

# 高頻度金融データに対する最尤型・ベイズ型推定法

荻原 哲平 数理・推論研究系 助教

## 【高頻度金融データの可能性と問題点】

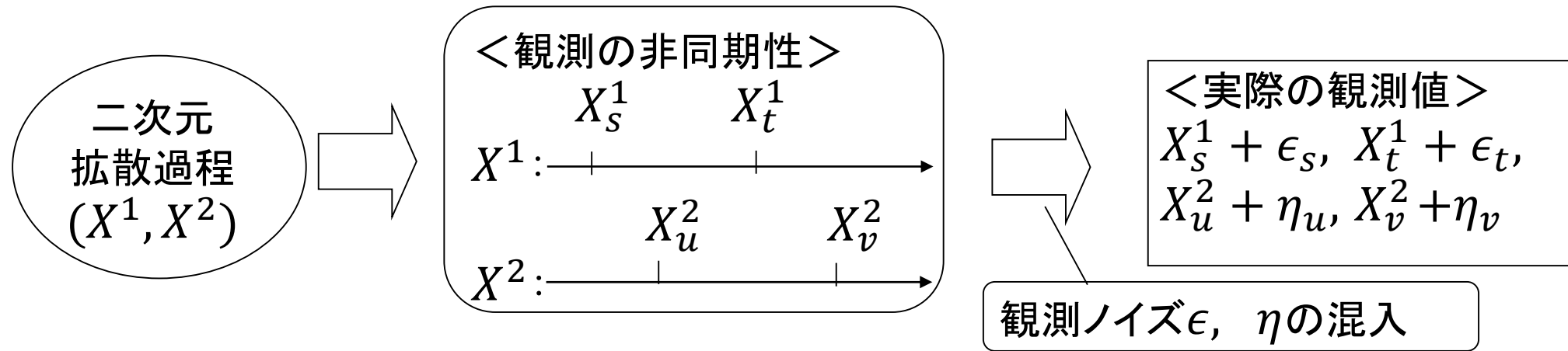
近年株式の全取引における取引価格や取引高の情報を記録した高頻度金融データが整備されており、その膨大な情報量から証券市場の解析の発展が期待される。しかし、そのデータの膨大さに加え、特有の複雑な構造を持つため統計解析は容易ではない。ここでは高頻度金融データの重要な問題である以下の二点を取り上げる。

### 1. マーケット・マイクロストラクチャー・ノイズ

高頻度金融データにおける観測ノイズの混入が実証研究より示唆されている

### 2. 非同期観測

異なる株式に対する観測時刻が異なる → 従来の統計解析手法の適用が困難



## 【高頻度金融データのモデリング】

二次元の潜在株価過程  $X_t = (X_t^1, X_t^2)$  は確率微分方程式

$$dX_t = \mu(t, X_t, \sigma_*)dt + b(t, X_t, \sigma_*)dW_t, \quad t \in [0, T]$$

を満たすとする。ただし、 $W_t$  は標準ブラウン運動、 $\mu, b$  は既知の関数、 $\sigma_*$  は  $d$  次元未知パラメータである。 $X_t^1, X_t^2$  の観測時刻はそれぞれ  $s_0^1, s_1^1, \dots, s_l^1$  と  $s_0^2, s_1^2, \dots, s_m^2$  で与えられるとする（非同期観測）。さらに平均0の独立同分布の観測ノイズ  $(\epsilon_i)_{0 \leq i \leq l}, (\eta_j)_{0 \leq j \leq m}$  に対して、観測データは

$$Y_i^1 = X_{s_i^1}^1 + \epsilon_i \quad (0 \leq i \leq l), \quad Y_j^2 = X_{s_j^2}^2 + \eta_j \quad (0 \leq j \leq m)$$

で与えられるとする。観測データから未知パラメータ  $\sigma_*$  を推定することで株価過程のボラティリティや共変動などのリスク量を計算することができる。

## 【近似尤度関数の構築】

観測区間  $[0, T]$  を等間隔の分割  $[u_0, u_1], [u_1, u_2], \dots, [u_{p-1}, u_p]$  に分けて、それぞれの部分区間で近似尤度関数を構築する。それぞれの区間  $[u_{k-1}, u_k]$  において潜在株価過程の増分  $X_t^j - X_s^j$  は、 $X_t^j - X_s^j \approx \mu_k^j(\sigma_*)(t - s) + b_k^j(\sigma_*)(W_t - W_s)$  と近似される。ただし、 $\mu_k^j(\sigma_*), b_k^j(\sigma_*)$  は区間  $[u_{k-1}, u_k]$  における  $\mu^j(t, X_t, \sigma_*), b^j(t, X_t, \sigma_*)$  の近似量である。観測ノイズを正規分布とみなすことにより、観測の増分

$$Z_k = \left( (Y_i^1 - Y_{i-1}^1)_{i: s_i^1 \in [u_{k-1}, u_k]}, (Y_j^2 - Y_{j-1}^2)_{j: s_j^2 \in [u_{k-1}, u_k]} \right)^T$$

は局所的に多変量正規分布で近似される。

$$S_k(\sigma, v_1, v_2) = \begin{pmatrix} L_{1,k}(\sigma) + v_1 M_{1,k} & L_{1,2,k}(\sigma) \\ L_{1,2,k}(\sigma) & L_{2,k}(\sigma) + v_2 M_{2,k} \end{pmatrix},$$

$$(M_{l,k})_{ij} = 2\delta_{ij} - 1\delta_{|i-j|=1}, \quad (L_{l,k}(\sigma))_{ij} = |b_k^l(\sigma)|^2 \delta_{ij} (s_i^l - s_{i-1}^l),$$

$$(L_{1,2,k}(\sigma))_{ij} = b_k^1(\sigma) \cdot b_k^2(\sigma) (s_i^1 \wedge s_j^2 - s_{i-1}^1 \vee s_{j-1}^2)_+$$

と定め、近似対数尤度関数  $H_n(\sigma, v_1, v_2)$  を

$$H_n(\sigma, v_1, v_2) = -\frac{1}{2} \sum_{k=2}^p \{ Z_k^T S_k^{-1}(\sigma, v_1, v_2) Z_k + \log \det S_k(\sigma, v_1, v_2) \}$$

と定める。 $\epsilon_0, \eta_0$  の分散をそれぞれ  $v_{1,*}, v_{2,*}$  とすると、 $Z_k$  の分散共分散行列は  $S_k(\sigma, v_{1,*}, v_{2,*})$  で近似される。

## 【最尤型・ベイズ型推定量の構築】

個々の観測の変動において観測ノイズの影響が支配的であるため、観測ノイズの分散  $v_{1,*}, v_{2,*}$  の推定量  $\hat{v}_{1,n}, \hat{v}_{2,n}$  の構築は容易である。例えば  $\hat{v}_{1,n} = \sum_i (Y_i^1 - Y_{i-1}^1)^2$  とすればよい。この時未知パラメータ  $\sigma_*$  の最尤型推定量  $\hat{\sigma}_n$  は

$$\hat{\sigma}_n = \operatorname{argmax}_{\sigma} H_n(\sigma, \hat{v}_{1,n}, \hat{v}_{2,n})$$

と定義される。また、事前分布  $\pi(\sigma)$  に対するベイズ型推定量  $\tilde{\sigma}_n$  は以下で定義される：

$$\tilde{\sigma}_n = \int \sigma \exp(H_n(\sigma, \hat{v}_{1,n}, \hat{v}_{2,n})) \pi(d\sigma) / \left( \int \exp(H_n(\sigma, \hat{v}_{1,n}, \hat{v}_{2,n})) \pi(d\sigma) \right)$$

## 【推定量の漸近的性質】

観測数のオーダーを  $b_n$  と書くと、提案推定量  $\hat{\sigma}_n, \tilde{\sigma}_n$  は一定の条件の下  $n \rightarrow \infty$  で以下のような性質を満たす：

$$b_n^{\frac{1}{4}}(\hat{\sigma}_n - \sigma_*) \rightarrow^d \Gamma^{-1/2} \zeta, \quad b_n^{\frac{1}{4}}(\tilde{\sigma}_n - \sigma_*) \rightarrow^d \Gamma^{-1/2} \zeta.$$

ただし、 $\Gamma = P - \lim_{n \rightarrow \infty} (-b_n^{-1/2} \partial_{\sigma}^2 H_n(\sigma_*, v_{1,*}, v_{2,*}))$ ,  $\zeta \sim N(0, I_d) \perp \Gamma$ .

また、上式において  $\Gamma^{-1}$  が推定誤差の漸近分散に対応するが、この漸近分散は最小であることも示される。つまり、**最尤型・ベイズ型推定量の推定誤差の漸近分散は任意の推定量の中で最良である。**

## 【シミュレーション分析】

シンプルなモデル

$$dX_t^1 = \sigma_{1,*} dW_t^1, \quad dX_t^2 = \sigma_{3,*} dW_t^1 + \sigma_{2,*} dW_t^2$$

に対して最尤型推定量のパフォーマンスをシミュレーションする。観測時刻  $s_i^j$  はパラメータ  $\lambda_j$  のポアソン過程より生成されるとし、観測ノイズは  $(\epsilon_i) \sim i.i.d. N(0, v_{1,*}), (\eta_i) \sim i.i.d. N(0, v_{2,*})$  とする。パラメータの値を  $T = 1, \lambda_1 = \lambda_2 = 1, \sigma_1 = 1, \sigma_2 = \sqrt{1 - 0.5^2}, \sigma_3 = 0.5$  とした時の最尤型推定量  $(\hat{\sigma}_{1,n}, \hat{\sigma}_{2,n}, \hat{\sigma}_{3,n})$  のパフォーマンスが以下ようになった。

	n	$\hat{\sigma}_{1,n}$	$\hat{\sigma}_{2,n}$	$\hat{\sigma}_{3,n}$
true values		1	0.866	0.5
$v_{1,*} = 0.001$	1000	0.972	0.849	0.493
		(0.047)	(0.047)	(0.061)
$v_{2,*} = 0.001$	5000	1.000	0.859	0.499
		(0.029)	(0.032)	(0.044)
$v_{1,*} = 0.005$	1000	0.990	0.865	0.488
		(0.069)	(0.066)	(0.095)
$v_{2,*} = 0.005$	5000	0.992	0.863	0.493
		(0.049)	(0.038)	(0.060)

また、潜在株価過程  $X_t^1, X_t^2$  の共変動を表す  $(X^1, X^2)_T$  の推定量  $T\hat{\sigma}_{1,n}\hat{\sigma}_{3,n}$  も計算できるため、そのパフォーマンスを既存の推定量と比較する。既存の推定量として、Christensen, Kinnebrock, and Podolskij (2010) の pre-averaged HY(PHY), Modulated Realized Covariance (MRC), Bibinger (2011) の Generalized Multiscale Estimator (GME) を用いた。

	n	MLE	PHY	GME	MRC
$v_1 = 0.001$	500	0.458	0.522	0.511	0.522
		(0.078)	(0.143)	(0.142)	(0.129)
$v_2 = 0.001$	1000	0.480	0.502	0.490	0.502
		(0.070)	(0.133)	(0.129)	(0.118)
	5000	0.500	0.504	0.503	0.506
		(0.049)	(0.088)	(0.077)	(0.079)
$v_1 = 0.005$	500	0.499	0.499	0.497	0.528
		(0.137)	(0.176)	(0.190)	(0.175)
$v_2 = 0.005$	1000	0.485	0.478	0.466	0.483
		(0.108)	(0.163)	(0.137)	(0.122)
	5000	0.490	0.488	0.489	0.492
		(0.072)	(0.103)	(0.095)	(0.086)

## 【日本株式市場における実証分析】

日本株式市場の高頻度金融データ(日経NEEDS)に対し、上記のシンプルなモデルを適用し、2013年4月の日内データのうち、前場(9:00~11:30)データの各銘柄ペア相関を計算した。

銘柄コード	企業名	東証17業種分類	時価総額ランキング	2013年4月の1日当たり取引件数
7201	日産自動車	自動車・輸送機	13	3380
7203	トヨタ自動車	自動車・輸送機	1	5195
7267	本田技研工業	自動車・輸送機	7	3077

