

# 確率及び統計/レポート 1

035743A : 比嘉雅樹

## 1 課題

DNA の配列は A・T・C・G の 4 種類の塩基で構成されている。

… ATGCCCGTCAATGTCA …

長さ 10 の配列に於いて「ATG」と「TAG」のパターンが含まれる確率を求めよ。「ATG」と「TAG」の順序はどちらが先でもよい。4 塩基の出現確率は等しいとする。

## 2 解答

まず、「ATG」と「TAG」をそれぞれ一つの塊として X、Y。その他の要素を とおく。X、Y をともに含むには

| | | | |

上記の | 部分のうち 2 つに X、Y が入ればいいので  ${}_5P_2$ 、同じ場所に 2 つ入るときは  $5 \times 2$  通り、他の 4 つの要素はそれぞれ A・T・C・G の 4 通りあるので  $4^4$ 、全部で

$$\{{}_5P_2 + (5 \times 2)\} \times 4^4 = 7680$$

しかし、それだと同じ配列を 2 回数えてしまうことがおきる。

(例)

ATG TAG TAGC

と

ATG TAG TAG C

同じ物を 2 回数えてはまずいので重複分を引く。

重複分には  $XXY$ 、 $YYX$  の 2 つの場合があり、それぞれ

$$\frac{4!}{2!} \times 4$$

の 48 通りあるので重複分は 96 通り、全事象は  $4^{10}$  なので

$$\begin{aligned} P(A) &= \frac{7680 - 96}{4^{10}} \\ &= \frac{237}{32768} \end{aligned}$$