

離散数理工学 第 10 回
離散確率論：確率的離散システムの解析

岡本 吉央
okamotoy@uec.ac.jp

電気通信大学

2016 年 1 月 5 日

最終更新：2016 年 1 月 4 日 15:11

- | | | |
|---|----------------------|---------|
| 1 | 数え上げの基礎：二項係数と二項定理 | (10/6) |
| ★ | 休講 (体育祭) | (10/13) |
| 2 | 数え上げの基礎：漸化式の立て方 | (10/20) |
| 3 | 数え上げの基礎：漸化式の解き方 (基礎) | (10/27) |
| ★ | 祝日で休み | (11/3) |
| 4 | 数え上げの基礎：漸化式の解き方 (発展) | (11/10) |
| 5 | 離散代数：整数と有限体 | (11/17) |
| 6 | 離散代数：多項式環 | (11/24) |
| 7 | 離散代数：多項式環による有限体の構成 | (12/1) |
| 8 | 離散代数：有限体の応用 | (12/8) |

スケジュール 後半 (予定)

- | | | |
|----|---------------------------|---------|
| 9 | 離散確率論：確率の復習と確率不等式 | (12/15) |
| ★ | 中間試験 | (12/22) |
| 10 | 離散確率論：確率的離散システムの解析 | (1/5) |
| 11 | 離散確率論：乱択データ構造とアルゴリズム (基礎) | (1/12) |
| 12 | 離散確率論：乱択データ構造とアルゴリズム (発展) | (1/19) |
| 13 | 離散確率論：マルコフ連鎖 (基礎) | (1/26) |
| 14 | 離散確率論：マルコフ連鎖 (発展) | (2/2) |
| ★ | 予備日 | (2/9) |
| ★ | 期末試験 | (2/16?) |

注意：予定の変更もありうる

今日の目標

典型的な確率的離散システムの解析ができるようになる

- ▶ 不公平な硬貨投げ
- ▶ クーポン収集問題
- ▶ 誕生日のパラドックス

目次

- ① 不公平な硬貨投げ
- ② クーポン収集問題
- ③ 誕生日のパラドックス
- ④ 今日のまとめ

不公平な硬貨投げ：設定

不公平な硬貨投げ

次のような硬貨 (コイン) を 1 つ投げる

- ▶ 表の出る確率 = p
- ▶ 裏の出る確率 = $1 - p$

ただし, $0 < p \leq 1$

典型的な問題：この硬貨を続けて何回か独立に投げる

- 1 n 回投げて, 表が n 回出る確率は?
- 2 n 回投げて, 表が一度も出ない確率は?
- 3 n 回投げて, 表が一度は出る確率は?
- 4 n 回投げて, 表が出る回数の期待値は?
- 5 表が出るまで投げ続けるとき, 投げる回数の期待値は?

不公平な硬貨投げ：表が出続ける確率は？

- ▶ $E_i = i$ 回目に表が出る (事象)
- ▶ このとき, E_1, \dots, E_n は互いに独立なので

$$\begin{aligned}\Pr(\text{表が } n \text{ 回出る}) &= \Pr(E_1 \text{ かつ } E_2 \text{ かつ } \dots \text{ かつ } E_n) \\ &= \Pr(E_1) \cdot \Pr(E_2) \cdot \dots \cdot \Pr(E_n) \\ &= p \cdot p \cdot \dots \cdot p \\ &= p^n\end{aligned}$$

不公平な硬貨投げ：表が一度も出ない確率は？

- ▶ $\overline{E}_i = i$ 回目に裏が出る (事象)
- ▶ このとき, $\overline{E}_1, \dots, \overline{E}_n$ は互いに独立なので

$$\begin{aligned}\Pr(n \text{ 回中, 表が一度も出ない}) &= \Pr(\overline{E}_1 \text{ かつ } \dots \text{ かつ } \overline{E}_n) \\ &= \Pr(\overline{E}_1) \cdots \cdots \Pr(\overline{E}_n) \\ &= (1 - p) \cdots \cdots (1 - p) \\ &= (1 - p)^n\end{aligned}$$

不公平な硬貨投げ：表が一度は出る確率は？

- ▶ 「表が一度は出る」という事象は
「表が一度も出ない」という事象の余事象
- ▶ したがって、

$$\begin{aligned}\Pr(n \text{ 回中, 表が一度は出る}) &= 1 - \Pr(n \text{ 回中, 表が一度も出ない}) \\ &= 1 - (1 - p)^n\end{aligned}$$

不公平な硬貨投げ：表が出る回数の期待値は？

- ▶ 次の確率変数を考える (事象 E_i の標示確率変数と呼ばれる)

$$X_i = \begin{cases} 1 & (E_i \text{ が生起する, つまり, } i \text{ 回目に表が出る}) \\ 0 & (E_i \text{ が生起しない, つまり, } i \text{ 回目に裏が出る}) \end{cases}$$

- ▶ このとき, $E[X_i] = 1 \cdot p + 0 \cdot (1 - p) = p$
- ▶ したがって,

$$\begin{aligned} E[n \text{ 回中, 表が出る回数}] &= E[X_1 + \cdots + X_n] \\ &= E[X_1] + \cdots + E[X_n] = pn \end{aligned}$$

不公平な硬貨投げ：表が出るまで投げ続けるとき、投げる回数の期待値は？

- ▶ $A_i = 1$ 回目から $i-1$ 回目まですべて裏で、 i 回目で表が出る (事象)
- ▶ このとき、

$$\begin{aligned} \Pr(A_i) &= \Pr(\overline{E_1} \text{ かつ } \dots \text{ かつ } \overline{E_{i-1}} \text{ かつ } E_i) \\ &= \Pr(\overline{E_1}) \cdot \dots \cdot \Pr(\overline{E_{i-1}}) \cdot \Pr(E_i) \\ &= (1-p)^{i-1} p \end{aligned}$$

- ▶ したがって、

$$\begin{aligned} \text{求める期待値} &= \sum_{i=1}^{\infty} i \cdot \Pr(A_i) \\ &= \sum_{i=1}^{\infty} i \cdot (1-p)^{i-1} p \\ &= \frac{1}{p} \quad (\text{詳細は演習問題}) \end{aligned}$$

不公平な硬貨投げ：表が出る回数が期待値から離れる確率は？

- ▶ 次の確率変数を考える (事象 E_i の**標示確率変数**と呼ばれる)

$$X_i = \begin{cases} 1 & (E_i \text{ が生起する, つまり, } i \text{ 回目に表が出る}) \\ 0 & (E_i \text{ が生起しない, つまり, } i \text{ 回目に裏が出る}) \end{cases}$$

- ▶ このとき,

$$\begin{aligned} E[n \text{ 回中, 表が出る回数}] &= E[X_1 + \cdots + X_n] \\ &= E[X_1] + \cdots + E[X_n] = pn \end{aligned}$$

次の確率はどれくらい小さいか？ (または大きいか？)

$$\Pr(X_1 + \cdots + X_n \geq 2pn)$$

不公平な硬貨投げ：マルコフの不等式

マルコフの不等式より

$$\Pr(X_1 + \cdots + X_n \geq 2pn) \leq \frac{E[X_1 + \cdots + X_n]}{2pn} = \frac{pn}{2pn} = \frac{1}{2}$$

「とても小さい」ということが証明できない

マルコフの不等式

(復習)

自然数値確率変数 $X \geq 0$ と正実数 $t > 0$ に対して、 $E[X]$ が存在するとき

$$\Pr(X \geq t) \leq \frac{E[X]}{t}$$

不公平な硬貨投げ：チェルノフ上界の技法

マルコフの不等式より

$$\begin{aligned}\Pr(X_1 + \cdots + X_n \geq 2pn) &= \Pr(2^{X_1 + \cdots + X_n} \geq 2^{2pn}) \\ &\leq \frac{E[2^{X_1 + \cdots + X_n}]}{2^{2pn}}\end{aligned}$$

よって、 $E[2^{X_1 + \cdots + X_n}]$ を知りたい

不公平な硬貨投げ：チェルノフ上界の技法 (2)

X_1, \dots, X_n は互いに独立なので, $2^{X_1}, \dots, 2^{X_n}$ も互いに独立であり,

$$\begin{aligned} E[2^{X_1 + \dots + X_n}] &= E\left[\prod_{i=1}^n 2^{X_i}\right] \\ &= \prod_{i=1}^n E[2^{X_i}] \quad \leftarrow \text{独立性を利用} \end{aligned}$$

ここで, 任意の i に対して

$$E[2^{X_i}] = 2^1 \cdot p + 2^0 \cdot (1 - p) = 2p + (1 - p) = 1 + p$$

ゆえに,

$$E[2^{X_1 + \dots + X_n}] = \prod_{i=1}^n E[2^{X_i}] = (1 + p)^n$$

不公平な硬貨投げ：チェルノフ上界の技法 (3)

まとめると,

$$\begin{aligned}\Pr(X_1 + \cdots + X_n \geq 2pn) &\leq \frac{\mathbb{E}[2^{X_1 + \cdots + X_n}]}{2^{2pn}} \\ &= \frac{(1+p)^n}{2^{2pn}} = \left(\frac{1+p}{4^p}\right)^n\end{aligned}$$

- ▶ 右辺は n が大きくなるにつれて小さくなる
- ▶ $p = 1/2$ のとき, 右辺 = $(3/4)^n$

不公平な硬貨投げ：チェルノフ上界の技法 (4)

疑問

- ▶ 疑問： X_i から 2^{X_i} を作ったが、「2」でないといけないのか？
- ▶ 回答：「2」でなくてもよい。1 より大きければよい

例えば、2ではなく、3にすると、

$$\begin{aligned} \Pr(X_1 + \dots + X_n \geq 2pn) &\leq \frac{E[3^{X_1 + \dots + X_n}]}{3^{2pn}} \\ &= \frac{(1 + 2p)^n}{3^{2pn}} = \left(\frac{1 + 2p}{9p}\right)^n \end{aligned}$$

$p = 1/2$ のとき、この右辺は $(2/3)^n$

チェルノフ上界の技法： X が独立確率変数の和であるとき

- ▶ $E[X]$ の代わりに $E[c^X]$ を考えて、マルコフの不等式 (など) を適用
- ▶ 上界ができる限り小さくなるように、定数 c を定める

目次

- ① 不公平な硬貨投げ
- ② **クーポン収集問題**
- ③ 誕生日のパラドックス
- ④ 今日のまとめ

クーポン収集問題

クーポン収集問題

設定

- ▶ 商品を買うと n 種類の景品 (クーポン) の中の 1 つが当たる
- ▶ 景品の集合 $N = \{1, \dots, n\}$
- ▶ どの景品 i に対しても, $\Pr(\text{景品 } i \text{ が当たる}) = \frac{1}{n}$ で,
これらは商品の間で同一であり, 互いに独立

問題

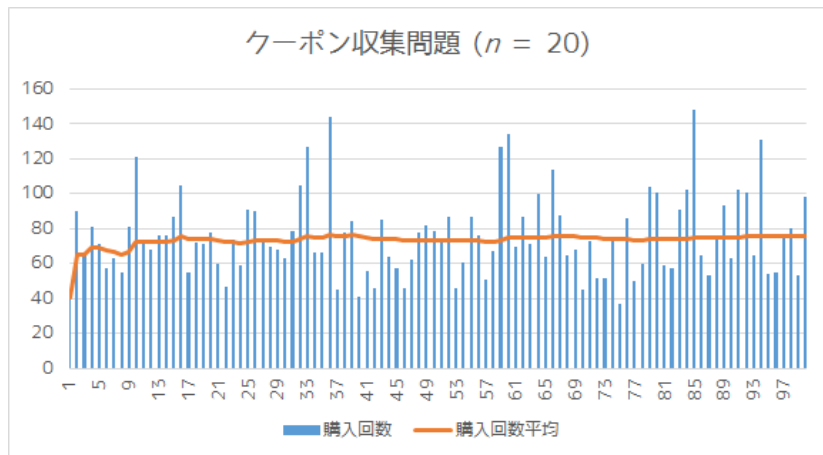
- ▶ 全種類の景品を集め切るまで, 何個商品を購入すればよいか?

注意: 購入商品数は確率変数なので, 答えたいものは

- ▶ 購入商品数の期待値
- ▶ 高確率で購入する商品数 (の上界)

クーポン収集問題：シミュレーション

景品数 20 の場合



10000 回の試行：購入商品数平均 = 72.0825

クーポン収集問題：期待値

考え方：商品を次々と買うとき，既にいくつ景品を持っているか考慮する

$$\text{▶ Pr(新しい景品が当たる | 既に景品を } j \text{ 個所持)} = \frac{n-j}{n}$$

ここで，次の確率変数を考える

$X_j =$ 景品を j 種類所持した瞬間から，
新しい景品が当たるまでに購入した商品の数

- ▶ 景品を j 種類所持しているとき，新しい景品が当たることは表が出る確率が $\frac{n-j}{n}$ である硬貨を投げて表が出ることとみなせる
- ▶ したがって， $E[X_j] = \frac{n}{n-j}$

クーポン収集問題：期待値 (続き)

- ▶ 購入商品数 = $X_0 + X_1 + \dots + X_{n-1}$ なので,

$$\begin{aligned} E[\text{購入商品数}] &= E[X_0 + X_1 + \dots + X_{n-1}] \\ &= E[X_0] + E[X_1] + \dots + E[X_{n-1}] \\ &= \frac{n}{n} + \frac{n}{n-1} + \dots + \frac{n}{1} \\ &= n \sum_{k=1}^n \frac{1}{k} \end{aligned}$$

調和数とは？

第 n 調和数とは、次で定義される数 H_n のこと

$$H_n = \sum_{k=1}^n \frac{1}{k}$$

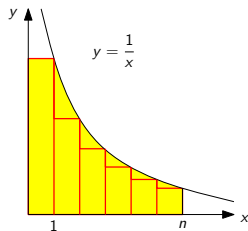
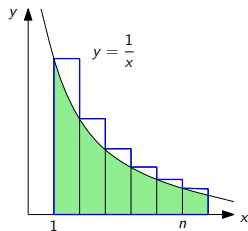
調和数の性質

調和数の上界と下界

任意の整数 $n \geq 1$ に対して

$$\ln(n+1) \leq H_n \leq 1 + \ln n$$

証明 : 演習問題 (ヒントは次の図)



帰結

$$H_n = \ln n + O(1)$$

クーポン収集問題：期待値から確率へ

- ▶ すなわち,

$$E[\text{購入商品数}] = nH_n = n \ln n + O(n)$$

- ▶ マルコフの不等式より

$$\Pr(\text{購入商品数} \geq 2nH_n) \leq \frac{E[\text{購入商品数}]}{2nH_n} = \frac{1}{2}$$

購入商品数が大きくなる確率に対して、もっと「きつい」上界が欲しい

クーポン収集問題：期待値から確率へ — 合併上界の利用 (1)

- ▶ $E_i = 2nH_n$ 回の商品購入で景品 i が得られない (事象)
- ▶ このとき、任意の $i \in \{1, \dots, n\}$ に対して、

$$\begin{aligned} \Pr(E_i) &= \left(\frac{n-1}{n}\right)^{2nH_n} = \left(1 - \frac{1}{n}\right)^{2nH_n} \\ &\leq \left(e^{-\frac{1}{n}}\right)^{2nH_n} = e^{-2H_n} \\ &\leq e^{-2\ln(n+1)} = \frac{1}{(n+1)^2} \end{aligned}$$

事実：有用な不等式

(第1回講義より)

任意の実数 x に対して

$$1 + x \leq e^x$$

クーポン収集問題：期待値から確率へ — 合併上界の利用 (2)

▶ したがって,

$$\begin{aligned}
 \Pr(\text{購入商品数} > 2nH_n) &= \Pr(E_1 \text{ または } E_2 \text{ または } \dots \text{ または } E_n) \\
 &\leq \sum_{i=1}^n \Pr(E_i) \\
 &\leq n \cdot \frac{1}{(n+1)^2} \leq \frac{n+1}{(n+1)^2} = \frac{1}{n+1}
 \end{aligned}$$

つまり, $\lim_{n \rightarrow \infty} \Pr(\text{購入商品数} > 2nH_n) = 0$

合併上界

事象 A, B に対して

$$\Pr(A \cup B) \leq \Pr(A) + \Pr(B)$$

クーポン収集問題：期待値から確率へ (続)

次が知られている (証明は省略：ポアソン近似とチェルノフ技法を使う)

エルデシュとレニイによる 1961 年の結果

任意の正実数 $c > 0$ に対して,

$$\lim_{n \rightarrow \infty} \Pr(\text{購入商品数} > n \ln n + cn) = 1 - e^{-e^{-c}},$$

$$\lim_{n \rightarrow \infty} \Pr(\text{購入商品数} < n \ln n + cn) = 1 - e^{-e^{-c}}$$

つまり購入商品数 (確率変数) は, その期待値の周りに集中している

クーポン収集問題：まとめ

クーポン収集問題

設定

- ▶ 商品を買うと n 種類の景品 (クーポン) の中の 1 つが当たる
- ▶ 景品の集合 $N = \{1, \dots, n\}$
- ▶ どの景品 i に対しても, $\Pr(\text{景品 } i \text{ が当たる}) = \frac{1}{n}$ で, これらは商品の間で同一であり, 互いに独立

問題

- ▶ すべての景品を集め切るまで, 何個商品を購入すればよいか?

回答

- ▶ 購入商品数の期待値は nH_n であり,
- ▶ $n \rightarrow \infty$ のとき, 購入商品数は高い確率で nH_n になる

目次

- ① 不公平な硬貨投げ
- ② クーポン収集問題
- ③ 誕生日のパラドックス
- ④ 今日のまとめ

誕生日のパラドックス：例

誕生日問題

10人いる部屋の中に、誕生日が同じ2人はいるか？
そのような2人がいる確率は？

仮定

- ▶ 1年は366日
- ▶ 人の誕生日がそれら366日の間に等確率で分布する

$$\Pr(i \text{ さんの誕生日が } j) = \frac{1}{366}$$

誕生日のパラドックス：計算

まず、10人の誕生日がすべて異なる確率を計算する

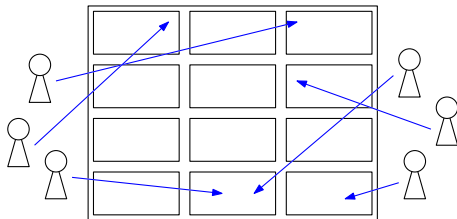
▶ 10人の誕生日がすべて異なる確率 = $\frac{366 \cdot 365 \cdot \dots \cdot 357}{366^{10}} \approx 0.883$

したがって

▶ 10人の中に誕生日の同じ人がいる確率 $\approx 1 - 0.883 = 0.117$

つまり、

▶ 11% ぐらいの確率で同じ誕生日の2人がいる



誕生日のパラドックス：計算 — 30 人の場合

まず、30 人の誕生日がすべて異なる確率を計算する

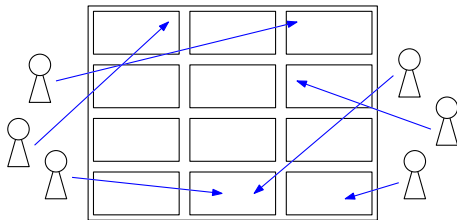
$$\blacktriangleright 30 \text{ 人の誕生日がすべて異なる確率} = \frac{366 \cdot 365 \cdot \dots \cdot 337}{366^{30}} \approx 0.295$$

したがって

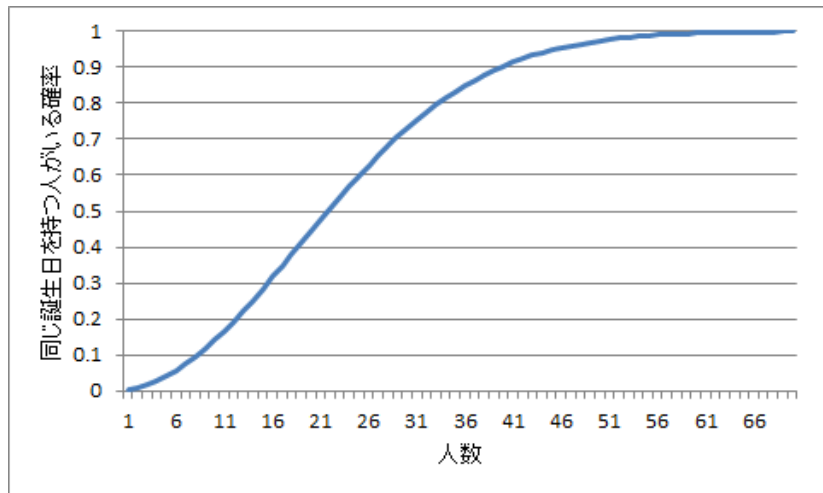
$$\blacktriangleright 30 \text{ 人の中に誕生日の同じ人がいる確率} \approx 1 - 0.295 = 0.705$$

つまり、

$$\blacktriangleright 70 \% \text{ ぐらいの確率で同じ誕生日の 2 人がいる}$$



誕生日のパラドックス：計算してみた



誕生日のパラドックス：一般化

設定

- ▶ $k = 1$ 年の日数
- ▶ $m =$ 部屋の人数
- ▶ $\Pr(i \text{ さんの誕生日が } j) = \frac{1}{k}$

問題

- 1 部屋の中に同じ誕生日の 2 人がいる確率は？
- 2 同じ誕生日の 2 人がいる確率が $\frac{1}{2}$ を超えるのはいつ？

誕生日のパラドックス：一般化

まず、 m 人の誕生日がすべて異なる確率を計算する

- ▶ m 人の誕生日がすべて異なる確率 $= \frac{k \cdot (k-1) \cdot \dots \cdot (k-m+1)}{k^m}$
- ▶ ここで、

$$\begin{aligned} \frac{k \cdot (k-1) \cdot \dots \cdot (k-m+1)}{k^m} &= \prod_{i=0}^{m-1} \frac{k-i}{k} = \prod_{i=0}^{m-1} \left(1 - \frac{i}{k}\right) \\ &\leq \prod_{i=0}^{m-1} e^{-\frac{i}{k}} = e^{\sum_{i=0}^{m-1} -\frac{i}{k}} = e^{-\frac{m(m-1)}{2k}} \end{aligned}$$

事実：有用な不等式

(第1回講義の復習)

任意の実数 x に対して

$$1 + x \leq e^x$$

誕生日のパラドックス：一般化 (2)

したがって,

- ▶ m 人の中に誕生日が同じ 2 人がいる確率 $\geq 1 - e^{-\frac{m(m-1)}{2k}}$
- ▶ $m \geq \sqrt{(2 \ln 2)k} + 1$ のとき, この右辺が $\frac{1}{2}$ 以上になる

なぜならば, $m \geq \sqrt{(2 \ln 2)k} + 1$ であるとき,

$$(m-1)^2 \geq (2 \ln 2)k$$

$$\therefore m(m-1) \geq (2 \ln 2)k$$

$$\therefore -\ln 2 \geq -\frac{m(m-1)}{2k}$$

$$\therefore \frac{1}{2} \geq e^{-\frac{m(m-1)}{2k}}$$

$$\therefore 1 - e^{-\frac{m(m-1)}{2k}} \geq \frac{1}{2} \quad \text{となるから}$$

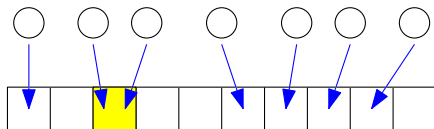
誕生日のパラドックス：ハッシュ値の衝突との関係

ハッシュ

(アルゴリズム論第一の復習)

ハッシュ関数は $N = \{1, \dots, n\}$ から $K = \{1, \dots, k\}$ への関数 h
(典型的には $k < n$)

- ▶ 性質： h が「よくかき混ぜる」関数であるとき
 $h(x) = h(y)$ であるならば、 $x = y$ である可能性が高い
- ▶ $x \neq y$ であるのに $h(x) = h(y)$ であるとき、
 x と y のハッシュ値が衝突 (好ましくない)



誕生日のパラドックス：ハッシュ値の衝突との関係 (続)

ハッシュ

(アルゴリズム論第一の復習)

ハッシュ関数は $N = \{1, \dots, n\}$ から $K = \{1, \dots, k\}$ への関数 h
(典型的には $k < n$)

- ▶ 性質： h が「よくかき混ぜる」関数であるとき
 $h(x) = h(y)$ であるならば、 $x = y$ である可能性が高い
- ▶ $x \neq y$ であるのに $h(x) = h(y)$ であるとき、
 x と y のハッシュ値が衝突 (好ましくない)

次の2つは同じであると見なせる

- ▶ 要素数 m の部分集合 $S \subseteq N$ にハッシュ値の衝突する2要素があるか？
- ▶ 1年が k 日の場合、 m 人の部屋の中に誕生日の同じ2人がいるか？

$\therefore m \geq \sqrt{(2 \ln 2)k + 1}$ のとき、そのような2要素の存在確率は $\frac{1}{2}$ 以上

目次

- ① 不公平な硬貨投げ
- ② クーポン収集問題
- ③ 誕生日のパラドックス
- ④ 今日のまとめ

今日の目標

今日の目標

典型的な確率的離散システムの解析ができるようになる

- ▶ 不公平な硬貨投げ
- ▶ クーポン収集問題
- ▶ 誕生日のパラドックス

残った時間の使い方

- ▶ 演習問題をやる
 - ▶ 相談推奨 (ひとりでやらない)
- ▶ 質問をする
 - ▶ 教員と TA は巡回
- ▶ 退室時, 小さな紙に感想など書いて提出する ← 重要
 - ▶ 内容は何でも OK
 - ▶ 匿名で OK

目次

- ① 不公平な硬貨投げ
- ② クーポン収集問題
- ③ 誕生日のパラドックス
- ④ 今日のまとめ