高頻度データ解析に現れる最大値統計量の分布の Gauss 型近似について

小池祐太 [*†‡]

本研究のもともとの動機は, 2 つの Brown 運動の間に時間差を許す相関関係が存在するかどうかを, それらの高頻度離散観測データに基づいて検定する問題にある. 大雑把に述べると, この問題の数学的設定は以下の通りである. 以下の 2 つの連続時間確率過程を区間 $[0,T]$ 上で離散観測する状況を考える:

$$X_t^1 = x_0^1 + \sigma_1 B_t^1, \qquad X_t^2 = x_0^2 + \sigma_2 B_{t-\vartheta}^2, \qquad t \in [0,T]. \tag{1}$$

ここに, $x_0^1, x_0^2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$, $\vartheta \in \mathbb{R}$ で, $B_t = (B_t^1, B_t^2)$ $(t \in \mathbb{R})$ は相関 $\rho \in (-1,1)$ をもつ両側 Brown 運動である. 各 $\nu = 1,2$ について, 確率過程 X^ν は時点 $0 \le t_0^\nu < t_1^\nu < \cdots < t_{n_\nu}^\nu \le T$ で観測されていると仮定する. 従って観測時刻は一般には非同期である. 我々の目的は, 観測データ $(X_{t_i^1}^1)_{i=0}^{n_1}$ および $(X_{t_j^2}^2)_{j=0}^{n_2}$ に基づいて次の統計的仮説検定問題を解くことである:

$$H_0 : \rho = 0 \qquad \text{vs} \qquad H_1 : \rho \neq 0. \tag{2}$$

モデル (1) は高頻度金融データのリード・ラグ効果をモデル化するために (より一般的な形で)Hoffmann *et al.* [12] によって導入されたものである. [12] はリード・ラグパラメーター ϑ を推定する問題を考えた. ϑ を推定するために, [12] は以下のコントラスト関数を導入した:

$$U_n(\theta) = \sum_{i,j} (X_{t_i^1}^1 - X_{t_{i-1}^1}^1)(X_{t_j^2}^2 - X_{t_{j-1}^2}^2) 1_{\{(t_{i-1}^1, t_i^1] \cap (t_{j-1}^2 - \theta, t_j^2 - \theta] \neq \emptyset\}}.$$

$U_n(\theta)$ は Hayashi & Yoshida [11] の方法に基づいて計算した, ラグ θ での X^1 と X^2 のリターン間の (標本) 交差共分散関数と考えられる. [12] は, 適当な正則条件の下, 有限集合 $\mathcal{G}_n$ を適切にとれば, $\underline{\rho \neq 0 \text{ である限り}}$

$$\widehat{\vartheta}_n = \arg\max_{\theta \in \mathcal{G}_n} |U_n(\theta)|$$

が ϑ の一致推定量となることを示した. $\rho = 0$ の場合 ϑ は明らかに識別不能となるため, 条件 $\rho \neq 0$ は必須である. 従って, なんらかの外生的理由で仮定 $\rho \neq 0$ が信用できる場合を除いて, ϑ の推定を実行する前に上の検定問題の帰無仮説を棄却する必要がある. 検定問題 (2) を解くための最も自然なアプローチは, $\max_{\theta \in \mathcal{G}_n} |U_n(\theta)|$ の値が大きすぎるときに帰無仮説を棄却する方法である. この方法を厳密に実行するためには, 帰無仮説の下での $\max_{\theta \in \mathcal{G}_n} |U_n(\theta)|$ の分布を近似する必要がある. 本研究の主要な目的の 1 つはこの問題に解答を与えることである. より一般に, 本研究では

[*] 首都大学東京大学院社会科学研究科経営学専攻
[†] CREST, Japan Science and Technology Agency
[‡] 統計数理研究所

高頻度データ解析に現れる最大値統計量の分布を近似する問題を考える. 最大値統計量は高頻度データ解析の多くの問題において自然に現れる. 例えば, 瞬間ボラティリティに代表される (日内) 時間変動をもつ統計量に対する一様信頼バンドの構成や, (日内の) 複数時点で検定を実行する際の FWER(family-wise error rate) のコントロール (cf. [1, 10]), ボラティリティの変化点問題 (cf. [2]), ジャンプの有無の検定 (cf. [14, 17]) などが挙げられる.

数学的観点から見ると, 本研究は近年発展した 2 つの異なる分野の研究に基礎をおいている. 第 1 の研究は Chernozhukov, Chetverikov & Kato [5, 6, 8, 9] による一連の研究であり, ここでは Chernozhukov-Chetverikov-Kato 理論, もしくは略して CCK 理論と呼ぶことにする. CCK 理論から得られる主要な結論の 1 つは, (高次元) 確率ベクトルの最大値の分布と Gauss 型ベクトルの最大値の分布の間の Kolmogorov 距離に対する上界であり, 我々の問題と明らかな関連をもつ. しかし, CCK 理論ではターゲットの確率ベクトルが独立な確率ベクトルの和の場合 [5, 6, 9] と Gauss 型の場合 [8, 9] を考察しているため, 我々の問題には直接適用できない. 実際, 我々の主要なターゲットである $(U_n(\theta))_{\theta\in\mathcal{G}_n}$ は, 観測時刻の非同期性のため, たとえ帰無仮説の下でも従属性をもつ確率ベクトルの和となってしまう. CCK 理論の従属性をもつ確率ベクトルの和への拡張はいくつか研究があるが (例えば [3, 4, 7, 18, 19]), 非同期性は非常に複雑な, 「非定常な」, 従属関係を生じさせるため, それらの結果を我々の問題に適用することはやはり難しいように見える. この観点からいうと, 本研究の目的は CCK 理論を我々の目的に適合するよう拡張することであるといえる. 実際, 我々の結果は [8] の結果のいくつかの一般化となっており, その適用にはそもそもターゲットの確率ベクトルが確率ベクトルの和の形に書かれている必要はなく, また Kolmogorov 距離の上界は上述の研究と比べて非常に単純な形で得られる.

CCK 理論の証明を読むと, ターゲットのベクトルに対する独立性/Gauss 性の仮定は Stein の方法の適用において決定的であることがわかる.*1 言い換えると, Stein の方法が有効に適用できるようなターゲットに対しては, CCK 理論を自然に拡張することが可能である. この観点は我々をもう一つの重要な理論, Malliavin 解析を我々の問題に適用するというアイディアに導く. 実際, Nourdin & Peccati [16] の仕事から始まる近年の研究は, 「Stein の方法と Malliavin 解析は非常に見事に組み合わさる」(Nourdin [15] の 3 ページ) ことを示している. 本研究ではこの文言が CCK 理論への応用においても引き続き正しいことを示す.

本研究の詳細はワーキングペーパー [13] にまとめてある.

参考文献

[1] Andersen, T. G., Bollerslev, T. & Dobrev, D. (2007). No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications. *J. Econometrics* **138**, 125–180.

[2] Bibinger, M., Jirak, M. & Vetter, M. (2017). Nonparametric change-point analysis of volatility. *Ann. Statist.* **45**, 1542–1578.

[3] Chang, J., Qiu, Y., Yao, Q. & Zou, T. (2017). On the statistical inference for large precision ma-

*1 独立性の仮定は最大モーメント不等式の導出においても役割を演じるが, これは主要な問題とは分離して考えることができる.

trices with dependent data. Working paper. Available at arXiv: https://arxiv.org/abs/1603.06663.

[4] Chen, X. (2017). Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. *Ann. Statist. (forthcoming).* Available at arXiv: https://arxiv.org/abs/1610.00032.

[5] Chernozhukov, V., Chetverikov, D. & Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41**, 2786–2819.

[6] Chernozhukov, V., Chetverikov, D. & Kato, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42**, 1564–1597.

[7] Chernozhukov, V., Chetverikov, D. & Kato, K. (2014). Testing many moment inequalities. Working paper. Available at arXiv: https://arxiv.org/abs/1312.7614v4.

[8] Chernozhukov, V., Chetverikov, D. & Kato, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162**, 47–70.

[9] Chernozhukov, V., Chetverikov, D. & Kato, K. (2016). Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. *Stochastic Process. Appl.* **126**, 3632–3651.

[10] Christensen, K., Oomen, R. & Renò, R. (2016). The drift burst hypothesis. Working paper. Available at SSRN: https://ssrn.com/abstract=2842535.

[11] Hayashi, T. & Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* **11**, 359–379.

[12] Hoffmann, M., Rosenbaum, M. & Yoshida, N. (2013). Estimation of the lead-lag parameter from non-synchronous data. *Bernoulli* **19**, 426–461.

[13] Koike, Y. (2017). Gaussian approximation of maxima of Wiener functionals and its application to high-frequency data. Working paper.

[14] Lee, S. S. & Mykland, P. A. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *The Review of Financial Studies* **21**, 2535–2563.

[15] Nourdin, I. (2013). Lectures on Gaussian approximations with Malliavin calculus. In C. Donati-Martin, A. Lejay & A. Rouault, eds., *Séminaire de probabilitiés XLV*, vol. 2078 of *Lecture Notes in Math.* Springer, pp. 3–89.

[16] Nourdin, I. & Peccati, G. (2009). Stein's method on Wiener chaos. *Probab. Theory Related Fields* **145**, 75–118.

[17] Palmes, C. & Woerner, J. H. C. (2016). A mathematical analysis of the Gumbel test for jumps in stochastic volatility models. *Stoch. Anal. Appl.* **34**, 852–881.

[18] Zhang, D. & Wu, W. B. (2015). Gaussian approximation for high dimensional time series. *Ann. Statist. (forthcoming).* Available at arXiv: https://arxiv.org/abs/1508.07036.

[19] Zhang, X. & Cheng, G. (2017). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli (forthcoming)* .