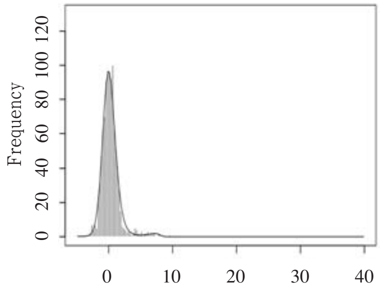
大規模データに対してベイジアンモデルを考えた場合，Efron (2011) で設定されている事前分布が超事前分布の役割を果たす。これらの事前情報は推定結果に大きく影響をも



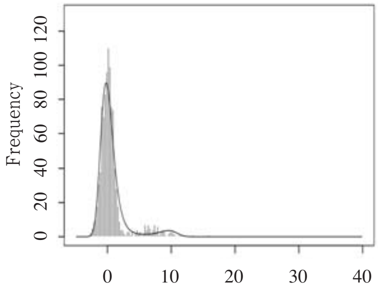
(a) non-null $\beta_j=1$



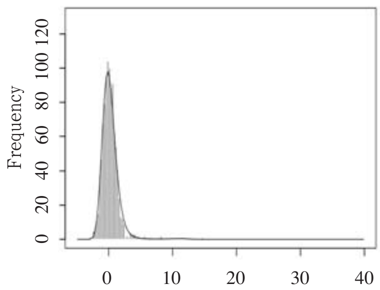
(b) non-null $\beta_j=3$



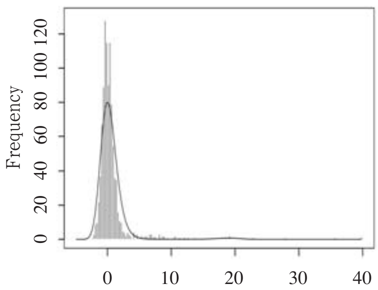
(c) non-null $\beta_j \overset{D}{\sim} N(1,\ 1)$



(d) non-null $\beta_j \overset{D}{\sim} N(3,\ 1)$

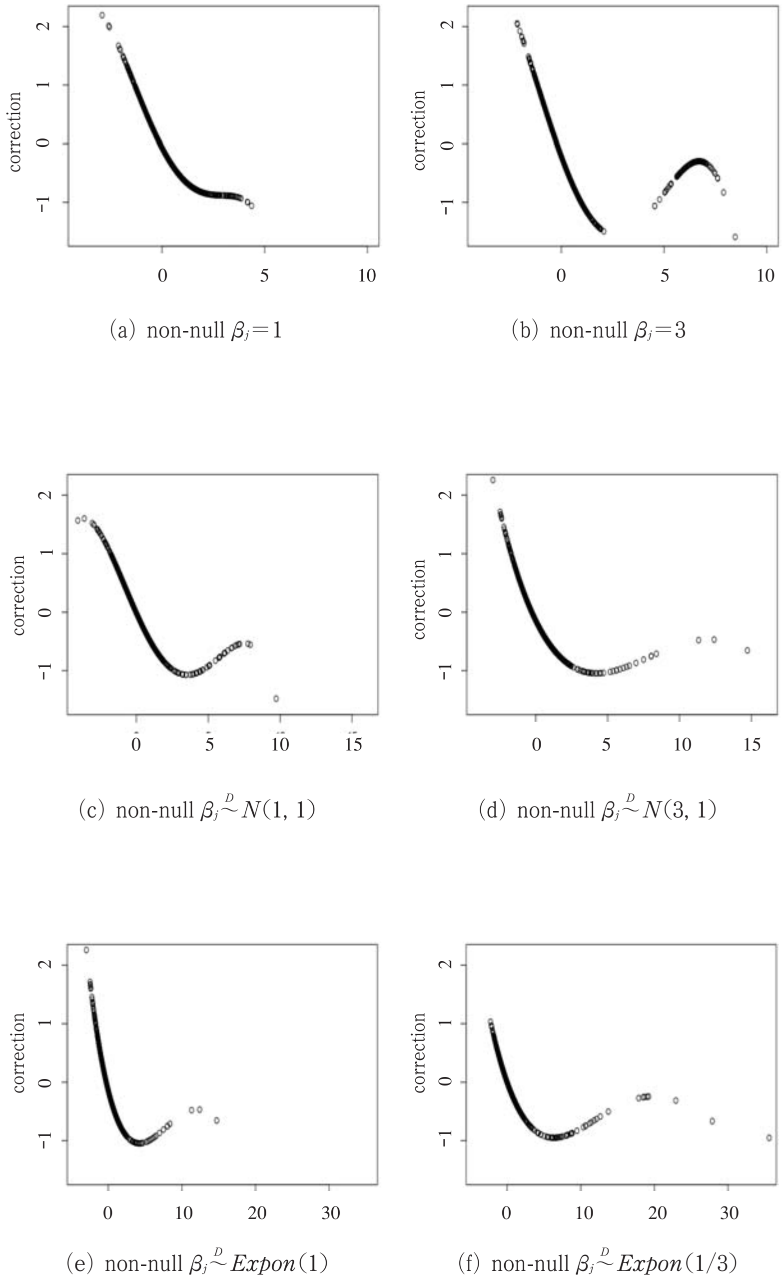


(e) non-null $\beta_j \overset{D}{\sim} Expon(1)$



(f) non-null $\beta_j \overset{D}{\sim} Expon(1/3)$

図３：頑強標準偏差を用いた事後的スコアのヒストグラムと推定された分布



(a) non-null $\beta_j=1$

(b) non-null $\beta_j=3$

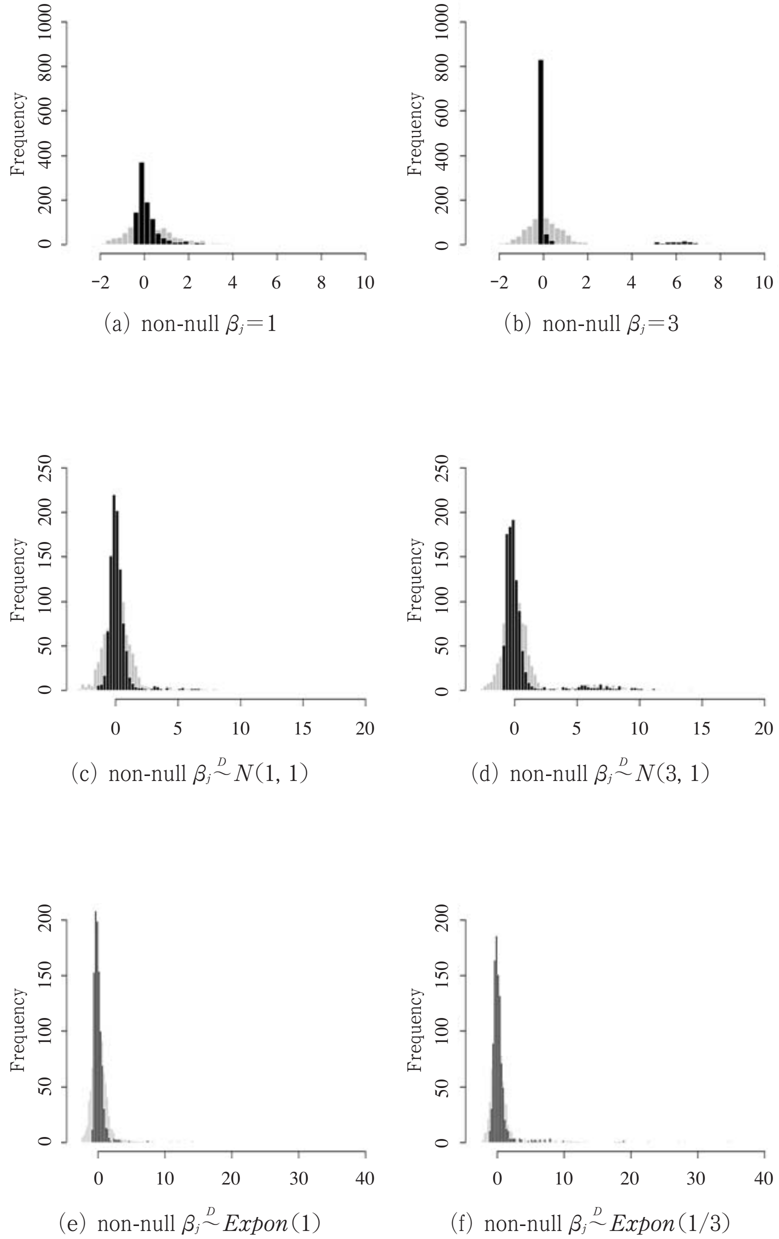
(c) non-null $\beta_j \overset{D}{\sim} N(1, 1)$

(d) non-null $\beta_j \overset{D}{\sim} N(3, 1)$

(e) non-null $\beta_j \overset{D}{\sim} Expon(1)$

(f) non-null $\beta_j \overset{D}{\sim} Expon(1/3)$

図４：経験ベイズ修正項



(a) non-null $\beta_j=1$　(b) non-null $\beta_j=3$

(c) non-null $\beta_j \overset{D}{\sim} N(1, 1)$　(d) non-null $\beta_j \overset{D}{\sim} N(3, 1)$

(e) non-null $\beta_j \overset{D}{\sim} Expon(1)$　(f) non-null $\beta_j \overset{D}{\sim} Expon(1/3)$

図５：修正前（light grey）と修正後（black）のヒストグラム

たらすため，その設定には慎重を要する。本稿では，ベイジアンモデルの事後的予測確率に基づいた局所的経験ベイズ修正の適用を行うことで，超事前分布を仮定することなく，経験ベイズ修正の適用を試みた。その結果，良好な修正結果が得られることが示された。

参考文献

1. Efron, B. (2010), *Large-scale inference empirical Bayes methods for estimation, testing, and prediction*, Cambridge University Press.
2. Efron, B. (2011), Tweedie's formula and selection bias. Journal of the American Statistical Association 105, 1602-1614.
3. Efron, B. and Morris, C. (1973), Stein's estimation rule and its competitors - an empirical Bayes approach. Journal of the American Statistical Association 68, 117-130.
4. Efron, B. and Morris C. (1975), Data analysis using Stein's estimator and its generalizations, Journal of the American Statistical Association, 70, 350, 311-319.
5. Efron, B. and Zhang, N. (2011), False discovery rates and copy number variation, Biometrika, 98, 2, 251-271.
6. Gottardo, R. et al. Bayesian robust inference for differential gene expression in microarrays with multiple samples. Biometrics 62, 10-18, 2006.
7. Hayashi, Y. et al. (2016), Robust Bayesian modelling for preprocessing large-scale data, unpublished manuscript.
8. Hogg, R. V. and Tanis, E. A. (2009), *Probability and Statistical Inference*, Pearson.
9. James, W. and Stein, C. (1961), Estimation with quadratic loss. Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, 361-379.
10. Koenker, R. and Gu, J. (2016), REBAYES: An R package for empirical Bayes mixture methods, http://jiayinggu.weebly.com/uploads/3/8/9/3/38937991/rebayes.pdf
11. Lindley, D. V. (1962), Discussion on Professor Stein's paper, Jouranl of Royal Statistical Society Series B 24 2 285-288.
12. Robbins, H. (1956), An empirical Bayes approach to statistics. In Proceedings of the Third Berkley Symposium on Mathematical Statistics and Probability, 1954-55, 1 157-163.
13. Müller, P., Parmigiani, G., and Rice, K. (2007) FDR and Bayesian multiple comparisons rules. In *Proceedings of the Valencia ISBA 8th World Meeting on Bayesian Statistics*. Oxford University Press.
14. Simon, N. and Simon, R. (2013), On estimating many means, selection bias, and the bootstrap. arXiv. 1311.3709v1
15. Tan, K. M., Simon, N. and Witten, D. (2015), Selection bias correction and effect size estimation under dependence. arXiv. 1405.4251v2
16. Wager S. (2014), A geometric approach to density estimation with additive noise, Statistica Sinca 24 533-554.
17. Ventrucci, M. and Scott, M. E. (2011), Multiple testing on standardized mortality ratios: a Bayesian hierarchical model for FDR estimation, Biostatistics, 12 1 55-67.