Answer on Question #43187-Math-Statistics and Probability

Your DNA code is composed of a series of four nucleotides: adenine, guanine, thymidine and cytosine (A, G, T and C, respectively).

1) What is the probability an individual has the following nucleotide sequence: “TATATA” at any particular position? (assuming independence).

2) What is the probability that an individual has k T’s in their DNA code at any particular position? (k can be any integer and you may assume independence). Here we're looking for the probability of k consecutive T's.

Solution

1) The probability an individual has the following nucleotide sequence: “TATATA” at any particular position is

$$P(T\ is\ first \& A\ is\ second \& T\ is\ third \& A\ is\ fourth \& T\ is\ fifth \& A\ is\ sixth).$$

And since "&" tells us to multiply probabilities:

$$P(T\ is\ first)\cdot P(A\ is\ second)\cdot P(T\ is\ third)\cdot P(A\ is\ fourth)\cdot P(T\ is\ fifth)\cdot P(A\ is\ sixth).$$

The probability of A or T is

$$P(A)=P(T)=\frac{1}{4}.$$

So, the probability an individual has the following nucleotide sequence: “TATATA” at any particular position is

$$P(\text{TATATA})=\overbrace{\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}}^{6}=\frac{1}{4^6}=\frac{1}{4096}.$$

2) Therefore the probability of k consecutive T's

$$P\left(\overbrace{\text{TTT}\ldots\text{T}}^{k}\right)=\overbrace{\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}\cdots\frac{1}{4}}^{k}=\frac{1}{4^k}.$$

Answer: 1) $\frac{1}{4096}$; 2) $\frac{1}{4^k}$.