ノンパラメトリック確率分布による確率的DEA法

01604524 大阪大学 ＊森田 浩 MORITA Hiroshi
大阪大学 羽場 洋介 HABA Yosuke

1．はじめに

確率的データによる効率性分析のための確率的 DEA 法には機会制約条件モデルがよく使われている[1]。生産可能集合を規定している条件を、機会制約条件に拡張したものである。そこでは制約条件の充足確率を算出するために、多くの場合、確率的データに正規性を仮定している。正規分布は2次モーメントまで与えると分布形を特定することができるため、機会制約条件は2次関数で表すことができ、平方根を含んだ2次計画問題として等価確定問題を導くことができる[2]。このとき問題となるのは、分布パラメータをどう与えるかである。これらは既知としているものが多いが、実際にデータから推定するとなれば、平均値は与えられるとしても、分散を推定するには相応のデータ数が必要となる。カルマンフィルターなどの他の統計手法によって求められた推定値を使って評価しようとする場合には、その推定誤差の分布をデータの分散とすることもできる[5]。いずれにしても、分散成分の推定には確率分布を仮定しており、そこではパラメータを推定している。このパラメトリックモデルはデータに対する仮定であって、効率性評価のためのパラメトリックモデルではないので、DEA のもっているノンパラメトリック性は失われない。しかしながら、推定したパラメータの値は効率性評価の結果に敏感に影響するため、その推定精度を高めることは重要であるが、必ずしも十分なデータがあるとは限らない。むしろ非常に少ないデータしかない場合の方が一般的であろう。このような状況に対して、確率分布形を仮定せず、与えられたデータのみに基づいてノンパラメトリックな分布を考え、確率的な効率性を評価するための DEA 法を考察する。

2．機会制約条件をもつ確率的DEA法

CCR モデルで考えることにすると、機会制約条件付き計画問題における確率最大化モデルは次のように与えられる。

$$
\begin{aligned}
\max \quad & \Pr\left\{\frac{u'\tilde{y}_o}{v'\tilde{x}_o} \ge 1\right\} \\
st \quad & \Pr\left\{\frac{u'\tilde{y}_j}{v'\tilde{x}_j} \le 1\right\} \ge \alpha_j, j = 1,\dots,n \\
& u \ge 0, v \ge 0
\end{aligned}
\tag{1}
$$

ただし、$(\tilde{x},\tilde{y})$は確率変数として与えられる入出力値、αは制約条件が満たされなければならない確率レベルを表している。問題(1)の最適値α^*を確率的効率性と呼ぶ。α^*は常にα_o以下の値をとるが、$\alpha^* = \alpha_o$のとき、DMU_oは**確率的に効率的**(stochastic efficient)であるという。

入力データは確定的とし、出力データのみに正規性$\tilde{y}_j \sim N(\bar{y}_j, \Sigma_j)$を仮定したとき、問題(1)の等価確定問題を導出すると、適当な変数変換により

$$
\begin{aligned}
\max \quad & \mu'\bar{y}_o - \nu'x_o \\
st \quad & \mu'\bar{y}_j - \nu'x_j - \Phi^{-1}(\alpha_j)\zeta_j \le 0, j = 1,\dots,n \\
& C_j\left(\zeta_j^2 - \mu'\Sigma_j\mu\right) \ge 0, j = 1,\dots,n \\
& \mu'\Sigma_o\mu \ge 1, \mu \ge 0, \nu \ge 0, \zeta \ge 0
\end{aligned}
\tag{2}
$$

のように2次計画問題とすることができるが、平均や分散の値が制約条件や目的関数に現れ、それらの推定精度が悪ければ、その影響は直接最適値に反映されることになる。

分布形を仮定せず、観測データのみからデータの不確実性を記述するのにブーツストラップ法[3]もあるが、数個といった少ないデータに適用するには十分ではない。一方、2つのデータのみがある場合には一様分布を考えていることになるが、同様のデータを区間データとみて解析した効率性分析法[4]もある。そこではデータのとり得るすべての可能性を考慮した上での効率値の区間を与えており、その区間幅は一般に非常に大きいものになってしまう。

3．ノンパラメトリック機会制約条件をもつ確率的DEA法

k個のデータ$x_1, x_2, \dots, x_k$が与えられたとき、正規分布を仮定するなら、その平均$\bar{x}$と分散s^2により、正規分布$N(\bar{x}, s^2)$に従う確率変数であるとみなせる。この平均$\bar{x}$と分散s^2は推定値であり、そ

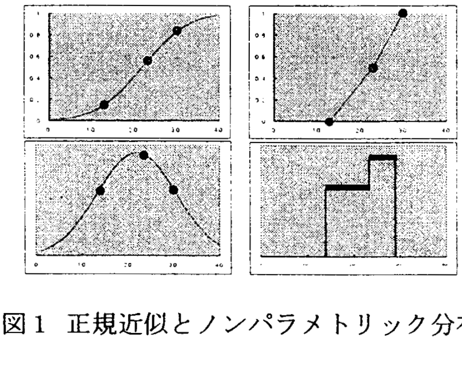

図1 正規近似とノンパラメトリック分布

のばらつきは $V(\bar{x}) = \sigma^2/k, V(s^2) = 2\sigma^4/(k-1)$ となり、データ数が十分でなければ、特に分散の推定精度が著しく劣る。

今、データが大きさの順に並べ替えられているとして、ノンパラメトリックな分布を

$$F(t) = \begin{cases} 0 & t < x_i \\ \dfrac{1}{k}\left(\dfrac{t - x_i}{x_{i+1} - x_i} + i - 1\right) & x_i \le t < x_{i+1} \\ 1 & t \ge x_{i+1} \end{cases} \quad (3)$$

で与える。図1に正規近似との対比を示す。データが2個ならば一様分布 $U[x_1, x_2]$ と一致する。

問題(1)を(3)式の確率分布を用いて展開し、最適値を計算することによって確率的効率性を求める。また超効率値の考え方を適用した問題(4)では、

$$\begin{aligned} \max \quad & \Pr\left\{\frac{u'\tilde{y}_o}{v'\tilde{x}_o} \ge 1\right\} \\ st \quad & \Pr\left\{\frac{u'\tilde{y}_j}{v'\tilde{x}_j} \le 1\right\} \ge \alpha_j, j \ne o \qquad (4) \\ & u \ge 0, v \ge 0 \end{aligned}$$

確率的効率性の値は α_o 以上となることができ、他の DMU が生産可能集合に確率 α_j 以上で含まれるときに効率的と評価される確率を求めることができる。

4. 簡単な数値例

1入力1出力の5つのDMUがあり、各入出力には2つずつの観測値があるものとしている。入出力データを表1と図2に示す。

表2 入出力データ

	A	B	C	D	E
入力	1, 2	2, 4	3, 5	4, 5	6, 7
出力	1, 2	3, 4	5, 7	1, 3	2, 3

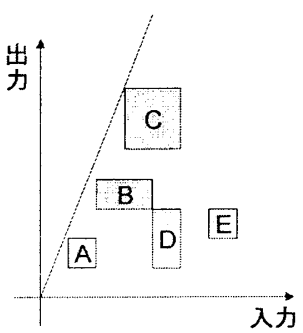

図2 入出力データの図

A, B, C は最も良い値をとるときには効率的となることがあるが、DとEはどのような値をとっても効率的となることはない。区間データとして解析したときの区間効率値の最大値と最小値、および提案したノンパラメトリック機会制約モデルによる効率性(目的関数値)を90%の確率条件のもとで求めたものを表3に示す。

表3 区間効率値と確率的効率性

	A	B	C	D	E
最大値	1	1	1	0.750	0.500
最小値	0.214	0.321	0.500	0.086	0.122
確率的効率性	0.17%	0.34%	10%	0%	0%

区間効率値では、その区間幅は最小のものでも0.5もあり、どのような判断をすればよいか難しい。確率的効率性は、10%となるものが確率的に効率的となり、Cのみが確率的に効率的である評価される。確率的超効率性の値は 37.8%となり、効率的と評価される確率が得られる。

5. 今後の課題

まだ手計算をしている段階で、入出力数やデータの観測数が増えた場合には対応できていない。データ数が5,6個もあれば正規分布近似すれば良いだろう。入出力数というより不確実データを持つ入出力数が問題となり、3項目以上あるだけでかなり煩雑になる。確率条件の水準を限定したりすることで計算の効率化も検討できるものと考えている。

参考文献

[1] W. W. Cooper, Z. Huang and S. X. Li, Satisficing DEA models under chance constraints, Annals of Operations Research, vol. 66, pp. 279-295 (1996).

[2] H. Morita and L. M. Seiford, Characteristics of stochastic DEA efficiency – Reliability and probability being efficient –, Journal of Operations Research Society of Japan, vol. 42, no. 4, pp. 389-404 (1999).

[3] L. Simar and P.W. Wilson, Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models, Management Science, vol. 44, pp. 49-61, (1998).

[4] 円谷友英, 前田豊, 田中英夫, 区間効率値によるDEAモデル, オペレーションズ・リサーチ, vol. 44, no. 8, pp. 425-434, (1999).

[5] T. Ueda and K. Hoshino, Estimation of efficiencies using Kalman filter and stochastic efficiency model, IFORS'02, Edinburgh, (2002).