

確率的変数選択による統計的劣化予測モデルの推定時間短縮効果

大阪大学大学院工学研究科 正会員 ○水谷大二郎 正会員 貝戸清之

1. はじめに

社会基盤施設のアセットマネジメントにおいて、点検データを用いた統計的劣化予測モデルの発展が著しい。それらのモデルの実装方法として、PMS や BMS といったスタンドアローンシステム内での実装が考え得る。その際、長時間のモデル推定はシステムを利用する実務者への負担を増大させると考えられる。一方、近年の統計的劣化予測モデルの発展に伴い、より複雑な劣化過程が推定可能であるが推定に時間を要するようなモデルも提案されている。特に、マルコフ連鎖モンテカルロ (MCMC) 法を用いて推定を行う場合、繰り返し計算が必要となる。これらの先端的なモデルの実装のためには、モデル推定時間の短縮が望ましい。しかしながら、統計的劣化予測モデルの推定時間に関する研究は、多少の事例²⁾はあるものの過去には行われていない。

本研究では、統計的劣化予測モデルの特性変数の選定プロセスにおける計算負荷に着目する。特性変数の候補として N 種類の変数を考慮する場合、考えられる特性変数の組み合わせ数は 2^N 通りとなり、その数は指数関数的に増加する。従来、 2^N 通り全てのモデルを推定し、最適な特性変数の組み合わせを決定していた (以下、総当たり法)。ここでは、特性変数の選定に対し、Stochastic search variable selection (SSVS)³⁾を適用し、その統計的劣化予測モデル推定における有用性を検証する。

2. Stochastic search variable selection

SSVS 法は George & McCulloch³⁾により提案されて以降、適用事例が蓄積されている^{4),5)}。SSVS 法では MCMC 法においてパラメータの事前分布を以下のように設定する。

$$\beta_{i,w} \sim (1-r_{i,w})N(0, \sigma_0^2) + r_{i,w}N(0, \sigma_1^2) \quad (11)$$

ここに、 $N(0, \sigma^2)$ は正規分布、 $\sigma_0^2 \ll \sigma_1^2$ とする。マルコフ劣化ハザードモデルを MCMC 法で推定する¹⁾ことを考え、添え字 i, w は健全度 i ($i=1, \dots, J$) の w ($w=1, \dots, W$) 番目の特性変数に対応するパラメータであ

表-1 サンプル諸元

サンプルサイズ		2,364				
点検年		2004, 2009				
		事後健全度				
		1	2	3	4	5
事前健全度	1	103	228	22	4	1
	2	-	1,329	446	37	5
	3	-	-	137	46	3
	4	-	-	-	2	1
	5	-	-	-	-	0

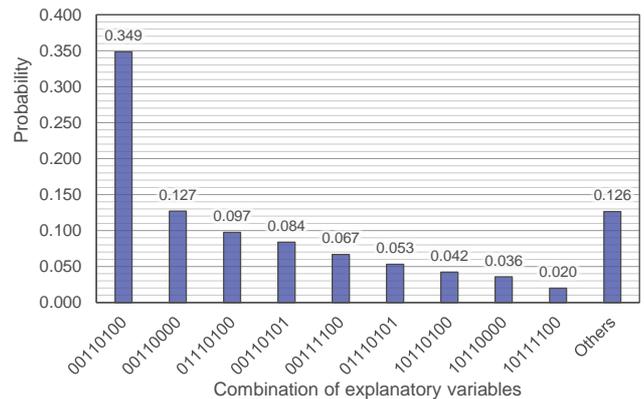


図-1 R のサンプリング結果

ることを表す。 $r_{i,w}$ は 0 か 1 を取る 2 値変数であり、パラメータ事後分布と同時に MCMC 法でサンプリングされる。 σ_0^2 を 0 付近の値に設定することにより、 $r_{i,w} = 0$ のとき $\beta_{i,w}$ も 0 付近の値になり、 $\beta_{i,w}$ を特性変数としてモデルに含むか否かが MCMC 法の中で確率的にサンプリングされる。

3. 実データを対象とした分析

表-1 に内訳を示すような、スイスのコミュニティ道路で獲得された実データに対して、マルコフ劣化ハザードモデルを MCMC 法で推定し、特性変数の組み合わせを SSVS 法、総当たり法のそれぞれで選定した。 σ_0^2 を 0.07、 σ_1^2 を 5 とした。SSVS 法を用いた MCMC 法において、50,000 回のパラメータセットのサンプリングを行った際の $\mathbf{R}=(r_{1,1}, r_{1,2}, r_{2,1}, \dots, r_{4,2})$ のサンプリング結果を集計して図-1 に示す。 $\mathbf{R}=(0,0,1,1,0,1,0,0)$ が 50,000 回中 17,077 回と最も多くサンプリングされ、ここでは、当該 17,077 サンプルを用いて算出した表-2 の統計量を推定結果として用いる。総当たり法では、 $2^{2 \times 4} = 256$ 通り全ての組み合わせを独立に MCMC 法を用いて推定し、i) 全てのパラメータの Geweke 検定統

キーワード：アセットマネジメント 統計的劣化予測モデル モデル推定時間 階層ベイズ推定 SSVS

連絡先 〒565-0871 吹田市山田丘 2-1 大阪大学大学院工学研究科 TEL: 06-6879-7622

表-2 SSVS 法による推定結果

i	統計量	$\beta_{i,0}$	$\beta_{i,1}$	$\beta_{i,2}$
1	期待値	-1.40	2.00×10^{-4}	2.00×10^{-4}
	90%信用区間	[-1.51,-1.29]	[-0.008,0.008]	[-0.008,0.008]
	Geweke 統計量	0.01	0.03	-0.03
2	期待値	-3.72	1.30	1.23
	90%信用区間	[-3.98,-3.43]	[0.69,1.85]	[0.75,1.64]
	Geweke 統計量	0.07	-0.10	0.02
3	期待値	-3.54	2.44×10^{-5}	1.25
	90%信用区間	[-3.95,-3.17]	[-0.008,0.008]	[0.40,2.12]
	Geweke 統計量	-0.09	-0.01	0.09
4	期待値	-3.08	2.00×10^{-5}	2.00×10^{-5}
	90%信用区間	[-3.63,-2.55]	[-0.008,0.008]	[-0.008,0.008]
	Geweke 統計量	-0.02	-0.02	-0.03

表-3 総当たり法による推定結果

i	統計量	$\beta_{i,0}$	$\beta_{i,1}$	$\beta_{i,2}$
1	期待値	-1.40	0	0
	90%信用区間	[-1.51,-1.29]	-	-
	Geweke 統計量	-0.02	-	-
2	期待値	-3.72	1.30	1.23
	90%信用区間	[-4.01,-3.45]	[0.74,1.87]	[0.82,1.66]
	Geweke 統計量	0.14	-0.16	-0.11
3	期待値	-3.61	0	1.43
	90%信用区間	[-4.03,-3.23]	-	[0.60,2.40]
	Geweke 統計量	0.15	-	-0.16
4	期待値	-3.07	0	0
	90%信用区間	[-3.65,-2.54]	-	-
	Geweke 統計量	0.03	-	-

計量 θ_i の絶対値が 1.96 以下 (有意水準 5%) であり,
 ii) 全てのパラメータの 90%信用区間が 0 を含まず,
 iii) 最小の WAIC⁷⁾ を取る特性変数の組み合わせを選定した. なお, 総当たり法での MCMC 法のループ回数はそれぞれ 5,000 回 (burn-in : 1,000 回) とした.

これらの結果から, SSVS 法による特性変数の選定結果 $R=(0,0,1,1,0,1,0,0)$ は, 総当たり法での選定結果と整合的であることが分かる. さらに, $r_{i,w}=1$ となる $\beta_{i,w}$ と $\beta_{i,0}$ の推定結果の SSVS 法と総当たり法での差異を TOST (two one-sided test)⁸⁾ により評価する. 本研究では, TOST において, SSVS 法と総当たり法でのパラメータ推定値の差 $\beta_{i,w}^{SSVS} - \beta_{i,w}^{Enum}$ を用いて, 以下の帰無仮説, 対立仮説を設定する.

$$H_0 : \beta_{i,w}^{SSVS} - \beta_{i,w}^{Enum} \leq -\theta_1 \text{ or } \theta_1 \leq \beta_{i,w}^{SSVS} - \beta_{i,w}^{Enum} \quad (2)$$

$$H_1 : -\theta_1 < \beta_{i,w}^{SSVS} - \beta_{i,w}^{Enum} < \theta_1 \quad (3)$$

$[-\theta_1, \theta_1]$ は Evidence interval と呼ばれ⁸⁾, SSVS 法と総当たり法でのパラメータ推定値の差を, 表-4 に示すように, 帰無仮説(2)を棄却できる最小の θ_1 を用いて評価する. 例えば, $\beta_{1,0}$ において $\beta_{1,w}^{SSVS} - \beta_{1,w}^{Enum}$ が [-0.003,0.003] から外れる確率は 5% 以下であると解釈できる. $\beta_{3,2}$ 以外では, SSVS 法と総当たり法でのパラメータ推定値は極めて近い値を取っていると判断できる. $\beta_{3,2}$ に関しては, 点検サンプルの更なる蓄積により, θ_1 の最小値が減少していくことが期待される.

SSVS 法によるモデル推定時間の減少効果を表-5 を用いて議論する. 全ての計算コードは MATLAB

表-4 TOST による評価(有意水準 5%)

パラメータ	θ_1 の最小値	パラメータ	θ_1 の最小値
$\beta_{1,0}$	0.003	$\beta_{3,0}$	0.080
$\beta_{2,0}$	0.007	$\beta_{3,2}$	0.195
$\beta_{2,1}$	0.008	$\beta_{4,0}$	0.026
$\beta_{2,2}$	0.013	-	-

表-5 モデル推定時間の比較

	平均計算時間 (秒/ループ)	MCMC 法 ループ総数	獲得できた サンプル数
SSVS	0.8854	50,000	17,077
総当たり	0.4710	$2^{8*5,000}$ =1,280,000	4,000

R2015b で書かれ, 標準的なスペックのノート PC (OS: Windows 8.1 Pro (64-bit), Processor: Intel(R) Core(TM) i7-4600U CPU @ 2.10GHz 2.69 GHz, Memory (RAM): 8.00GB) を用いて計算を行った. SSVS 法を用いることによる, 1 ループあたりの計算時間の増加を考慮しても, SSVS 法を用いた方が, 短い時間で, 最適な特性変数の組み合わせを持つモデルのパラメータの事後分布を十分な数獲得できることが分かる.

4. おわりに

本研究では, SSVS 法の統計的劣化予測モデル推定への適用可能性を議論した. 具体的には, SSVS 法を用いた場合においても, 総当たり法を用いたモデル推定結果と極めて近いパラメータ推定値を短い計算時間で獲得できることを示した.

【参考文献】

- 1) 貝戸清之, 小林潔司: マルコフ劣化ハザードモデルのベイズ推定, 土木学会論文集 A, Vol.63, No2, pp.336-355, 2007.
- 2) 水谷大二郎, 坂口創, 貝戸清之, 小林潔司: 準モンテカルロ法を用いた多段階ワイブル劣化ハザードモデルのベイズ推定, 土木学会論文集 F4 (建設マネジメント) 特集号, Vol.71, No.4, pp.I_23-I_34, 2015.
- 3) George, E.I., McCulloch, R.E.: Variable Selection via Gibbs Sampling. *Journal of the American Statistical Association*, No.88, Vol.423, pp.881-889, 1993.
- 4) Ntzoufras, I., Forster, J.J., Dellaportas, P.: Stochastic search variable selection for log-linear models. *Journal of Statistical Computation and Simulation*, No.68, Vol.1, pp.23-37, 2000.
- 5) 数実浩佑, 松岡弘大, 貝戸清之, 徳永宗正, 曾我部正道: SSVS-VAR モデルを用いた列車走行時の鉄道橋梁振動性状分析法, 土木学会論文集 A2 (応用力学) 特集号, Vol.69, No.2, pp.I_719-I_729, 2013.
- 6) Geweke, J.: *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, In *Bayesian Statistics 4*. pp. 169-193, 1992.
- 7) Watanabe, S.: Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory, *Journal of Machine Learning Research*, No.11, pp.3571-3594, 2010.
- 8) Schuirmann, D.J.: A comparison of the Two One-Sided Tests Procedure and the Power Approach for assessing the equivalence of average bioavailability, *Journal of Pharmacokinetics and Biopharmaceutics*, No.15, Vol.6, pp.657-680, 1987.