

関数関係モデルによる大気中 CoPCB の発生源に関する推測

統計数理研究所 柏木 宣久
東京都環境局環境改善部 佐藤 博
芳住登紀子

1. はじめに

関数関係モデルあるいは構造関係モデルでは、誤差等のため「真」の値を直接観測できない変数あるいは変量の間に関数関係が仮定される (Kendall and Stuart (1979), Chap.29)。初等的な物理法則をはじめ、こうした設定を必要とする場面は多い。しかしながら、主成分分析等の特別な場合を除き、これらのモデルが実際に利用されるのは希であった。推定が常に可能とは限らないからである。

今、 η と ξ を変数とし、両者に

$$\xi = \theta\eta$$

という関数関係を仮定する。ただし、 θ は媒介変数である。そして、両変数の観測値には誤差が含まれており、以下により定義される変量 X と Z の値を観測しているとする。

$$\begin{aligned}x_i &= \eta_i + u_i, & u_i &\sim \text{NID}(0, \sigma_u^2), \\z_i &= \xi_i + w_i, & w_i &\sim \text{NID}(0, \sigma_w^2), \\i &= 1, \dots, n.\end{aligned}$$

この時、尤度は以下のように書ける。

$$L \propto \sigma_u^{-n} \sigma_w^{-n} \exp \left\{ -\frac{1}{2\sigma_u^2} \sum_i (x_i - \eta_i)^2 - \frac{1}{2\sigma_w^2} \sum_i (z_i - \theta\eta_i)^2 \right\}.$$

ここで、 $\hat{\eta}_i = x_i$ という推定値を考えると、 $\hat{\sigma}_u^2 = 0$ となり、更に $L = \infty$ となる。従って、 θ の推定は不定となる。

上述の推定問題は、誤差分散に対し適切な仮定、例えば $\sigma_u^2 = \sigma_w^2$ という仮定を設定すれば、ほとんどの場合、解消される。ただし、常に許容される普遍的な仮定は存在しない。誤差分散に対する仮定は推定に大きな影響を与える可能性があるため、仮定を設定できるのは先験情報が存在する場合に限られる。そのため、関数関係モデルあるいは構造関係モデルは一般的なモデルとはならなかった。

本報告では、大気中のコプラナーポリ塩化ビフェニール (CoPCB) の発生源寄与率を推定するための関数関係モデルについて述べる。発生源寄与率を推定するためには、大気および発生源における CoPCB の濃度間に関数関係を仮定する必要があるが、濃度の観測値には誤差が含まれて

いるため、モデルは必然的に関数関係モデルとなる。この関数関係モデルを利用可能と判断したのは、観測値が比率の形で与えられており、全ての観測値が同様の測定方法により得られていたからである。

2. データ

CoPCB データの解析は、ダイオキシン類データの解析の雛形として企画された。本節では、先ずダイオキシン類について説明した後、本報告で用いたデータを提示する。

ここで言うダイオキシン類とは、ポリ塩化ジベンゾパラジオキシン (PCDD)、ポリ塩化ジベンゾフラン (PCDF)、および CoPCB の3種類の有機化合物を指す。何れも2個のベンゼン環からなる芳香族で、種類の違いはベンゼン環を橋渡しする酸素原子の数に起因する。ダイオキシン類には、ベンゼン環に付く塩素原子の数と位置により、多くの異性体が存在する。例えば、PCDD で 75 種、PCDF で 135 種ある。これらは分離分析法により識別可能である。各異性体の毒性は同一ではなく、毒性が最も強いとされる PCDD の異性体である 2,3,7,8-TCDD の毒性を 1 とする毒性等価係数 (TEF) が定められている。一般にダイオキシン濃度として公表される値は TEF に基づく毒性当量 (TEQ) である場合が多い。

表 1. 異性体名と TEF

#	異性体名	TEF
77	3,3',4,4'-T4CB	0.0001
81	3,4,4',5-T4CB	0.0001
105	2,3,3',4,4'-P5CB	0.0001
114	2,3,4,4',5-P5CB	0.0005
118	2,3',4,4',5-P5CB	0.0001
123	2',3,4,4',5-P5CB	0.0001
126	3,3',4,4',5-P5CB	0.1
156	2,3,3',4,4',5-H6CB	0.0005
157	2,3,3',4,4',5'-H6CB	0.0005
167	2,3',4,4',5,5'-H6CB	0.00001
169	3,3',4,4',5,5'-H6CB	0.01
189	2,3,3',4,4',5,5'-H7CB	0.0001

表 2. 異性体構成比のデータ

#	焼却	製品	大気
77	0.10405	0.04316	0.09745
81	0.03716	0.01271	0.01816
105	0.07184	0.19873	0.19283
114	0.03303	0.02279	0.02293
118	0.14203	0.49518	0.50576
123	0.01982	0.03155	0.06338
126	0.14368	0.00329	0.01855
156	0.10983	0.07581	0.03488
157	0.05615	0.01402	0.01136
167	0.04955	0.09378	0.01643
169	0.08258	0.00044	0.00604
189	0.15029	0.00855	0.01224
合計	1.0	1.0	1.0

CoPCB にも、他のダイオキシン類と同様、多くの異性体がある。その内、本報告で用いたのは、表 1 に示す TEF が定められている 12 種の異性体である。東京都では様々な環境におけるこれら異性体の濃度を継続的に調査しており、大気中からも有意な濃度が検出されている。その主な発生源と考えられるのが、かつての PCB 製品からの蒸散であり、もうひとつが焼却炉からの排気である。これらの発生源が大気中の濃度にどの程度寄与しているかが分かれば、環境改善策

を策定する上で有用な情報となる。そこで、焼却、製品、大気における代表的な異性体構成比をデータとし、発生源寄与率の推定を試みた。使用したデータを表2に示す。ただし、今回のデータはあくまで代表的な値（ある種の平均値）であるため、いくつかの論理的矛盾を抱えている。すなわち、焼却と製品のみを発生源とし、移流・拡散過程での種々の状態変化を無視できるとすれば、大気中濃度は焼却と製品の間位置するはずである。しかしながら、5つの異性体でこの条件が満たされていない。従って、今回の解析結果は必ずしも精密な結果とは言えないが、少なくともひとつの目安は得られると考える。

3. 方法

先ず、モデルを提示する。今、 ζ_i 、 η_i および ξ_i を、各々、焼却、製品および大気における i 番目の異性体の構成比とする。ただし、 $1 \leq i \leq n$ であり（今の場合、 $n = 12$ ）、 $\sum \zeta_i = \sum \eta_i = \sum \xi_i = 1$ である。焼却と製品に由来する各異性体が混合して大気における異性体構成比が決まるとすれば、変数間に

$$\xi_i = \theta \zeta_i + (1 - \theta) \eta_i$$

という関数関係を仮定できる。ただし、 θ は混合比率である。そして、異性体構成比の観測値には誤差が含まれており、以下により定義される変量 X_i 、 Y_i および Z_i の値を観測しているとする。

$$\begin{aligned} x_i &= \zeta_i + u_i, \\ y_i &= \eta_i + v_i, \\ z_i &= \xi_i + w_i = \theta \zeta_i + (1 - \theta) \eta_i + w_i. \end{aligned}$$

ただし、 $E(u_i) = E(v_i) = E(w_i) = 0$ 、 $\text{cov}(u_i, v_j) = \text{cov}(u_i, w_j) = \text{cov}(v_i, w_j) = 0$ および

$$\begin{aligned} \text{cov}(u_i, u_j) &= \begin{cases} \rho^2 x_i(1 - x_i) & i = j \\ -\rho^2 x_i x_j & i \neq j \end{cases}, \\ \text{cov}(v_i, v_j) &= \begin{cases} \rho^2 y_i(1 - y_i) & i = j \\ -\rho^2 y_i y_j & i \neq j \end{cases}, \\ \text{cov}(w_i, w_j) &= \begin{cases} \rho^2 z_i(1 - z_i) & i = j \\ -\rho^2 z_i z_j & i \neq j \end{cases} \end{aligned}$$

であり、 ρ^2 は未知乗数である。最後に、推定を簡単にするため、各変量は正規分布に従うと仮定する。ちなみに、ロジットモデルや多項分布の利用も考えられるが、これらについては状況の推移を見ながら検討していく。

仮定したモデルの場合、第1節で述べた推定問題は生じない。観測値が比率の形で与えられるため、共分散の相対的な大きさを観測値を用いて定義し、全ての観測値が同様の測定方法により得られていたため、共分散の絶対的な大きさを調節するための乗数を共通にしたからである。より精密には、共分散の相対的な大きさは ζ_i 等の変数を用いて定義すべきであるが、そうすると第

1 節で述べた推定問題が生じてしまう。経験によれば、変数の代わりに観測値を用いて定義しても、観測値がある程度信頼できる場合には、推定に際し大きな偏りが生じる可能性は低い。

さて、仮定したモデルに対する尤度は以下の様に見える。

$$L \propto \rho^{-3(n-1)} |U|^{-1/2} |V|^{-1/2} |W|^{-1/2} \exp \left[-\frac{1}{2\rho^2} \left\{ (\mathbf{x} - \boldsymbol{\zeta})' U^{-1} (\mathbf{x} - \boldsymbol{\zeta}) + (\mathbf{y} - \boldsymbol{\eta})' V^{-1} (\mathbf{y} - \boldsymbol{\eta}) + (\mathbf{z} - \theta \boldsymbol{\zeta} - (1 - \theta) \boldsymbol{\eta})' W^{-1} (\mathbf{z} - \theta \boldsymbol{\zeta} - (1 - \theta) \boldsymbol{\eta}) \right\} \right]$$

ただし、 $\mathbf{x}$ 、 $\mathbf{y}$ 、 $\mathbf{z}$ 、 $\boldsymbol{\zeta}$ 、 $\boldsymbol{\eta}$ および $\boldsymbol{\xi}$ は、 $\mathbf{a} = (a_1, \dots, a_{n-1})'$ のように、同一記号で書かれる変数の最初の $n-1$ 個を要素とする縦ベクトルであり、 $\rho^2 U$ 、 $\rho^2 V$ および $\rho^2 W$ は対応する共分散行列を表す。未知変数は、この尤度を最大にするように推定すればよい。ちなみに、 $\boldsymbol{\eta}$ の最尤推定値は

$$\hat{\boldsymbol{\eta}} = \left\{ V^{-1} + (1 - \theta)^2 W^{-1} - \theta^2 (1 - \theta)^2 W^{-1} (U^{-1} + \theta^2 W^{-1})^{-1} W^{-1} \right\}^{-1} \times \left\{ V^{-1} \mathbf{y} + (1 - \theta) W^{-1} \mathbf{z} - \theta (1 - \theta) W^{-1} (U^{-1} + \theta^2 W^{-1})^{-1} (U^{-1} \mathbf{x} + \theta W^{-1} \mathbf{z}) \right\}$$

と書け、 $\boldsymbol{\zeta}$ の最尤推定値は

$$\hat{\boldsymbol{\zeta}} = (U^{-1} + \theta^2 W^{-1})^{-1} \left[U^{-1} \mathbf{x} + \theta W^{-1} \{ \mathbf{z} - (1 - \theta) \hat{\boldsymbol{\eta}} \} \right]$$

と書け、 ρ^2 の最尤推定値は

$$\hat{\rho}^2 = \frac{1}{3(n-1)} \left\{ (\mathbf{x} - \hat{\boldsymbol{\zeta}})' U^{-1} (\mathbf{x} - \hat{\boldsymbol{\zeta}}) + (\mathbf{y} - \hat{\boldsymbol{\eta}})' V^{-1} (\mathbf{y} - \hat{\boldsymbol{\eta}}) + (\mathbf{z} - \theta \hat{\boldsymbol{\zeta}} - (1 - \theta) \hat{\boldsymbol{\eta}})' W^{-1} (\mathbf{z} - \theta \hat{\boldsymbol{\zeta}} - (1 - \theta) \hat{\boldsymbol{\eta}}) \right\}$$

と書ける。ただし、 θ の最尤推定値は陽に書けないので、数値的方法により推定する。

4. 結果

本節では、前節で述べた方法を表 2 のデータに適用した結果について述べる。

最尤推定の結果、混合比率は $\hat{\theta} = 0.0533$ 、共分散の大きさを調整する乗数は $\hat{\rho} = 0.0595$ と推定された。すなわち、大気中の CoPCB の約 5% が焼却炉からの排気が原因であり、約 95% がかつての PCB 製品からの蒸散が原因と推定された。焼却、製品、大気における異性体構成比の推定値を表 3 に示す。

表３．異性体構成比の推定値

#	焼却	製品	大気
77	0.10620	0.06198	0.06434
81	0.03744	0.01529	0.01647
105	0.07181	0.21098	0.20356
114	0.03300	0.02396	0.02444
118	0.14221	0.53984	0.51864
123	0.02020	0.04461	0.04331
126	0.14664	0.00471	0.01228
156	0.10722	0.04901	0.05211
157	0.05550	0.01208	0.01440
167	0.04744	0.02936	0.03033
169	0.08356	0.00056	0.00498
189	0.14878	0.00761	0.01513
合計	1.0	1.0	1.0

ちなみに、焼却由来を 5.33%、製品由来を 94.67% とした後、異性体構成比の推定値に TEF を掛け、TEQ に換算すると、表４の結果が得られる。そして、この TEQ 換算値を、焼却と製品の合計が 1 となるよう比率に直すと、表５の結果が得られる。これにより、毒性当量で見ると、大気中の CoPCB の約 60% が焼却炉からの排気が原因であり、約 40% がかつての PCB 製品からの蒸散が原因と推定される。

表４．TEQ 換算値

#	焼却	製品
77	0.00000057	0.00000587
81	0.00000020	0.00000145
105	0.00000038	0.00001997
114	0.00000088	0.00001134
118	0.00000076	0.00005111
123	0.00000011	0.00000422
126	0.00078161	0.00044629
156	0.00000286	0.00002320
157	0.00000148	0.00000572
167	0.00000003	0.00000028
169	0.00004454	0.00000531
189	0.00000079	0.00000072
合計	0.00083419	0.00057548

表５．TEQ 換算値の比率

#	焼却	製品
77	0.00040154	0.00416262
81	0.00014156	0.00102651
105	0.00027153	0.01416880
114	0.00062385	0.00804643
118	0.00053768	0.03625421
123	0.00007639	0.00299618
126	0.55446140	0.31659417
156	0.00202696	0.01645666
157	0.00104920	0.00405700
167	0.00001794	0.00019720
169	0.03159330	0.00376538
189	0.00056255	0.00051095
合計	0.59176390	0.40823610

5. 信頼区間

第3節で述べたように、 θ の最尤推定値は陽に書けない。しかも、他の変数の最尤推定値は全て θ の関数として書かれているため、少なくとも現時点では、各推定量の信頼区間を陽に求めるのは困難である。そこで、信頼区間をパラメトリックブートストラップにより求める。

パラメトリックブートストラップを実行するには、推定値を基に表2に相当するデータを乱数により生成して第3節の方法により推定する手続きを繰り返せばよい。データの生成には、本来であれば、第3節のモデルに基づく多変量正規乱数を用いるべきであるが、正規分布の仮定はあくまで近似に過ぎない。そこで、ここでは、表3の値を多項確率とする多項乱数を発生させ、各項の生起比率をデータとする。一回の試行で発生させる多項乱数の数、すなわち、標本の大きさ m は、第3節のモデルと多項確率 p_i の最尤推定量 $\hat{p}_i$ の性質

$$\text{var}(\hat{p}_i) = \frac{p_i(1-p_i)}{m}$$

を考慮し、 $m = 1/\hat{\rho}^2$ により定める。今の場合、 $m \approx 282$ となる。更に、生起比率が0の項が存在すると、推定に支障が生じるため、各項に予め比率 0.0001 を割り当てておき、最終的に比率の合計が1となるよう調整する。

上述のパラメトリックブートストラップを 10000 回試行した結果、混合比率の平均値は 0.042、標準偏差は 0.029 となった。標準偏差は相当大きく見積もられたが、データの大きさが極めて小さい現状では、当然の結果と言える。今後、データを整備し、再検討していく予定である。なお、参考のため、異性体構成比の平均値と標準偏差を表6および7に示す。

表6. 平均値

#	焼却	製品	大気
77	0.10585	0.06165	0.06354
81	0.03748	0.01399	0.01498
105	0.07193	0.21220	0.20626
114	0.03325	0.02262	0.02308
118	0.14219	0.54619	0.52912
123	0.02016	0.04363	0.04265
126	0.14684	0.00438	0.01036
156	0.10726	0.04868	0.05118
157	0.05540	0.01092	0.01282
167	0.04740	0.02827	0.02909
169	0.08297	0.00032	0.00380
189	0.14928	0.00716	0.01312
合計	1.0	1.0	1.0

表7. 標準偏差

#	焼却	製品	大気
77	0.01823	0.01111	0.01061
81	0.01123	0.00589	0.00568
105	0.01535	0.01786	0.01728
114	0.01066	0.00718	0.00688
118	0.02070	0.02250	0.02282
123	0.00840	0.00933	0.00896
126	0.02081	0.00382	0.00506
156	0.01834	0.00998	0.00955
157	0.01355	0.00544	0.00527
167	0.01269	0.00784	0.00749
169	0.01601	0.00121	0.00284
189	0.02099	0.00479	0.00558