判別分析における丘本先生の貢献と最近の発展

中央大　藤越康祝

丘本先生 (以下では敬称を略す) の研究成果は, 統計的推測理論, 判別分析, 主成分・因子分析, 数量化理論, など統計学の幅広い分野に及んでいる．ここでは, 判別分析の分野における主要な成果を紹介し, 次にそれらの成果に関連して, 漸近展開の誤差評価, 高次元枠組みでの漸近近似, さらには, 高次元特有の判別法, など最近の発展を概観する.

丘本は判別分析の分野において, ５編の論文を発表している.

Okamoto, M. (1961). Discrimination for variance matrices. *Osaka Math. J.*, **13**, 1-39.

Okamoto, M. (1963). An asymptotic expansion for the distribution of the linear discriminant function. *Ann. Math. Statist.*, **34**, 1286-1301.

Memon, A.Z. and Okamoto, M. (1970). The classification statistic W^* in covariate discriminant analysis. *Ann. Math. Statist.*, **41**,1491-1499.

Memon, A. Z. and Okamoto, M. (1971). Asymptotic expansion of the distribution of the Z statistic in discriminant analysis. *J. Multivariate Anal.*, **1**, 294-307.

Sedransk, N. and Okamoto, M. (1971). Estimation of the probabilities of misclassification for a linear discriminant function in the univariate case. *Ann. Inst. Statist. Math.*, **23**, 419-435.

これらは２群の場合の判別問題である．Okamoto (1961) では, 両群の平均ベクトルは等しいが共分散行列が異なる場合の判別問題を扱っている．母集団パラメータが既知の場合におけるベイズ法, ミニマックス法などを与え, それらの誤判別確率を評価している．母集団パラメータが未知の場合には推定量を代入する方法を提案し, １卵双生児と２卵双生児とを 10 個の特性量で分類する問題に適用している．これは学位論文である.

Okamoto (1963) では, 線形判別関数 W の漸近展開を導出している．これは, 丘本の代表作の１つである．当時, このような漸近展開の導出は, 代表的多変量検定の尤度比統計量の帰無分布と平均ベクトルの有意性に関するホテリング統計量に対して与えられていたに過ぎなかった．前者は Box (1949) により, 後者は Siotani (1957), Ito (1960) によって求められた．丘本による方法

は, 塩谷等によって用いられた微分作用素の方法を拡充発展させるものであるが, 判別の分野での漸近展開として最初の成果であると共に, 得られた結果の有用性も高く, その後の発展に大きく影響を与えている．一般に漸近展開の導出の多くは特性関数を展開し, それを形式的に反転させる方法によっていて, 誤差項のオーダー評価はこの時点では厳密には与えられていなかった.

線形判別関数の漸近展開と関連して, 2次判別統計量 Z, および, 共変量がある場合の線形判別関数 W^* の漸近展開も導出している (Memon and Okamoto (1970), Memon and Okamoto (1971)). 判別分析におけるこの他の研究成果として, Sedransk and Okamoto (1971) において, 1次元の場合における線形判別関数 W の条件付誤判別確率の推定問題を論じている.

このような丘本による成果と関連して, Anderson (1973) は, スチューデント化された線形判別関数の分布の漸近展開を与えている．そこでは, 漸近展開の妥当性, すなわち, 残差項のオーダー評価が証明されていることを注意したい. さらに, 丘本による W の漸近展開の妥当性も同様に示せることにも言及している．Siotani and Wang (1977) では, W と Z に対して, O_3 の項までの漸近展開を与えている.

一般に, 大標本の枠組での近似の精度は次元が大きくなるにつれて悪くなる．これを克服する方法として, 標本数に加え次元数も大きいとした高次元漸近的枠組のもでの近似が研究されている. たとえば, ロシアの研究者 Deev (1970), Rudys (1972) などによる線形判別関数に関する結果がある．線形判別関数や2次判別関数を含むクラスについては Fujikoshi and Seo (1998) によって数値的に調べられている. 次元 p が大きいとして得られる近似公式の精度は著しく良く, また, p が小さいときにも有効げあることは注目されてよい. さらに, Fujikoshi (2000) では, 高次元漸近近似に対して誤差限界を導出している.

$p > n$ の場合には, 標本共分散行列 S が特異になるため, 線形判別関数 W を直接利用することはできなくなる．このため, 何らかの修正が必要である. Loh (1997) は, Friedman (1989) の正則化判別解析法の考えに沿って, 適応型リッジ判別関数について考察している．大標本のもとで誤判別確率を漸近的に評価しているが, 高次元の枠組みでの評価は与えれていない. Srivastava (2007) は, S の逆行列の代わりにムーアペンローズ逆行列を用いることを提案している．高次元小標本データの高次元漸近的挙動を調べ, その結果を高次元判別理論に応用する研究も行われている (Hall et al. (2005)).