離散数理工学　第 9 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

2017 年 12 月 19 日

最終更新：2017 年 12 月 18 日　14:05

スケジュール 前半 (予定)

1. 数え上げの基礎：二項係数と二項定理 (10/3)
2. 数え上げの基礎：漸化式の立て方 (10/10)

⋆ 休講 (体育祭) (10/17)

3. 数え上げの基礎：漸化式の解き方 (基礎) (10/24)

⋆ 休講 (出張) (10/31)

4. 数え上げの基礎：漸化式の解き方 (発展) (11/7)
5. 離散代数：対称群と置換群 (11/14)
6. 離散代数：有限群 (11/21)
7. 離散代数：有限群の応用 (11/28)

スケジュール 後半 (予定)

8. 離散確率論：確率の復習と確率不等式 (12/5)
- ⋆ 中間試験 (12/12)
9. 離散確率論：確率的離散システムの解析 (12/19)
10. 離散確率論：乱択データ構造とアルゴリズム (基礎) (1/9)
11. 離散確率論：乱択データ構造とアルゴリズム (発展) (1/16)
12. 離散確率論：マルコフ連鎖 (基礎) (1/23)
13. 離散確率論：マルコフ連鎖 (発展) (1/30)
- ⋆ 予備日 (2/6)
- ⋆ 期末試験 (2/13?)

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

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，$E_1, \ldots, E_n$ は互いに独立なので

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

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，$\overline{E_1}, \ldots, \overline{E_n}$ は互いに独立なので

$$\begin{aligned}
\Pr(n\text{ 回中，表が一度も出ない}) &= \Pr(\overline{E_1}\text{ かつ } \cdots \text{ かつ } \overline{E_n}) \\
&= \Pr(\overline{E_1}) \cdot \cdots \cdot \Pr(\overline{E_n}) \\
&= (1-p) \cdot \cdots \cdot (1-p) \\
&= (1-p)^n
\end{aligned}$$

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，$\mathrm{E}[X_i] = 1 \cdot p + 0 \cdot (1-p) = p$
- したがって，

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

(補足) 不公平な硬貨投げ：表が出る回数の期待値は？

⇝ 標示確率変数を使わなかったら⋯

- $F_j = n$ 回の中で j 回表が出る (事象)
- このとき，$\Pr(F_j) = \binom{n}{j} p^j (1-p)^{n-j}$
- したがって，

$$
\begin{aligned}
\mathrm{E}[n \text{ 回中，表が出る回数}] &= \sum_{j=0}^{n} j \cdot \Pr(F_j) = \sum_{j=0}^{n} j \cdot \binom{n}{j} p^j (1-p)^{n-j} \\
&= np(p + (1-p))^{n-1} = pn
\end{aligned}
$$

ここで (演習問題)

$$
nx(x+y)^{n-1} = \sum_{j=0}^{n} j \binom{n}{j} x^j y^{n-j}
$$

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，

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

- したがって，

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

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，

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

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

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

「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，$\mathrm{E}[X]$ が存在するとき

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

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

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

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

不公平な硬貨投げ：チェルノフ上界の技法 (2)

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

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

ここで，任意の i に対して

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

ゆえに，

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

不公平な硬貨投げ：チェルノフ上界の技法 (3)

まとめると，

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

- 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，

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

$p = 1/2$ 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

- $\mathrm{E}[X]$ の代わりに $\mathrm{E}[c^X]$ 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

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 $N = \{1, \ldots, n\}$
- どの景品 i に対しても，$\Pr(\text{景品 } i \text{ が当たる}) = \dfrac{1}{n}$ で，
  これらは商品の間で同一であり，互いに独立

問題

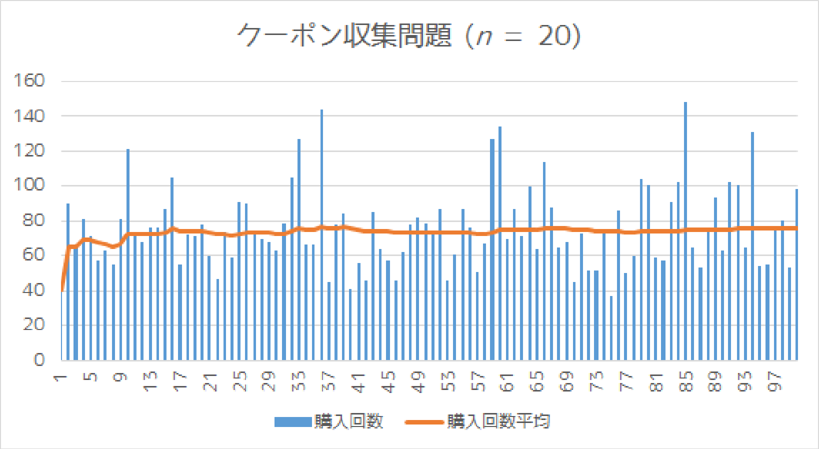
- 全種類の景品を集め切るまで，何個商品を購入すればよいか？

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

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

ここで，次の確率変数を考える

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

- 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 $\dfrac{n-j}{n}$ 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，$\mathrm{E}[X_j] = \dfrac{n}{n-j}$

クーポン収集問題：期待値 (続き)

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

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

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

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

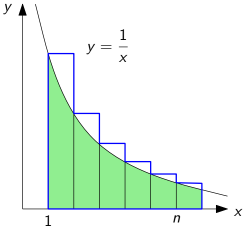
調和数の性質

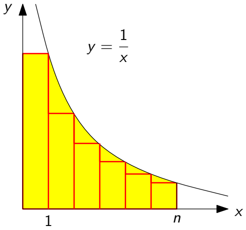
調和数の上界と下界

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，
$$\mathrm{E}[\text{購入商品数}] = nH_n = n \ln n + O(n)$$
- マルコフの不等式より
$$\Pr(\text{購入商品数} \geq 2nH_n) \leq \frac{\mathrm{E}[\text{購入商品数}]}{2nH_n} = \frac{1}{2}$$

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 $i \in \{1, \ldots, n\}$ に対して，

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

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

$$1 + x \leq \mathrm{e}^x$$

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，

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

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

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)

エルデシュとレニィによる 1961 年の結果

任意の正実数 $c > 0$ に対して，

$$\begin{aligned}\lim_{n\to\infty} \Pr(\text{購入商品数} > n\ln n + cn) &= 1 - e^{-e^{-c}},\\ \lim_{n\to\infty} \Pr(\text{購入商品数} < n\ln n + cn) &= 1 - e^{-e^{c}}\end{aligned}$$

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 $N = \{1, \ldots, n\}$
- どの景品 i に対しても，$\Pr(\text{景品 } i \text{ が当たる}) = \dfrac{1}{n}$ 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，
- $n \to \infty$ のとき，購入商品数は高い確率で nH_n になる

目次

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

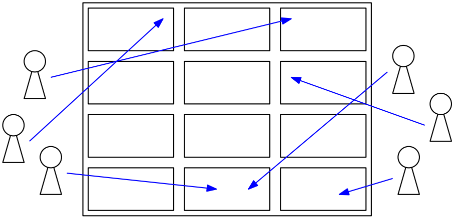
- 10 人の誕生日がすべて異なる確率 $= \dfrac{366 \cdot 365 \cdot \cdots \cdot 357}{366^{10}} \approx 0.883$

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

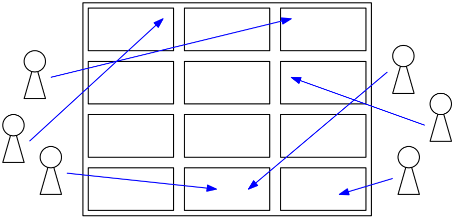
- 30 人の誕生日がすべて異なる確率 $= \dfrac{366 \cdot 365 \cdot \cdots \cdot 337}{366^{30}} \approx 0.295$

したがって

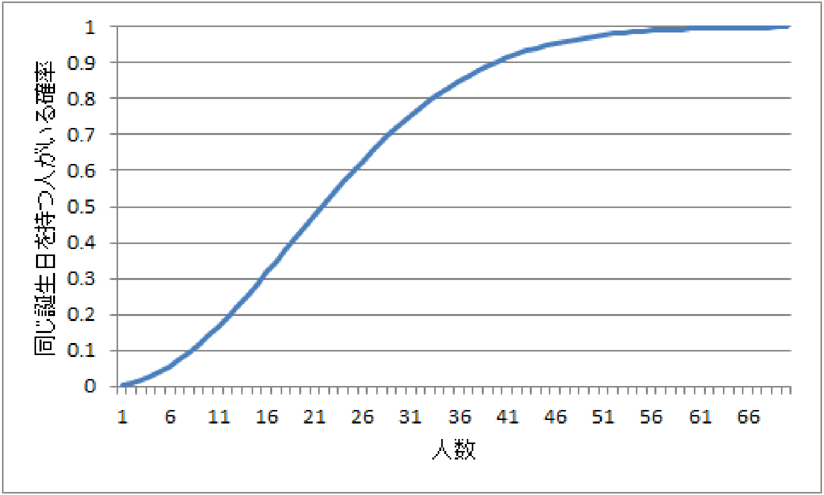
- 30 人の中に誕生日の同じ人がいる確率 $\approx 1 - 0.295 = 0.705$

つまり，

- 70 % 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
- $\Pr(i$ さんの誕生日が $j) = \dfrac{1}{k}$

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 $= \dfrac{k\cdot(k-1)\cdot\cdots\cdot(k-m+1)}{k^m}$
- ここで，

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

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

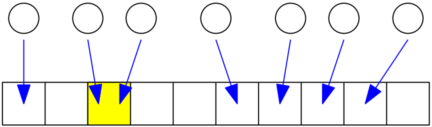
$$
\begin{array}{rcll}
\text{なぜならば，}\quad m & \geq & \sqrt{(2\ln 2)k} + 1 & \text{であるとき，} \\
(m-1)^2 & \geq & (2\ln 2)k & \\
\therefore \quad m(m-1) & \geq & (2\ln 2)k & \\
\therefore \quad -\ln 2 & \geq & -\dfrac{m(m-1)}{2k} & \\
\therefore \quad \dfrac{1}{2} & \geq & e^{-\frac{m(m-1)}{2k}} & \\
\therefore \quad 1 - e^{-\frac{m(m-1)}{2k}} & \geq & \dfrac{1}{2} & \text{となるから}
\end{array}
$$

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 $N = \{1, \ldots, n\}$ から $K = \{1, \ldots, k\}$ 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 $N = \{1, \ldots, n\}$ から $K = \{1, \ldots, k\}$ 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

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