

正規分布と指数分布の畳み込みの母数推定について¹

On Parameter Estimation of the Convolution of Normal and Exponential Distributions

穴 沢 務

1 はじめに

Y , Z は互いに独立な連続型確率変数であり, それぞれ次のような分布関数を持つとする:

$$F_Y(y) = \Phi\left(\frac{y-\mu}{\sigma}\right), \text{ 但し } \Phi(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt,$$

$$F_Z(z) = \begin{cases} 1 - e^{-z/\lambda} & \text{if } z \geq 0 \\ 0 & \text{if } z < 0. \end{cases}$$

このとき, $X=Y+Z$ の分布関数は次のような畳み込みで表現できる:

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{\infty} F_Y(x-z) dF_Z(z) \\ &= \int_0^{\infty} \Phi\left(\frac{x-z-\mu}{\sigma}\right) dF_Z(z). \end{aligned} \quad (1)$$

本稿では, X に対する標本 $x = (x_1, x_2, \dots, x_n)$ に基づいて,

¹本稿執筆のきっかけとなる問題提起をしていただいた本学助手平沢尚毅氏に, 心より感謝いたします. また本学教授兼岩龍二氏との週一回の数学セミナーがなければ, 本稿の完成はなかったことを付け加えます. なお, 本稿で紹介するシミュレーション実験は, すべて本学情報処理センターのワークステーションDEC 3000/400 (ホスト名GEMINI) 上で実行しました.

3つの母数 μ, σ, λ を推定する方法について検討する。

上記の F_x に従う確率変数の例として、コンピュータやワープロに付属するキーボードの打鍵時間間隔が考えられる。1つのキーを打ってから次のキーを打つまでに、ユーザは

1. ディスプレイに表示される文字の知覚
2. 内容の認識と次の打鍵について何らかの短い思考（もしくは迷い）
3. 打鍵動作

を行うであろう。もし、1と3を加えた時間が正規分布、2が指数分布に従い、しかも互いに独立であると仮定できるならば、打鍵時間間隔は F_x によってモデル化できる。実際に母数を推定するには、1, 2が共に心理的現象である故にそれぞれの時間を別々に計測することができないので、打鍵時間間隔に基づいて推定する以外にない。

この類のモデルはCardら[1]によって最初に詳しく研究された。木村ら[3]、[4]は、 F_x を Ψ 分布と呼び、日本語ワープロ利用者の思考時間を、母数の異なる複数の Ψ 分布の混合（mixture）でモデル化した。彼らはそれらの論文の中で、母数と混合比率を同時に推定するための独特の方法を採用している。

本稿では、木村らのような Ψ 分布の混合については考えず、最も単純な F_x に対して、1つの新たな母数推定法を提案する。もちろん、木村らの方法を F_x に応用することもできるが、その方法との比較は別の機会に譲る。我々の方法は、分布の距離を表す1つの評価関数を近似することで実現する。本稿では、その近似をどこまで粗くできるか（すなわち、どこまで計算を簡略化できるか）に主眼を置き、他の方法との比較は行わない。第2章では、 F_x の母数を推定するためのよく知られた方法とその欠点を述べ、さらに、我々が提案する推定法を紹介する。第3章では、シミュレーション実験の結果から、評価関数の近似がかなり粗い場合でも、良好な推定ができることを示す。

2 推定方法

2.1 よく知られた推定法

このような点推定問題において、まず考えなければならないのが最尤推定法である。\$X\$の密度関数を\$f_X(x)\$とすると、

$$f_X(x) = \frac{1}{\lambda} \int_0^{\infty} \phi\left(\frac{x-z-\mu}{\sigma}\right) e^{-z/\lambda} dz, \quad \text{但し } \phi(x) \equiv \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

と書ける。このとき尤度関数を

$$L(\mu, \sigma, \lambda; \mathbf{x}) = \prod_{i=1}^n f_X(x_i)$$

とおき、この\$L\$（あるいは多くの場合\$\log L\$）を最大にするような\$\mu, \sigma, \lambda\$を求めればよい。しかし\$f_X\$は積分形になっているので、3つのパラメータの最尤推定量を\$\{x_i\}\$の関数として陽に表現することは困難である。

一方、計算が単純で、直感的にわかり易い方法として、モーメント法がある。それは、\$X\$の3次までのモーメントをそれぞれ\$\mu, \sigma, \lambda\$の関数として表し、

$$E(X) = \bar{x} \equiv \frac{1}{n} \sum_{i=1}^n x_i, \quad E(X - E(X))^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2,$$

$$E(X - E(X))^3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3,$$

を解いてその解を推定量としよう、というものである。各モーメントは積率母関数

$$m_X(t) \equiv E(e^{Xt}) = m_Y(t)m_Z(t) \quad (Y \text{ と } Z \text{ の独立性より})$$

の性質を利用すれば容易に得られる。この場合は、

$$E(X) = \mu + \lambda, \quad E(X - E(X))^2 = \sigma^2 + \lambda^2, \quad E(X - E(X))^3 = 2\lambda^3$$

である。ところが、この方法では極めてしばしば、\$\sigma^2\$の推定値が負になってしまう。

2. 2 提案する方法

本稿で提案する方法は、本質的には、評価関数として F_X と経験分布関数 G_n とのある種の距離を用い、その評価関数を最小にするような μ, σ, λ の値を求める、というものである。但し、その評価関数は複雑な積分形をしているので、適当な近似式を求めて計算の簡略化を図っている。

我々が採用する評価関数は、次のようなクラメル=フォン・ミーゼス型距離である：

$$E_{F_X} [(G_n - F_X)^2] \equiv \int_{-\infty}^{\infty} \{G_n(x) - F_X(x)\}^2 dF_X(x), \quad (2)$$

但し

$$G_n(x) = \begin{cases} \frac{i}{n} & \text{if } x_{(i)} < x < x_{(i+1)} \quad (i = 0, 1, \dots, n) \\ \frac{i - \frac{1}{2}}{n} & \text{if } x = x_{(i)} \quad (i = 1, \dots, n), \end{cases}$$

$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ は標本順序統計量、 $x_{(0)} = -\infty$ 、 $x_{(n+1)} = \infty$ とする。

付記： 経験分布関数のこのような定義は、通常とはやや異なる。Loève [5] (p186) は、分布関数 F の不連続点 x の値を

$$\frac{F(x-0) + F(x+0)}{2}$$

と定義することを「正規化する」と呼んでいる。その呼称を踏襲すれば、上の G_n は「正規化された」経験分布関数と呼べるであろう。本稿では一貫して、連続関数を近似するときには「正規化された」階段関数を用いる。

(2) 式の右辺の測度関数 F_X を G_n で置き換えれば、

$$\begin{aligned} E_{F_X} [(G_n - F_X)^2] &\approx \int_{-\infty}^{\infty} \{G_n(x) - F_X(x)\}^2 dG_n(x) \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{i - \frac{1}{2}}{n} - F_X(x_{(i)}) \right\}^2 \end{aligned} \quad (3)$$

と近似できる. さらに F_x も (1) のように積分形をしているから, その測度関数 F_z を次のような階段関数で置き換える:

$$F_1(x; m) = \begin{cases} \frac{j}{m} & \text{if } \xi_j < x < \xi_{j+1} \quad (j = 0, 1, \dots, m) \\ \frac{j - \frac{1}{2}}{m} & \text{if } x = \xi_j \quad (j = 1, \dots, m), \end{cases}$$

但し

$$\xi_j = F_z^{-1} \left(\frac{j - \frac{1}{2}}{m} \right) = -\lambda \log \left(1 - \frac{j - \frac{1}{2}}{m} \right) \quad (4)$$

m は適当に固定された自然数である. 以後, この m のことを近似レベルと呼ぼう. このとき F_x は次のように近似できる:

$$\begin{aligned} F_X(x) &\approx \int_0^\infty \Phi \left(\frac{x - \mu - z}{\sigma} \right) dF_1(z; m) \\ &= \frac{1}{m} \sum_{j=1}^m \Phi \left(\frac{x - \mu - \xi_j}{\sigma} \right). \end{aligned} \quad (5)$$

したがって, 評価関数 $E_{F_x} [(G_n - F_x)^2]$ の近似式として, (3), (4), (5) より,

$$R_1 = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{i - \frac{1}{2}}{n} - \frac{1}{m} \sum_{j=1}^m \Phi \left(\frac{x_{(i)} - \mu + \lambda \log(1 - \frac{j - \frac{1}{2}}{m})}{\sigma} \right) \right\}^2$$

が得られる. 我々は, (μ, σ, λ) の一つの推定量として, R_1 を最小にする $(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$ を考える. その精度を確かめるためのシミュレーション実験の結果を, 第 3.2 節に示す.

ところで, 先に F_z の近似に用いた F_1 は特にある基準を最適化して得られたものではない. そこで, もう一つの近似式として, より一般的な階段関数

$$F_2(x; m) = \begin{cases} p_j & \text{if } \eta_j < x < \eta_{j+1} \quad (j = 0, 1, \dots, m) \\ \frac{p_{j-1} + p_j}{2} & \text{if } x = \eta_j \quad (j = 1, \dots, m) \end{cases}$$

$$\left(\begin{array}{l} \text{但し } \eta_j = F_Z^{-1} \left(\frac{p_{j-1} + p_j}{2} \right) = -\lambda \log \left(1 - \frac{p_{j-1} + p_j}{2} \right), \\ 0 = p_0 < p_1 < \dots < p_m = 1 \end{array} \right)$$

のクラス \mathcal{F} を考え、 L_1 -ノルム

$$\begin{aligned} & \int_{-\infty}^{\infty} |F_Z(x) - F_2(x; m)| dx \\ &= \sum_{j=1}^m \left\{ \int_{-\lambda \log(1-p_{j-1})}^{-\lambda \log(1-\frac{p_{j-1}+p_j}{2})} (1 - e^{-\frac{x}{\lambda}} - p_{j-1}) dx + \int_{-\lambda \log(1-\frac{p_{j-1}+p_j}{2})}^{-\lambda \log(1-p_j)} (p_j - 1 + e^{-\frac{x}{\lambda}}) dx \right\} \\ &= 2\lambda \left\{ \sum_{j=1}^{m-1} (1-p_j) \log(1-p_j) - \sum_{j=1}^m \left(1 - \frac{p_{j-1}+p_j}{2} \right) \log \left(1 - \frac{p_{j-1}+p_j}{2} \right) \right\} \end{aligned}$$

を最小にする $F_2^* \in \mathcal{F}$ を採用することもできる。 F_2^* を達成する p_j の値を p_j^* ($j=1, \dots, m-1$) とすれば、評価関数 $E_{F_X}[(G_n - F_X)^2]$ は、

$$R_2 = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{i - \frac{1}{2}}{n} - \sum_{j=1}^m (p_j^* - p_{j-1}^*) \Phi \left(\frac{x_{(i)} - \mu + \lambda \log(1 - \frac{p_{j-1}^* + p_j^*}{2})}{\sigma} \right) \right\}^2$$

($P_0^* = 0, P_m^* = 1$) と近似できる。 R_2 を最小化して得られる推定量 $\tilde{\mu}, \tilde{\sigma}, \tilde{\lambda}$ の精度については、第3.3節で検討する。

本稿では、十分小さい近似レベル m に対する推定量 $(\hat{\mu}, \hat{\sigma}, \hat{\lambda}), (\tilde{\mu}, \tilde{\sigma}, \tilde{\lambda})$ の RMSE (Root Mean Square Error) に注目する。 m を小さくするにつれて、 $\min_{\mu, \sigma, \lambda} R_1$ や $\min_{\mu, \sigma, \lambda} R_2$ の値は大きくなることが予想されるが、我々の目的から見れば、そのことは大きな問題ではない。第3.2節、3.3節では、 m が小さくなるにつれて μ, λ の推定量の RMSE も小さくなるという、逆説的な結果が示される。

なお、 σ の推定量の RMSE は、 m が小さくなるにしたがって大きくなる。その原因と修正方法については、第3.4節で具体的に述べる。

3 シミュレーション実験

3. 1 実験の概要

本稿では、母集団として次の5つのケースを考える。

1. $\mu = 0.5, \sigma = 0.1, \lambda = 0.5$
2. $\mu = 0.5, \sigma = 0.1, \lambda = 0.25$ (ケース1よりも正規分布に近い)
3. $\mu = 0.5, \sigma = 0.1, \lambda = 0.05$ (ケース2よりも正規分布に近い)
4. $\mu = 0.5, \sigma = 0.05, \lambda = 0.5$ (ケース1よりも指数分布に近い)
5. $\mu = 0.5, \sigma = 0.01, \lambda = 0.5$ (ケース4よりも指数分布に近い)

これらの各ケースに対して、標本サイズを $n = 500$ に固定して、近似レベル $m = 2, 3, 4, 5, 10, 20, 50, 100$ のそれぞれにおける $(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$, $(\bar{\mu}, \bar{\sigma}, \bar{\lambda})$ の RMSE と bias を、標本数 (実験回数) $N = 500$ で推定した。例えば、 R_1 による、ある m における $\hat{\mu}$ の RMSE と bias は、それぞれ次のように推定される：

$$\text{RMSE}(\hat{\mu}) = \sqrt{\frac{1}{N} \sum_{k=1}^N (\hat{\mu}_k - \mu_0)^2},$$

$$\text{bias}(\hat{\mu}) = \frac{1}{N} \sum_{k=1}^N \hat{\mu}_k - \mu_0,$$

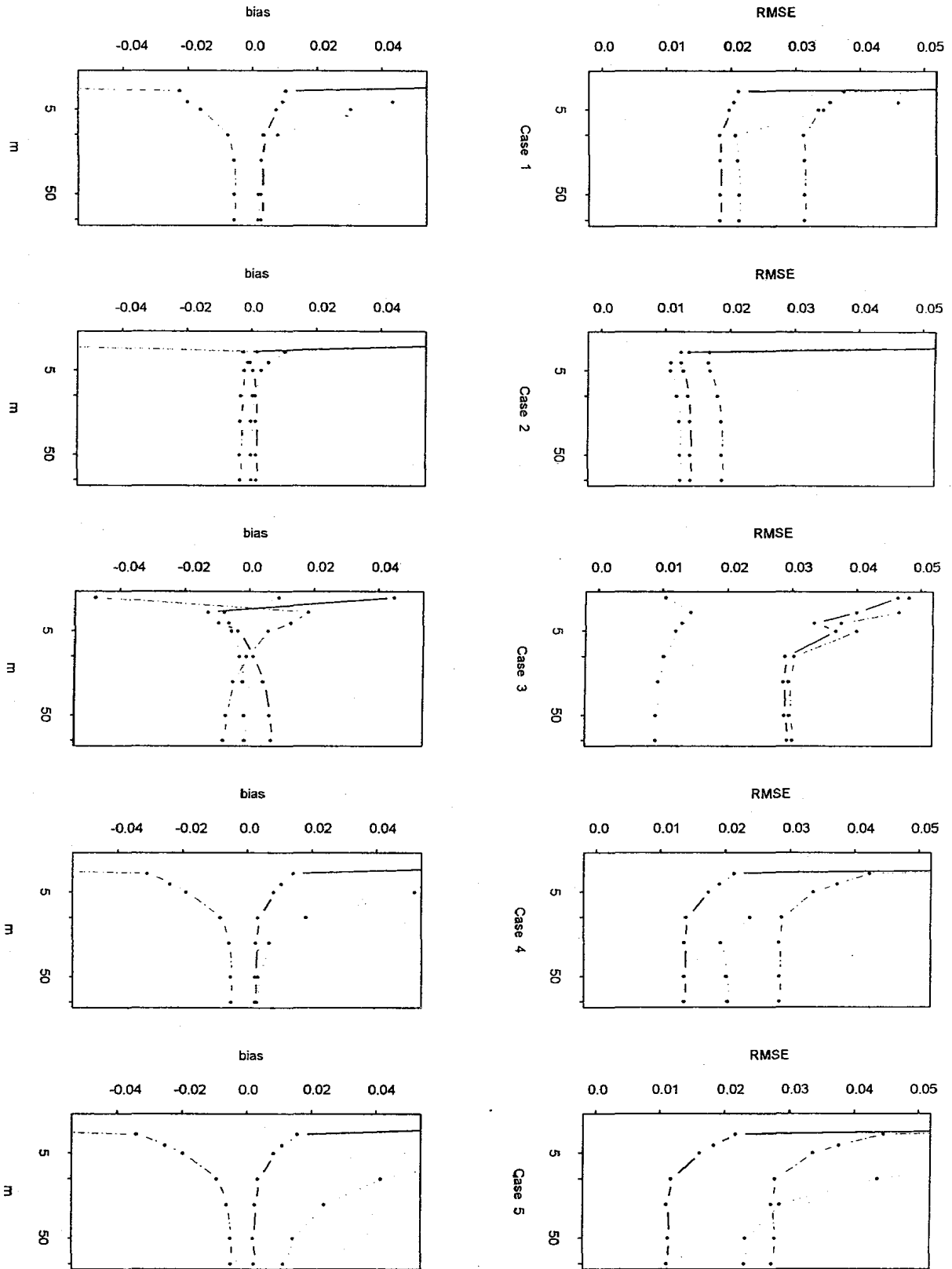
但し $\hat{\mu}_k$ は第 k 標本に基づく μ の推定値、 μ_0 は μ の真の値である。なお、非線形最小自乗法や標準正規分布関数の算法は、Press ら [6] にしたがった。

3. 2 近似式 R_1 の最小化による推定

評価関数の近似式 R_1 を最小化して得られる $(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$ の RMSE, bias の推定値を図1にプロットする。上段は RMSE 値, 下段は bias 値で、各図とも実線は μ , 点線は σ , 破線は λ にそれぞれ対応する。

RMSE が最小になる m は、ケースごとに、さらには母数ごとに異なる。まず $\hat{\mu}, \hat{\lambda}$ について見ると、(ケース2を除いて) $m = 10$ の RMSE 値が最小ま

図1: R_1 最小化による推定量のRMSE, bias



たは最小値+0.001の範囲にあり, $m \leq 5$ ではRMSE値が急激に大きくなる. biasの絶対値も同様に $m \leq 5$ で急に大きくなる. ケース2だけはRMSEの最小値は $3 \leq m \leq 5$ にある.

$\hat{\sigma}$ のRMSEは, (ケース3を除き)小さな m に対してかなり大きな値を示す. 特に母分布が指数分布に近づくほど, RMSE値が急増する m の値が大きくなる. また, 同じ m においてbiasも急増している. 実は, RMSE値急増の主な原因はbiasの急増であることが確かめられている. 推定量の分散($\text{RMSE}^2 - \text{bias}^2$)は, m にあまり大きな影響を受けない.

なお, どのケースにおいても, $\hat{\lambda}$ には(十分大きな m に対してさえ)負のbiasがあることがわかる.

3. 3 近似式 R_2 の最小化による推定

評価関数の近似式 R_2 による $(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$ のRMSE, biasの推定値を図2にプロットする. 図の見方は前節と同様である.

$\hat{\mu}$, $\hat{\lambda}$ について見ると, 前節の結果に比べて, より小さい m に対しても小さいRMSE値を示している. 多くのケースで, $m=5, 10$ のRMSE値が最小または最小値+0.001の範囲にある(ケース2の $\tilde{\lambda}$ に至っては, $m=2$ でRMSE値が最小になっている). $m=5, 10$ に対する $\hat{\mu}$, $\hat{\sigma}$, $\hat{\lambda}$, $\tilde{\lambda}$ のRMSE値を表1に示す. この表より, μ , λ の推定に際しては, 多くの場合, 評価関数の近似式として R_1 よりも R_2 を採用した方が良いことがわかる.

図2: R_2 最小化による推定量のRMSE, bias

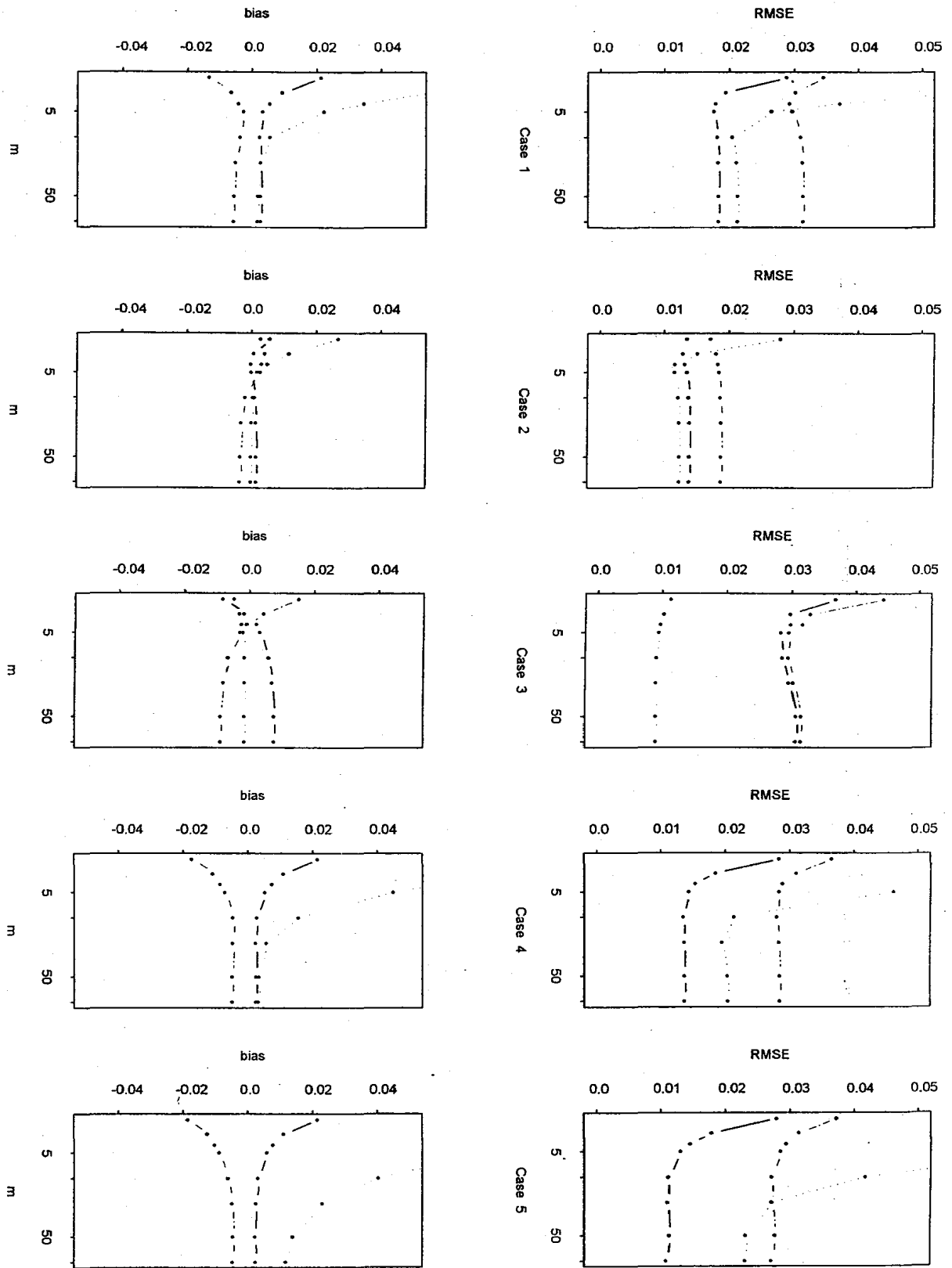


表1 : $m = 5, 10$ に対する $\hat{\mu}$, $\tilde{\mu}$, $\hat{\lambda}$, $\tilde{\lambda}$ の RMSE

		ケース1	ケース2	ケース3	ケース4	ケース5
$m = 5$	$\hat{\mu}$	0.0199	0.0130	0.0369	0.0176	0.0164
	$\tilde{\mu}$	0.0178	0.0136	0.0285	0.0144	0.0132
	$\hat{\lambda}$	0.0337	0.0171	0.0402	0.0338	0.0338
	$\tilde{\lambda}$	0.0299	0.0186	0.0298	0.0284	0.0287
$m = 10$	$\hat{\mu}$	0.0185	0.0137	0.0292	0.0141	0.0119
	$\tilde{\mu}$	0.0183	0.0140	0.0288	0.0136	0.0113
	$\hat{\lambda}$	0.0315	0.0183	0.0305	0.0290	0.0280
	$\tilde{\lambda}$	0.0313	0.0188	0.0297	0.0281	0.0273

σ については、前節と同様に、小さな m に対してかなり大きな RMSE 値を示している。

3. 4 σ の再推定

小さな m に対する σ の推定量に大きな bias があり、それ故に RMSE が大きくなることには、直感的にも明かな理由がある。近似式 (5) は σ の値が 0 に近づくほど階段関数に近くなるので、 F_x の値が 1 に近づくにつれて、近似が粗くなる。(5) の μ, σ, λ を変数として F_x にフィットさせようとするれば、自然に σ の値は真の値よりも大きくなる。 σ の過大推定を防ぐには、 m の値を大きくする以外にないように見える。

しかし、 μ, λ については、小さな m についても良好な推定量が得られているのだから、それらを利用して、比較的少ない手間で σ を再推定する方法がないだろうか、という疑問が生じる。その答として、我々は次のような方法を提案する： $F_x(\hat{\mu})$ の近似値としては、 $G_n(\hat{\mu})$ と、(5)より

$$\frac{1}{m} \sum_{j=1}^m \Phi \left(-\frac{\xi_j}{\sigma} \right)$$

(但し $m=100$ とし, ξ_i は(4)にて $\lambda = \tilde{\lambda}$ とおいて求める) が考えられる. この2つの近似値を等号で結び, その方程式の σ に関する解を, 新たな推定量 σ^* としよう.

表2では, $m=5, 10$ における σ^* の RMSE と, $m \geq 5$ に対する $\tilde{\sigma}$ の RMSE が比較されている. この表より, ケース4, 5, すなわち指数分布に近い場合においては, σ^* の RMSE はどの $\tilde{\sigma}$ の RMSE よりも ($m=100$ に対してすら) 小さい値を示している.

表2: σ^* ($m=5, 10$) と $\tilde{\sigma}$ ($m \geq 5$) の RMSE

		ケース1	ケース2	ケース3	ケース4	ケース5
σ^*	$m=5$	0.0205	0.0145	0.0275	0.0179	0.0232
	$m=10$	0.0209	0.0145	0.0447	0.0165	0.0189
$\tilde{\sigma}$	$m=5$	0.0267	0.0118	0.0096	0.0464	0.0758
	$m=10$	0.0206	0.0123	0.0093	0.0214	0.0420
	$m=20$	0.0213	0.0125	0.0093	0.0196	0.0274
	$m=50$	0.0215	0.0125	0.0092	0.0205	0.0233
	$m=100$	0.0215	0.0125	0.0092	0.0205	0.0232

4 おわりに

本稿で触れることのできなかつた今後の課題と, シミュレーション実験に関する反省をもって, 本稿の結語としたい.

- 第3.2節で述べたように, 我々の推定量には本質的な bias が潜んでいる. この bias を, 例えば Efron [2] のブートストラップ法などによって除去できるかどうかは, たいへん大きな課題と言えよう.
- RMSE を最小にするような m は分布や母数によって幅がある. よって, 最適な m を選択するためのルールを考える必要がある.
- 本稿では, 標本サイズを $m=500$ に固定したが, 大きい n に対する推定量

の一致性や, n と最適な m との依存関係についても調べる必要がある。

- ケース 5 においては, 時折 λ の推定値が負になる。 μ, λ を推定するだけならば, そのことが直接計算上のトラブルに結びつかないと思われるが, 変数変換によってその事態を防げば, より小さい RMSE を得られかもしれない。

参 考 文 献

- [1] Card, S. K., Moran, T. P. and Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- [2] Efron, B. (1979). Bootstrap methods: another look at the jack-knife. *Ann. Statist.* 7, 1-26.
- [3] 木村, 粕川, 谷越 (1986). 「ワープロ利用者の思考時間に関する統計的模型」, 情報処理学会研究報告 (日本語文書処理研究会 11-2).
- [4] 木村, 粕川 (1987). 「ワープロ利用者の思考時間に関する統計的模型の精緻化」, 情報処理学会研究報告 (文書処理とヒューマンインターフェース 14-4).
- [5] Loève, M. (1963). *Probability Theory*, Third edition. D. Van Nostrand Company.
- [6] Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1988). *Numerical Recipes in C*. Cambridge University Press. (邦訳: 丹慶, 奥村, 佐藤, 小林訳 (1993). 『Numerical Recipes in C [日本語版]』, 技術評論社.)