Algorithm rna mutation insertion deletion

Recall that S is defined as the set of all RNA strands, nonempty strings made of the bases in $B = \{\texttt{A}, \texttt{U}, \texttt{G}, \texttt{C}\}$. We define the functions

$$mutation : S \times \mathbb{Z}^+ \times B \to S \qquad insertion : S \times \mathbb{Z}^+ \times B \to S$$

$$deletion : \{s \in S \mid rnalen(s) > 1\} \times \mathbb{Z}^+ \to S \qquad \text{with rules}$$

```
1   procedure  mutation(b_1 ··· b_n : a RNA strand , k : a positive integer , b : an element of B)
2   for  i := 1  to  n
3     if  i = k
4       c_i := b
5     else
6       c_i := b_i
7   return  c_1 ··· c_n  {The return value is a RNA strand made of the c_i values}
```

```
1   procedure  insertion(b_1 ··· b_n : a RNA strand , k : a positive integer , b : an element of B)
2   if  k > n
3     for  i := 1  to  n
4       c_i := b_i
5     c_{n+1} := b
6   else
7     for  i := 1  to  k − 1
8       c_i := b_i
9     c_k := b
10    for  i := k + 1  to  n + 1
11      c_i := b_{i−1}
12  return  c_1 ··· c_{n+1}  {The return value is a RNA strand made of the c_i values}
```

```
1   procedure  deletion(b_1 ··· b_n : a RNA strand with n > 1 , k : a positive integer)
2   if  k > n
3     m := n
4     for  i := 1  to  n
5       c_i := b_i
6   else
7     m := n − 1
8     for  i := 1  to  k − 1
9       c_i := b_i
10    for  i := k  to  n − 1
11      c_i := b_{i+1}
12  return  c_1 ··· c_m  {The return value is a RNA strand made of the c_i values}
```

Alternating quantifiers order rna examples

Alternating nested quantifiers

$$\forall s \in S \; \exists n \in \mathbb{N} \; (\; basecount(\; (s, \mathtt{U})\;) = n \;)$$

In English: For each strand, there is a nonnnegative integer that counts the number of occurrences of $\mathtt{U}$ in that strand.

$$\exists n \in \mathbb{N} \; \forall s \in S \; (\; basecount(\; (s, \mathtt{U})\;) = n \;)$$

In English: There is a nonnnegative integer that counts the number of occurrences of $\mathtt{U}$ in every strand.

Are these statements true or false?

$$\forall s \in S \; \exists b \in B \; (\; basecount(\; (s,b)\;) = 3 \;)$$

In English: For each RNA strand there is a base that occurs 3 times in this strand.

Write the negation and use De Morgan's law to find a logically equivalent version where the negation is applied only to the BC predicate (not next to a quantifier).

Is the original statement **True** or **False**?

Alternating quantifiers strategies rna examples

Alternating nested quantifiers

$$\forall s \in S \; \exists b \in B \; (\; basecount(\; (s,b)\;) = 3 \;)$$

In English: For each RNA strand there is a base that occurs 3 times in this strand.

Write the negation and use De Morgan's law to find a logically equivalent version where the negation is applied only to the BC predicate (not next to a quantifier).

Is the original statement **True** or **False**?

$$\exists s \in S \; \forall b \in B \; \exists n \in \mathbb{N} \; (\; basecount(\; (s,b)\;) = n \;)$$

In English: There is an RNA strand so that for each base there is some nonnegative integer that counts the number of occurrences of that base in this strand.

Write the negation and use De Morgan's law to find a logically equivalent version where the negation is applied only to the BC predicate (not next to a quantifier).

Is the original statement **True** or **False**?

Alternating quantifiers proofs rna examples

Which proof strategies could be used to prove each of the following statements?

Hint: first translate the statements to English and identify the main logical structure.

$\forall s \in S \ (\ rnalen(s) > 0 \ )$

$\forall b \in B \ \exists s \in S \ (\ basecount(\ (s,b)\ ) \ > 0 \ )$

$\forall s \in S \ \exists b \in B \ (\ basecount(\ (s,b)\ ) > 0 \ )$

$\exists s \in S \ (rnalen(s) = basecount(\ (s, \mathtt{A})\ )$

$\forall s \in S \ (rnalen(s) \geq basecount(\ (s, \mathtt{A})\ ))$

Defining functions more examples

Let's practice with functions related to some of our applications so far.

Recall: We model the collection of user ratings of the four movies Dune, Oppenheimer, Barbie, Nimona as the set $\{-1, 0, 1\}^4$. One function that compares pairs of ratings is

$$d_0 : \{-1, 0, 1\}^4 \times \{-1, 0, 1\}^4 \to \mathbb{R}$$

given by

$$d_0(\, (\, (x_1, x_2, x_3, x_4), (y_1, y_2, y_3, y_4) \,) \,) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + (x_4 - y_4)^2}$$

Notice: any ordered pair of ratings is an okay input to d_0.

Notice: there are (at most)
$$(3 \cdot 3 \cdot 3 \cdot 3) \cdot (3 \cdot 3 \cdot 3 \cdot 3) = 3^8 = 6561$$
many pairs of ratings. There are therefore lots and lots of real numbers that are not the output of d_0.

Recall: RNA is made up of strands of four different bases that encode genomic information in specific ways. The bases are elements of the set $B = \{\mathtt{A}, \mathtt{C}, \mathtt{U}, \mathtt{G}\}$. The set of RNA strands S is defined (recursively) by:

Basis Step:	$\mathtt{A} \in S, \mathtt{C} \in S, \mathtt{U} \in S, \mathtt{G} \in S$
Recursive Step:	If $s \in S$ and $b \in B$, then $sb \in S$

where sb is string concatenation.

Pro-tip: informal definitions sometime use $\cdots$ to indicate "continue the pattern". Often, to make this pattern precise we use recursive definitions.

Name	Domain	Codomain	Rule	Example
$rnalen$	S	$\mathbb{Z}^+$		
			Basis Step: If $b \in B$ then $rnalen(b) = 1$ Recursive Step: If $s \in S$ and $b \in B$, then $rnalen(sb) = 1 + rnalen(s)$	$rnalen(\texttt{AC}) \overset{\text{rec step}}{=} 1 + rnalen(\texttt{A})$ $\overset{\text{basis step}}{=} 1 + 1 = 2$
$basecount$	$S \times B$	$\mathbb{N}$		
			Basis Step: If $b_1 \in B, b_2 \in B$ then $basecount(\ (b_1, b_2)\ ) =$ $\begin{cases} 1 & \text{when } b_1 = b_2 \\ 0 & \text{when } b_1 \neq b_2 \end{cases}$ Recursive Step: If $s \in S, b_1 \in B, b_2 \in B$ $basecount(\ (sb_1, b_2)\ ) =$ $\begin{cases} 1 + basecount(\ (s, b_2)\ ) & \text{when } b_1 = b_2 \\ basecount(\ (s, b_2)\ ) & \text{when } b_1 \neq b_2 \end{cases}$	$basecount(\ (\texttt{ACU}, \texttt{C})\ ) =$
"2 to the power of"	$\mathbb{N}$	$\mathbb{N}$		
			Basis Step: $2^0 = 1$ Recursive Step: If $n \in \mathbb{N}, 2^{n+1} =$	
"b to the power of i"	$\mathbb{Z}^+ \times \mathbb{N}$	$\mathbb{N}$		
			Basis Step: $b^0 = 1$ Recursive Step: If $i \in \mathbb{N}, b^{i+1} = b \cdot b^i$	

$2^0 = 1 \quad 2^1 = 2 \quad 2^2 = 4 \quad 2^3 = 8 \quad 2^4 = 16 \quad 2^5 = 32 \quad 2^6 = 64 \quad 2^7 = 128 \quad 2^8 = 256 \quad 2^9 = 512 \quad 2^{10} = 1024$

Defining sets

To define sets:

To define a set using **roster method**, explicitly list its elements. That is, start with { then list elements of the set separated by commas and close with }.

To define a set using **set builder definition**, either form "The set of all x from the universe U such that x is ..." by writing
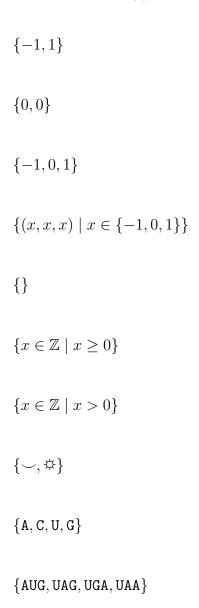
$$\{x \in U \mid ...x...\}$$

or form "the collection of all outputs of some operation when the input ranges over the universe U" by writing

$$\{...x... \mid x \in U\}$$

We use the symbol $\in$ as "is an element of" to indicate membership in a set.

Example sets: For each of the following, identify whether it's defined using the roster method or set builder notation and give an example element.

Can we infer the data type of the example element from the notation?

$\{-1, 1\}$

$\{0, 0\}$

$\{-1, 0, 1\}$

$\{(x, x, x) \mid x \in \{-1, 0, 1\}\}$

$\{\}$

$\{x \in \mathbb{Z} \mid x \geq 0\}$

$\{x \in \mathbb{Z} \mid x > 0\}$

$\{\smile, \maltese\}$

$\{\mathtt{A}, \mathtt{C}, \mathtt{U}, \mathtt{G}\}$

$\{\mathtt{AUG}, \mathtt{UAG}, \mathtt{UGA}, \mathtt{UAA}\}$

Rna motivation

RNA is made up of strands of four different bases that encode genomic information in specific ways. The bases are elements of the set $B = \{\mathtt{A}, \mathtt{C}, \mathtt{U}, \mathtt{G}\}$. Strands are ordered nonempty finite sequences of bases.

Formally, to define the set of all RNA strands, we need more than roster method or set builder descriptions.

Set recursive examples

Definition The set of nonnegative integers $\mathbb{N}$ is defined (recursively) by:

Basis Step:
Recursive Step:

Examples:

Definition The set of all integers $\mathbb{Z}$ is defined (recursively) by:

Basis Step:
Recursive Step:

Examples:

Definition The set of RNA strands S is defined (recursively) by:

Basis Step: $\quad \mathtt{A} \in S, \mathtt{C} \in S, \mathtt{U} \in S, \mathtt{G} \in S$
Recursive Step: $\quad$ If $s \in S$ and $b \in B$, then $sb \in S$

where sb is string concatenation.

Examples:

Definition The set of bitstrings (strings of 0s and 1s) is defined (recursively) by:

Basis Step:
Recursive Step:

Notation: We call the set of bitstrings $\{0,1\}^*$ and we say this is the set of all strings over $\{0,1\}$.

Examples: