

# 解答例

問題1(1), (2) 授業で配布した資料を参考にしてください

## 問題2

Cormen et al. “Introduction to Algorithms, third edition” (MIT Press)  
9. Medians and Order Statistics. 9.3 Selection in worst-case linear time  
191ページ

日本語の解説も書いておきました。



問題 3 (3)

	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>
<b>S</b>	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	<b>T</b>	<b>\$</b>
<i>ISA<sub>1</sub>[i]</i>	4	10	10	4	10	10	4	10	10	4	11	0
<i>ISA<sub>2</sub>[i]</i>	3	10	7	3	10	7	3	10	7	4	11	0

*ISA<sub>1</sub>[SA<sub>1</sub>[i]+1]*

*ISA<sub>2</sub>[SA<sub>2</sub>[i]+2]*

<i>i</i>	<i>SA<sub>1</sub>[i]</i>	<i>ISA<sub>1</sub>[SA<sub>1</sub>[i]]</i>	この値で ソート	
0	11	\$ 0		
1	0	C 4	G	10
2	3	C 4	G	10
3	6	C 4	G	10
4	9	C 4	T	11
5	1	G 10	G	10
6	2	G 10	C	4
7	4	G 10	G	10
8	5	G 10	C	4
9	7	G 10	G	10
10	8	G 10	C	4
11	10	T 11		

<i>i</i>	<i>SA<sub>2</sub>[i]</i>	<i>ISA<sub>2</sub>[SA<sub>2</sub>[i]]</i>	この値で ソート	
0	11	\$ 0		
1	0	C G 3	G C	7
2	3	C G 3	G C	7
3	6	C G 3	G C	7
4	9	C T 4		
5	2	G C 7	G G	10
6	5	G C 7	G G	10
7	8	G C 7	T \$	11
8	1	G G 10	C G	3
9	4	G G 10	C G	3
10	7	G G 10	C T	4
11	10	T \$ 11		

<i>i</i>	<i>SA<sub>4</sub>[i]</i>				
0	11				
1	0	C	G	G	C
2	3	C	G	G	C
3	6	C	G	G	C
4	9				
5	2	G	C	G	G
6	5	G	C	G	G
7	8	G	C	T	\$
8	1	G	G	C	G
9	4	G	G	C	G
10	7	G	G	C	T
11	10				

問題 3 (3)

	0	1	2	3	4	5	6	7	8	9	10	11
<i>S</i>	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	<b>T</b>	<b>\$</b>
<i>ISA</i> <sub>4</sub> [ <i>i</i> ]	3	9	6	3	9	6	3	10	7	4	11	0

<i>i</i>	<i>SA</i> <sub>4</sub> [ <i>i</i> ]	<i>ISA</i> <sub>4</sub> [ <i>SA</i> <sub>4</sub> [ <i>i</i> ]]										
		<i>ISA</i> <sub>4</sub> [ <i>SA</i> <sub>4</sub> [ <i>i</i> ]+4]										
0	11	この値で ソート										
1	0	<b>\$</b>	0									
2	3	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	3	<b>G</b>	<b>G</b>	<b>C</b>	<b>G</b>	9	
3	6	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	3	<b>G</b>	<b>G</b>	<b>C</b>	<b>T</b>	10	
4	9	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	3	<b>T</b>	<b>\$</b>	11			
5	2	<b>C</b>	<b>T</b>	<b>\$</b>	4							
6	5	<b>G</b>	<b>C</b>	<b>G</b>	<b>G</b>	6	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	3	
7	8	<b>G</b>	<b>C</b>	<b>G</b>	<b>G</b>	6	<b>C</b>	<b>T</b>	<b>\$</b>	4		
8	1	<b>G</b>	<b>C</b>	<b>T</b>	<b>\$</b>	7						
9	4	<b>G</b>	<b>G</b>	<b>C</b>	<b>G</b>	9	<b>G</b>	<b>C</b>	<b>G</b>	<b>G</b>	6	
10	7	<b>G</b>	<b>G</b>	<b>C</b>	<b>G</b>	9	<b>G</b>	<b>C</b>	<b>T</b>	<b>\$</b>	7	
11	10	<b>G</b>	<b>G</b>	<b>C</b>	<b>T</b>	10						
		<b>T</b>	<b>\$</b>	11								

<i>i</i>	<i>SA</i> <sub>8</sub> [ <i>i</i> ]	<i>ISA</i> <sub>8</sub> [ <i>SA</i> <sub>8</sub> [ <i>i</i> ]]										
0	11	<b>\$</b>										
1	0	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>C</b>	<b>G</b>	
2	3	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	<b>T</b>			
3	6	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	<b>T</b>	<b>\$</b>					
4	9	<b>C</b>	<b>T</b>									
5	2	<b>G</b>	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>			
6	5	<b>G</b>	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	<b>T</b>	<b>\$</b>				
7	8	<b>G</b>	<b>C</b>	<b>T</b>	<b>\$</b>							
8	1	<b>G</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>G</b>			
9	4	<b>G</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	<b>T</b>	<b>\$</b>			
10	7	<b>G</b>	<b>G</b>	<b>C</b>	<b>T</b>							
11	10	<b>T</b>	<b>\$</b>									

問題 3 (4) 授業で配布した資料を参考にしてください

問題 4 (1)

	0	1	2	3	4	5	6	7	8	9	10	11
<i>S</i>	C	G	G	C	G	G	C	G	G	C	T	\$
	S	L	L	S	L	L	S	L	L	S	L	S
				*			*			*		*

アルゴリズムは、  
授業で配布した  
資料を参考にして  
ください

問題 4 (2)

Step 2  
Sort L-types

Step 3  
Sort S-types

	\$	C		G				T
SA:	{11}<{-1	03<06<09}<{-1	-1	-1	-1	-1	-1	-1}<{-1}
	@^	^		^				^
	{11}<{-1	03<06<09}<{-1	-1	-1	-1	-1	-1	-1}<{10}
	^	^	@	^				^
	{11}<{-1	03<06<09}<{02	<05<08<01	<04	-1			<{10}
	^	^	@	^				^
	{11}<{-1	03<06<09}<{02	<05<08<01	<04	<07			<{10}
	^	^	@	^	^	@^		^
	{11}<{-1	03	06	09	<{02	<05<08<01	<04<07	<{10}
	^	^	^	^	@	^	^	^
	{11}<{-1	03	06	09	<{02	<05<08<01	<04<07	<{10}
	^	^	^	^	@	^	^	^
	{11}<{-1	03	06	09	<{02	<05<08<01	<04<07	<{10}
	^	^	^	^	@	^	^	^
	{11}<{00	<03<06<09}<{02	<05<08<01	<04<07	<{10}			

*i* SA[*i*]

0	11
1	0
2	3
3	6
4	9
5	2
6	5
7	8
8	1
9	4
10	7
11	10

# 問題 4 (3)

	0	1	2	3	4	5	6	7	8	9	10	11
<i>S</i>	C	G	G	C	G	G	C	G	G	C	T	\$
	S	L	L	S	L	L	S	L	L	S	L	S
			*			*			*		*	

Step 2  
Sort L-types

SA:

\$	C			G						T
{11}<{-1	03=06=09}<{-1	-1	-1	-1	-1	-1	-1	-1	-1	<{-1}
@^	^			^						^
{11}<{-1	03=06=09}<{-1	-1	-1	-1	-1	-1	-1	-1	-1	<{10}
^	^	@		^						^
{11}<{-1	03=06=09}<{02	-1	-1	-1	-1	-1	-1	-1	-1	<{10}
^	^	@		^						^
{11}<{-1	03=06=09}<{02=05=08	-1	-1	-1	-1	-1	-1	-1	-1	<{10}
^	^	@		^						^
{11}<{-1	03=06=09}<{02=05=08<01=04	-1	-1	-1	-1	-1	-1	-1	-1	<{10}
^	^	@		^						^
{11}<{-1	03=06=09}<{02=05=08<01=04=07	-1	-1	-1	-1	-1	-1	-1	-1	<{10}
^	^	@		^					@^	^

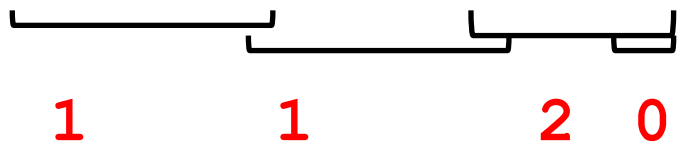
11<03=06<09

Step 3  
Sort S-types

{11}<{-1	03	06	09	<{02=05=08<01=04=07	<{10}
^	^	^	^	@^	^
{11}<{-1	03	06<09	<{02=05=08<01=04=07	<{10}	
^	^	^	@	^	^
{11}<{-1	03=06<09	<{02=05=08<01=04=07	<{10}		
^	^	@	^	^	
{11}<{00=03=06<09	<{02=05=08<01=04=07	<{10}			

## 問題 4 (4)

	0	1	2	3	4	5	6	7	8	9	10	11
<i>S</i>	C	G	G	C	G	G	C	G	G	C	T	\$
	S	L	L	S	L	L	S	L	L	S	L	S
			*				*			*		*



LMS substrings

$$11 < 03 = 06 < 09$$

- Order LMS substrings **03 06 09 11** according to the above ranking to yield **1 1 2 0**.
- Treat **1 1 2 0** as a string, and generate its suffix array using SA-IS to order LMS suffixes.
- 線形時間の証明は講義資料を参考にして下さい。

問題 5 (1)

*Occ*

問題 5 (2)

$BWT[i]$  は  $S$  で  $SA[i]-1$  番目の文字. その suffix のランキングは  $T[i]$  を与えるので

$$T[i] = ISA[SA[i]-1]$$

$SA$  を利用せずに  $T$  を  $O(n)$  時間で計算する方法

$$T[i] = C[BWT[i]] + Occ(BWT[i], i) - 1$$

$x$	<b>\$</b>	<b>C</b>	<b>G</b>	<b>T</b>
$C[x]$	0	1	5	11

$i$	$SA[i]$	$SA[i]-1$	$BWT$	<b>C</b>	<b>G</b>	<b>T</b>	$T$
0	11	10	<b>T</b>	0	0	1	11
1	0	11	<b>\$</b>	0	0	1	0
2	3	2	<b>G</b>	0	1	1	5
3	6	5	<b>G</b>	0	2	1	6
4	9	8	<b>G</b>	0	3	1	7
5	2	1	<b>G</b>	0	4	1	8
6	5	4	<b>G</b>	0	5	1	9
7	8	7	<b>G</b>	0	6	1	10
8	1	0	<b>C</b>	1	6	1	1
9	4	3	<b>C</b>	2	6	1	2
10	7	6	<b>C</b>	3	6	1	3
11	10	9	<b>C</b>	4	6	1	4

	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>
$S$	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>G</b>	<b>C</b>	<b>T</b>	<b>\$</b>
$ISA[i]$	1	8	5	2	9	6	3	10	7	4	11	0



問題 5 (3)

*Occ*

命題:  $\{\}$  は空の配列.

$$lb(\{\}) = 0$$

$$ub(\{\}) = |S| - 1$$

$$Occ(x, -1) = 0$$

$$lb(xW) = C(x) + Occ(x, lb(W) - 1)$$

$$ub(xW) = C(x) + Occ(x, ub(W)) - 1$$

$$lb(\{\}) = 0$$

$$ub(\{\}) = 11$$

$$lb(\mathbf{C}) = C(\mathbf{C}) + Occ(\mathbf{C}, lb(\{\}) - 1) = 1$$

$$ub(\mathbf{C}) = C(\mathbf{C}) + Occ(\mathbf{C}, ub(\{\})) - 1 = 4$$

$$lb(\mathbf{GC}) = C(\mathbf{G}) + Occ(\mathbf{G}, lb(\mathbf{C}) - 1) = 5$$

$$ub(\mathbf{GC}) = C(\mathbf{G}) + Occ(\mathbf{G}, ub(\mathbf{C})) - 1 = 7$$

$$lb(\mathbf{GGC}) = C(\mathbf{G}) + Occ(\mathbf{G}, lb(\mathbf{GC}) - 1) = 8$$

$$ub(\mathbf{GGC}) = C(\mathbf{G}) + Occ(\mathbf{G}, ub(\mathbf{GC})) - 1 = 10$$

$$lb(\mathbf{CGGC}) = C(\mathbf{C}) + Occ(\mathbf{C}, lb(\mathbf{GGC}) - 1) = 1$$

$$ub(\mathbf{CGGC}) = C(\mathbf{C}) + Occ(\mathbf{C}, ub(\mathbf{GGC})) - 1 = 3$$

*Occ* の間引き方は配布資料を参考に

		<b>C</b>	<b>G</b>	<b>T</b>	<i>T</i>
0	<b>T</b>	0	0	1	11
1	<b>\$</b>	0	0	1	0
2	<b>G</b>	0	1	1	5
3	<b>G</b>	0	2	1	6
4	<b>G</b>	0	3	1	7
5	<b>G</b>	0	4	1	8
6	<b>G</b>	0	5	1	9
7	<b>G</b>	0	6	1	10
8	<b>C</b>	1	6	1	1
9	<b>C</b>	2	6	1	2
10	<b>C</b>	3	6	1	3
11	<b>C</b>	4	6	1	4

<i>x</i>	<b>\$</b>	<b>C</b>	<b>G</b>	<b>T</b>	<b>CGGC</b>
$C[x]$	0	1	5	11	



## 問題6

(1) 以下の条件で sensitivity を上げることができるが、同時に偽陽性率も上がる  
①  $q$  を長くする ② seed 長  $k$  を短くする

(2)

$$\sum_{m=0}^q \sum_{i=k}^q T[q][m][i] M^{q-m} (1-M)^m$$

(3) 講義資料

(4) ミスマッチ数4個の領域を seed を使って絞り込むようにする。  
長さ10、ミスマッチ数2個のウィンドウを2つ使う場合が有利 なぜなら

長さ7のウィンドウを3つ使う場合、ミスマッチが1個のウィンドウがどこかに存在するものの、この場合の seed hits 数の期待値は

$$(1+3 \times 7) \times 3 \times (|T|/4^7) = 66 \times |T|/4^7 \quad \textcircled{1}$$

一方、長さ10、ミスマッチ数2個のウィンドウを2つ使う場合の期待値は

$$(1+3^1 \times {}_{10}C_1 + 3^2 \times {}_{10}C_2) \times 2 \times |T|/4^{10} \\ = (1+30+405) \times |T|/4^{10} = 436 \times |T|/4^{10} \quad \textcircled{2}$$

① > ②