学生と挑む「誕生日のパラドックス」

茂木快治 [1]

[1] 神戸大学大学院経済学研究科准教授

神戸大学大学院経済学研究科

第 629 回経済学会例会

2024 年 1 月 17 日

報告者の研究プロジェクトの全体像

1. 時系列分析 (MIDAS, グレンジャー因果性, TAR など).
2. ミクロ計量経済学 (欠測データ分析, 因果推論など).
3. その他 (誕生日のパラドックス, J-REIT など).

- 本日は誕生日のパラドックス (birthday paradox; BDP) に関する研究・教育の成果と課題を報告する.
- ここ数年, 複数の学部生と取り組んでいる研究テーマ.
- 主要成果物: Motegi & Woo (2023, CSTM), Motegi & Hayashi (2023, WP).

目次

誕生日のパラドックスとは

- カレンダーサイズを N とする (例: $N = 365$ 日).
- クラスサイズを K とする (例: $K = 35$ 人).
- K 人のうち 1 人は自分であるとする.
- 自分を含め, 全員の誕生日はランダムに決まるとする ($1/N$ ずつの確率).
- 問 **1**: 自分と同じ誕生日の人がクラス内に少なくとも 1 人いる確率 $P_1(N, K)$ を求めなさい.
- 問 **2**: 誕生日の同じペアがクラス内に少なくとも 1 組いる確率 $P_2(N, K)$ を求めなさい.

誕生日のパラドックスとは

- 問 1 の解答: 自分と同じ誕生日の人がクラス内に少なくとも 1 人いる確率は,

$$P_1(N,K) = 1 - \frac{N(N-1)^{K-1}}{N^K}.$$

- 問 2 の解答: 誕生日の同じペアがクラス内に少なくとも 1 組いる確率は,

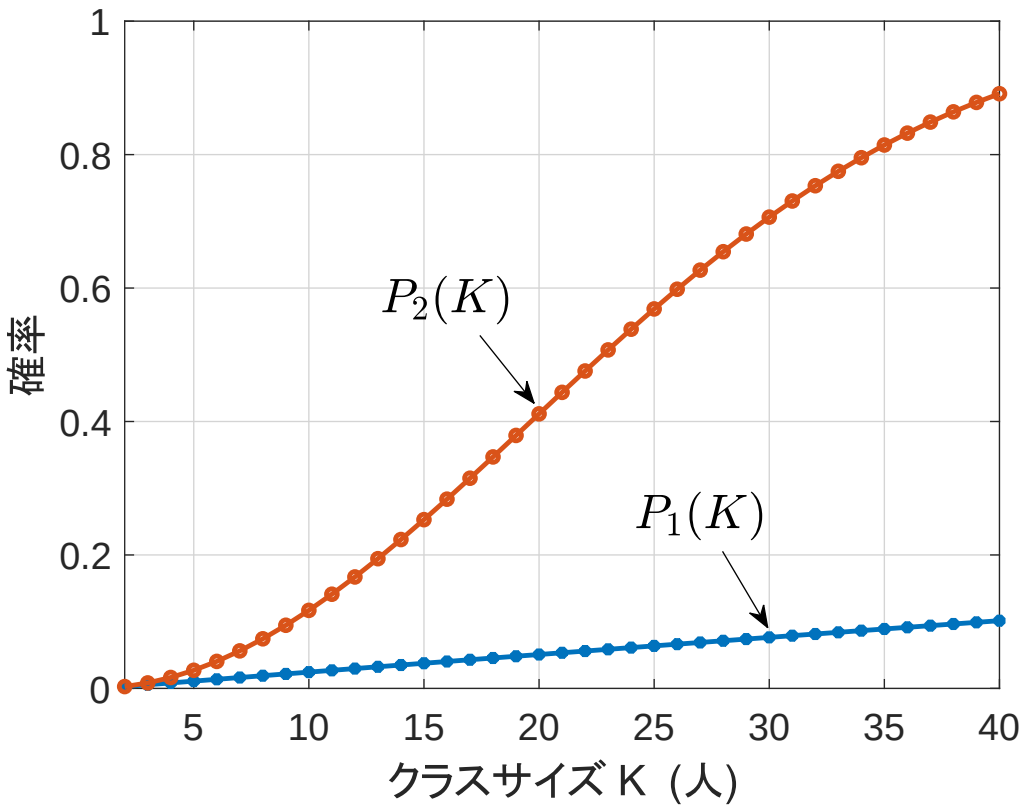
$$P_2(N,K) = 1 - \frac{N(N-1)\cdots(N-K+1)}{N^K} = 1 - \frac{{}_N\mathrm{P}_K}{N^K}.$$

(順列アプローチと呼ぶ)

- $N = 365$ と定め, $P_1(K) \equiv P_1(365,K)$ と $P_2(K) \equiv P_2(365,K)$ を図示してみよう.

誕生日のパラドックスとは

図 1: 問 1 解答と問 2 解答の図示



- クラスサイズ K の増加とともに, $P_1(K)$ は緩やかに上昇し, $P_2(K)$ は急速に上昇する.

誕生日のパラドックスとは

- BDP の本質: 「自分」と「誰か」の誕生日が一致する確率 $P_1(N, K)$ は予想どおり低く，「誰か」と「誰か」の誕生日が一致する確率 $P_2(N, K)$ は意外なほど高い.
- BDP の歴史は少なくとも Mosteller (1962) まで遡る．今日では藪 (2012, コラム 4-5) で紹介されるなど，有名な逆説である.
- BDP の応用分野:
  1. マルコフ連鎖 (Kim, Montenegro, Peres, and Tetali, 2010).
  2. 暗号学 (Bellare and Kohno, 2004; Suzuki, Tonien, Kurosawa, and Toyota, 2006)．誕生日攻撃 (birthday attack).
- BDP のサーベイ論文: DasGupta (2005).

問2再考

- 改めて問 2 の解答を考える:

$$P(N,K) = 1 - Q(N,K), \qquad Q(N,K) = \frac{{}_N\mathrm{P}_K}{N^K}.$$

ただし,

$P(N,K) \equiv$ 少なくとも 1 組のペアの誕生日が揃っている確率,

$Q(N,K) \equiv$ 全員の誕生日が異なっている確率.

- 問 **2′**: 次のアプローチは問 2 の別解として成立するか?

$$Q^{ap}(N,K) \equiv \left(\frac{{}_N\mathrm{P}_2}{N^2}\right)^{{}_K\mathrm{C}_2}.$$

- 「任意のペアが異なる誕生日をもつ確率が ${}_N\mathrm{P}_2/N^2$, ペアの組み合わせが ${}_K\mathrm{C}_2$ 通り, よって $Q(N,K) = Q^{ap}(N,K)$」

べき乗近似

- 一般に, $Q(N,K) \neq Q^{ap}(N,K)$ である!
- $Q^{ap}(N,K)$ は $Q(N,K)$ のべき乗近似 (exponentiation approximation) であり, $Q(N,K)$ の厳密解ではない.
- 誤差が生じる理由: ペアの独立性が成り立たないから
- $K=3$ の例: 個人 1 と個人 2 の誕生日が異なり, かつ個人 1 と個人 3 の誕生日が異なるという条件の下で, 個人 2 と個人 3 の誕生日が異なる確率は ${}_N\mathrm{P}_2/N^2$ よりも小さくなる.

  (∵ 個人 2, 3 ともに個人 1 の誕生日を避ける必要があるため, 個人 2, 3 の間で誕生日が重なる確率が高まる.)

ペアワイズ・アプローチ

- Motegi & Woo (2023) のリサーチ・クエスチョン: 近似解ではなく厳密解を与えるペアワイズ・アプローチは作れるか?
- ペアを標準的な順序で一列に並べる:

	$\mathcal{B}_1$	$\mathcal{B}_2$	$\mathcal{B}_3$	$\mathcal{B}_4$	$\mathcal{B}_5$	$\mathcal{B}_6$
$K=3$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	–	–	–
$K=4$	$\{1,2\}$	$\{1,3\}$	$\{1,4\}$	$\{2,3\}$	$\{2,4\}$	$\{3,4\}$

- ペア $\mathcal{B}$ において誕生日の衝突 (collision) が生じないという事象を $D_{\mathcal{B}}$ と書く.

ペアワイズ・アプローチ

- 条件付きペアワイズ非衝突確率 (conditional pairwise probability of non-collision):

$$Q_i(N,K) \equiv \Pr\left(D_{\mathcal{B}_i} \,\middle|\, \bigcap_{j=1}^{i-1} D_{\mathcal{B}_j} \right), \quad i \in \{1, \ldots, {}_K\mathrm{C}_2\}.$$

- 確率の乗法定理を繰り返し用いると, 次式が厳密に成り立つ:

$$Q(N,K) = \prod_{i=1}^{{}_K\mathrm{C}_2} Q_i(N,K).$$

- べき乗近似は, 誤って (あるいは単純化のため意図的に) $Q_i(N,K) = \Pr(D_{\mathcal{B}_i})$, $\forall i$ と仮定したケースと解釈できる

ペアワイズ・アプローチ

定理 (Theorem 1, Motegi & Woo, 2023)

$$Q_i(N,K) = \frac{N-1-\sum_{\ell=1}^{K-2}\ell \times \mathbf{1}\{i \in \mathcal{M}_\ell(K)\}}{N-\sum_{\ell=1}^{K-2}\ell \times \mathbf{1}\{i \in \mathcal{M}_\ell(K)\}},$$

$$\forall i \in \{1,\ldots,{}_K C_2\},$$

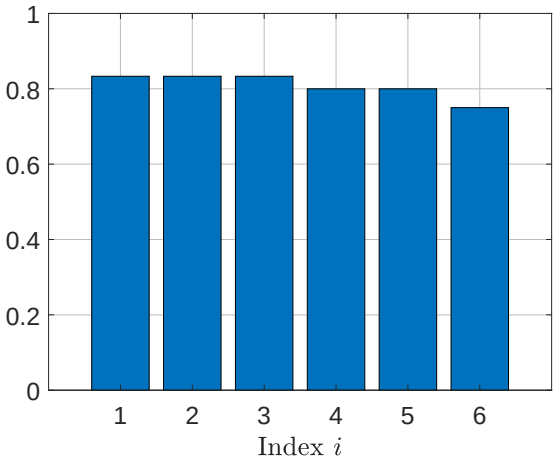
ただし, $\mathcal{M}_1(K) = \{K,\ldots,2K-3\}$,

$$\mathcal{M}_\ell(K) = \left\{K+\sum_{h=1}^{\ell-1}(K-1-h),\ldots,2K-3+\sum_{h=1}^{\ell-1}(K-2-h)\right\},$$

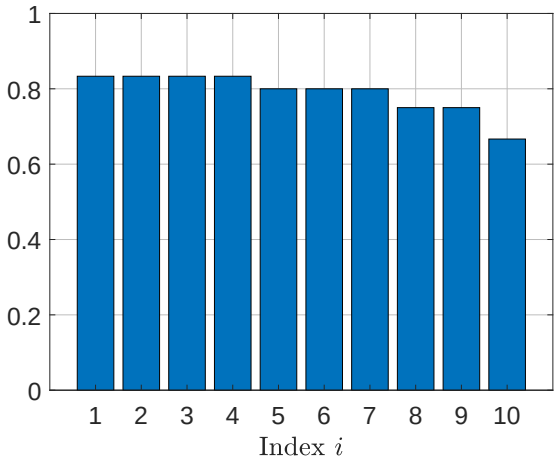
$$\forall \ell \in \{2,\ldots,K-2\}.$$

ペアワイズ・アプローチ: 数値例

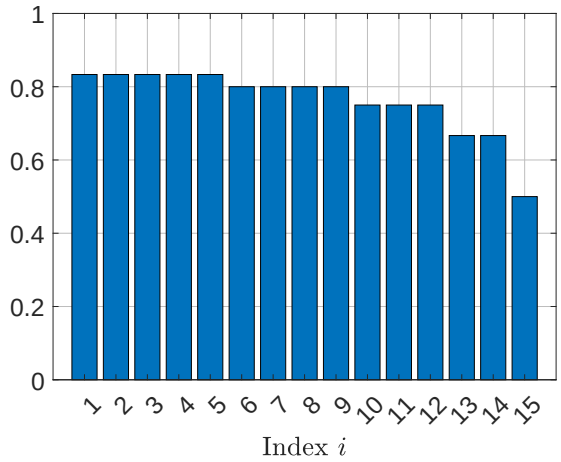
図 2: 条件付きペアワイズ非衝突確率 $Q_i(N, K)$ $(N = 6)$



$K = 4$



$K = 5$

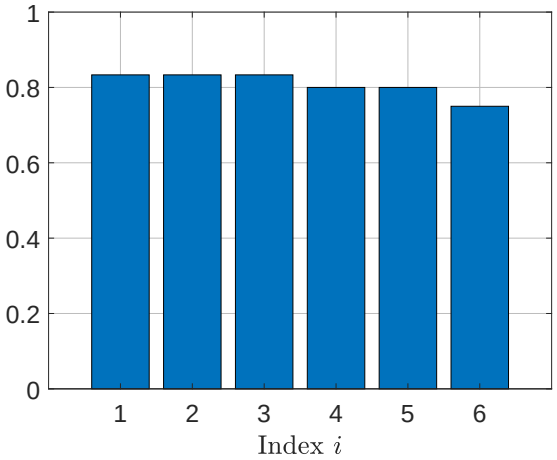


$K = 6$

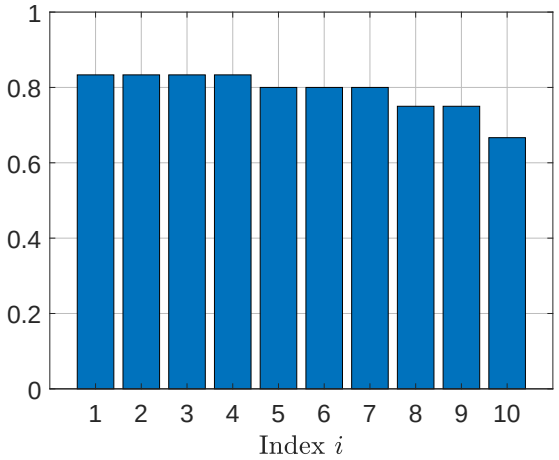
- ペアが進むにつれ, $Q_i(N, K)$ は階段状に減少する.
- ペアの第 1 構成員が入れ替わるときのみ減少が生じる.
- 厳密に単調減少するのではない点に意外性がある.

ペアワイズ・アプローチ: 数値例

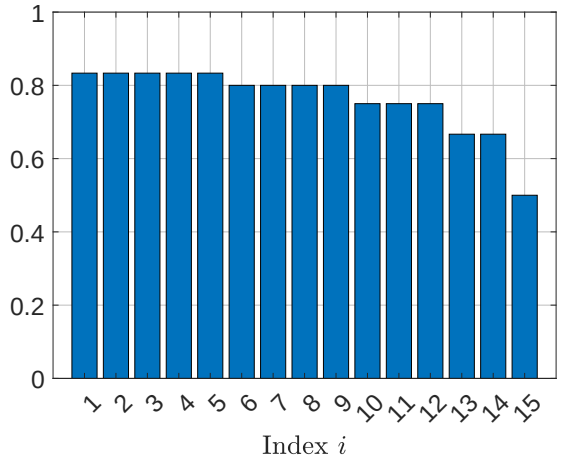
図 2: 条件付きペアワイズ非衝突確率 $Q_i(N,K)$ $(N=6)$



$K=4$



$K=5$



$K=6$

- べき乗近似は, 1 本目の棒が最後までずっと並ぶと仮定しているに等しい. よって, べき乗近似は真の確率を過大推定する. つまり, すべての (N,K) について $Q^{ap}(N,K) > Q(N,K)$.

グループワイズ・アプローチ

- Motegi & Hayashi (2023) のリサーチ・クエスチョン: 2 人 1 組のペアではなく, J 人 1 組のグループで考えたらどうなる?
- $J = 2$ のときペアワイズ・アプローチと合流し, $J = K$ のとき順列アプローチと合流するはず. 一般の J では何が起こる?
- グループを標準的な順序で一列に並べる:

$J = 3$	$\mathcal{B}_1$	$\mathcal{B}_2$	$\mathcal{B}_3$	$\mathcal{B}_4$	$\mathcal{B}_5$	$\cdots$	$\mathcal{B}_{10}$
$K = 4$	1	1	1	2	–	$\cdots$	–
	2	2	3	3	–	$\cdots$	–
	3	4	4	4	–	$\cdots$	–
$K = 5$	1	1	1	1	1	$\cdots$	3
	2	2	2	3	3	$\cdots$	4
	3	4	5	4	5	$\cdots$	5

グループワイズ・アプローチ

- グループ $\mathcal{B}$ において誕生日の衝突が生じないという事象を $D_{\mathcal{B}}$ と書く.
- 条件付きグループワイズ非衝突確率:

$$Q_i(N,K,J) \equiv \Pr\left(D_{\mathcal{B}_i} \,\middle|\, \bigcap_{j=1}^{i-1} D_{\mathcal{B}_j} \right), \quad i \in \{1,\ldots,{}_K\mathsf{C}_J\}.$$

- 確率の乗法定理を繰り返し用いると, 次式が厳密に成り立つ:

$$Q(N,K) = \prod_{i=1}^{{}_K\mathsf{C}_J} Q_i(N,K,J).$$

グループワイズ・アプローチ

定理 (Theorem 1, Motegi & Hayashi, 2023)

$$Q_1(N,K,J) = \frac{{}_N P_J}{N^J},$$

$$Q_i(N,K,J) = \frac{N-J+1}{N}, \qquad \forall i \in \{2,\ldots,K-J+1\},$$

$$Q_i(N,K,J) = \frac{N-J+1-\sum_{\ell=1}^{K-J} \ell \times \mathbf{1}\{i \in \mathcal{M}_\ell(K,J)\}}{N-J+2-\sum_{\ell=1}^{K-J} \ell \times \mathbf{1}\{i \in \mathcal{M}_\ell(K,J)\}},$$

$$\forall i \in \{K-J+2,\ldots,i^*(K-J)\},$$

$$Q_i(N,K,J) = 1, \qquad \forall i \in \{i^*(K-J)+1,\ldots,{}_K C_J\}.$$

グループワイズ・アプローチ

定理 (Theorem 1, Motegi & Hayashi, 2023)

ただし, $i^*(K-J) = {}_{K-J+2}C_2$ であり,

$$\mathcal{M}_1(K,J) = \{K-J+2, \ldots, 2(K-J)+1\},$$

$$\mathcal{M}_\ell(K,J) = \left\{ K-J+2+\sum_{h=1}^{\ell-1}(K-J-h+1),\ \ldots,\right.$$

$$\left. 2(K-J)+1+\sum_{h=1}^{\ell-1}(K-J-h) \right\},$$

$$\ell \in \{2, \ldots, K-J\}.$$

- $J=2$ のときペアワイズ・アプローチと合流し, $J=K$ のとき順列アプローチと合流する.

グループワイズ・アプローチ: $\{Q_i\}$ の形状

定理 (Theorem 2, Motegi & Hayashi, 2023)

$J = 2$ のとき, そしてそのときに限り, 条件付きグループワイズ非衝突確率 $\{Q_i(N, K, J)\}_{i=1}^{K C_J}$ はグループインデックス i に関して単調減少になる.

定理 (Theorem 3, Motegi & Hayashi, 2023)

$J \geq 3$ と仮定する. $K = J + 1$ のとき, そしてそのときに限り, $\{Q_i(N, K, J)\}_{i=1}^{K C_J}$ は i に関して単調増加になる.

グループワイズ・アプローチ: $\{Q_i\}$ の形状

定理 (Theorem 4, Motegi & Hayashi, 2023)

$J \geq 3$ と仮定する. 次の 2 つのステートメントは同値である:

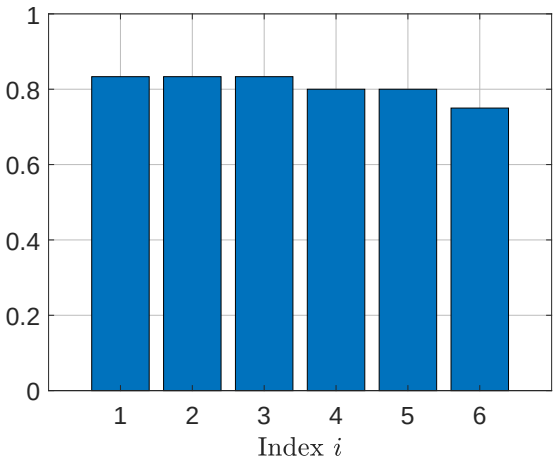
1. $K > J + 1$.
2. $\{Q_i(N, K, J)\}_{i=1}^{i^*(K-J)}$ は $i \in \{1, \ldots, K - J + 2\}$ の範囲で単調増加, $i \in \{K - J + 2, \ldots, i^*(K - J)\}$ の範囲で単調減少となり, $i \in \{K - J + 2, \ldots, 2(K - J) + 1\}$ の範囲で最大値 $Q_i(N, K, J) = (N - J)/(N - J + 1)$ をとる.

グループワイズ・アプローチ: $\{Q_i\}$ の形状

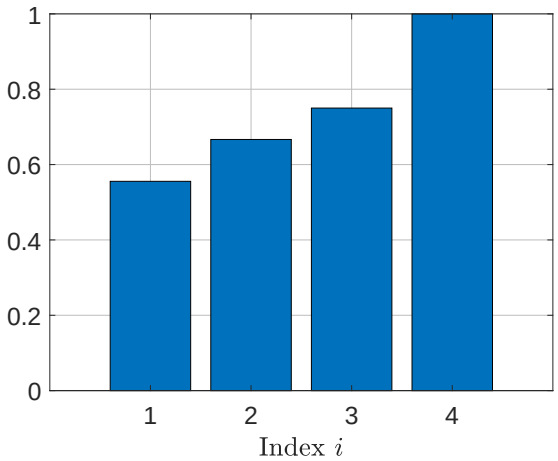
- 直観: グループ構成員に係る非衝突制約と, 非構成員に係る非衝突制約のバランスが, $\{Q_i(N,K,J)\}_{i=1}^{K\mathsf{C}_J}$ の形状を決める.
- 構成員に係る非衝突制約は $Q_i(N,K,J)$ の押し上げ要因, 非構成員に係る非衝突制約は $Q_i(N,K,J)$ の押し下げ要因となる.
- $J=2$ のときは押し下げ要因が支配的となり, $\{Q_i(N,K,J)\}$ は単調減少する (Theorem 2).
- $J\geq 3$ かつ $K=J+1$ のときは押し上げ要因が支配的となり, $\{Q_i(N,K,J)\}$ は単調増加する (Theorem 3).
- $J\geq 3$ かつ $K>J+1$ のときは, $i\in\{1,\ldots,K-J+2\}$ の範囲では押し上げ要因, $i\in\{K-J+2,\ldots,i^*(K-J)\}$ の範囲では押し下げ要因が支配的となり, $\{Q_i(N,K,J)\}_{i=1}^{i^*(K-J)}$ は山型 (hump-shaped) となる (Theorem 4).

グループワイズ・アプローチ: 数値例

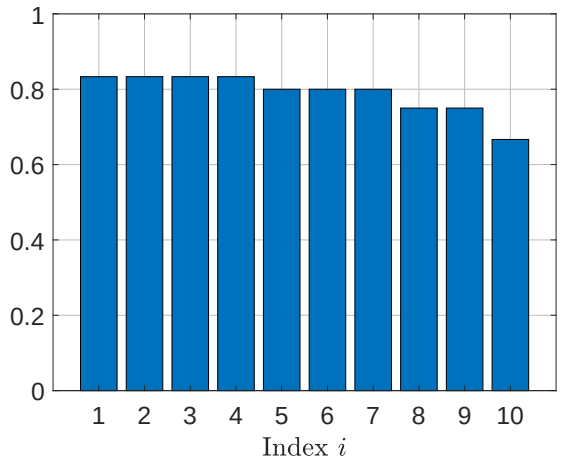
図 3: 条件付きグループワイズ非衝突確率 $Q_i(N, K, J)$ $(N = 6)$



$(K, J) = (4, 2)$



$(K, J) = (4, 3)$

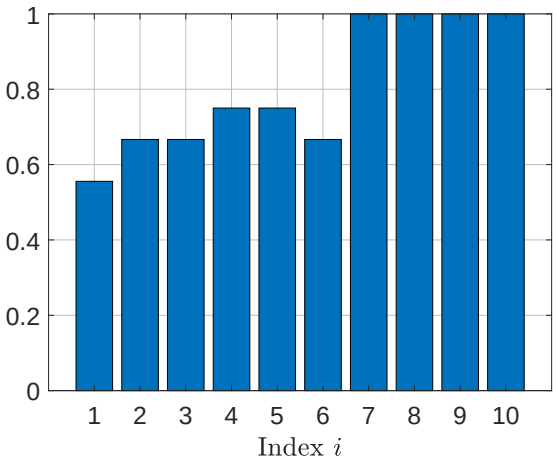


$(K, J) = (5, 2)$

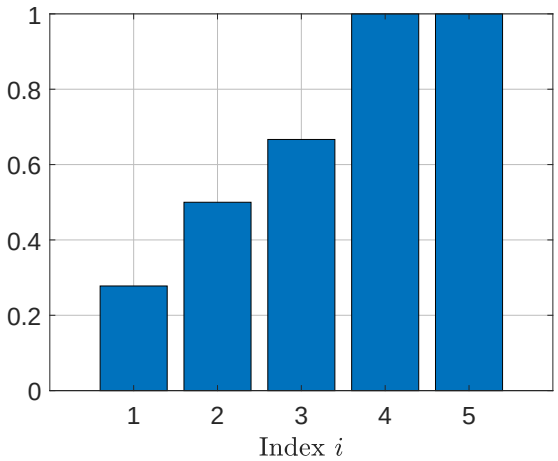
- Theorems 2-3 と整合的な結果が得られている.

グループワイズ・アプローチ: 数値例

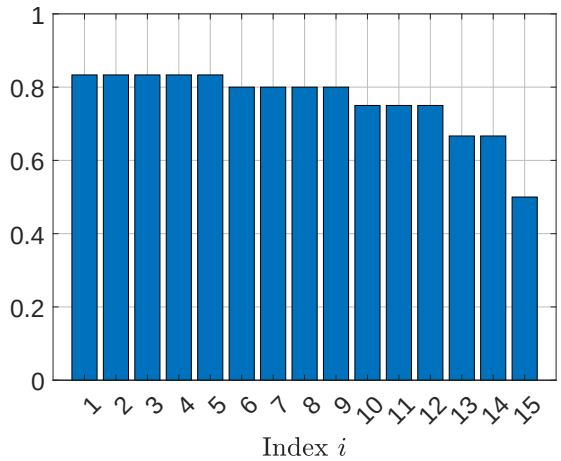
図 4: 条件付きグループワイズ非衝突確率 $Q_i(N, K, J)$ $(N = 6)$



$(K, J) = (5, 3)$



$(K, J) = (5, 4)$

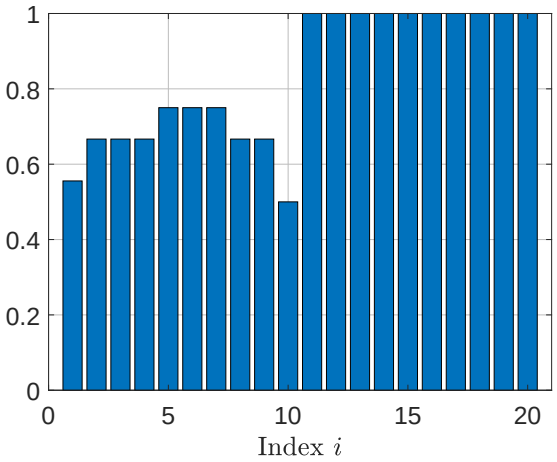


$(K, J) = (6, 2)$

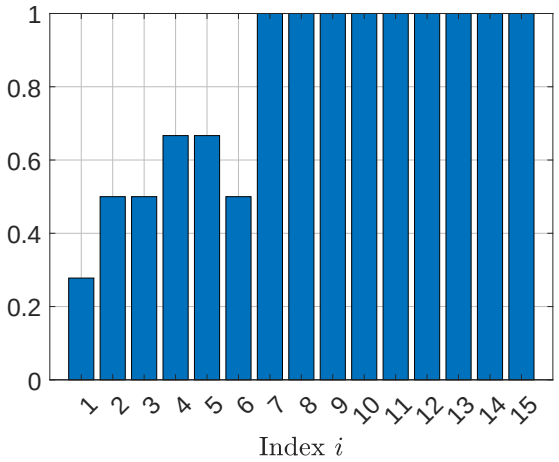
- Theorems 2-4 と整合的な結果が得られている.

グループワイズ・アプローチ: 数値例

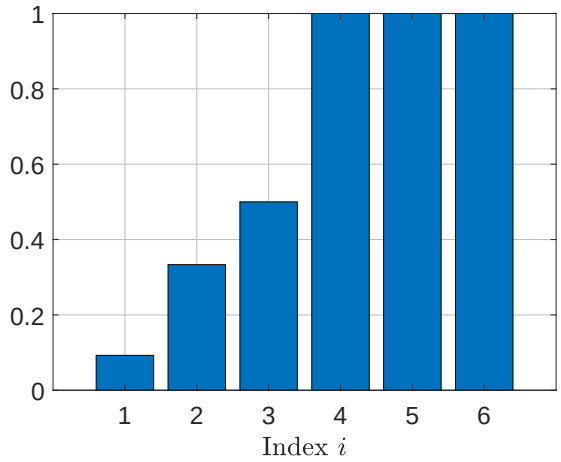
図 5: 条件付きグループワイズ非衝突確率 $Q_i(N, K, J)$ $(N = 6)$



$(K, J) = (6, 3)$



$(K, J) = (6, 4)$



$(K, J) = (6, 5)$

- Theorems 3-4 と整合的な結果が得られている.
- $i \in \{1, \ldots, i^*(K - J)\}$ の範囲では $Q_i(N, K, J) < 1$,
  $i \in \{i^*(K - J) + 1, \ldots, {}_K\mathrm{C}_J\}$ の範囲では $Q_i(N, K, J) = 1$.

グループワイズ・アプローチ: べき乗近似

- べき乗近似は, 1 本目の棒が最後までずっと並ぶと仮定しているに等しい.
- Theorem 2 より, $J = 2$ のときは 1 本目の棒が最大値をとるので, べき乗近似は真の非衝突確率 $Q(N, K)$ を過大推定する.
- Theorem 3 より, $J \geq 3$ かつ $K = J + 1$ のときは 1 本目の棒が最小値をとるので, べき乗近似は $Q(N, K)$ を過小推定する.
- $J \geq 3$ かつ $K > J + 1$ のときは 2 通り考えられる:
  1. 1 本目の棒が最小値をとる場合, べき乗近似は $Q(N, K)$ を過小推定する (例: $(N, K, J) = (6, 6, 4)$).
  2. 1 本目の棒が最小値をとらない場合, べき乗近似の誤差の符号は未確定 (例: $(N, K, J) = (6, 6, 3)$).

グループワイズ・アプローチ: べき乗近似

予想

任意のカレンダーサイズ $N \in \{5, 6, \ldots, \}$, クラスサイズ $K \in \{5, 6, \ldots, N\}$, グループサイズ $J \in \{3, 4, \ldots, K-2\}$ について, べき乗近似は真の非衝突確率 $Q(N, K)$ を過小推定する:

$$Q(N, K) > \{Q_1(N, K, J)\}^{K C_J}.$$

- $Q(N, K) = {}_N\mathrm{P}_K / N^K$, $Q_1(N, K, J) = {}_N\mathrm{P}_J / N^J$.
- 一見容易に見えるが, 実際に証明しようとすると難解.
- 茂木・林が数学的帰納法, 背理法, 両辺の対数変換など様々なアプローチを試みたが, いまだ証明に至らず.
- 数値計算によると, 上記の不等式は強く成り立つ.
- 予想の証明に関する助言・協力大歓迎.

グループワイズ・アプローチ: べき乗近似

- 前頁の予想不等式は次のとおり書き換えられる:

$$Q(N,K) > \{Q(N,J)\}^{{}_K\mathrm{C}_J}.$$

- 仮定より $K \geq J+2$ なので, $0 < Q(N,K) < Q(N,J) < 1$ が成り立つ.
- $Q(N,J)$ を十分な回数べき乗すると, 不等号の向きが入れ替わる. その回数が ${}_K\mathrm{C}_J$ 回以下であることを示せば証明完了.
- この命題はグループワイズ・アプローチの枠を越え, BDP 全体に関わる重要な命題である (異なる 2 つのクラスサイズの下での非衝突確率の比率の特徴づけ).

研究と教育の融合: Motegi & Woo (2023)

- 2021 年度前期「研究指導 I」において, 単なる教育目的で BDP の問題演習を行った.
- 茂木は学生時代からその時点に至るまで, べき乗近似 $Q^{ap}(N,K)$ が厳密解を与えると勘違いしていた.
- 履修者の Sejun Woo 君 (当時学部 3 年生) が順列アプローチに基づいて真の厳密解 $Q(N,K)$ を導き, 茂木案 $Q^{ap}(N,K)$ との数値の不一致を指摘.
- 二人で検討した結果, $Q^{ap}(N,K)$ は厳密解ではなく近似解であることが判明.
- 近似誤差を特徴づけるべく共同研究を始め, Motegi & Woo (2023) 執筆.

研究と教育の融合: Motegi & Hayashi (2023)

- 2023 年度前期「初年次セミナー・基礎演習」において, BDP および Motegi & Woo (2023) を紹介.
- 履修者の林蒼馬君 (学部 1 年生) が, 論文初見にもかかわらず, ペアワイズ・アプローチからグループワイズ・アプローチへの拡張可能性を指摘.
- グループワイズ・アプローチという発送は, 茂木にとって聞いたことも考えたこともなかった.
- 茂木による下調べの結果, 有望な研究課題であることが判明 ($\{Q_i(N, K, K)\}$ が単調減少となるのは $J = 2$ のときのみ).
- 夏休みに共同研究実施, Motegi & Hayashi (2023) 執筆.

研究と教育の融合

- BDP の特長: 高校数学の知識と紙と鉛筆だけで研究を始められるため, 多くの学生にとって取り掛かりやすい.
- 研究体験を通じ, 学生の意識が「勉強」から「研究」へと変わり, 授業履修が「目的」から「手段」へと変わる (例: 数学, 英語, プログラミング, アカデミック・ライティング).
- 早い段階での研究体験は, 学生の交換留学, 卒業研究, キャリア選択など多方面でプラスの効果を生むと期待される.
- 参考: Woo 君は不動産経済学に関する立派な卒業論文を完成させた後, 慶應義塾大学大学院経営管理研究科へ進学.

まとめ

- 誕生日のパラドックス (birthday paradox; BDP) という歴史ある問題に対して, グループワイズ・アプローチという新たな接近方法を提案した.
- グループワイズ・アプローチを採用すると, 古典的な順列アプローチでは得られない条件付き確率に関する興味深い知見を得ることができる.
- BDP は学生にとって取り掛かりやすく, 英文学術雑誌掲載も目指せるテーマであるため, 教育と研究の相乗効果が大きい.
- 待望: 茂木・林予想 (pp. 26-27) を証明できる方.

参考文献

- 藪友良 著, 『入門実践する統計学』, 東洋経済新報社, 2012 年.
- Bellare, M. and T. Kohno (2004). "Hash function balance and its impact on birthday attacks," in *Advances in Cryptology – EUROCRYPT 2004*, ed. by C. Cachin and J. Camenisch, pp. 401-418, Springer-Verlag Berlin Heidelberg, Germany.
- DasGupta, A. (2005). "The matching, birthday and the strong birthday problem: A contemporary review," *Journal of Statistical Planning and Inference*, vol. 130, pp. 377–389.
- Kim, J. H., R. Montenegro, Y. Peres, and P. Tetali (2010). "A birthday paradox for Markov chains with an optimal bound for collision in the Pollard Rho algorithm for discrete logarithm," *Annals of Applied Probability*, vol. 20, pp. 495–521.

参考文献

- Mosteller, F. (1962). "Understanding the birthday problem," *Mathematics Teacher*, vol. 55, pp. 322–325.
- Motegi, K. and S. Hayashi (2023). "A groupwise approach to the birthday paradox," SSRN Working Paper No. 4593602.
- Motegi, K. and S. Woo (2023). "A note on the exponentiation approximation of the birthday paradox". *Communications in Statistics – Theory and Methods*, DOI: 10.1080/03610926.2023.2245086
- Suzuki, K., D. Tonien, K. Kurosawa, and K. Toyota (2006). "Birthday paradox for multi-collisions," in *Information Security and Cryptology – ICISC 2006*, ed. by M. S. Rhee and B. Lee, pp. 29–40, Springer-Verlag Berlin Heidelberg, Germany.