

第IV 部門

長期記憶性を考慮したヘイズ災害モデルの構築

京都大学工学部
京都大学大学院学生会員 ○牛山 泰喜
正会員 小林 潔司

1 はじめに

マレーシアにおいては1990年以降現在に至るまで「ヘイズ」と呼ばれる大気汚染がほぼ毎年観測されている。主な原因は隣国インドネシアのスマトラ島における焼畑農業による森林火災であり、乾季となる5～10月を中心に南西モンスーンに乗ってマレーシア国内に大気汚染物質が飛来することで生じる越境汚染である。最近では2013年に高濃度の汚染物質が観測され非常事態宣言が発令されるなど、深刻な問題となっている。そこで、本研究ではヘイズ災害の先行指標となる火災情報と大気汚染の状況を表す観測データを用いてヘイズの状況を予測する状態変数モデルを構築する。既往研究により、マレーシアのヘイズは1) ヘイズ発生段階における大気汚染データの急激な上昇、2) 大気汚染データの長期記憶性の存在が示唆されている。また、最近マレーシア国内で発生した火災件数が増加傾向にあり、国外だけでなく国内の火災の影響を考慮することが必要になる。

2 状態変数モデル

マレーシア国内に設けられたK箇所の観測局(以下、地点と呼ぶ)においてPM10濃度が観測される。観測局においてPM10に関する観測値の記録が開始された時点をもとに $t = 0$ とする離散的時間軸 $t = 0, 1, 2, \dots$ を導入する。記録が開始された時点以前に関しては、時間軸とは逆方向に $t = -1, -2, \dots$ と記述することとする。大気汚染濃度をS個のカテゴリー(以下、深刻度と呼ぶ)に離散化する。各時点の深刻度は確率変数であり、状態変数 s_t^k ($k = 1, \dots, K, t = \dots, -1, 0, 1, \dots$)を用いて表現する。ただし、 $s_t^k = 1, \dots, S$ であり、sの値が大きくなるほど、大気中のヘイズ濃度が深刻化していることを意味する。期間 $t \in [0, T]$ において、深刻度に関する観測値が得られており、地点kにおいて時点tに観測された深刻度を($s_t^k = 1, \dots, S$)と表す。記号「 $\sim$ 」は観測値であることを意味する。地点kにおいて観

測された深刻度ベクトルを $\bar{s}^k = (\bar{s}_k^0, \dots, \bar{s}_k^T)$ で表現する。

ヘイズ災害の場合、森林火災の数や規模、モンスーン風等の外生的要因により深刻度が変化する。過去の時点に大気中に滞留したPM10は長期にわたって影響を及ぼす。いま、地点kの時点 $t + 1$ における深刻度 s_{t+1}^k は、時点tに至るまでの過去の状態変数 $\mathbf{s}_t^k = (s_t^k, s_{t-1}^k, \dots)$ 、及び火災地点数を表す外生変数ベクトル $\mathbf{v}_t^k = (v_t^k, v_{t-1}^k, \dots)$ の影響を受けて実現する。いま、時点tまでの状態変数ベクトル $\mathbf{s}_t^k$ 、および、外生変数ベクトル $\mathbf{v}_t^k$ を与件としたときに、時点 $t + 1$ に深刻度 $s_{t+1}^k = n$ ($n = 1, \dots, S$)が生起する条件付き確率は次式で表される。

$$P(s_{t+1}^k = n | \mathbf{s}_t^k, \mathbf{v}_t^k) = \pi_{\mathbf{s}_t^k, n}^k(\mathbf{v}_t^k, \boldsymbol{\beta}^k) \quad (1-a)$$

$$\sum_{n=1}^S \pi_{\mathbf{s}_t^k, n}^k(\mathbf{v}_t^k, \boldsymbol{\beta}^k) = 1 \quad (1-b)$$

ただし、 $\boldsymbol{\beta}^k$ は地点kに固有な未知パラメータベクトルである。いま、時点tに至るまでの状態変数 $\mathbf{s}_t^k$ と外生変数ベクトル $\mathbf{v}_t^k = (v_t^k, v_{t-1}^k, \dots)$ を与件とする。このとき、時点 $t + 1$ において状態 $s_{t+1}^k = n$ が生起する条件付き確率を

$$P(s_{t+1}^k = n | \mathbf{s}_t^k, \mathbf{v}_t^k) = \frac{\exp(\xi_{k,t}^{x,n} + \xi_{k,t}^{y,n} + \eta_{k,t}^n)}{\sum_{n=1}^S \exp(\xi_{k,t}^{x,n} + \xi_{k,t}^{y,n} + \eta_{k,t}^n)} \quad (2)$$

と特定化する。ただし、

$$\xi_{k,t}^{x,n} = \beta_k^{x,n} (1 - B)^{d^x} v_t^k \quad (3-a)$$

$$\xi_{k,t}^{y,n} = \beta_k^{y,n} (1 - B)^{d^y} v_{t-\Delta_y}^k \quad (3-b)$$

$$\eta_{k,t}^n = \beta_k^{z,n} (1 - B)^{d^z} s_t^k \quad (3-c)$$

である。ここで、 Δ_y は国外火災発生時点から国内で大気汚染物質濃度が上昇し始めるまでのラグである。ただし、 $\beta_k^{a,n}$ ($a = x, y, z$)は未知パラメータ、 d^a ($a = x, y, z$)はそれぞれk地点における国内外の

火災地点数、深刻度の差分操作を支配する実差分作用素パラメータである。また、 B は $Bs_t^k = s_{t-1}^k$ を満たすラグ作用素である。このとき、実数差分作用素 $(1-B)^d$ は2項級数展開により、

$$(1-B)^d = \sum_{j=0}^{\infty} \binom{d}{j} (-B)^j \quad (4)$$

と表せる。ここで、 d は任意の実数、 j は非負整数である。この実数差分作用素を利用することにより、長期の過去の値の影響まで考慮することが可能となる。なお、実数差分作用素のパラメータの値によって過去の値の影響の強さは係数が幾何級数的に小さくなることによって重み付けされる。これにより、長期過去の影響を無視することなく状態推移確率を表現することが可能である。

3 初期値問題

以上の議論では、観測時点 t に対して、過去の状態変数、外生変数の実現値ベクトル $\bar{s}_t^k, \bar{v}_t^k$ に関する情報が入手可能であるという仮定の下で議論した。しかし、これらのベクトルの中で、時点 $t=0$ より過去の時点における実現値 $\bar{s}_0^k, \bar{v}_0^k$ に関しては入手可能ではない。MCMC 手法を用いて、パラメータベクトル ψ を推計する場合、このように入手可能でない情報を潜在変数を用いて表現することを試みる。このことを説明するために、時点 $t=0$ において式 (2) に含まれる変数 $\xi_t^{x,k}, \xi_t^{y,k}, \eta_t^k$ を具体的に展開する。すなわち、式 (3-a)-(3-c) より

$$\xi_0^{x,k} = \beta_k^{x,n} \left\{ \sum_{n=0}^{\infty} a_{\tau}^x \bar{v}_{-n}^{x,k} \right\} \quad (5-a)$$

$$\xi_0^{y,k} = \beta_k^{y,n} \left\{ \sum_{n=0}^{\infty} a_{\tau}^y \bar{v}_{-n-\Delta_y}^{y,k} \right\} \quad (5-b)$$

$$\eta_0^k = \beta_k^{z,n} \left\{ \sum_{n=0}^{\infty} b_{\tau} s_{-n}^k \right\} \quad (5-c)$$

を得る。ここに、 $a_{\tau}^x, a_{\tau}^y, b_{\tau}$ は、それぞれ式 (4) で表される実数差分作用素の2項級数展開により得られる後退作用素 B に係る係数であり、たとえば a_{τ}^x について

$$a_{\tau}^x = \frac{1}{n!} \prod_{j=0}^{n-1} (j - d^x) \quad (6)$$

と表される。同様に a_{τ}^y, b_{τ} についても求めることができる。ここで、各観測地点における観測開始時点以前に発生した国内、国外の火災と深刻度の過去の推移による長期記憶の効果をそれぞれ $\Omega_k^x, \Omega_k^y, \Theta_k$

とすると、時点 $t=0$ での状態推移確率における $\xi_0^{x,k}, \xi_0^{y,k}, \eta_0^k$ とそれぞれ対応する。時間の経過により過去時点の観測値には $a_{\tau}^x, a_{\tau}^y, b_{\tau}$ によって比重が加わることから、 t 期における $\xi_t^{x,k}, \xi_t^{y,k}, \eta_t^k$ は以下のように表される。

$$\xi_t^{x,k} = \beta_k^{x,n} \left\{ \sum_{\tau=0}^{t-1} a_{\tau}^x \bar{v}_{t-n}^{x,k} \right\} + a_t \Omega_k^x \quad (7)$$

$$\xi_t^{y,k} = \beta_k^{y,n} \left\{ \sum_{\tau=0}^{t-1} a_{\tau}^y \bar{v}_{t-n-\Delta_y}^{y,k} \right\} + a_t \Omega_k^y \quad (8)$$

$$\eta_t^k = \beta_k^{z,n} \left\{ \sum_{\tau=0}^{t-1} b_{\tau} \bar{s}_{t-n}^k \right\} + b_t \Theta_k \quad (9)$$

観測を開始する初期時点では、過去の観測データが得られないことによる過去の総合的な観測値を $\Omega_k^x, \Omega_k^y, \Theta_k$ によって補完する。観測データが蓄積するにしたがって、 $\Omega_k^x, \Omega_k^y, \Theta_k$ が状態推移確率に対して持つ影響力は小さくなる。

4 尤度関数

以上で定義された状態推移確率を用いてすべての観測地点 K の観測時刻 $t \in [0, T]$ の観測値が得られる尤度を定式化すると $\Omega_k^x, \Omega_k^y, \Theta_k$ を与件とした深刻度の初期値分布 f を用いて以下の式が得られる。

$$\mathcal{L}(\bar{s}, \phi) = \prod_{k=1}^K \prod_{r=0}^{T-1} \pi_{\bar{s}_r^k, \bar{s}_{r+1}^k} (v_r^k, \beta) \cdot f(s_0^k = \bar{s}_0^k | \Omega_k^x, \Omega_k^y, \Theta_k) \quad (10)$$

尤度を最大化するようなパラメータを推計することにより、地域ごとのヘイズの傾向をつかみ、将来のヘイズを予測することが可能となる。

5 おわりに

本研究では観測可能な大気汚染物質濃度データ、国内外のホットスポットデータを用いて将来時刻におけるヘイズを予測するためのモデルの構築とそれによる尤度関数の定式化を行った。状態変数モデルは既往研究にみられるようなモデルと異なり、長期過去の観測値の影響を考慮することができるだけでなく、状態推移によってヘイズ期の汚染物質濃度の急激な上昇を再現することができる。また、既往研究にみられるモデルと比較しても構造がより簡単であるため、計算負荷を小さく抑えることが期待できる。しかしながら、予測の精度についてはまだ評価できていないため、実際のデータを用いて推計することにより確かめる必要がある。