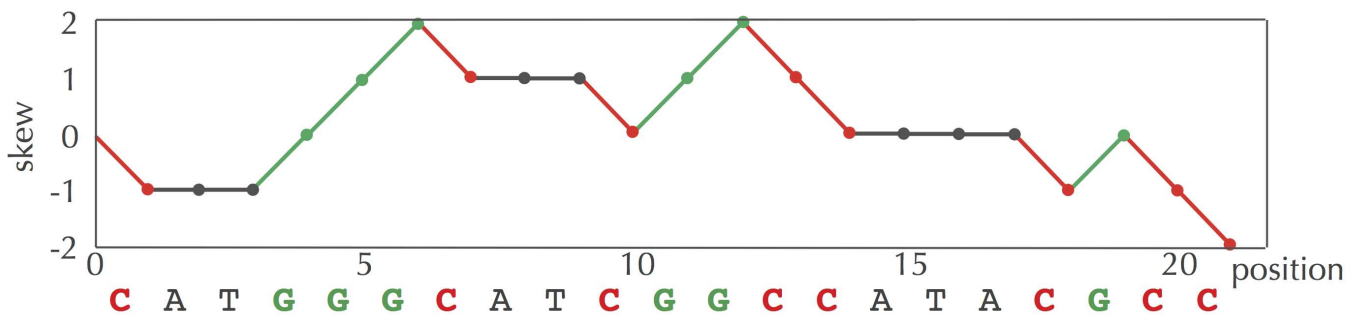


## 1.7 Peculiar Statistics of the Forward and Reverse Half-Strands

7 out of 11 steps passed 0 out of 5 points received

Since we don't know the location of *ori* in a circular genome, let's linearize it (i.e., select an arbitrary position and pretend that the genome begins here), resulting in a linear string *Genome*. We define  $Skew_i(Genome)$  as the difference between the total number of occurrences of G and the total number of occurrences of C in the first  $i$  nucleotides of *Genome*. The **skew diagram** is defined by plotting  $Skew_i(Genome)$  (as  $i$  ranges from 0 to  $|Genome|$ ), where  $Skew_0(Genome)$  is set equal to zero. The figure below shows a skew diagram for the DNA string CATGGGCATCGGCCATACGCC.

Note that we can compute  $Skew_{i+1}(Genome)$  from  $Skew_i(Genome)$  according to the nucleotide in position  $i$  of *Genome*. If this nucleotide is **G**, then  $Skew_{i+1}(Genome) = Skew_i(Genome) + 1$ ; if this nucleotide is **C**, then  $Skew_{i+1}(Genome) = Skew_i(Genome) - 1$ ; otherwise,  $Skew_{i+1}(Genome) = Skew_i(Genome)$ .



**Exercise Break:** Give all values of  $Skew_i(GAGCCACCGCGATA)$  for  $i$  ranging from 0 to 14.

**Sample Input:**

CATGGGCATCGGCCATACGCC

**Sample Output:**

0 -1 -1 -1 0 1 2 1 1 1 0 1 2 1 0 0 0 0 -1 0 -1 -2

To solve this problem please visit  
<https://stepik.org/lesson/240220/step/8>

