3つのパラメータをもつ汚染近傍とそのロバスト推測への応用

M2013SS001 安藤周平
指導教員：木村美善

1　はじめに

統計的分析を行う多くの場合，標本は仮定されたモデルからずれており，せいぜい近似的に満たされるだけである．従って，そうした仮定からの「ずれ」や「乖離」があったり，標本が汚染されていても，その影響力を小さくして，モデル分布のもとでの良さがさほど失われず，安心して使用できるロバスト推測の方法が望ましい．モデル分布からの「ずれ」や「乖離」を表現するための分布近傍として，Kakiuchi and Kimura (2012) が提案した 3 つのパラメータをもつ (c_1, c_2, γ)-汚染近傍（以後 (c_1, c_2, γ)-近傍と省略）があるが，これは従来用いられてきた各種分布近傍を特殊な場合として含む新たな近傍であり，そのロバスト推測への応用が期待されている（[3] 参照）．本論文では，(c_1, c_2, γ)-近傍とそのロバスト推定・検定への応用について考察する．そして，正規分布の平均のミニマックス検定と回帰推定量（s-，τ-，CM-推定量）の最大漸近バイアスについてシミュレーション評価を行う．

2　ロバスト推測

$\mathbb{R}$ を実数直線，$\mathscr{B}$ を $\mathbb{R}$ の部分集合からなるボレル集合族，$\mathscr{M}$ を $(\mathbb{R}, \mathscr{B})$ 上の確率分布の全体とする．$X_1, \ldots, X_n$ を分布 G に従う互いに独立な標本とする．G が分布 F° に近似的に等しいことがわかっているとき，F° の未知の特性値に関する推測を行う場合，ロバスト推測理論では，F° の近傍 $\mathscr{P}$ を導入し，$G \in \mathscr{P}$ として，その特性値に関する統計的推測を行う．ロバスト推測で達成すべきこととして，Huber は次のように主張している（[7] 参照）．

1. Efficiency: 仮定されたモデル分布のもとで，高い効率をもつ．
2. Stability: 仮定されたモデル分布からのわずかな「ずれ」や「乖離」があっても，安定した振る舞いをする．
3. Breakdown: 仮定されたモデル分布からのある程度大きな「ずれ」や「乖離」があっても極端に効率が下がらず，破綻しない．

3　(c_1, c_2, γ)-近傍

(c_1, c_2, γ)-近傍は次のように定義される：

$$\begin{aligned}\mathscr{P}_{c_1,c_2,\gamma}(F^\circ) =& \{G \in \mathscr{M} | c_1 F^\circ\{A\} \le G\{A\} \\ &\le c_2 F^\circ\{A\} + \gamma, {}^\forall A \in \mathscr{B}\}\end{aligned} \tag{1}$$

ただし，c_1, c_2, γ はパラメータで $0 \le c_1 \le 1-\gamma \le c_2 < \infty$，$0 \le \gamma < 1$ を満たす．この (c_1, c_2, γ)-近傍は

$$\begin{aligned}\mathscr{P}_{c_1,c_2,\gamma}(F^\circ) = \{G \in \mathscr{M} | G\{A\} \le \min(c_2 F^\circ\{A\} + \gamma, \\ c_1 F^\circ\{A\} + 1 - c_1), {}^\forall A \in \mathscr{B}\}\end{aligned} \tag{2}$$

のように表現することができる．したがって

$$v_h\{A\} = \begin{cases} h(F^\circ\{A\}) & (\phi \ne A \in \mathscr{B} \text{ のとき}) \\ 0 & (A = \phi \text{のとき}) \end{cases}$$

$$h(t) = \min(c_2 t + \gamma, c_1 t + 1 - c_1),\ 0 \le t \le 1$$

と定義すると，Bednaruski(1981) により v_h は特殊容量となり，次式が成り立つ．

$$\mathscr{P}_{c_1,c_2,\gamma}(F^\circ) = \{G \in \mathscr{M} | G\{A\} \le v_h\{A\}, {}^\forall A \in \mathscr{B}\}. \tag{3}$$

4　(c_1, c_2, γ)-近傍の特徴づけ

$\mathscr{M}_c (\subset \mathscr{M})$ は $(\mathbb{R}, \mathscr{B})$ 上の絶対連続な分布からなる集合，$F^\circ \in \mathscr{M}_c$ の密度関数を f° とする．このとき

$$\begin{aligned}\mathscr{P}_{c_1,c_2,\gamma}(F^\circ) = \{G = (1-\gamma)F + \gamma K \in \mathscr{M} \\ | F \in \mathscr{F}_{c1,c2,\gamma}(F^\circ), K \in \mathscr{M}\},\end{aligned} \tag{4}$$

と表される．ただし

$$\mathscr{F}_{c_1,c_2,\gamma}(F^\circ) = \left\{ F \in \mathscr{M}_c \mid \frac{c_1}{1-\gamma} f^\circ \le f \le \frac{c_2}{1-\gamma} f^\circ \right\}$$

であり，F の密度関数を f とする．

$\mathscr{F}_{c_1,c_2,\gamma}(F^\circ)$ における最小分布 F_L° と最大分布 F_R° はそれぞれ

$$F_L^\circ(x) = \begin{cases} \frac{c_2}{1-\gamma} F^\circ(x), & x \le x_L \\ \frac{c_1}{1-\gamma} F^\circ(x) + \left(1 - \frac{c_1}{1-\gamma}\right), & x > x_L \end{cases} \tag{5}$$

$$F_R^\circ(x) = \begin{cases} \frac{c_1}{1-\gamma} F^\circ(x), & x \le x_R \\ \frac{c_2}{1-\gamma} F^\circ(x) + \left(1 - \frac{c_2}{1-\gamma}\right), & x > x_R \end{cases} \tag{6}$$

によって与えられる．ここで

$$x_L = (F^\circ)^{-1} \left(\frac{1-\gamma-c_1}{c_2-c_1} \right), \tag{7}$$

$$x_R = (F^\circ)^{-1} \left(\frac{c_2-1+\gamma}{c_2-c_1} \right). \tag{8}$$

5　正規分布の平均のロバスト検定

$X_1, \ldots, X_n$ を $F_\mu = N(\mu, 1)$ に近似的に従う独立な標本とするとき，(c_1, c_2, γ)-近傍を用いて $\boldsymbol{X} = (X_1, \ldots, X_n)$ に基づく μ のロバスト検定問題を次のように定義する（[1] 参照）．

$$\begin{aligned} &H_0: \pounds(\boldsymbol{X}) \in \mathscr{P}^n_{c_1,c_2,\gamma}(F_0) \\ &H_1: \pounds(\boldsymbol{X}) \in \mathscr{P}^n_{c_1,c_2,\gamma}(F_\mu) \quad (0 < \mu_1 \le \mu) \end{aligned} \tag{9}$$

ここで，$\pounds(\boldsymbol{X})$ は $\boldsymbol{X} = (X_1, \ldots, X_n)$ の確率分布，$\mathscr{P}_{c_1,c_2,\gamma}(F_\mu)$ は F_μ の (c_1, c_2, γ)-近傍，$\mathscr{P}^n_{c_1,c_2,\gamma}(F_\mu)$ は

$\mathscr{P}_{c_1,c_2,\gamma}(F_\mu)$ の n 重積，μ_1 は指定された値を表す．この検定問題に対する検定 ψ の最大の大きさ $\alpha_\psi(\mu_0)$ と最小検出力 $\beta_\psi(\mu)$ は

$$\begin{aligned}\alpha_\psi(0) &= \sup\{E_{G_n}[\psi(\boldsymbol{X})] : G_n \in \mathscr{P}^n_{c_1,c_2,\gamma}(F_0)\},\\ \beta_\psi(\mu) &= \inf\{E_{G_n}[\psi(\boldsymbol{X})] : G_n \in \mathscr{P}^n_{c_1,c_2,\gamma}(F_\mu)\}\end{aligned} \tag{10}$$

によって定義される．ψ が水準 $\alpha(0 < \alpha < 1)$ であるとは，$\alpha_\psi(0) \leq \alpha$ を満たすことをいう．Ψ_α を水準 α 検定の全体からなる集合とするとき，$\psi^* \in \Psi_\alpha$ が水準 α のミニマックス検定であるとは

$$\beta_{\psi^*}(\mu) = \sup\{\beta_\psi(\mu) : \psi \in \Psi_\alpha\} \tag{11}$$

を満たすことをいう．

6 ミニマックス検定

π が特殊容量 v_h の v_0 に対する Radon-Nikodym 導関数 $\pi \in \frac{dv_h}{dv_0}$ であるとは $\forall t \geq 0$ に対して

$$tv_0(\pi > t) + v_\mu(\pi \leq t) = \inf\{tv_0(A) + v_\mu(A^c) \ : \ A \in \mathscr{B}\} \tag{12}$$

を満たすことをいう．また，(Q_0, Q_μ) が分布の対 $(\mathscr{P}_{c_1,c_2,\gamma}(F_{\mu_0}), \mathscr{P}_{c_1,c_2,\gamma}(F_\mu))$ に対する最も不利な分布の対であるとは $Q_0 \in \mathscr{P}_{c_1,c_2,\gamma}(F_{\mu_0}), Q_\mu \in \mathscr{P}_{c_1,c_2,\gamma}(F_\mu)$ であり，$\forall t \geq 0$ に対して

$$\begin{aligned}Q_0(\pi > t) &= \sup\{P(\pi > t) : P \in \mathscr{P}_{c_1,c_2,\gamma}(F_{\mu_0})\}\\ &= v_0(\pi > t)\\ Q_\mu(\pi \leq t) &= \sup\{P(\pi \leq t) : P \in \mathscr{P}_{c_1,c_2,\gamma}(F_\mu)\}\\ &= v_\mu(\pi \leq t)\end{aligned} \tag{13}$$

が成り立つことをいう．ここで，π は Q_μ の Q_0 に対する Radon-Nikodym 導関数 $\pi \in \frac{dQ_\mu}{dQ_0}$ である．v_μ の v_0 に対する Randon-Nikodym 導関数 π はまた Q_μ の Q_0 に対する Radon-Nikodym 導関数であることにも注意する．$(\mathscr{P}_{c_1,c_2,\gamma}(F_{\mu_0}), \mathscr{P}_{c_1,c_2,\gamma}(F_\mu))$ に対する最も不利な分布の対 (Q_0, Q_μ) が存在するとき，$(\mathscr{P}^n_{c_1,c_2,\gamma}(F_{\mu_0}), \mathscr{P}^n_{c_1,c_2,\gamma}(F_\mu))$ に対する最も不利な分布の対は (Q_0^n, Q_μ^n) により与えられる．ただし，Q_0^n, Q_μ^n は Q_0 , Q_μ の n 重積である．

また，Q_μ^n の Q_0^n に対する Radon-Nikodym 導関数は $\pi_n(\boldsymbol{x}) = \prod_{i=1}^n \pi(x_i)$ となる．このとき，検定問題 (9) に対する水準 α ミニマックス検定は

$$\varphi^*(\boldsymbol{x}) = \begin{cases} 0 & (\pi_n(\boldsymbol{x}) < \lambda^*_\alpha)\\ \xi & (\pi_n(\boldsymbol{x}) = \lambda^*_\alpha)\\ 1 & (\pi_n(\boldsymbol{x}) > \lambda^*_\alpha)\end{cases} \tag{14}$$

により与えられる．ただし，$E_{Q_0^n}[\varphi^*(\boldsymbol{X})] = \alpha$.

7 Radon-Nikodym 導関数の導出

特殊容量 v_μ の v_0 に対する Radon-Nykodym 導関数 π を導出する．f_μ を $F_\mu = N(\mu, 1)$ の密度導関数とすると

$$\frac{dF_\mu}{dF_0}(x) = \frac{f_\mu(x)}{f_0(x)} = \frac{f_0(x-\mu)}{f_0(x)} \tag{15}$$

となる．よって，$\frac{dF_\mu}{dF_0}(X)$ の F_0 と F_μ の下での分布関数を G_μ，K_μ とすると

$$\begin{aligned}G_\mu(x) &= F_0\left(\frac{dF_\mu}{dF_0}(X) \leq x\right) = F_0(\frac{1}{\mu}\log x + \frac{1}{2}\mu)\\ K_\mu(x) &= F_\mu\left(\frac{dF_\mu}{dF_0}(X) \leq x\right) = F_0(\frac{1}{\mu}\log x - \frac{1}{2}\mu)\end{aligned} \tag{16}$$

となる．いま，(12) の下限を与える集合を A_t，すなわち

$$tv_0(A_t) + v_\mu(A_t^c) = \inf\{tv_0(A) + v_\mu(A^c) \ : \ A \in \mathscr{B}\} \tag{17}$$

とすると (12) の右辺は $A = \phi$ のとき 1，$A = \Omega$ のとき t であるから

$$tv_0(A_t) + v_\mu(A_t^c) \leq \min(t, 1) \tag{18}$$

である．集合

$$D = \{t \mid tv_0(A_t) + v_\mu(A_t^c) < \min(t, 1)\tag{19}$$

は端点 Δ_1,Δ_2 を持つ区間となる．ただし Δ_1,Δ_2 は

$$\begin{aligned}\Delta_1 v_0(A_{\Delta_1}) + v_\mu(A^c_{\Delta_1}) &= \Delta_1\\ \Delta_2 v_0(A_{\Delta_2}) + v_\mu(A^c_{\Delta_2}) &= 1\end{aligned} \tag{20}$$

を満たす定数である．集合 A_t は

$$A_t = \begin{cases} \Omega & (0 \leq t \leq \Delta_1)\\ \left\{\frac{dF_\mu}{dF_0} > w(t)\right\} & (\Delta_1 \leq t \leq \Delta_2)\\ \phi & (\Delta_2 \leq t \leq \infty)\end{cases} \tag{21}$$

のように表すことができる．ここで，$w(t)$ は t の非減少関数である．

8 シミュレーション

$c_1 = 0.8$，$c_2 = 1.2$，$\gamma = 0.05$，$\mu = 2$ の場合を考える．このとき π は

$$\pi = \begin{cases} \Delta_1, & 0 \leq \frac{dF_\mu}{dF_0} < \left(\frac{c_1}{c_2}\right)\Delta_1\\ \left(\frac{c_2}{c_1}\right)\frac{dF_\mu}{dF_0}, & \left(\frac{c_1}{c_2}\right)\Delta_1 \leq \frac{dF_\mu}{dF_0} < \left(\frac{c_1}{c_2}\right)a_1\\ a_1, & \left(\frac{c_1}{c_2}\right)a_1 \leq \frac{dF_\mu}{dF_0} < a_1\\ \frac{dF_\mu}{dF_0}, & a_1 \leq \frac{dF_\mu}{dF_0} < a_2\\ a_2, & a_2 \leq \frac{dF_\mu}{dF_0} < \left(\frac{c_2}{c_1}\right)a_2\\ \left(\frac{c_1}{c_2}\right)\frac{dF_\mu}{dF_0}, & \left(\frac{c_2}{c_1}\right)a_2 \leq \frac{dF_\mu}{dF_0} < \left(\frac{c_2}{c_1}\right)\Delta_2\\ \Delta_2, & \left(\frac{c_2}{c_1}\right)\Delta_2 \leq \frac{dF_\mu}{dF_0} < \infty\end{cases} \tag{22}$$

となる．ただし，$\Delta_1 = 0.1931$，$\Delta_2 = 5.1772$，$a_1 = 0.3096$，$a_2 = 3.2303$ である．図 1 はこの π のグラフである．$(\mathscr{P}_{c_1,c_2,\gamma}(F_0), \mathscr{P}_{c_1,c_2,\gamma}(F_\mu))$ に対する最も不利な分布の対 (Q_0^*, Q_μ^*) を構成する．構成にあたっては Kakiuchi and Kimura (2012) の Theorem 3.1 の結果，すなわち，$\mathscr{P}_{c_1,c_2,\gamma}(F_\mu)$ の任意の要素が

$$\mathscr{F}_{c_1,c_2,\gamma} = \left\{F \mid \frac{c_1}{1-\gamma}f_\mu \leq f \leq \frac{c_2}{1-\gamma}\right\} \tag{23}$$

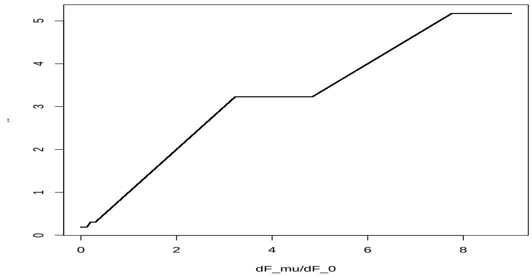

図 1　πのグラフ

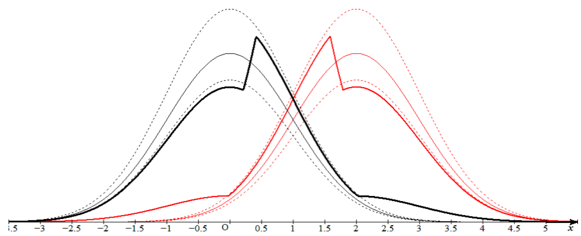

図 2　q_0^*，q_μ^* のグラフ

の要素の γ-汚染で表現されることを利用し，2 段階で構成する．$Q_0 \in \mathscr{F}_{c_1,c_2,\gamma}(F_0)$，$Q_\mu \in \mathscr{F}_{c_1,c_2,\gamma}(F_\mu)$ であり，これらの Radon-Nikodym 導関数 $\frac{dQ_\mu}{dQ_0}$ が $\frac{dF_\mu}{dF_0}(x) = \frac{f_\mu(x)}{f_0(x)}$ の小さいところと大きいところを除いて π に等しくなるような分布の対 (Q_0, Q_μ) を構成する．次に，Q_0 と Q_μ の γ-汚染近傍の対に対する最も不利な分布の対 Q_0^* と Q_μ^* を求めると，これが $(\mathscr{P}_{c_1,c_2\gamma}(F_0), \mathscr{P}_{c_1,c_2\gamma}(F_\mu))$ に対する最も不利な分布の対になる（[8] 参照）．

$Q_0 \in \mathscr{F}_{c_1,c_2,\gamma}(F_0)$，$Q_\mu \in \mathscr{F}_{c_1,c_2,\gamma}(F_\mu)$ であるためには，その密度関数 q_0，q_μ が次の条件を満たすことが必要十分である．

条件 1-1 $\int_{-\infty}^{\infty} q_0(x)dx = 1$.

条件 1-2 $\int_{-\infty}^{\infty} q_\mu(x)dx = 1$.

条件 2-1 $\frac{c1}{1-\gamma} f_0(x) \le q_0(x) \le \frac{c2}{1-\gamma} f_0(x)$.

条件 2-2 $\frac{c1}{1-\gamma} f_0(x-\mu) \le q_\mu(x) \le \frac{c2}{1-\gamma} f_0(x-\mu)$.

この q_0, q_μ から次の q_0^*, q_μ^* を構成すると $Q_0^* \in \mathscr{P}_{c_1,c_2,\gamma}(F_0)$，$Q_\mu^* \in \mathscr{P}_{c_1,c_2,\gamma}(F_\mu)$，$\pi(x) = \frac{dQ_\mu^*}{dQ_0^*}(x) = \frac{q_\mu^*}{q_0^*}$ が成り立ち，Q_0^*，Q_μ^* は $(\mathscr{P}_{c_1,c_2,\gamma}(F_0), \mathscr{P}_{c_1,c_2,\gamma}(F_\mu))$ の最も不利な分布の対になる．

$$q_0^* = \begin{cases} c_1 f_0(x), & 0 \le \frac{f_0(x-\mu)}{f_0(x)} \le \left(\frac{c_1}{c_2}\right)\Delta_1 \\ c_1 f_0(x), & \left(\frac{c_1}{c_2}\right)\Delta_1 \le \frac{f_0(x-\mu)}{f_0(x)} \le \left(\frac{c_1}{c_2}\right) a_1 \\ \frac{c_2}{a_1} f_0(x-\mu), & \left(\frac{c_1}{c_2}\right) a_1 \le \frac{f_0(x-\mu)}{f_0(x)} \le a_1 \\ c_2 f_0(x), & a_1 \le \frac{f_0(x-\mu)}{f_0(x)} \le a_2 \\ c_2 f_0(x), & a_2 \le \frac{f_0(x-\mu)}{f_0(x)} \le \left(\frac{c_2}{c_1}\right) a_2 \\ c_2 f_0(x), & \left(\frac{c_2}{c_1}\right) a_2 \le \frac{f_0(x-\mu)}{f_0(x)} \le \left(\frac{c_2}{c_1}\right)\Delta_2 \\ \frac{c_1}{\Delta_2} f_0(x-\mu) & \left(\frac{c_2}{c_1}\right)\Delta_2 \le \frac{f_0(x-\mu)}{f_0(x)} < \infty \end{cases} \tag{24}$$

$$q_\mu^* = \begin{cases} \Delta_1 c_1 f_0(x), & 0 \le \frac{f_0(x-\mu)}{f_0(x)} \le \left(\frac{c_1}{c_2}\right)\Delta_1 \\ c_2 f_0(x-\mu), & \left(\frac{c_1}{c_2}\right)\Delta_1 \le \frac{f_0(x-\mu)}{f_0(x)} \le \left(\frac{c_1}{c_2}\right) a_1 \\ c_2 f_0(x-\mu), & \left(\frac{c_1}{c_2}\right) a_1 \le \frac{f_0(x-\mu)}{f_0(x)} \le a_1 \\ c_2 f_0(x-\mu), & a_1 \le \frac{f_0(x-\mu)}{f_0(x)} \le a_2 \\ a_2 c_2 f_0(x), & a_2 \le \frac{f_0(x-\mu)}{f_0(x)} \le \left(\frac{c_2}{c_1}\right) a_2 \\ c_1 f_0(x-\mu), & \left(\frac{c_2}{c_1}\right) a_2 \le \frac{f_0(x-\mu)}{f_0(x)} \le \left(\frac{c_2}{c_1}\right)\Delta_2 \\ c_1 f_0(x-\mu), & \left(\frac{c_2}{c_1}\right)\Delta_2 \le \frac{f_0(x-\mu)}{f_0(x)} < \infty. \end{cases} \tag{25}$$

図 2 は q_0^*，q_μ^* のグラフである．

ミニマックス検定 φ^* の棄却点 λ_α^* と最小検出力 $\beta_{\varphi^*}(\mu)$ を求める．また，パラメトリックな場合の最強力検定との比較を行う．ここで CN は正規分布の混合分布 $(1-\gamma)f_0(x-\mu) + \gamma f_0(x-\mu+\eta)$ である．表 1 は $\gamma = 0.05$, $\eta = 10$ のときの結果である． ただし，ミニマックス検定

表 1　検定の比較

	ミニマックス検定				最強力検定	
n	λ_α^*	$P_{Q_\mu^*}$	$P_{N(\mu,1)}$	P_{CN}	$P_{N(\mu,1)}$	P_{CN}
1	5.177	0.389	0.492	0.466	0.638	0.607
2	4.058	0.568	0.709	0.642	0.881	0.797
3	3.684	0.698	0.854	0.773	0.965	0.828
4	2.665	0.811	0.932	0.868	0.990	0.805
5	2.011	0.878	0.969	0.925	0.997	0.780
10	0.262	0.989	0.999	0.996	0.999	0.899

に対する $P_{Q_\mu^*}$ は $Q_\mu^{*n}(\pi_n \ge \lambda_\alpha^*)$，$P_{N(\mu,1)}$ は $P_{N(\mu,1)}^n(\pi_n \ge \lambda_\alpha^*)$，$P_{CN}$ は $P_{CN}^n(\pi_n \ge \lambda_\alpha^*)$ であり，最強力検定に対する $P_{N(\mu,1)}$ は $P_{N(\mu,1)}^n(\sqrt{n}\bar{\mathbf{X}} \ge \lambda_\alpha)$，$P_{CN}$ は $P_{CN}^n(\sqrt{n}\bar{\mathbf{X}} \ge \lambda_\alpha)$ である．表 2 は $c_1 = 0.8$，$c_2 = 1.2$，$\gamma = 0.1$ で混合分布の η を動かしたときの比較である． $\eta = 7$ あたりで

表 2　$c_1 = 0.8, c_2 = 1.2, \gamma = 0.1$ での汚染による影響

η	$P_{CN}^n(\pi_n \ge \lambda_\alpha^*)$	$P_{CN}^n(\sqrt{n}\bar{\mathbf{X}} \ge \lambda_\alpha)$
1	0.915	0.995
2	0.889	0.988
3	0.877	0.973
4	0.876	0.950
5	0.874	0.919
6	0.856	0.884
7	0.875	0.849
8	0.873	0.817
9	0.855	0.794
10	0.875	0.783

検出力の逆転が起こっている．ミニマックス検定は正規分布の下では，最強力検定と比べると検出力は高くないが汚染が入っていても安定している，しかし，最強力検定は汚染が入ると検出力が下がり，ミニマックス検定の検出力の方が高くなる．

9　ロバスト回帰推定量の最大バイアス

線形モデルを

$$y = \alpha_0 \ + \ \boldsymbol{\theta}_0' \boldsymbol{x} + u \tag{26}$$

とする．ここで $\boldsymbol{x} = (x_1, \ldots, x_p)'$ を p 次元ユークリッド空間 R^p の値をとる確率ベクトル，α_0 を真の定数項，

$\boldsymbol{\theta}_0 = (\theta_{10}, \ldots, \theta_{p0})'$ を真の回帰母数ベクトル，u を誤差で $\boldsymbol{x}$ とは独立な確率変数とする．また，F_0 を u の分布関数，G_0 を $\boldsymbol{x}$ の分布関数，H_0 を $(y, \boldsymbol{x})$ の分布関数とする．このとき，

$$H_0(y, \boldsymbol{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} F_0(y - \alpha_0 - \boldsymbol{\theta}_0' \boldsymbol{s}) dG_0(\boldsymbol{s}) \quad (27)$$

と表すことができる．$\boldsymbol{T}$ を R^{p+1} 上の分布関数 H からなる空間上で定義された R^p の値をとる汎関数とする．H_n を H からの標本 $(y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n)$ の経験分布とするとき，$\boldsymbol{T}_n = \boldsymbol{T}(H_n)$ は $\boldsymbol{\theta}_0$ の推定量とみなすことができる．$\boldsymbol{T}$ が回帰アフィン共変であるときは，一般性を失うことなく，$\alpha_0 = 0$，$\boldsymbol{\theta}_0 = \boldsymbol{0}$ と仮定できるので，$\boldsymbol{T}$ の H における漸近バイアスは

$$b(\boldsymbol{T}, H) = ||\boldsymbol{T}(H)|| \quad (28)$$

となる．よって，$\boldsymbol{T}$ の (c,γ)-近傍 $\mathscr{P}_{c,\gamma}(H_0)$ $(= \mathscr{P}_{0,c,\gamma}(H_0))$ 上での最大漸近バイアスは

$$B_{\boldsymbol{T}}(c, \gamma) = \sup_{H \in \mathscr{P}_{c,\gamma}(H_0)} ||\boldsymbol{T}(H)|| \quad (29)$$

で与えられる．広いクラスの汎関数 $\boldsymbol{T}$ が

$$[T_0, \boldsymbol{T}(H)] = \arg\min_{\alpha, \boldsymbol{\theta}} J(F_{H,\alpha,\boldsymbol{\theta}}) \quad (30)$$

により定義される．ここで，J はロバスト損失関数であり，$F_{H,\alpha,\boldsymbol{\theta}}$ は H の下での $|y_i - \alpha_0 - \boldsymbol{\theta}' \boldsymbol{x}|$ の分布である．この $\boldsymbol{T}$ のクラスには，S 推定量，τ 推定量や CM 推定量などの代表的な推定量が含まれている（[5] 参照）．Ando and Kimura (2015) はこのクラスの $\boldsymbol{T}$ の最大漸近バイアス $B_{\boldsymbol{T}}(c,\gamma)$ の上界 $\bar{B}_{\boldsymbol{T}}(c,\gamma)$ と下界 $\underline{B}_{\boldsymbol{T}}(c,\gamma)$ に対する一般式を与えた．この上界と下界は γ-汚染，すなわち，$c = 1-\gamma$ の場合には一致し，最大漸近バイアスに等しくなる．そして，特に重要な場合として，$H_0 = N_{p+1}(\boldsymbol{0}, \boldsymbol{I})$ のときに，代表的な 3 つのロバスト回帰推定量である S 推定量，τ 推定量，CM 推定量の最大漸近バイアス $B_S(c,\gamma)$，$B_\tau(c,\gamma)$，$B_{CM}(c,\gamma)$ の上界と下界を導いている．本研究では，損失関数 J に用いられる ρ 関数が Huber score function と Tukey の biweight function の場合に，これらの回帰推定量の最大漸近バイアスの上界と下界の値をシミュレーションにより求め，c と γ を変数とする数表を作成した．その一部が表 3，表 4，表 5，表 6 である．3 つの回帰推定量は破綻点が 0.5，漸近効率が 0.95 となるように ρ 関数の定数を調整している．γ が比較的小さく，c が $1-\gamma$ よりもそれほど大きくなければ，上界と下界の差は大きくなく，十分に実用的である．

表 3 $\bar{B}_\tau$（Tukey biweight function）

$c\backslash\gamma$	0	0.01	0.02	0.05	0.1
1	0	0.485	0.705	1.242	2.148
1.1	0.824	1.016	1.2	1.747	2.766
1.2	1.188	1.391	1.591	2.205	3.366
1.5	2.084	2.367	2.649	3.519	5.161
2	3.48	3.925	4.367	5.722	8.234

表 4 $\underline{B}_\tau$（Tukey biweight function）

$c\backslash\gamma$	0	0.01	0.02	0.05	0.1
1	0	0.382	0.557	0.956	1.529
1.1	0	0.407	0.588	0.992	1.561
1.2	0	0.417	0.602	1.009	1.578
1.5	0	0.441	0.634	1.056	1.633
2	0	0.463	0.665	1.1	1.781

表 5 $\bar{B}_{CM}$（Tukey biweight function）

$c\backslash\gamma$	0	0.01	0.02	0.05	0.1
1	0	0.432	0.625	1.093	1.889
1.1	0.817	0.959	1.099	1.536	2.395
1.2	1.18	1.319	1.462	1.927	2.882
1.5	2.078	2.249	2.429	3.034	4.324
2	3.479	3.725	3.986	4.873	6.784

表 6 $\underline{B}_{CM}$（Tukey biweight function）

$c\backslash\gamma$	0	0.01	0.02	0.05	0.1
1	0	0.319	0.462	0.789	1.277
1.1	0	0.319	0.462	0.789	1.277
1.2	0	0.319	0.462	0.789	1.277
1.5	0	0.319	0.462	0.789	1.277
2	0	0.319	0.463	0.984	1.893

参考文献

[1] 安藤周平・木村美善：(c_1, c_2, γ)-汚染の下での正規分布の平均のロバスト検定，統計関連学会連合大会講演報告集，197, 2015.

[2] Ando M. and Kimura.：The Maximum Asymptotic Bias of Robust Regression Estimate over (c,γ)- Contamination Neighborhoods，Technical Report of the Nanzan Academic Society，2015.

[3] 板東 宜彦：3 つのパラメータをもつ分布近傍の下でのメディアンのロバスト推測，神戸大学大学院工学研究科修士論文，2010.

[4] Bednarski. T：On solutions of minimax tests problems for special capacities, Z. Wahrscheinlichkeitstheorie, 1981.

[5] Berrendero, J. R. ,Mendes, B. V. M., and Tyler, D.E. : On the maximum bias function of MM-estimates and constrained M-estimates of regression, Ann.Statist, **35**, 13-40, 2007.

[6] Kakiuchi I. and Kimura M. : Robust nonparametric inference for the median under a new neighborhood of distributions, Technical Report of the Nanzan Academic Society, 2012.

[7] Huber, P.J., Ronchetti, E.M. : *Robust Statistics*, Wiley, 2009.

[8] Huber, P.J. : A robust version of the probability ratio test, *Ann.Math.Statist*, 1965.