The Value of Excess Supply in Spatial Matching Market

Akbarpour, M., Alimohammadi, Y., Li, S., & Saberi, A. (2021).
ArXiv. /abs/2104.03219

理論談話会 2023/05/11(Thu)

黛 風雅 / Mayuzumi Fuga

サマリー・良かった点・悪かった点

【空間上での動的マッチング問題】
ドライバーがランダムに分布する空間で利用者が逐次的に発生する状況を扱う.

【確率的需要に対し動的な戦略ではなく供給量のみで対応】
ドライバー供給の増加による影響を定量化し，需給が一致している状況での事後合理性を満たすような最適マッチングのサービスレベルを達成するために必要な過剰供給の程度を明らかに．また，極めて単純な貪欲アルゴリズムでこれが達成されることを強調した.

⭕ ランダムウォークによる到着過程を前提に，ドライバー・利用者の位置を順序統計として解析的に扱いやすい変数として処理していたが，順序統計に従う確率という概念を扱っていることが新鮮だった.

▲ 実NWへの応用性(空間が一次元，実データでの検証がされていない)

新規性・有用性・信頼度

- 新規性

タクシー/バス待ちのような待ち行列形成においては一般的な客/タクシーの到着過程の設定を，マッチング問題に持ち込んでいる

- 有用性

行動モデル的な移動需要の表現を入れるのが難しいが，マッチングにかかる費用に代表させるような簡略化をすれば可能かもしれない．

- 信頼度

上界をとるときの数学的根拠は(難しいながら)書かれているが，物理空間上での解釈が併記されていない点が苦しい．

構成

1. 概要

コンセプト

Dynamic Spatial Matching

特徴

1. 利用リクエストが逐次的に発生し，その都度処理する必要がある
2. 利用者・ドライバーは地理的な位置関係に基づき，より近い相手とマッチングしたい

需給バランスと効率性

- 計画的に自然な方向性としては
  ① 需給をバランスさせて，② そのうえでマッチングを最適化する
- 余計にドライバーを増やすことで利用者便益を高めることも，また自然なアプローチではあるが その計画に必要な費用はすぐに明確にわかるが，便益は捉え難い＆アルゴリズム次第となる

想定するプラットフォーム

設定

システム (プラットフォーマー)

- あるドライバーと利用者をマッチングさせ，システムから取り除く
- 全ての利用者にサービスを提供する必要があり，その総費用を最小化させたい
- <u>費用はドライバー・利用者間の距離と待ち時間で算出する</u>

Supply side

- n ドライバー,
- $[0, l]$ での一様分布

Demand side

- n 利用者
- $[0, l]$ での一様分布
- リクエスト発生後には
  単位時間あたり c の費用が発生

概要

最適方策

最適方策は
1. マッチングされていない利用者とドライバーの位置
2. 利用者の到着過程
3. プラットフォームが将来についてどの程度把握しているか

に依存する

ベンチマーク: あらゆるアルゴリズムと同じかそれ以上に優れた「全知」アルゴリズムを想定

A. 利用者の位置と到着時刻を完璧に予測する
B. ドライバーとライダーをマッチングさせ，事後的な総コストを最小にする。
C. 待ち時間を生まない <- 全知アルゴリズムは到着した利用者を即座にマッチングさせる

バランス市場 n 利用者, n ドライバー

全知アルゴリズムの期待費用 $\Theta(l\sqrt{n})$

→

アンバランス市場 n 利用者, $n(1+\epsilon)$ ドライバー

全知アルゴリズムの期待費用 $\Theta(l)$

緩和問題に対するアルゴリズム

緩和問題

仮定 ①

~~利用者の位置と到着時刻を完璧に予測する~~
>>> 将来を予測できない，あるいは計算能力に限界がある

貪欲アルゴリズムの適用: 到着した利用者と最も近くにいるドライバーとマッチング
- 将来の利用者についての予測を必要としない
- $n(1+\varepsilon)$ ドライバーがいれば，期待費用が $O\left(l\log^3(n)\right)$ と計算される
- n ドライバーの際の全知アルゴリズムの場合の期待費用 $(\Theta(l\sqrt{n}))$ よりもよい性能

仮定 ②

~~全ての利用者をマッチングさせなくてはならない~~
>>> 市場のサイズ n に依存するペナルティ ν のもとで, 放置することもできる

$\nu < ln^{-\frac{1}{2}}$ 十分大きな n に対し ⟶ プラットフォームは全ての利用者を放置することが最適に

したがって $\nu \geq ln^{-\frac{1}{2}+\delta}$ ※$\delta > 0$ ⟶ 全知アルゴリズムの期待費用は $\Omega(ln^{\delta})$ と計算できる
>>> 依然として n の多項式費用で抑えられている

数値実験

貪欲アルゴリズム vs. 全知アルゴリズム

Question: ある定数 n に対して, どの程度の過剰供給量があれば貪欲アルゴリズムによるマッチングでの期待費用は需給が一致した状態での全知アルゴリズムによるマッチングでの期待費用よりも小さくすることができるか？

$$\begin{cases} n = 25 \\ n = 100 \\ n = 1000 \end{cases} \quad ▶ \quad \begin{cases} \text{1 excess driver} \\ \text{4 excess drivers} \\ \text{13 excess drivers} \end{cases}$$

主張

バランスは目指されるべきか？

- 現実のマッチングプラットフォームでは、需給バランスを目的としたポリシーが採用される

Uber

市場にバランスをもたらすことが重要な焦点である
"The main focus is trying to bring balance to the marketplace."

ダイナミックプライシングこそリアルタイムに需給をバランスさせる技術である
"Dynamic pricing is the main technology that allows us to maintain market balance in real-time."

lyft

- 数値実験から得られた結果は過剰供給の重要性を明らかにし，これらの格言が文字通りに受け取られるべきでないことを示唆した

- きっちり需給バランスをとることを目指なくても，
  - わずかな過剰供給でパフォーマンスを大幅に向上させることができる
  - 単純なアルゴリズムでもこれらの利点を実現することができる

2. モデル設定と主な結果

プラットフォームの概要

基本的な設定

$R = \{r_1, \ldots, r_n\}$ 利用者を表す点の集合

$D = \{d_1, \ldots, d_m\}$ ドライバーを表す点の集合

それぞれ独立かつ一様分布の乱数
$[0, l]$ かつ $m \geq n$

$m = n$ … バランス市場

$m > n$ … アンバランス市場

※ 一様分布は必ずしも条件とはならない

Matching M

順序付きペア (r, d) s.t. $r \in R, d \in D$ の集合
ライダーやドライバーは多くとも M の1ペアに属する= one-to-one マッチング

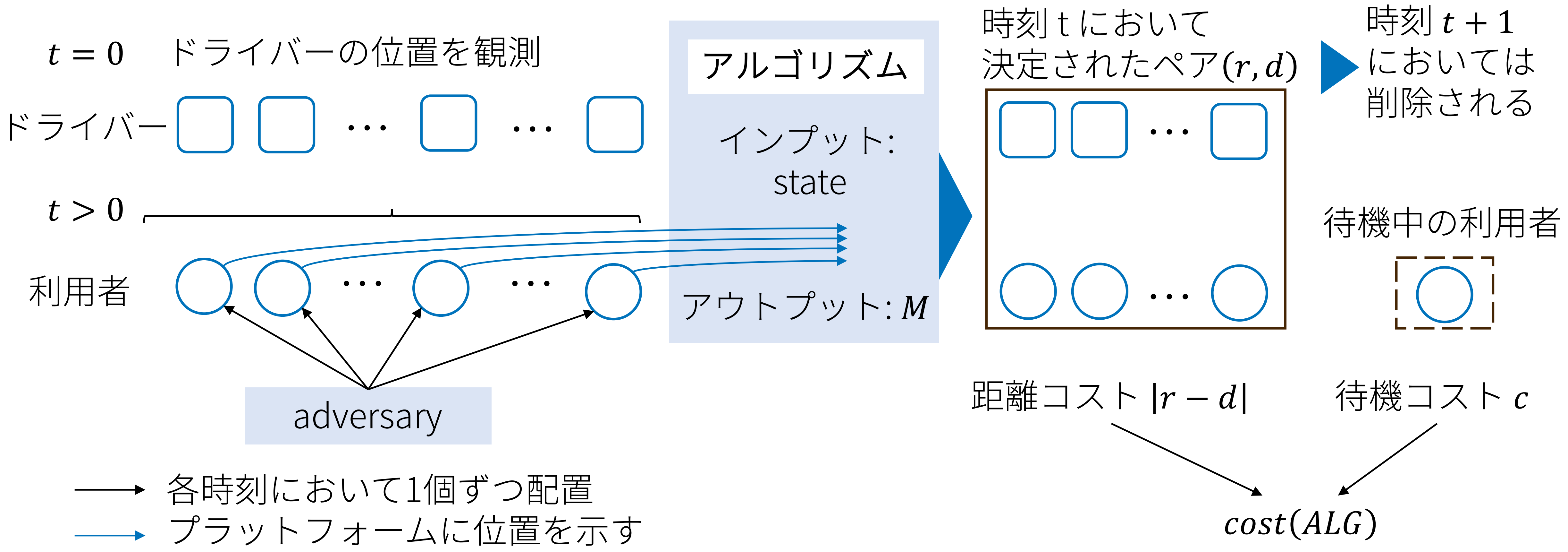
$cost(ALG)$

ALGで示されるアルゴリズムによる期待費用
全ペアに関する距離費用 $|r - d|$ と 単位時間あたりの待機費用 c で計算

マッチングシステム

イメージ

システムのstate = 全てのドライバーの位置と，時刻 t までに到着している利用者の位置



比較する３つのアルゴリズム

最適アルゴリズム

- 利用者の到着順が敵対的であることを把握している
- 残りのドライバーと新たに到着した利用者全員の位置を考慮してマッチング

全知アルゴリズム (ベンチマーク)

- 将来の利用者が到着する場所に関する情報を取得出来る
- すべての利用者の位置と到着時刻を事前に考慮することができる
- 利用者とドライバーを事後最適性を満たすようににマッチングさせる
  >> 「最適アルゴリズム」よりも優れた性能を発揮
- $cost(OMN)$ でその費用を表す

貪欲アルゴリズム

- 各利用者を最も近いドライバーとマッチングさせる
- $cost(Greedy)$ でその費用を表す

全知アルゴリズムの費用計算

線形計画問題としての定式化

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{n}\sum_{j=1}^{m} x_{(i,j)} \left| r_i - d_j \right| \\
\text{s.t.} \quad & \sum_{j=1}^{n} x_{(i,j)} = 1, \ \ \forall r_i \in R, \\
& \sum_{i=1}^{n} x_{(i,j)} \leq 1, \ \ \forall d_j \in D, \\
& x_{(i,j)} \in \{0,1\}
\end{aligned}
$$

全知アルゴリズムは
最適化問題として容易に定式化できる
∵オフライン・アルゴリズムに相当する

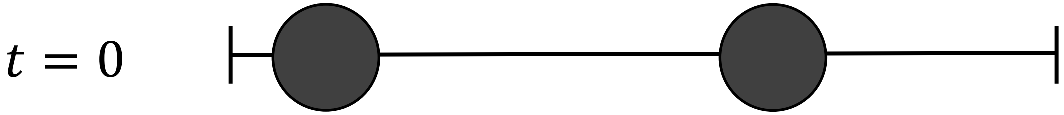
- 待機コスト c
- 利用者の到着過程(ランダム/敵対的)

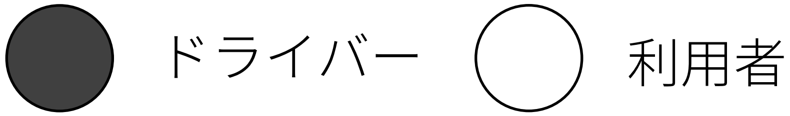
に依存しない

マッチング完了したドライバーのシステム排除による負の外部性

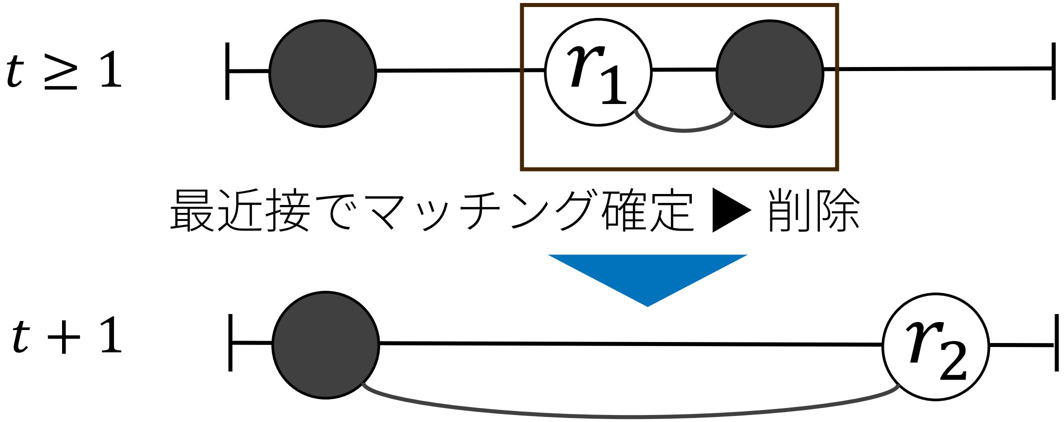
- 全知アルゴリズムと同様に，待機コストは発生しない(最近接でマッチングを確定させるため)
- システム上のドライバーを削除することに伴う外部性があることを無視する。

仮定　以下のようなドライバーの分布で
r_1 が $t \geq 1$ において先に到着，
r_2 が $t+1$ で次に到着





貪欲アルゴリズムの出力



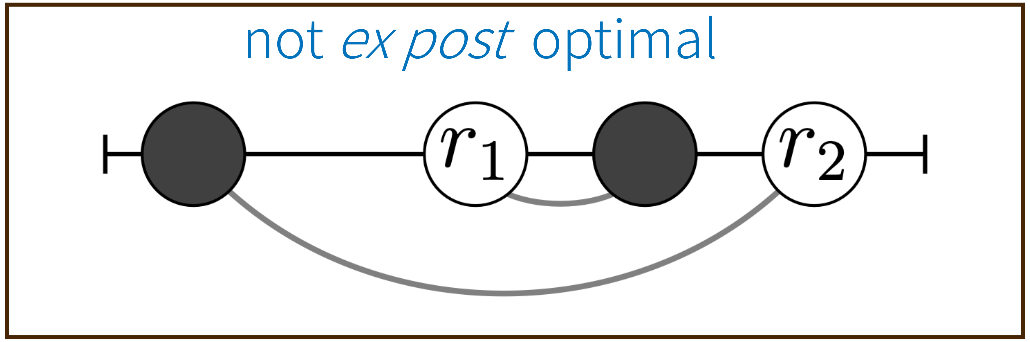
<u>残されたドライバーのうち</u>最近接でマッチング確定

貪欲アルゴリズム

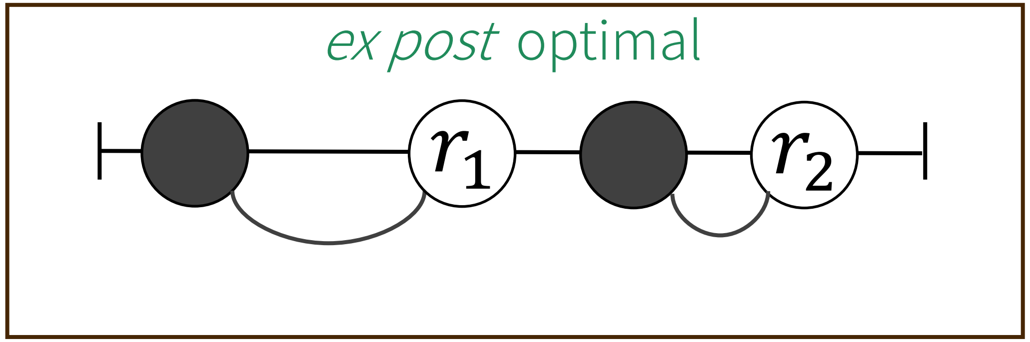
事後効率的(*ex-post* optimal)なマッチングの達成

- アルゴリズムが"patience"を備えていたら，将来的な利用者の到着可能性を考慮して何ステップか待った方がよい場合がある
  - これによって全知アルゴリズムと同様の事後効率的なマッチングを得られる
  - しかし，その分待機コストが生じることになる(サーチ理論との関連性がありそう)

貪欲アルゴリズムによる出力



全知 (or patient) アルゴリズムによる出力



実際にUberはこのようなpatienceを導入したバッチ処理を行っている

Theorem 2.1

$R = \{r_1, \ldots, r_n\}, D = \{d_1, \ldots, d_m\}$ をそれぞれ [0, l] での一様分布から乱択された点の集合とする.

1. バランス市場では, 全ての $n \geq 1$ において以下を満たすような $C > 0$ が存在する.

$$\mathbb{E}[cost(OMN)] \geq Cl\sqrt{n}\ .$$

2. アンバランス市場では，ある $\epsilon > 0$ が存在し $m \geq (1+\epsilon)n$ を満たすとする．このとき，全ての $n \geq 4\epsilon^{-1}$ において以下を満たすような $C_\epsilon > 0$ が存在する.

$$\mathbb{E}[cost(OMN)] \leq C_\epsilon l.$$

ここで，期待値は区間 $[0, l]$ 間の R, D に対するものである.

アンバランス市場での貪欲アルゴリズム

Theorem 2.2

ある $\epsilon > 0$ が存在し $m \geq (1+\epsilon)n$ を満たすようなアンバランス市場を考える．このとき十分大きな n に対して以下を満たすような $C_\epsilon > 0$ が存在する．

$$\mathbb{E}[cost(Greedy)] \leq C_\epsilon l \log^3(n) ,$$

ここで，期待値は区間 $[0, l]$ 間の R, D に対するものである．

- n ドライバーでの全知アルゴリズムでは

cf. $$\mathbb{E}[cost(OMN)] \geq Cl\sqrt{n}$$

- $n(1+\epsilon)$ ドライバーでの全知アルゴリズムでは

cf. $$\mathbb{E}[cost(OMN)] \leq C_\epsilon l$$

筆者は $\mathbb{E}[cost(OMN)]$ より much betterとしているが実際にはほぼ変わらないのでは？

ちなみに $n = e^3 \approx 20.0855$
$(\log(n))^3 = 27$

= ほぼ n のオーダーとして考えられる

分布を一般化した場合

- 利用者とドライバーが単位区間上に独立に存在しているとして，累積分布関数 $F:[0,1] \to [0,1]$が存在し $F(0) = 0, F(1) = 1$ を満たすとする．
- そのような累積分布関数 $F(x)$ に従うランダムな変数 x を仮定すると， $F(x)$ は[0,1] 上に一様分布することになる
- 利用者 r_i とドライバー d_j のマッチングにかかる一般化距離を $\left|F(r_i) - F\left(d_j\right)\right|$ として，$cost_F(\cdot)$ を一般化距離で算出したアルゴリズムの総費用とすればよい

一般化分布での命題

分布の一般形についての命題 Corollary 2.3, 2.4は Theorem 2.1, 2.2から直接的に以下のように導かれる．(証明略)

Corollary 2.3.

F がリプシッツ連続であれば, バランス市場において全ての$n \geq 1$ について以下を満たすような$C > 0$ が存在

$$\mathbb{E}[cost(OMN)] \geq C\sqrt{n}$$

Corollary 2.4.

F^{-1} がリプシッツ連続であれば, ある$\epsilon > 0$ が$m \geq (1+\epsilon)n$ を満たすようなアンバランス市場において，十分大きなn に対して 以下を満たすような$C_\epsilon > 0$ が存在

$$\mathbb{E}[cost(Greedy_F)] \leq C_\epsilon \log^3(n)$$

マッチング制約の緩和

遠方のマッチング費用の高い利用者を無視する場合

- 緩和された問題では，マッチングされていない利用者を無視することが，それにより生じるペナルティの多寡によっては可能になるとする．
- 例えばペナルティが $\nu = \frac{l}{n}$ であれば，，
  全利用者を無視するコストが $n\frac{l}{n} = l$ となり，マッチング費用を下回る
- 実際には十分大きな n について $\nu < \frac{l}{n^{1/2}}$ であればマッチングが発生しなくなる．

▶ ある $\boldsymbol{\delta > 0}$ が存在し，$\boldsymbol{\nu \geq \frac{l}{n^{1/2-\delta}}}$ であるような
ペナルティの時のみを想定すればよい

マッチング制約の緩和

緩和問題の整数計画法による定式化

cf. 全知アルゴリズム

$$\text{minimize} \quad \sum_{i=1}^{n}\sum_{j=1}^{m} x_{(i,j)}|r_i - d_j|$$

$$\text{s.t.} \quad \sum_{j=1}^{n} x_{(i,j)} = 1, \quad \forall r_i \in R,$$

$$\sum_{i=1}^{n} x_{(i,j)} \leq 1, \quad \forall d_j \in D,$$

$$x_{(i,j)} \in \{0,1\}$$

M^* : Minimum matching with penalty (2.6)

$$\text{minimize} \quad \sum_{i=1}^{n}\sum_{j=1}^{m} x_{(i,j)}|r_i - d_j| + \sum_{i=1}^{n} \nu \left(1 - \sum_{j=1}^{m} x_{(i,j)}\right)$$

※ ペナルティ項

$$\text{s.t.} \quad \sum_{j=1}^{n} x_{(i,j)} \leq 1, \quad \forall r_i \in R,$$

※ 緩和により $\sum_{j=1}^{n} x_{(i,j)} = 0$ がありうることに

$$\sum_{i=1}^{n} x_{(i,j)} \leq 1, \quad \forall d_j \in D,$$

$$x_{(i,j)} \in \{0,1\}$$

ペナルティ導入時の費用計算

Theorem 2.5

整数 $n \geq 0$ *と* l *を仮定する．* $R = \{r_1, \dots, r_n\}, D = \{d_1, \dots, d_m\}$ *をそれぞれ [0, l] での一様分布から乱択された点の集合とする．このとき* M^* *について，*

1. $\nu = \frac{l}{n}$, *であれば* $C_1, C_2 > 0$ *sが存在し，全ての* $n \geq 1$ *について*

$$C_1 l\sqrt{n} \geq \mathbb{E}[cost(M^*)] \geq C_2 l\sqrt{n}\ .$$

2. *ある* $\delta > 0$ *が存在し,* $\nu \geq \frac{l}{n^{1/2-\delta}}$ *であれば,*
十分大きな n *に対して以下を満たすような* $C > 0$ *が存在*

$$\mathbb{E}[cost(M^*)] \geq Cln^{\delta}\ .$$

$\Omega(n^{\delta})$ の下限を与える

ここで，期待値は区間 $[0, l]$ *間の* R, D *に対するものである.*

▶ 遠方の利用者を無視する選択肢があっても，少なくとも n についての多項式費用がかかる

数値実験

異なるnでの全知アルゴリズムと貪欲アルゴリズムの比較

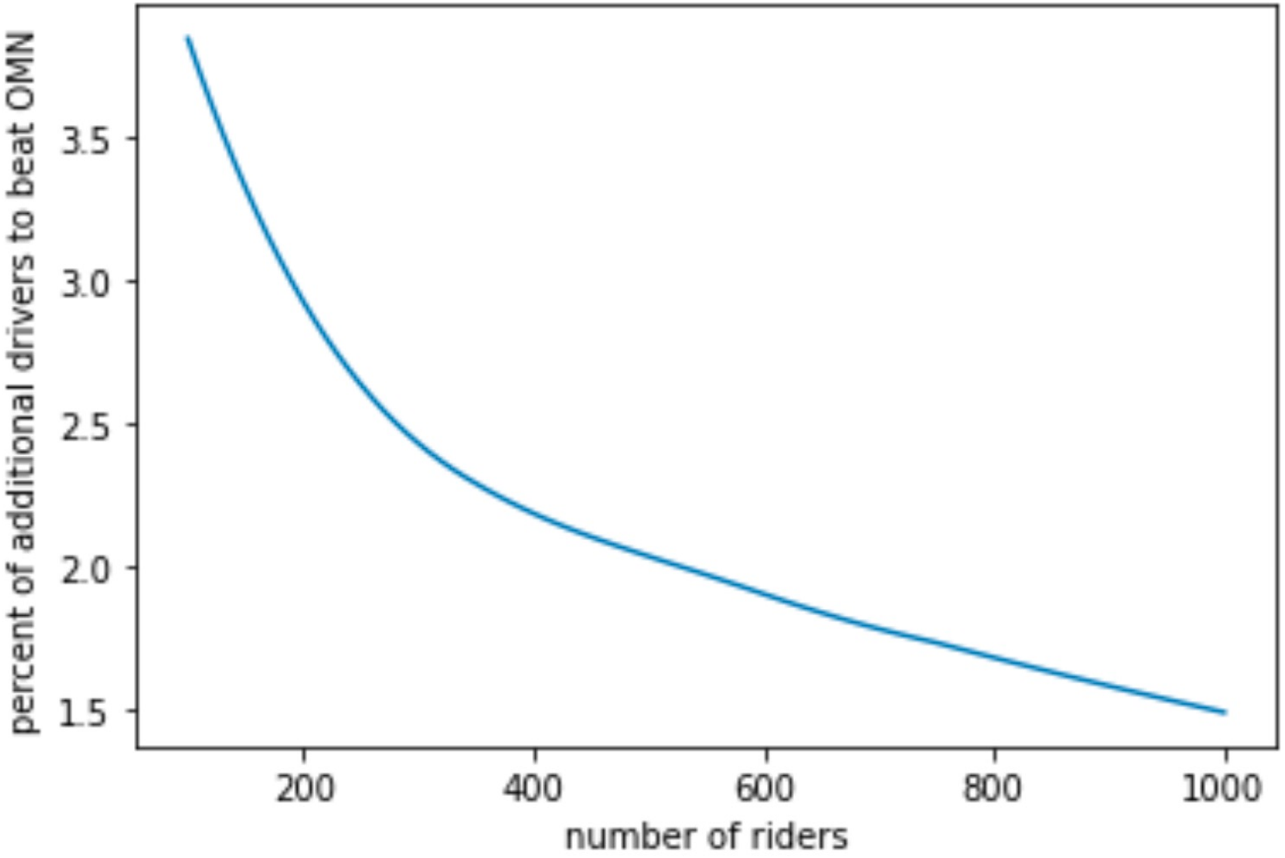


図. 貪欲アルゴリズムが全知アルゴリズムを上回るために必要な追加ドライバーの割合

$n = 25$ → 1 excess driver

$n = 100$ → 4 excess drivers

$n = 1000$ → 13 excess drivers

ドライバー数が増えるに従い,
過剰供給が必要とされる割合は減少

3. 需給がアンバランスな市場

Sec 3.1: ドライバー・利用者系列のランダムウォーク表現
Sec 3.2: 全知アルゴリズムの費用に関するTheorem 2.1の証明
Sec 3.3: 貪欲アルゴリズムの費用に関するTheorem 2.2の証明

ランダムウォーク表現

ランダムウォークによるドライバーと利用者の系列

- ランダムウォーク表現により，区間 $[0, l]$ を以下の i, ii を満たすような $m - n + 1$ 個のサブ区間に分割可能になる(このサブ区間をスライスと呼ぶ)
  i. 各スライスにおいて利用者よりも多いドライバーが存在する
  ii. 各スライス内の利用者とドライバーの総数は指数関数的に減衰する

- 集合 $R = \{r_1, \dots, r_n\}$, $D = \{d_1, \dots, d_m\}$ について $W(0) = 0$ であるようなウォーク $W\colon [0, l] \to \mathbb{Z}$ が存在し, $y > x$ について (3.1) 式が成り立つとする

$$W(y) - W(x) = |D \cap [x, y]| - |R \cap [x, y]| \tag{3.1}$$

- ウォークが最後に i の値をとる時点を脱出時刻(exit time) γ_i とすると (3.2) 式のようにかける．

$$\gamma_i = \sup_{t \geq 0}\{t\colon W(t) - W(0) \leq i\}. \tag{3.2}$$

- $D \cap [0, l] = m, R \cap [0, l] = n$ なので，明らかに $W(l) - W(0) = m - n$ が得られる
  => 合計で $m - n + 1$ 個の脱出時刻が存在する．つまり $i \in \{0, 1, \dots, m - n\}$, where $\gamma_{m-n} = l$

ランダムウォーク表現

スライス集合

- $[0, l]$ を分割する $m - n + 1$ 個のサブ区間 は $[0, \gamma_0], [\gamma_0, \gamma_1], \dots, [\gamma_{m-n-1}, \gamma_{m-n}]$ と書ける
- W は $x \in D$ において $hops\ up$ し， $x \in R$ において $hops\ down$ するという表現を導入する
  （一般的には点の移動は「ジャンプ」として書いている資料が多い）
- 確率変数 H_T が区間 $[0, T]$ 間で発生するホップ数を表すとする，

$$H_T = |\{x \le T : x \in R\ or\ x \in D\}| \tag{3.3}$$

- また，以下のように $\widehat{H_i}$ を定義する

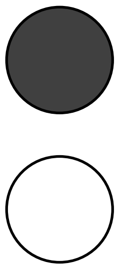
$$\begin{aligned} \widehat{H_0} &= H_{\gamma_0}, \\ \widehat{H_i} &= H_{\gamma_i} - H_{\gamma_{i-1}} - 1, \qquad \text{for } 0 \le i \le m - n + 1 \end{aligned} \tag{3.4}$$

▶ $\frac{\widehat{H_i}}{2}$ の利用者と $\frac{\widehat{H_i}}{2} + 1$ のドライバーがスライス $[\gamma_{i-1}, \gamma_i]$ に存在する

- 各ステップにおける独立なホップの確率が定数で与えられる場合は脱出時刻間のホップ数が独立かつ期待値が定数かつ有界である(Janson 1986)ことが知られている
- 今回は W の系列変化に応じてホップ数の確率は変動するが，$\widehat{H_i}$ に上界を与えることは可能

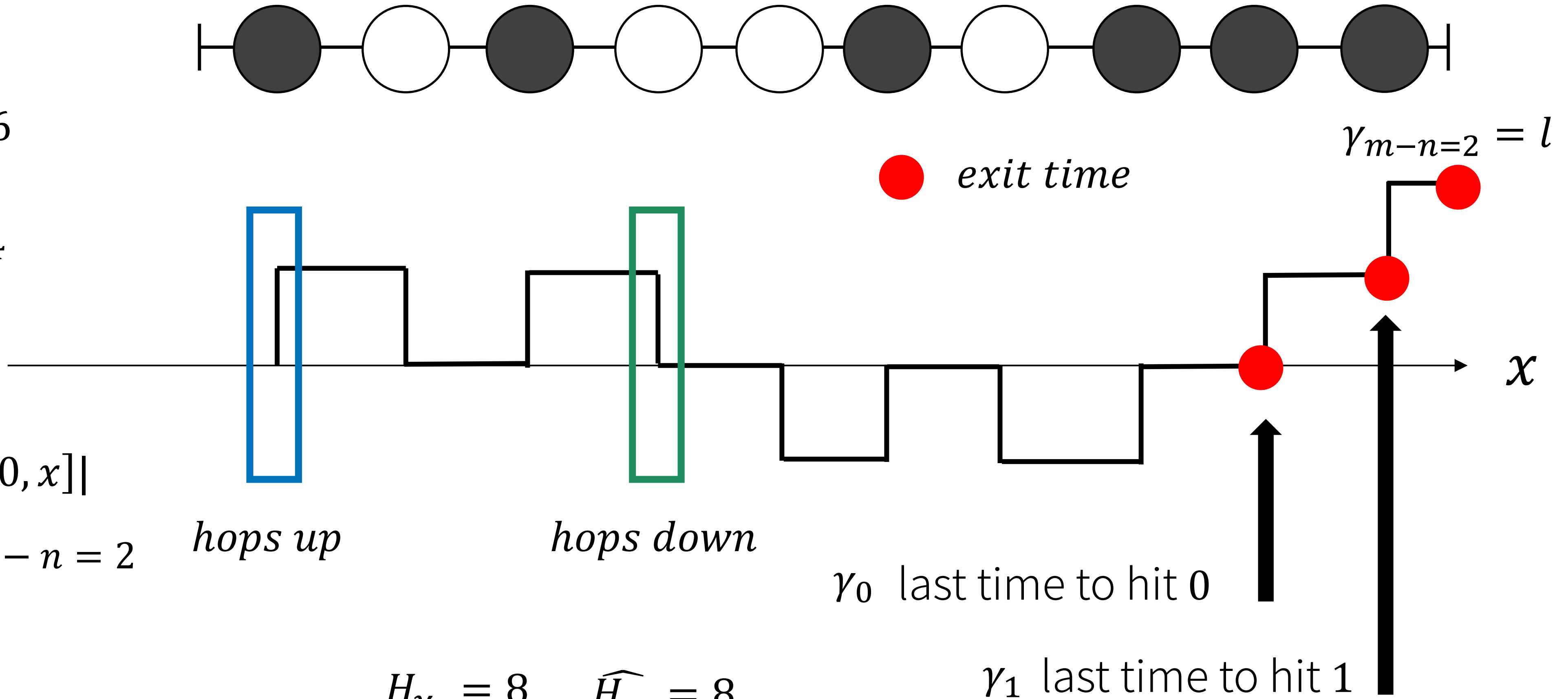
ランダムウォーク中のスライスの図解

簡単な例



driver $m = 6$

rider $n = 4$



ウォーク

$W(x) - W(0)$
$= |D \cap [0,x]| - |R \cap [0,x]|$

※ $W(0) = 0, W_{max} = m - n = 2$

脱出時刻

$$\gamma_i = \sup_{t \geq 0}\{t : W(t) - W(0) \leq i\}$$

最後に i に到達する時刻

$H_{\gamma_0} = 8$	$\widehat{H_{\gamma_0}} = 8$
$H_{\gamma_1} = 9$	$\widehat{H_{\gamma_1}} = 0$
$H_{\gamma_1} = 10$	$\widehat{H_{\gamma_2}} = 0$

スライス $[\gamma_{i-1}, \gamma_i]$ に
$\frac{\hat{H}_i}{2}$ の利用者と
$\frac{\hat{H}_i}{2} + 1$ のドライバーを確認できる

Lemma 3.1

$R = \{r_1, \dots, r_n\}, D = \{d_1, \dots, d_m\}$ *をそれぞれ* $[0, l]$ *での一様分布から乱択された点の集合する.*
(3.4)式により定義された $\widehat{H_i}$ について以下が成り立つ.

i. $i = 0$ について

$$\mathbb{P}\left(\widehat{H_0} = 2k\right) = \frac{m-n}{m+n-2k} \frac{\binom{2k}{k}\binom{m+n-2k}{n-k}}{\binom{m+n}{n}}$$

ii. $i \geq 1$ について

$$\mathbb{P}\left(\widehat{H_i} = 2k\right) = \frac{\binom{2k}{k}}{k+1} \frac{\binom{m+n-2k}{n-k}}{\binom{m+n}{n}}$$

iii. $m \geq (1+\epsilon)n$ かつ $n \geq 4\epsilon^{-1}$ であるとき，十分小さな $\epsilon > 0$ に関して，全ての $i \geq 1$ かつ $k \leq n$ で以下が成り立つようなある $C > 0$ が存在する

$$\mathbb{P}\left(\widehat{H_i} = 2k\right) \leq \frac{C}{k\sqrt{k}}\left(1 - \frac{\epsilon^2}{18}\right)^k,$$

$$and\ for\ i = 0, \qquad \mathbb{P}\left(\widehat{H_0} = 2k\right) \leq \frac{C}{\sqrt{k}}\left(1 - \frac{\epsilon^2}{18}\right)^k.$$

Lemma 3.1の証明

Part i

i. $i=0$ について $\mathbb{P}(\widehat{H_0}=2k)=\dfrac{m-n}{m+n-2k}\dfrac{\binom{2k}{k}\binom{m+n-2k}{n-k}}{\binom{m+n}{n}}$

- W が格子上で $(0,0)$ から $(n+m,m-n)$ の経路 Pに推移するホップを表現する
- $(+1,+1)$ の推移をhops-up，$(+1,-1)$ の推移をhops-downとする
- $[0,\gamma_0]$ の場合のみ，ドライバーと利用者数の総数が一致する
  したがって $\widehat{H}_0=2\mathrm{k}$ であるとき，$(0,0)$ から $(2k,0)$ への経路数は $\binom{2k}{k}$ で得られる

- 次に，$x>\gamma_0$ であれば $W(x)-W(\gamma_0)>0$ であり，$(2k,0)$ から $(n+m,m-n)$ への格子上の経路では 0 に当たることはない．
- $a\geq b$ の条件下での $(0,0)$ から $(a+b,a-b)$ 経路数はKern and Walter (1978)で示されている通り，

$$\frac{a-b}{a+b}\binom{a+b}{b} \tag{3.5}$$

Lemma 3.1の証明

Part i

i. $i=0$ について $\mathbb{P}(\widehat{H}_0 = 2k) = \dfrac{m-n}{m+n-2k} \dfrac{\binom{2k}{k}\binom{m+n-2k}{n-k}}{\binom{m+n}{n}}$

- $a = m-k$， $b = n-k$ とすれば

$$\frac{a-b}{a+b}\binom{a+b}{b} = \frac{(m-k)-(n-k)}{(m-k)+(n-k)}\binom{(m-k)+(n-k)}{n-k} = \frac{m-n}{m+n-2k}\binom{m+n-2k}{n-k}$$

- 経路数の合計は単純に $\binom{m+n}{n}$ で与えられるので，以上の結果を用いると

lattice paths from $(0,0)$ to $(2k,0)$

$$\mathbb{P}(\widehat{H}_0 = 2k) = \boxed{\frac{m-n}{m+n-2k}\binom{m+n-2k}{n-k}} \frac{\boxed{\binom{2k}{k}}}{\boxed{\binom{m+n}{n}}}$$

lattice paths from $(2k,0)$ to $(n+m, m-n)$ total number of paths

Lemma 3.1の証明

Part ii

ii.　$i \geq 1$ について　　$\mathbb{P}(\widehat{H_i} = 2k) = \dfrac{\binom{2k}{k}}{k+1} \dfrac{\binom{m+n-2k}{n-k}}{\binom{m+n}{n}}$

- 格子上経路 P を３つに分割する ▶ $[0, \gamma_{i-1}], [\gamma_{i-1}, \gamma_i], [\gamma_i, l]$

$$(0,0) \xrightarrow[P_0]{} \left(H_{\gamma_{i-1}}, i-1\right) \xrightarrow[P_1]{} \left(H_{\gamma_{i-1}} + 2k + 1, i\right) \xrightarrow[P_2]{} (m+n, m-n)$$

o_{i-1}　　o_i

ここ以降，i に触れない

o_{i-1} から o_i への経路数　$\dfrac{1}{k+1}\binom{2k+1}{k}$ $\cdots a = k+1, b = k$ のときに対応している

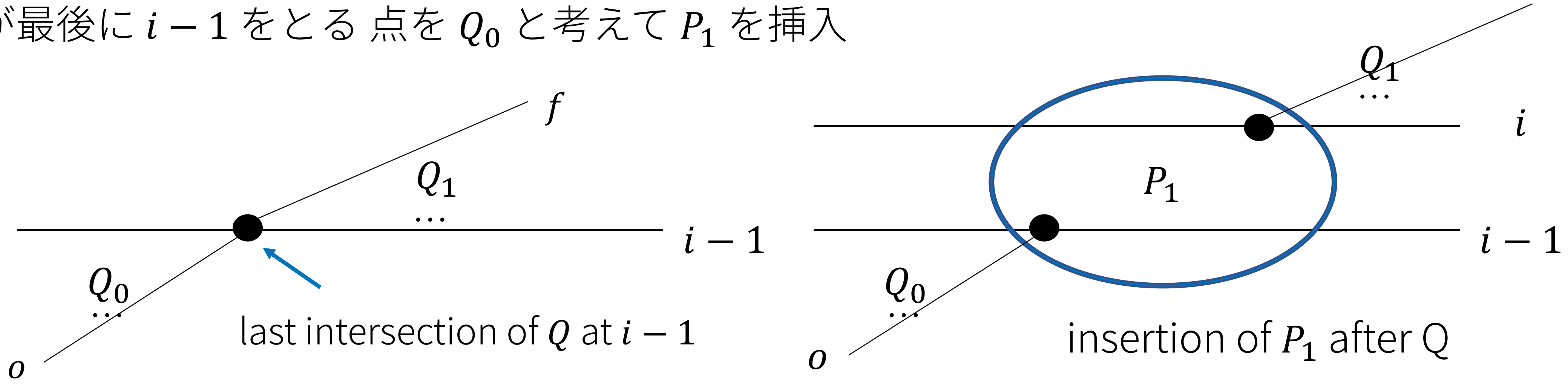
- $[0, \gamma_{i-1}]$ と $[\gamma_i, l]$ については，それぞれに含まれるホップ数に制約はないから，P_0 と P_2 は繋げて経路を考えても問題ないことになる．
- このとき，P_1 を取り除いて $o = (0,0)$ から $f = (m+n-2k-1, m-n-1)$ 経路を考えればよい

Lemma 3.1の証明

Part ii

ii. $i \geq 1$ について $\mathbb{P}(\widehat{H}_i = 2k) = \frac{\binom{2k}{k}}{k+1}\frac{\binom{m+n-2k}{n-k}}{\binom{m+n}{n}}$

- o から f への任意の経路を Q として，
  Q が最後に $i-1$ をとる 点を Q_0 と考えて P_1 を挿入



$$\mathbb{P}(\widehat{H}_i = 2k) = \frac{\boxed{\binom{2k+1}{k}}}{\boxed{k+1}}\frac{\boxed{\binom{m+n-2k-1}{n-k}}}{\boxed{\binom{m+n}{n}}}$$

lattice paths (P_1) from o_{i-1} to o_i

lattice paths from o to f

total number of paths

Lemma 3.1の証明

Part iii

iii. $m \geq (1+\epsilon)n$ かつ $n \geq 4\epsilon^{-1}$ であるとき，十分小さな $\epsilon > 0$ に関して，全ての $i \geq 1$ かつ $k \leq n$ で以下が成り立つようなある $C > 0$ が存在する

$$\mathbb{P}\left(\widehat{H_i} = 2k\right) \leq \frac{C}{k\sqrt{k}}\left(1 - \frac{\epsilon^2}{18}\right)^k,$$

$$and\ for\ i = 0, \qquad \mathbb{P}\left(\widehat{H_0} = 2k\right) \leq \frac{C}{\sqrt{k}}\left(1 - \frac{\epsilon^2}{18}\right)^k$$

- スターリングの公式によって，階乗を指数関数で近似すると， $C > C' > 0$ を満たす定数が存在し，

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \qquad C'\frac{2^{2k}}{\sqrt{k}} \leq \binom{2k}{k} \leq C\frac{2^{2k}}{\sqrt{k}}$$

$$\binom{2k}{k} = \frac{(2k)!}{k!\,k!} = \frac{(2k)!}{(k!)^2}$$

$$(2k)! \sim \sqrt{4\pi k}\left(\frac{2k}{e}\right)^{2k}$$

$$(k!)^2 \sim \left(\sqrt{2\pi k}\left(\frac{k}{e}\right)^k\right)^2 = 2\pi k\left(\frac{k}{e}\right)^{2k}$$

$$\frac{(2k)!}{(k!)^2} \sim \frac{\sqrt{4\pi k}\left(\frac{2k}{e}\right)^{2k}}{2\pi k\left(\frac{k}{e}\right)^{2k}} = \frac{2^{2k}}{\sqrt{k}}\underbrace{\frac{1}{\sqrt{\pi}}}_{\text{定数}}$$

Part iii

$\mathbb{P}(\widehat{H_i} = 2k) = \dfrac{\binom{2k}{k}}{k+1}\dfrac{\binom{m+n-2k}{n-k}}{\binom{m+n}{n}}$ に上界を与えるために， $\dfrac{(2k)!}{(k!)^2} \sim \dfrac{2^{2k}}{\sqrt{k}}\dfrac{1}{\sqrt{\pi}}$ を利用し展開して整理

$$
\begin{aligned}
2^{2k}\frac{\binom{m+n-2k}{n-k}}{\binom{m+n}{n}} &= \frac{m-k}{m+n-2k}\prod_{i=1}^{k}\left(1-\frac{m-n-1}{m+n-2k+2i-1}\right)\left(1+\frac{m-n}{m+n-2k+2i}\right) \\
&\leq \prod_{i=1}^{k}\left(1-\frac{(m-n-1)(m-n)}{2(m+n-2k+2i-1)(m+n-2k+2i)}\right),
\end{aligned}
$$

以上と， $\dfrac{m-n-1}{m+n-2k+2i-1}\dfrac{m-n}{m+n-2k+2i} \geq \dfrac{(m-n)^2}{(m+n)^2} \geq \dfrac{\epsilon^2}{18}$ を合わせると，

$$
2^{2k}\frac{\binom{m+n-2k}{n-k}}{\binom{m+n}{n}} \leq \left(1-\frac{\epsilon^2}{18}\right)^k \tag{3.6}
$$

Lemma 3.1の証明

Part iii

- (3.6)式を用いて，Part ii の結果にあてはめると以下を得られる．

$$\mathbb{P}\left(\widehat{H_i} = 2k\right) \leq \frac{C}{k\sqrt{k}}\left(1 - \frac{\epsilon^2}{18}\right)^k ,$$

$$and\ for\ i = 0, \qquad \mathbb{P}\left(\widehat{H_0} = 2k\right) \leq \frac{C}{\sqrt{k}}\left(1 - \frac{\epsilon^2}{18}\right)^k$$

スライス中のホップ数分布

Corollary 3.2.

- R, D をTheorem 2.1 のものとする．ある$\epsilon > 0$について$m \geq (1+\epsilon)n$であるとする． (3.4)式で定義した $\widehat{H_i}$について考える．このときある$\alpha > 0$が存在し以下が成り立つ

$$\mathbb{P}\left(\boxed{\max_{0\leq i\leq m-n+1} \widehat{H_i} \geq} \alpha \log n\right) \leq \frac{1}{n}.$$

$[\gamma_{i-1}, \gamma_i]$ 間のホップ数

証明　Lemma 3.1 より以下が得られる　$\mathbb{P}(\widehat{H_i} \geq \alpha \log n) \leq \sum_{k\geq\alpha \log n} \frac{C}{\sqrt{k}}\left(1-\frac{\epsilon^2}{18}\right)^k$

$\alpha = \left\lceil\frac{-3}{\log(1-\epsilon^2/18)}\right\rceil$ とすると，

$$e^{k\log\left(\frac{1-\epsilon^2}{18}\right)} \leq e^{(a\log(\mathrm{n}))\left(\log\left(\frac{1-\epsilon^2}{18}\right)\right)} = \exp\left(\left(\frac{-3}{\log(1-\epsilon^2/18)}\right)(\log n)\left(\log\left(\frac{1-\epsilon^2}{18}\right)\right)\right) = \exp(-3\log(n)) = n^{-3} \leq n^{-2}$$

ブールの公式(Union bound) $P\left(\bigcup_i A_i\right) \leq \sum_i P(A_i)$ をこれに当てはめ，命題3.2が得られる

ref.確率過程としてのランダムウォーク

確率過程

- 時間的に変化する現象の時点tでの確率変数をX_tとする．
- この時点tをパラメータとする確率変数の列（集合）$\{X_t\}$を確率過程と呼ぶ
- 確率過程には事象の起こった回数/ステップ単位によって変化する過程(離散時間型, $t \in \mathbb{N}$)と時刻によって変化する過程(連続時間型, $t \in \mathbb{R}$)の2種類ある

▶ グラフ上のランダムウォーク表現は離散時間型，
これを時間幅を無限小にすることで連続時間型にしたものがブラウン運動

ランダムウォークとマルコフ性

- マルコフ性とは…将来の状態が過去に依らず現在の状態のみによるという性質
- マルコフ性を持つ確率過程をマルコフ過程といい，取りうる状態が離散的であればマルコフ連鎖という
  ランダムウォークはマルコフ連鎖/ブラウン運動はマルコフ過程となる

ORにおける活用 … 特に待ち行列モデル

- 待ち行列形成における客/タクシーの到着過程がランダムであるとき， 到着の点列はポアソン過程に従う
- 到着間隔は指数分布に従い，単位時間の到着数はポアソン分布に従う
  (ポアソン過程は指数分布に従う間隔で１ずつ増加する確率過程)

アンバランス市場での貪欲アルゴリズム

モチベーション

- オンラインモデルとしての定式化を行い，貪欲アルゴリズムの費用に対し上界を与えたい
  - 最近接のペアでマッチングを行い，システムから取り除くアルゴリズム

ALGORITHM 1: Greedy

Input: $D = \{d_1, \ldots, d_m\}$.
$M = \emptyset$
while *a new rider r arrives* **do**
  $d^* =$ the closest unmatched driver to r
  $M = M \cup (r, d^*)$
end
return M

- 高い確率で貪欲アルゴリズムが各利用者を$O(\log^2(n))$の最近傍のドライバーとマッチングさせることができる，というTheorem 2.2より強い結果を以て示す

順序統計に基づく点間距離の分布

Proposition 3.4.

一様分布からの乱択によって得られる点群が区間 $[0, l]$ 間で $d_{(1)} \leq d_{(2)} \leq \cdots \leq d_{(m)}$ で昇順に並べ替えられるとする．このとき，

$$\mathbb{P}\left(\max_{0 \leq i \leq m}\left(d_{(i)} - d_{(i-1)}\right) \geq 2\frac{l\log(m)}{m}\right) \leq \frac{1}{m}.$$

証明　$l = 1$ について示してからスケールを合わせる．簡単のため $d_{(0)} = 0, d_{(m+1)} = 1$ とする．
m 個の標準一様分布について， $d_{(i)} - d_{(i-1)} \sim Beta(1, \mathrm{m})$ が従うことが知られている．
このとき各 $m + 1 \geq i \geq 1$ について以下が成り立つ．

$$\mathbb{P}\left(d_{(i)} - d_{(i-1)} \geq t\right) \leq (1 - t)^m$$

したがって，再びブールの公式により

$$\mathbb{P}\left(\max_i\left(d_{(i)} - d_{(i-1)}\right) \geq \frac{\log(m)}{m}\right) \leq m\left(1 - \frac{\log(m)}{m}\right)^m \leq \frac{1}{m}$$

ドライバーに関する事象の定義

- Corollary 3.2.よりスライス $\Gamma = \{[0, \gamma_0], \dots, [\gamma_{m-n-1}, \gamma_{m-n}]\}$ に存在するドライバー数が高々 $\alpha \log(n)$ であるという事象を定義

$$\mathcal{A}_1 = \left\{R \cup D\colon \max_{0 \le i \le m-n+1} \widehat{H_i} \le \alpha \log n\right\}$$

- ドライバー集合の位置に関する順序統計量 $d_{(1)} \le d_{(2)} \le \cdots \le d_{(m)}$ からドライバー間距離が高々 $2l\log(n)$ である市場を定義

$$\mathcal{A}_2 = \left\{D\colon \max_{0 \le i \le m} \left(d_{(i)} - d_{(i-1)}\right) \le \frac{2l\log(n)}{n}\, n\right\}$$

- M_G を貪欲アルゴリズムが返すマッチングとする．

- このとき $R \cup D$ のインスタンスが $\mathcal{A}_1$ と $\mathcal{A}_2$ の双方を満たすときのみを考えれば十分となる
↑開区間 $I_{(r,d)} = (\min(r,d), \max(r,d))$ がペア (r,d) によってカバーされるとき，以下が満たされることが保証されていれさえいれば．(3.8)式の成立の妥当性は証明略．

$$\left|\left\{S \in \Gamma\colon S \cap I_{(r,d)} \neq \emptyset\right\}\right| \le 2\alpha \log n + 4, \qquad \text{for each pair } (r,d) \in M_G, \tag{3.8}$$

M_Gにより確定されるマッチング費用

マッチング費用の上界の算出

- 各利用者のマッチング対象とみなすことが出来るドライバー数の上界

$$\alpha \log n \times (2\alpha \log n + 4)$$

$\mathcal{A}_1$ が与える最もホップ数の多いスライスにおけるエージェント数の上界

(3.8)式が与える最も離れたペア間をまたぐスライス区間数

- マッチングペア間の距離費用の上界は，上のマッチング費用合計にさらにドライバー間距離 $\mathcal{A}_2$ の上界を乗じて利用者数 n で除した値として以下のように得られる

$$|d - r| \leq 4\alpha l(\alpha \log(n) + 2) \frac{\log^2(n)}{n}$$

- $\mathcal{A}_1, \mathcal{A}_2$ が満たされている条件下での全マッチングによる総費用上界はペア数 n を乗じ，ある正定数α'を用いて以下のように書ける

$$\mathbb{E}[cost(M_G)|\mathcal{A}_1, \mathcal{A}_2] = \alpha' \log^3(n)$$

M_Gにより確定されるマッチング費用

マッチング費用の上界の算出

- Proposition 3.4 と Corollary 3.2 はそれぞれ $\frac{1}{m}, \frac{1}{\mathrm{n}}$ の確率での条件であった．
- これらが満たされない場合においては，ドライバーが相当遠隔地にあるうえでスライス中のホップ数に大きな偏りがあるということを示すことになる．
  - このとき$(1-\mathbb{P}(\mathcal{A}_1, \mathcal{A}_2))$のマッチング費用は最大値の ln を上界として与えるとする．

- 以上より，Theorem 2.2 で示されている上界の妥当性を証明することが出来る

$$
\begin{aligned}
\mathbb{E}[cost(M_G)] &= \mathbb{E}[cost(M_G)|\mathcal{A}_1, \mathcal{A}_2]\mathbb{P}(\mathcal{A}_1, \mathcal{A}_2) + ln\big(1-\mathbb{P}(\mathcal{A}_1, \mathcal{A}_2)\big) \\
&\leq \alpha' \log^3(n) + \left\{1-\left(1-\frac{1}{\mathrm{n}}\right)\left(1-\frac{1}{\mathrm{m}}\right)\right\} ln \\
&\leq \alpha' \log^3(n) + \left\{1-\left(1-\frac{1}{\mathrm{n}}\right)^2\right\} ln \\
&\leq \alpha' \log^3(n) + \left\{1-\left(1-\frac{2}{\mathrm{n}}\right)\right\} ln = \alpha' \log^3(n) + 2l
\end{aligned}
$$

所感

- 必要な定理・命題をどのような筋道を立てて設定した上で論文を書きあげているかの想像が立ちづらかった．解ける問題に変形するという点で，確率過程/統計的な型が身についていないと書けない種類の論文だと感じた．
- 先日の留学生ゼミで扱ったRath&Chow(2022)でも価値関数近似のためにOD間交通量に幾何ブラウン過程をあてはめることで解析的に計算できる上界を利用するという手法が採られていたため，OR的な数値計算に欠かせないものであると実感できた．
- 点における時系列的な変量の観測からの推定問題/予測問題はオイラー的であり監視カメラや駅施設のIoTから得られ，複数観測地点間での相関性を評価する上では(PPのような)ラグランジュ的手法が必要になる．Jeanがバス待ち行列を扱っていたが，ネットワーク設計において点の方から問題を記述することで，行動(需要)モデルが精緻になるのではないか．