データ構造と
アルゴリズム I

第5回

予定 (I)

4/14: 第1回：イントロダクション
4/21: 第2回：アルゴリズムの解析
4/28: 第3回：関数の増加
5/2 (火，金曜日の講義日):
　第4回：漸化式，確率論的解析
5/12: 小テスト
5/19: 第5回：乱択アルゴリズム
5/26: 休講
6/2: 前半まとめテスト

C.2 確率

- 確率は標本空間 S によって定義される.
- S の要素を根元事象という.
- 例：2枚の硬貨を投げる試行の場合
  $S = \{HH, HT, TH, TT\}$
- 事象：標本空間の部分集合
  - 2枚のうち1枚が表である事象は $\{HT, TH\}$
- S に関する確率分布 $\Pr\{S\}$: 事象から実数への写像 --- 事象に対して，事象が生じる確率を与える.

確率公理

1. 任意の事象 A に対して $\Pr\{A\} \geq 0$
2. $\Pr\{S\}=1$
3. 互いに素な任意の2つの事象 A, B に対し
   $\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\}$

$\Pr\{\emptyset\} = 0$

$A \subseteq B$のとき $\Pr\{A\} \leq \Pr\{B\}$

$\Pr\{\overline{A}\} = 1 - \Pr\{A\}$

$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\}$$
$$\leq \Pr\{A\} + \Pr\{B\}$$

条件付確率

- 2枚のコインのうち少なくとも1枚が表であることがわかったときに，両方とも表の確率は?
- A: 両方が表である事象
- B: 少なくとも一方が表である事象

$$\Pr\{A \mid B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}} \quad (\Pr\{B\} \neq 0 \text{のとき})$$

$$= \frac{1/4}{3/4} = \frac{1}{3}$$

HH	HT
TH	TT

一般に　$\Pr\{A \mid B\} \cdot \Pr\{B\} = \Pr\{A \cap B\}$

$\Pr\{A\} \cdot \Pr\{B\} = \Pr\{A \cap B\}$　または

$\Pr\{A \mid B\} = \Pr\{A\}$　のとき

事象 A, B は独立 (independent) であるという.

Bayes (ベイズ) の定理

条件付確率の定義より

$$\Pr\{A \cap B\} = \Pr\{B\} \cdot \Pr\{A \mid B\}$$
$$= \Pr\{A\} \cdot \Pr\{B \mid A\}$$
$$\Pr\{B\} = \Pr\{B \cap A\} + \Pr\{B \cap \overline{A}\}$$
$$= \Pr\{A\}\Pr\{B \mid A\} + \Pr\{\overline{A}\}\Pr\{B \mid \overline{A}\}$$
$$\Pr\{A \mid B\} = \frac{\Pr\{A\} \cdot \Pr\{B \mid A\}}{\Pr\{B\}}$$
$$= \frac{\Pr\{A\} \cdot \Pr\{B \mid A\}}{\Pr\{A\}\Pr\{B \mid A\} + \Pr\{\overline{A}\}\Pr\{B \mid \overline{A}\}}$$

演習：事後確率

- x: 公正な硬貨 (表の確率が1/2)
- y: インチキな硬貨 (表の確率が1)
- xかyをランダムに選択し，それを2回振る.
- 2回とも表のときに，選んだ硬貨がyである確率は?
  - ベイズの定理を用いて求める
- A: y が選択される事象
- B: 硬貨が2回とも表である事象
  - 求める確率は Pr(A|B)

事後確率

- A: y が選択される事象
- B: 硬貨が2回とも表である事象

$$\Pr\{A \mid B\} = \frac{\Pr\{B \mid A\}\Pr\{A\}}{\Pr\{B \mid A\}\Pr\{A\} + \Pr\{B \mid \neg A\}\Pr\{\neg A\}}$$
$$= \frac{1 \cdot (1/2)}{1 \cdot (1/2) + (1/4) \cdot (1/2)}$$
$$= \frac{4}{5}$$

三囚人のパラドックス

- 三人の囚人A, B, Cがいる．当初は三人とも死刑の予定だったが，恩赦により一人だけ死刑を免れることになった (誰が恩赦になるかはランダムに選ばれていると仮定).
- Aが看守に聞いた：
  - B, Cのどちらかは必ず死刑なんだから，奴らの内から，死刑になる奴の名前を一人教えてくれないか? それを聞いたって別に俺に関する情報が増える訳じゃないし.
- 看守は，Cは死刑になると答えた.
- その晩，Aは考えた．今までは俺が助かる確率は1/3だった．ところが，今はCが死刑になるのは確実なので，俺が助かるか，Bが助かるかのどちらかだ．じゃあ，俺が助かる確率は1/2に増えてる!
- Aの考えは正しいか?

三囚人のパラドックス

- Pr{A恩赦}=Pr{B恩赦}=Pr{C恩赦}=1/3
- 求めたいのは，Pr{A恩赦|看守C死刑}
- Pr{A恩赦|看守C死刑}

= Pr{看守C死刑|A恩赦}･Pr{A恩赦}／

(Pr{看守C死刑|A恩赦}･Pr{A恩赦} --- ?*1/3

+ Pr{看守C死刑|B恩赦}･Pr{B恩赦} --- 1*1/3

+ Pr{看守C死刑|C恩赦}･Pr{C恩赦}) --- 0

三囚人のパラドックス

	A恩赦	B恩赦	C恩赦
Bが死刑と看守が言う	?	0	1/3
Cが死刑と看守が言う	?	1/3	0

看守の戦略

- Aが恩赦の時に，看守がB,Cのどちらを選んで答えるかの戦略が影響する．これが分からないと Pr{A恩赦|看守C死刑}が求まらない

case 1：看守はB,Cからランダムに選ぶ

Aが恩赦の確率は1/3，情報は増えない

case 2: 共に死刑なら看守はBを優先して答える

すなわち，Aが恩赦なら看守はCが死刑と言うことは決してない．よって，Cが死刑と言ったということは，Aが恩赦である確率は0 ☹

case 3: 看守はCを優先して答える

すなわち，Aが恩赦なら看守はBと答えることは決してない．Aが恩赦である確率は1/2 ☺

175

モンティ・ホール問題

- モンティ・ホール氏が司会者を務めるアメリカのクイズ番組でのエピソード
- 三つのドアがあり，一つのドアの中には賞品の高級車が，二つのドアの中にはヤギがいる
- 参加者はドアAを選んだ
- ここで，モンティはドアCを開けて，中にヤギがいるのを見せて，参加者に聞いた．「今なら，ドアAからBにスイッチしても良いですよ．どうしますか?」
- あなたが参加者だったらどうする?

176

どれが正しい?

意見1: AもBも当たりの確率は1/2，よってスイッチしてもしなくても構わない

意見2：最初の状態では，Aが当たりの確率は1/3，BかCのどちらかが当たりの確率は2/3，しかし，今はCが外れであることが分かったので，Bが当たりの確率は2/3に増えている．よって絶対スイッチすべき!

意見3: きっと何か裏がある．スイッチすべきでない!

177

条件付き確率を考えると?

- モンティは当たりのドアを開けることはないとする

	Aに車	Bに車	Cに車
モンティがCを開ける	?	1/3	0
モンティがBを開ける	?	0	1/3

178

モンティの戦略

Aに車がある時に，モンティがB,Cのどちらを選ぶかの戦略が影響する

case 1：モンティはB,Cからランダムに選ぶ

Aに車がある確率は1/3，Bは2/3，スイッチすべき

case 2: モンティはBを優先して開ける

すなわち，Aが当たりならモンティはCを開けることは決してない．よって，Cを開けたということは，Aに車がある確率は0, Bは1，スイッチすべき

case 3: モンティはCを優先して開ける

すなわち，Aが当たりならモンティはBを開けることは決してない．Aに車がある確率は1/2，Bも1/2，どちらでも良い

179

モンティの戦略（続き）

- 今までのケースだと，スイッチして損をすることはない．しかし，モンティがいつもドアを開けるとは限らない場合には?

case 4: モンティはAが当たりの場合にのみ，残りのどちらかのドアを開けてスイッチを促す

Aが当たりの確率は1, Bが当たりの確率は0, スイッチしてはいけない！

180

情報科学の立場からの考察

- センサやフィルタからの入力を解析して，環境を推定するのはよくあること
- その際に，センサやフィルタの特性を考慮することは必須
- これらの問題（三囚人，モンティホール問題）の解は，センサ／フィルタ（看守or モンティ）の特性を考慮しないと得られない
- 世の中にあるパズルの本やweb上でのモンティホール問題の解説のほとんどは，この点を無視しており，（情報科学的には）間違っている．

181

離散確率変数

- 離散確率変数 X: 標本空間 S (有限または可算無限) から実数への関数
  - 例えば，さいころを二個振る場合，標本空間は36個の要素 (1,1), (1,2), …からなる．確率変数Xとして，例えばさいころの大きいほうの値とすると，Xのとり得る値は1,2,…,6.
- 確率変数 X と実数 x に対して，事象 $X = x$ を $\{s \in S: X(s) = x\}$ と定義する: X=2は{(1,2), (2,1), (2,2)}.

$$\Pr\{X = x\} = \sum_{\{s \in S: X(s) = x\}} \Pr\{s\}$$

$$\Pr\{X = x\} \geq 0, \sum_x \Pr\{X = x\} = 1$$

- 関数 $f(x) = \Pr\{X = x\}$ を確率変数 X の確率密度関数 (probability density function) という

182

確率変数の期待値

- 離散確率変数 X の期待値 (expected value, expectation, mean) μ_X または μ

$$\mathrm{E}[X] = \sum_x x \Pr\{X = x\}$$

- E[X+Y] = E[X]+E[Y]
  ---X, Yが独立でなくても成立
- E[aX] = aE[X] (期待値の線形性)
- 新たな確率変数 $g(X)$ に対し

$$\mathrm{E}[g(X)] = \sum_x g(x) \Pr\{X = x\}$$

183

指標確率変数

- 標本空間Sと事象Aが与えられているとき，事象Aに関する指標確率変数X_A=I[A]を以下のように定義する
  - I[A]=1 if Aが生じるとき
  - I[A]=0 if Aが生じないとき
- X_A=I[A]が成立するとき，E[X_A]=Pr[A]が成立する
- 指標確率変数と期待値の線形性を使うと，期待値の計算を容易にすることができる
  - 求めたい期待値を指標確率変数の和で表現する

184

ベルヌーイ試行 (Bernoulli trial):

- 2通りの結果しかない試行を独立に繰り返す試行
  - 成功確率 p
  - 失敗確率 $q = 1-p$
- 例：コインのトス

185

2項分布

- ベルヌーイ試行を n 回繰り返したときに何回成功するか?
- 確率変数 X を，n 回の試行中の成功回数とする

$$\Pr\{X = k\} = \binom{n}{k} p^k q^{n-k}$$

2項分布 (binomial distribution)

- $\Pr\{X=k\}$ は, $(p+q)^n$ の展開式の k 番目の項

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1^n = 1$$

186

二項分布の平均値

$\mathrm{E}[X]$を求めたい (Xはn回中の成功回数)
$X_i = \mathrm{I}[\text{Success-i-th}]$(i 回目の試行で成功するかどうかを表す指標確率変数) とすると，期待値の線形性より

$$\mathrm{E}[X] = \mathrm{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathrm{E}[X_i] = \sum_{i=1}^{n} p = np$$

187

クイズ：期待効用は?

- 壷の中に100円玉が49枚，50円玉が99枚入っている
- 壷をよく振って，一枚ずつコインを取り出していく
- 一方のコインがすべてなくなった時点で，残りのコインを全部もらえる
- 得られる金額の期待値は?
- もしお金を払って，このゲームに参加するならいくら払う?

188

解答

- H_iを，i枚目の100円玉が残るという事象，F_iをi枚目の50円玉が残るという事象とすると，期待値は $E[\sum_{i=1}^{49} 100 \times I[H_i] + \sum_{i=1}^{99} 50 \times I[F_i]]$.
- $I[H_i]$は，i枚目の100円玉と99枚の50円玉を合わせた100枚のコイン中で，その100円玉が最後に選ばれた場合．よって，$E[I[H_i]] = 1/100$. $I[F_i]$は，i枚目の50円玉と49枚の100円玉を合わせた50枚のコイン中で，その50円玉が最後に選ばれた場合．よって，$E[I[F_i]] = 1/50$.
- もらえる金額の期待値は，期待値の線形性より，
$\sum_{i=1}^{49} 100 \times E[I[H_i]] + \sum_{i=1}^{99} 50 \times E[I[F_i]]$
$= 49 + 99 = 148.$

189

誕生日のパラドックス

- k人の人が集まっているとき，同じ誕生日の人がいる確率pを考える.
- 双子／うるう年等を無視すれば，明らかに366人いれば確実に1
- kがどのくらいになると，p=1に近くなるか?

190

誕生日のパラドックス

- k人から選ばれたペア (i, j) に関して，二人の誕生日が同じ場合に1, そうでない場合に0となる指標確率変数X_{ij}を考える
- n=365として，$E[X_{ij}]=1/n$が成立する
- 同じ誕生日の人の数を表す確率変数をXとすると，

$$X = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} X_{ij},$$

$$E[X] = E[\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} X_{ij}] = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} E[X_{ij}] = \binom{k}{2}\frac{1}{n} = \frac{k(k-1)}{2n}$$

- $k(k-1)>2n$, すなわちk=28ぐらいで，上記の期待値は1以上
- 一方，自分と同じ誕生日の人がいる確率はk=253で初めて0.5を超える
- 自分にとって めったにないこと≠集団としてめったにないこと

191

秘書雇用問題

- 新しい秘書を雇いたい
- 1からnまでの，n人の候補がいる
- まず1を雇用する
- 次に2を面接し，1よりも良ければ，1を解雇して2を雇う．そうでなければ2を断る
- 以下，nまで繰り返し
- 平均的に，何人の秘書を雇用することになるか?

192

入力に対する仮定

- 雇用する秘書の数は，明らかに秘書がどんな順番で来るかに依存
- 一番運が良い場合：良い順に来る．最初の一人を雇用し，後は全部断れる．
- 最悪の場合：悪い順に来る．全員雇用してしまう．
- 平均的な場合：ランダムな順序で到着する

193

ランダムという仮定

- 正しいとは限らない．
- 人材を紹介する会社が，雇用毎にマージンを取るなら，悪い順に紹介して来るかも知れない．
- 一般に，アルゴリズムの入力はユーザに依存し，ユーザの入力はなんらかの偏りがあることが通例．
- 運の悪い偏りがあっても性能を保証する方法 --- 乱択アルゴリズム

194

乱択アルゴリズム

- アルゴリズム中で，入力が必ずランダムになるように，乱数 (さいころ) を振って調整する．
- 雇用問題なら，紹介会社にリストを提出させて，リストの順をランダムに並び替える．
- 悪意のある紹介会社／偏りのある入力であっても，結果がランダムであることが保証できる．

195

秘書雇用問題の解析

- i番目の候補が雇用されるという事象に対応する指標確率変数をX_iとする．iが雇用されるなら1，されないなら0.
- 雇用される秘書の数は，$X=X_1+X_2+\ldots+X_n$
- iが雇用されるのは，iが1からiまでの候補中で最も優れている場合である．
- 順番がランダムなら，この確率は1/iである

$$E[X]=E[\sum_{i=1}^{n}X_i]=\sum_{i=1}^{n}E[X_i]=\sum_{i=1}^{n}\frac{1}{i}\cong \ln n$$

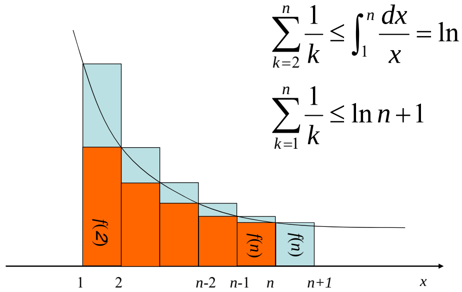
196

調和関数の評価

$$\sum_{k=1}^{n}\frac{1}{k}\geq\int_1^{n+1}\frac{dx}{x}=\ln(n+1)$$

$$\sum_{k=2}^{n}\frac{1}{k}\leq\int_1^{n}\frac{dx}{x}=\ln n$$

$$\sum_{k=1}^{n}\frac{1}{k}\leq\ln n+1$$



197

お見合い問題／オンライン秘書雇用問題

- 秘書を雇うのではなく，順にお見合いする場合を考える．
- お見合いをした直後に，その相手と結婚するかしないかを決める．
- 一度断った相手とは結婚できない．
- 結婚すると決めたら，後のお見合いはできない．
- n人の候補がいるとして，最も望ましい相手と結婚できる確率を最大化するには?

198

お見合い問題

- 以下の方法を考える
- 最初のk人は無条件に断る.
- k+1人目から，過去最高なら結婚する.
- kをどう選べば，最高の相手と結婚できる確率が最適となるか?
- 候補はランダムに来ると仮定する

199

簡単な例

- 3人の場合で，1が最高，2がその次，3が最悪とする.
- 起こりうる順番は，
  - (1,2,3), (1,3,2)
  - (3,2,1), (2,3,1)
  - (3,1,2), (2,1,3)の6通り
- 0人スキップするのは，最初の一人を選ぶのと同じ．1と結婚できる確率は1/3.
- 2人スキップするのは，最後の一人を選ぶのと同じ．1と結婚できる確率は1/3
- 1人スキップする場合は，1と結婚できる確率は1/2
- 1人スキップするのが最適

200

簡単な例(続き)

- 4人の場合は，一人スキップか，二人スキップが最適
- 一人スキップの場合:1が一人目ならアウト
  - 1が二人目ならOK, 確率1/4
  - 1が三人目の場合，最初の二人で，より良い方が一番目ならOK, 確率1/4*1/2=1/8
  - 1が四人目の場合，最初の三人で，一番良い人が一番目ならOK, 確率1/4*1/3=1/12
  - 合計は1/4+1/8+1/12=11/24
- 二人スキップの場合:1が最初の二人に入っていたらアウト
  - 1が三人目ならOK, 確率1/4
  - 1が四人目なら，最初の三人で，一番良い人が二番目以内ならOK, 確率1/4*2/3=1/6
  - 合計は1/4+1/6=5/12
- 一人スキップが良い --- 大体n/3スキップするのが良さそう?

201

お見合い問題(続き)

- i番目に最高の相手がいて，かつ，その相手と結婚できるという事象をS_iとする.
- 最高の相手と結婚できる確率は，$Pr(S_{k+1})+Pr(S_{k+2})+...+Pr(S_n)$となる.
- i≧kに関して，Pr(Si)=1/n ×k/(i-1)となる
  - 最高の相手はランダムにちらばっているので，i番目にいる確率は1/n, また，i番目の最高の相手と結婚できるためには，1からi-1番目までで，最も良い相手が，k番目までに入っている必要があり，この確率はk/(i-1) --- 条件付き確率と独立性

202

お見合い問題(続き)

$$\sum_{i=k+1}^{n} \frac{k}{n(i-1)} = \frac{k}{n} \sum_{i=k+1}^{n} \frac{1}{i-1} \cong \frac{k}{n}(\ln n - \ln k)$$

この式を微分すると, $\frac{1}{n}(\ln n - \ln k - 1)$

$k = n/e$ の時最大.

n=9だと，9/2.718...≒3.3

203