



Title	観測誤差の判別関数におよぼす影響について
Author(s)	中西, 寛子; Nakanishi, Hiroko; 河口, 至商 他
Citation	北海道大學工學部研究報告, 117, 93-99
Issue Date	1984-01-31
Doc URL	<a href="https://hdl.handle.net/2115/41822">https://hdl.handle.net/2115/41822</a>
Type	departmental bulletin paper
File Information	117_93-100.pdf



## 観測誤差の判別関数におよぼす影響について

中西 寛子 河口 至 商  
(昭和58年9月30日受理)

### On the Effect of the Contaminations in the Data on the Discriminant Function

Hiroko NAKANISHI and Michiaki KAWAGUCHI  
(Received September 30, 1983)

#### Abstract

When a set of data is given, it is unusual to expect that all of the data show the right values, some of which are the contaminating data. Because of the difficulty to distinguish the right data from the contaminating data, commonly all of the data with the contaminations are used for any statistical analysis. In the discriminant analysis, a mean and a variance are estimated by the given data as they are, and the discriminant function is solved by these estimations.

In this paper, we study the effect of the contaminations in the data on the solution of the discriminant function, i. e. the position of the discriminant point. We propose the use of the trimmed mean and the trimmed variance for the mean and the variance, and we investigate the effect of the trimmed way on the result of the discriminant function. This trimmed mean and variance are well known as robust estimations to the outliers. We can find that the trimmed way gives the better solution for the discriminant function with the contaminations in the data.

#### 1. はじめに

実験, 観測により得られた個々のデータが常に正しい値を示すことはまれであり, 実験器具や観測者が異なるなどの理由で誤った値をもつデータが含まれることがある。しかしながら, 「全データのうち, どのデータが観測誤差であるか」という判断を正確にすることは非常に困難であるため, 多くの場合, 得られたデータをすべてそのまま用いて種々の統計解析が行われる。判別分析を行う場合においても, 観測誤差をも含めてすべてのデータから平均, 分散を導き, 判別関数を解くことが多い。本論文では, 観測誤差が判別分析の結果, すなわち判別点の位置におよぼす影響について調べた。また, 外れ値や異常値に対して抵抗力のあるトリム平均, トリム分散を用いて判別関数を解くことを提案し, その効果についても考察を行った。なお, ここで取り扱っているデータは1次元のものであり, 2群の判別分析にのみ限定している。

## 2. 判別分析の各手法

ここでは、LDF (Linear Discriminant Function, 線型判別関数), QDF (Quadratic Discriminant Function, 2次判別関数), トリム-LDF, トリム-QDF について説明をする。なお、判別分析についての詳しい内容に関しては参考文献1)を参照されたい。

群  $G_i$  の平均  $\mu_i$ , 分散  $\sigma_i^2$  ( $i=1, 2$ ) に対し, LDF, QDF は各々, 次の様に定義される。

$$L(x) = (x - \mu_2)^2 - (x - \mu_1)^2,$$

$$Q(x) = (x - \mu_2)^2 \sigma_2^{-2} - (x - \mu_1)^2 \sigma_1^{-2} + \ln(\sigma_1^2 / \sigma_2^2).$$

$L(x_L) = 0$ ,  $Q(x_Q) = 0$  となる点  $x_L$ ,  $x_Q$  を各々 LDF, QDF の判別点と呼ぶ。式より,  $L(x)$  は  $Q(x)$  の特別な場合, すなわち,  $\sigma_1 = \sigma_2$  の場合に相当することがわかる。2群の分散が明らかに異なる場合, 一般に式  $Q(x)$  が用いられ, その他の場合は  $L(x)$  が用いられる。

次に,  $\alpha\%$ -トリム平均,  $\beta\%$ -トリム分散について述べる。これらの統計量は外れ値や異常値をもつデータに対してよりロバストな推定値(平均, 分散)を得るために考え出されたものである<sup>23)</sup>

得られたデータを大きさの順に並べ, 両端より  $\alpha\%$ ずつデータを取り除いた残りで計算された平均を  $\alpha\%$ -トリム平均と呼ぶ。両端より  $\beta\%$ ずつデータを取り除き,  $\alpha\%$ -トリム平均のまわりの2次のモーメントを  $\beta\%$ -トリム分散と呼ぶ。一般に  $\beta$  は  $\alpha$  より小さい値が選ばれる。各群  $G_i$  の  $\alpha\%$ -トリム平均  $\mu_{(T)i}$ ,  $\beta\%$ -トリム分散  $\sigma_{(T)i}^2$  ( $i=1, 2$ ) に対し, トリム-LDF, トリム-QDF は各々,

$$L_{(T)}(x) = (x - \mu_{(T)2})^2 - (x - \mu_{(T)1})^2,$$

$$Q_{(T)}(x) = (x - \mu_{(T)2})^2 \sigma_{(T)2}^{-2} - (x - \mu_{(T)1})^2 \sigma_{(T)1}^{-2} + \ln(\sigma_{(T)1}^2 / \sigma_{(T)2}^2)$$

と定義する。 $L_{(T)}(x_{TL}) = 0$ ,  $Q_{(T)}(x_{TQ}) = 0$  となる点  $x_{TL}$ ,  $x_{TQ}$  を各々トリム-LDF, トリム-QDF の判別点と呼ぶ。

## 3. 観測誤差を含むデータの分布

正しく観測されたデータが従う密度関数を  $f(x)$ , 全データに対する正しいデータの割合を  $1-p$  とし, 観測誤差が従う密度関数を  $f_i(x)$ , 全データに対するその割合を  $p_i$  とすると, ある観測により得られたデータが従う密度関数  $h(x)$  は次の様に考えることができる<sup>4)</sup>。このように作られた分布を一般に混合分布と呼ぶ。

$$h(x) = (1-p)f(x) + \sum_i p_i f_i(x)$$

ただし,  $\sum_i p_i = p$  とする。

今, 特に  $i=1$  のときを考える。密度関数  $f(x)$ ,  $f_1(x)$  が異なる正規分布  $N(\mu, \sigma^2)$ ,  $N(\mu^*, \sigma^{*2})$  であるとき, 密度関数  $h(x)$  は, 標準正規分布の密度関数  $\phi(x)$  を用いて表すと,

$$h(x) = (1-p) \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) + p \frac{1}{\sigma^*} \phi\left(\frac{x-\mu^*}{\sigma^*}\right)$$

となり, 密度関数  $h(x)$  の平均, 分散は各々,

$$\tilde{\mu} = (1-p)\mu + p\mu^*$$

$$\tilde{\sigma}^2 = (1-p)\sigma^2 + p\sigma^{*2} + (1-p)p(\mu - \mu^*)^2$$

となる。また,  $\alpha\%$ -トリム平均,  $\beta\%$ -トリム分散も各々, 次の様に正確に計算される。

$$\tilde{\mu}_{(T)} = \frac{1}{1-2\alpha} \left[ (1-p) \left\{ \mu \left( \Phi \left( \frac{b-\mu}{\sigma} \right) - \Phi \left( \frac{a-\mu}{\sigma} \right) \right) + \frac{\sigma}{\sqrt{2\pi}} \left( -\exp \left( -\frac{(b-\mu)^2}{2\sigma^2} \right) + \exp \left( -\frac{(a-\mu)^2}{2\sigma^2} \right) \right) \right\} + p \left\{ \mu^* \left( \Phi \left( \frac{b-\mu^*}{\sigma^*} \right) - \Phi \left( \frac{a-\mu^*}{\sigma^*} \right) \right) + \frac{\sigma^*}{\sqrt{2\pi}} \left( -\exp \left( -\frac{(b-\mu^*)^2}{2\sigma^{*2}} \right) + \exp \left( -\frac{(a-\mu^*)^2}{2\sigma^{*2}} \right) \right) \right\} \right]$$

ここで、 $\Phi(x)$ は標準正規分布の分布関数を示し、 $a$ は小さいほうから $\alpha\%$ となる点、 $b$ は大きいほうから $\alpha\%$ となる点を示す。

$$\tilde{\sigma}_{(T)}^2 = \frac{1}{1-2\beta} \left[ (1-p) \left\{ -\frac{\sigma}{\sqrt{2\pi}} \left( \left( \frac{d-\mu}{\sigma} \right) \exp \left( -\frac{(d-\mu)^2}{2\sigma^2} \right) - \left( \frac{c-\mu}{\sigma} \right) \exp \left( -\frac{(c-\mu)^2}{2\sigma^2} \right) \right) - \frac{2\sigma(\mu-\tilde{\mu}_{(T)})}{\sqrt{2\pi}} \left( \exp \left( -\frac{(d-\mu)^2}{2\sigma^2} \right) - \exp \left( -\frac{(c-\mu)^2}{2\sigma^2} \right) \right) + (\sigma^2 + (\mu - \tilde{\mu}_{(T)})^2) \left( \Phi \left( \frac{d-\mu}{\sigma} \right) - \Phi \left( \frac{c-\mu}{\sigma} \right) \right) \right\} + p \left\{ -\frac{\sigma^*}{\sqrt{2\pi}} \left( \left( \frac{d-\mu^*}{\sigma^*} \right) \exp \left( -\frac{(d-\mu^*)^2}{2\sigma^{*2}} \right) - \left( \frac{c-\mu^*}{\sigma^*} \right) \exp \left( -\frac{(c-\mu^*)^2}{2\sigma^{*2}} \right) \right) - \frac{2\sigma^*(\mu^*-\tilde{\mu}_{(T)})}{\sqrt{2\pi}} \left( \exp \left( -\frac{(d-\mu^*)^2}{2\sigma^{*2}} \right) - \exp \left( -\frac{(c-\mu^*)^2}{2\sigma^{*2}} \right) \right) + (\sigma^{*2} + (\mu^* - \tilde{\mu}_{(T)})^2) \left( \Phi \left( \frac{d-\mu^*}{\sigma^*} \right) - \Phi \left( \frac{c-\mu^*}{\sigma^*} \right) \right) \right\} \right]$$

ここで、 $c, d$ は各々小さいほうから $\beta\%$ となる点、大きいほうから $\beta\%$ となる点を示す。

#### 4. 実験方法の説明

群 $G_1$ の各データはすべて正しく観測されたものとして扱い、その分布は常に正規分布 $N(0, 1)$ とする。群 $G_2$ は観測誤差が含まれるとし、その密度関数や割合は表1に示すような各パラメータ $\mu, \sigma^2, p, \mu^*, \sigma^{*2}$ によって決まるものとする。これらのパラメータを変化させて作られた群 $G_2$ の混合分布と群 $G_1$ の分布、すなわち正規分布 $N(0, 1)$ に対して4手法LDF, QDF, トリム-LDF, トリム-QDFを用いた判別点 $X_L, X_Q, X_{TL}, X_{TQ}$ を求め、最適な判別点 $X_{opt}$ との絶対誤差を調べる。ここで、最適な判別点 $X_{opt}$ とは、群 $G_2$ のデータがすべて正しく観測された場合、すなわち、観測誤差の割合 $p$ が0のときに誤判別の確率を最小とする判別点のことである。また、QDF, トリム-QDFにおいては、一般に、2点の判別点が導かれるが、多くの場合、その一方の点は無視できるほど最適な判別点より離れた値を取るため、QDF, トリム-QDFでは意味ある方の判別点を各々 $X_Q, X_{TQ}$ とする。トリム-LDF, トリム-QDFの2手法で取り除く割合 $\alpha\%, \beta\%$ については $(\alpha, \beta) = (10, 5)$ と $(20, 10)$ の2通り行った。トリム-QDF

表1 群 $G_2$ のパラメータの取る値

$\mu$	: 1.0, 2.0
$\sigma^2$	: 1.0, 0.5
$\mu^*$	: 3.0, 2.0, 1.0, 0.0, -1.0 (if $\mu = 1.0$ ) 4.0, 3.0, 2.0, 1.0, 0.0 (if $\mu = 2.0$ )
$\sigma^{*2}$	: 0.5, 1.0, 1.5, 2.0
$p$	: 0.1, 0.2
$(\alpha, \beta)$	: (10, 5), (20, 10)

やトリム-LDFでは、群  $G_1$  についてもトリム平均、トリム分散が用いられることは言うまでもない。

5. 実験結果とまとめ

表2-1では群  $G_2$  のパラメータ  $\mu, \sigma^2, p$  が各々、1.0, 1.0, 0.1 のときについての結果を示した。また、 $\alpha, \beta$  は10と5である。このときの最適な判別点  $X_{opt}$  は0.5で、群  $G_1$  の10%-トリム平均は0.5%-トリム分散は0.62である。本表では参考としてデータから観測される平均  $\bar{\mu}$  と分散  $\bar{\sigma}^2$ 、また、10%-トリム平均  $\bar{\mu}_{(T)}$ 、5%-トリム分散  $\bar{\sigma}^2_{(T)}$  を記した。カッコ内の数字は最適な判別点  $X_{opt}$  と各手法の判別点  $X_L, X_Q, X_{TQ}$  との絶対誤差である。表2-2は  $\mu=1.0, \sigma^2=0.5$ 、表2-3は  $\mu=2.0, \sigma^2=1.0$ 、表2-4は  $\mu=2.0, \sigma^2=0.5$  の場合について同様に結果を示した。

トリム平均やトリム分散を用いた場合、用いなかった場合と比較してどの程度良くなったか、または悪くなったかを調べるため次のような計算を行った。たとえば、表2-1の1行目に関して、LDFを用いたときの絶対誤差(0.10)より、トリム-LDFにおける絶対誤差(0.09)を引いて値0.01を導き、QDFについても同様に、値0.06を導いた。すべてのデータに対してこれらの計算を行い、その頻度をまとめたものが表3である。正值はトリム平均、トリム分散を用いたために良くなったことを示し、負値は悪くなったことを示す。

これらの結果をまとめると次のようになる。

- 1) トリム平均、トリム分散を用いなかった場合、 $\mu=1.0, \sigma^2=1.0$  のとき LDF のほうが QDF よりよい(表2-1参照)。 $\mu=1.0, \sigma^2=0.5$  のとき QDF のほうが LDF よりよい(表2-2

表2-1 LDF, QDF, トリム-LDF, トリム-QDFによる判別点(1)

$(\mu, \sigma^2)=(1.0, 1.0), p=0.1, (\alpha, \beta)=(10, 5), X_{opt}=0.5$   
 $G_1: 10\%-\text{TRIMMED MEAN}=0.0, 5\%-\text{TRIMMED VAR.}=0.62$

$\mu^*$	$\sigma^{*2}$	$\bar{\mu}$	$\bar{\sigma}^2$	$X_L$	$X_Q$	$\bar{\mu}_{(T)}$	$\bar{\sigma}^2_{(T)}$	$X_{TL}$	$X_{TQ}$
3.	0.5	1.20	1.31	0.60(0.10)	0.69(0.19)	1.17	0.83	0.59(0.09)	0.63(0.13)
	1.0		1.36		0.70(0.20)	1.16	0.81	0.58(0.08)	0.62(0.12)
	1.5		1.41		0.72(0.22)	1.15	0.80	0.57(0.07)	0.62(0.12)
	2.0		1.46		0.73(0.23)	1.14	0.80	0.57(0.07)	0.61(0.11)
2.	0.5	1.10	1.04	0.55(0.05)	0.56(0.06)	1.11	0.66	0.55(0.05)	0.56(0.06)
	1.0		1.09		0.58(0.08)	1.09	0.68	0.55(0.05)	0.56(0.06)
	1.5		1.14		0.60(0.10)	1.08	0.69	0.54(0.04)	0.56(0.06)
	2.0		1.19		0.61(0.11)	1.08	0.70	0.54(0.04)	0.56(0.06)
1.	0.5	1.00	0.95	0.50(0.00)	0.48(0.02)	1.00	0.58	0.50(0.00)	0.49(0.01)
	1.0		1.00		0.50(0.00)	1.00	0.62	0.50(0.00)	0.50(0.00)
	1.5		1.05		0.52(0.02)	1.00	0.65	0.50(0.00)	0.51(0.01)
	2.0		1.10		0.54(0.04)	1.00	0.67	0.50(0.00)	0.51(0.01)
0.	0.5	0.90	1.04	0.45(0.05)	0.47(0.03)	0.89	0.66	0.45(0.05)	0.46(0.04)
	1.0		1.09		0.49(0.01)	0.91	0.68	0.45(0.05)	0.47(0.03)
	1.5		1.14		0.51(0.01)	0.92	0.69	0.46(0.04)	0.48(0.02)
	2.0		1.19		0.54(0.04)	0.92	0.70	0.46(0.04)	0.49(0.01)
-1.	0.5	0.80	1.31	0.40(0.10)	0.56(0.06)	0.83	0.83	0.41(0.09)	0.50(0.00)
	1.0		1.36		0.58(0.08)	0.84	0.81	0.42(0.08)	0.50(0.00)
	1.5		1.41		0.61(0.11)	0.85	0.80	0.43(0.07)	0.50(0.00)
	2.0		1.46		0.63(0.13)	0.86	0.80	0.43(0.07)	0.50(0.00)

表 2-2 LDF, QDF, トリム-LDF, トリム-QDFによる判別点(2)

$(\mu, \sigma^2) = (1.0, 0.5)$ ,  $p = 0.1$ ,  $(\alpha, \beta) = (10, 5)$ ,  $X_{\text{opt}} = 0.36$   
 $G_1$ : 10%-TRIMMED MEAN = 0.0, 5%-TRIMMED VAR. = 0.62

$\mu^*$	$\sigma^{*2}$	$\hat{\mu}$	$\hat{\sigma}^2$	$X_L$	$X_Q$	$\hat{\mu}_{(T)}$	$\hat{\sigma}^2_{(T)}$	$X_{TL}$	$X_{TQ}$
3.	0.5	1.20	0.86	0.60(0.24)	0.57(0.21)	1.13	0.48	0.57(0.21)	0.54(0.18)
	1.0		0.91		0.58(0.22)	1.12	0.46	0.56(0.20)	0.53(0.17)
	1.5		0.96		0.59(0.23)	1.11	0.45	0.56(0.20)	0.53(0.17)
	2.0		1.01		0.60(0.24)	1.11	0.45	0.55(0.19)	0.52(0.16)
2.	0.5	1.10	0.59	0.55(0.19)	0.45(0.09)	1.09	0.36	0.54(0.18)	0.50(0.14)
	1.0		0.64		0.45(0.09)	1.08	0.37	0.54(0.18)	0.49(0.13)
	1.5		0.69		0.46(0.10)	1.07	0.37	0.53(0.17)	0.49(0.13)
	2.0		0.74		0.48(0.12)	1.06	0.38	0.53(0.17)	0.49(0.13)
1.	0.5	1.00	0.50	0.50(0.14)	0.36(0.00)	1.00	0.31	0.50(0.14)	0.44(0.08)
	1.0		0.55		0.37(0.01)	1.00	0.33	0.50(0.14)	0.44(0.08)
	1.5		0.60		0.38(0.02)	1.00	0.35	0.50(0.14)	0.44(0.08)
	2.0		0.65		0.39(0.03)	1.00	0.36	0.50(0.14)	0.44(0.08)
0.	0.5	0.90	0.59	0.45(0.09)	0.30(0.06)	0.91	0.36	0.46(0.10)	0.38(0.02)
	1.0		0.64		0.31(0.05)	0.92	0.37	0.46(0.10)	0.39(0.03)
	1.5		0.69		0.33(0.03)	0.93	0.37	0.47(0.11)	0.40(0.04)
	2.0		0.74		0.34(0.02)	0.94	0.38	0.47(0.11)	0.40(0.04)
-1.	0.5	0.80	0.86	0.40(0.04)	0.33(0.03)	0.87	0.48	0.43(0.07)	0.38(0.02)
	1.0		0.91		0.35(0.01)	0.88	0.46	0.44(0.08)	0.38(0.02)
	1.5		0.96		0.38(0.02)	0.89	0.45	0.44(0.08)	0.38(0.02)
	2.0		1.01		0.41(0.05)	0.89	0.45	0.45(0.09)	0.39(0.03)

表 2-3 LDF, QDF, トリム-LDF, トリム-QDFによる判別点(3)

$(\mu, \sigma^2) = (2.0, 1.0)$ ,  $p = 0.1$ ,  $(\alpha, \beta) = (10, 5)$ ,  $X_{\text{opt}} = 1.0$   
 $G_1$ : 10%-TRIMMED MEAN = 0.0, 5%-TRIMMED VAR. = 0.62

$\mu^*$	$\sigma^{*2}$	$\hat{\mu}$	$\hat{\sigma}^2$	$X_L$	$X_Q$	$\hat{\mu}_{(T)}$	$\hat{\sigma}^2_{(T)}$	$X_{TL}$	$X_{TQ}$
4.	0.5	2.20	1.31	1.10(0.10)	1.10(0.10)	2.17	0.83	1.09(0.09)	1.06(0.06)
	1.0		1.36		1.10(0.10)	2.16	0.81	1.08(0.08)	1.05(0.05)
	1.5		1.41		1.10(0.10)	2.15	0.80	1.07(0.07)	1.05(0.05)
	2.0		1.46		1.10(0.10)	2.14	0.80	1.07(0.07)	1.04(0.04)
3.	0.5	2.10	1.04	1.05(0.05)	1.05(0.05)	2.11	0.66	1.05(0.05)	1.05(0.05)
	1.0		1.09		1.05(0.05)	2.09	0.68	1.05(0.05)	1.04(0.04)
	1.5		1.14		1.05(0.05)	2.08	0.69	1.04(0.04)	1.03(0.03)
	2.0		1.19		1.05(0.05)	2.08	0.70	1.04(0.04)	1.03(0.03)
2.	0.5	2.00	0.95	1.00(0.00)	1.00(0.00)	2.00	0.58	1.00(0.00)	1.01(0.01)
	1.0		1.00		1.00(0.00)	2.00	0.62	1.00(0.00)	1.00(0.00)
	1.5		1.05		1.00(0.00)	2.00	0.65	1.00(0.00)	1.00(0.00)
	2.0		1.10		1.00(0.00)	2.00	0.67	1.00(0.00)	0.99(0.01)
1.	0.5	1.90	1.04	0.95(0.05)	0.95(0.05)	1.89	0.66	0.95(0.05)	0.94(0.06)
	1.0		1.09		0.95(0.05)	1.91	0.68	0.95(0.05)	0.95(0.05)
	1.5		1.14		0.96(0.04)	1.92	0.69	0.96(0.04)	0.95(0.05)
	2.0		1.19		0.96(0.04)	1.92	0.70	0.96(0.04)	0.95(0.05)
0.	0.5	1.80	1.31	0.90(0.10)	0.93(0.07)	1.83	0.83	0.91(0.09)	0.90(0.10)
	1.0		1.36		0.93(0.07)	1.84	0.81	0.92(0.08)	0.91(0.09)
	1.5		1.41		0.94(0.06)	1.85	0.80	0.93(0.07)	0.92(0.08)
	2.0		1.46		0.94(0.06)	1.86	0.80	0.93(0.07)	0.92(0.08)

表2-4 LDF, QDF, トリム-LDF, トリム-QDFによる判別点(4)

$(\mu, \sigma^2) = (2.0, 0.5)$ ,  $p = 1.0$ ,  $(\alpha, \beta) = (10, 5)$ ,  $X_{opt} = 1.05$   
 $G_1: 10\% - \text{TRIMMED MEAN} = 0.0$ ,  $5\% - \text{TRIMMED VAR.} = 0.62$

$\mu^*$	$\sigma^{*2}$	$\hat{\mu}$	$\hat{\sigma}^2$	$X_L$	$X_Q$	$\hat{\mu}_{(T)}$	$\hat{\sigma}^2_{(T)}$	$X_{TL}$	$X_{TQ}$
4.	0.5	2.20	0.86	1.10(0.05)	1.11(0.06)	2.13	0.48	1.07(0.02)	1.10(0.05)
	1.0		0.91		1.11(0.06)	2.12	0.46	1.06(0.01)	1.10(0.05)
	1.5		0.96		1.10(0.05)	2.11	0.45	1.06(0.01)	1.10(0.05)
	2.0		1.01		1.10(0.05)	2.11	0.45	1.05(0.00)	1.10(0.05)
3.	0.5	2.10	0.59	1.05(0.00)	1.09(0.04)	2.09	0.36	1.04(0.01)	1.12(0.07)
	1.0		0.64		1.08(0.03)	2.08	0.37	1.04(0.01)	1.11(0.06)
	1.5		0.69		1.07(0.02)	2.07	0.37	1.03(0.02)	1.11(0.06)
	2.0		0.74		1.07(0.02)	2.06	0.38	1.03(0.02)	1.10(0.05)
2.	0.5	2.00	0.50	1.00(0.05)	1.05(0.00)	2.00	0.31	1.00(0.05)	1.10(0.05)
	1.0		0.55		1.04(0.01)	2.00	0.33	1.00(0.05)	1.08(0.03)
	1.5		0.60		1.03(0.02)	2.00	0.35	1.00(0.05)	1.08(0.03)
	2.0		0.65		1.02(0.03)	2.00	0.36	1.00(0.05)	1.07(0.02)
1.	0.5	1.90	0.59	0.95(0.10)	0.97(0.08)	1.91	0.36	0.96(0.09)	1.02(0.03)
	1.0		0.64		0.96(0.09)	1.92	0.37	0.96(0.09)	1.02(0.03)
	1.5		0.69		0.96(0.09)	1.93	0.37	0.97(0.08)	1.03(0.02)
	2.0		0.74		0.95(0.10)	1.94	0.38	0.97(0.08)	1.03(0.02)
0.	0.5	1.80	0.86	0.90(0.15)	0.90(0.15)	1.87	0.48	0.93(0.12)	0.96(0.09)
	1.0		0.91		0.90(0.15)	1.88	0.46	0.94(0.11)	0.97(0.08)
	1.5		0.96		0.90(0.15)	1.89	0.45	0.94(0.11)	0.97(0.08)
	2.0		1.01		0.90(0.15)	1.89	0.45	0.95(0.10)	0.98(0.07)

表3 トリム平均, トリム分散を用いた場合の効果

$(\alpha, \beta)$	(10, 5)				(20, 10)				
	0.1		0.2		0.1		0.2		
	LDF/QDF	LDF	QDF	LDF	QDF	LDF	QDF	LDF	QDF
0.10以上			4		10		8	3	19
0.09					4		4	3	3
0.08			5		2		1	3	4
0.07			3	3	5		3	3	4
0.06			6	4	5		4	4	2
0.05		3	5	5	7	9	3	5	2
0.04		6	3	12	3	7	1	11	7
0.03		11	3	10	2	8	4	4	3
0.02		8	5	7	8	10	5	13	2
0.01		16	6	11	4	10	2	7	4
0.00		24	9	20	7	24	5	16	2
-0.01		4	12	5	9	3	11	5	3
-0.02		4	6	1	3	5	9		10
-0.03		1	5	2	1	3	3	1	4
-0.04		2	2		3	1	3	2	
-0.05		1	3		4	3	6		2
-0.06			1		2		1		1
-0.07			1				2		3
-0.08			1		1		1		2
-0.09							1		1
-0.10以下							3		2

- 参照)。  $\mu=2.0$  のときは2手法にあまり差はないが QDF のほうが少しよい(表 2-3, 2-4 参照)。
- 2) 観測誤差の割り合い  $p$  が 0.1 のときより 0.2 のほうが両手法とも悪くなる。つまり、最適な判別点  $X_{opt}$  からの絶対誤差が大きくなる。
  - 3) トリム平均, トリム分散を用いたほうが用いないときよりも良くなる傾向を示している(表 3 参照)。
  - 4) トリム平均, トリム分散を用いた場合,  $\mu=1.0$ ,  $\sigma^2=1.0$  のときトリム-LDF とトリム-QDF の間にあまり差はないがトリム-LDF のほうが少しよい(表 2-1 参照)。  $\mu=1.0$ ,  $\sigma^2=0.5$  のときトリム-QDF のほうがトリム-LDF よりよい(表 2-2 参照)。  $\mu=2.0$  のとき2手法にほとんど差はないがトリム-QDF のほうが少しよい(表 2-3, 2-4 参照)。
  - 5) トリム平均, トリム分散を用いた場合,  $p=0.1$  より  $p=0.2$  のほうが両手法とも悪くなる。
  - 6) トリムの割り合い  $(\alpha, \beta)$  に関しては,  $(\alpha, \beta)=(10, 5)$  より  $(20, 10)$  のほうが良いことが多い(表 3 参照)。

## 6. お わ り に

これらの結果より, 観測誤差が含まれる場合は, 得られた全データより平均, 分散をそのまま用いるのではなく, トリム平均, トリム分散を用いたほうが一般に良くなることがわかった。トリムの割り合い  $\alpha, \beta$  に関しては今後, より詳しく調べなければならないと思う。また, トリム-LDF を用いるべきか, トリム-QDF を用いるべきかという問題についても簡単に述べることはできないが, 2群の平均の差が小さく, 分散がほぼ同じ場合はトリム-LDF のほうが良いと思われる。

今回, ロバストな平均, 分散を得るためにトリム法を用いたが他にも多くの手法<sup>3)</sup>があり, それらとの関係を見ることも重要であると思われる。

## 参 考 文 献

- 1) 河口至商: 多変量解析入門 I (昭 48), p. 79, 森北出版
- 2) R. ニャナデシカン: 統計的多変量データ解析 (昭 54), p. 114, 日科技連
- 3) Launer, R. L. and Wilkinson, G. N.: Robustness in statistics (1979), p. 1, Academic Press
- 4) Hawkins, D. M.: Identification of Outliers (1980), p. 2, Chapman and Hall