

確率・統計レポート 1

035760A : 横田敏明

問題

DNA の配列は , ATGC の 4 種類の塩基の組み合わせで構成されている .

...ATGTCCATATAGA...

長さ 10 の DNA 配列に於て「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良い . 4 塩基の出現確率は等しいとする .

解答

全体の数は 4^{10} 通りである .

「ATG」「TAG」をそれぞれ p, q とおき , その他の 4 つを r_n とおく .

p, q, r_1, r_2, r_3, r_4 の順列は $6!$ 通りあるので . 重複を含めて

$$6!4^4 \quad (1)$$

通りあるが , 3 つ隣り合った r_n 中に p, q が含まれる場合を引いて ,

$$6!4^4 - 4!2 \quad (2)$$

通りである . よって , 求める確率は .

$$\frac{6!4^4 - 4!2}{4^{10}} = \frac{11517}{65536} \quad (3)$$