

Introduction à la programmation

Timothée Poisot

Résumé

L'objectif de cette séance est l'initiation à la programmation. Les principales structures d'algorithmes et les opérateurs logiques seront abordés.

Table des matières

1	Programmer, pourquoi ?	1
2	Boucles	2
2.1	Boucles de type <i>for</i>	2
2.2	Boucles de type <i>while</i>	3
3	Tests	4
3.1	Expressions conditionnelles	4
3.2	Manipulation des valeurs booléennes	4
4	Fonctions	4
4.1	Généralités	5
4.1.1	Tout est fonction	5
4.1.2	Le <i>scope</i>	5
4.2	La commande <i>return</i>	5
5	Mise en application	5
5.1	Test par permutation	5
6	Solution des mises en application	7
6.1	Test par permutation	7

1 Programmer, pourquoi ?

Dans les séances précédentes, nous avons utilisé des fichiers `.R` pour sauvegarder des listes d'instructions. Nous avons aussi chargé et manipulé des jeux de données. Il est souvent nécessaire d'automatiser tout ou partie de ce processus, ce qui implique de faire appel à de la programmation.

L'objectif de cette séance est de

2 Boucles

Les boucles permettent parcourir une liste, ou de répéter une série d'instructions, dans des conditions bien définies ; c'est une des structures de base de l'algorithmique. R propose deux types de boucles, les boucles `for` et les boucles `while`. En français, on peut les résumer par “jusqu’à ce que” et “tant que”.

2.1 Boucles de type *for*

Une boucle `for` permet de répéter un bloc d'instructions un nombre prédéfini de fois, ou d'exécuter des commandes pour chaque élément d'un tableau de données. La syntaxe de base est la suivante :

```
> for (step in c(1:10)) cat(step)
```

```
12345678910
```

On peut bien spécifier plusieurs instructions qui doivent être exécutées à chaque *itération* (étapes de la boucle) en utilisant les accolades :

```
> for (step in c(1:3))
{
  cat(step)
  print(summary(rnorm(100,mean=step)))
}
```

```
1  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-1.6610 0.3997  1.0430  0.9598  1.4520  4.0090
2  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.9893 1.5040  2.1830  2.1170  2.7960  4.1410
3  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.7288  2.3870  3.0350  3.0230  3.5850  5.5500
```

Cette commande affiche le numéro de l'itération en cours (`cat(step)`), puis affiche les informations de base (`summary`) sur une distribution normale (`rnorm`) centrée sur `step`. Les boucles `for` peuvent contenir des instructions aussi longues que souhaité.

Une autre application des boucles `for` est de parcourir un objet. Par exemple, on peut souhaiter, pour chaque élément d'un objet, afficher sa valeur. R permet de réaliser ce genre d'opérations, avec la syntaxe suivante :

```
> vect = c('a', 'b', 'c', 'd')
> for (val in vect) cat(val)

abcd
```

Pour chaque élément du vecteur `vect`, que l'on nomme `val` pour pouvoir y accéder pendant les itérations, R va afficher la valeur que l'élément contient.

2.2 Boucles de type *while*

Les boucles de type `while`, littéralement *pendant*, permettent de répéter une série d'instructions tant qu'une condition n'a pas été atteinte. Pour cette raison, il faut bien prendre en compte le fait que mal utilisées, ces boucles peuvent ne jamais stopper. Il faut donc faire particulièrement attention à la condition qui est évaluée à chaque itération.

Un exemple simple d'utilisation d'une boucle `while` est le calcul d'une factorielle. On veut calculer $n!$, ce qui se fait simplement en multipliant l'ensemble des $1 \leq k \leq n$.

```
> n = 5
> k = n
> while(k > 1){
  k = k - 1
  n = n*k
}
> print(n)

[1] 120
```

On remarquera que dans la parenthèse après `while` se trouve un test logique ; les tests sont abordés dans la partie suivante.

3 Tests

3.1 Expressions conditionnelles

```
> a = 4
> if (a == 4){
  print('yep')
}
[1] "yep"
```

3.2 Manipulation des valeurs booléennes

4 Fonctions

L'utilisation des fonctions va permettre de gagner du temps dans la programmation. Comprendre le principe des fonctions dépasse de beaucoup le cadre de R, et mérite qu'on s'y arrête. Qu'est-ce qu'une fonction ? Une série d'instructions qui vont, à partir d'arguments, renvoyer un résultat. En quoi est-ce différent des scripts que nous avons utilisé jusqu'ici ? Écrire une fonction revient en quelque sorte à 'expliquer' le code une fois, et R se charge ensuite de redonner la bonne valeur aux arguments.

Le parallèle le plus évident est celui des fonctions mathématiques : si $f(x) = x^2$, on peut calculer $f(x)$ pour tout x , parce qu'on sait quoi faire. Ça devient donc très avantageux si on doit calculer $f(x)$ un grand nombre de fois. En écrivant uniquement des scripts, pour calculer la valeur de beaucoup de x^2 , on aurait écrit :

```
> 1^2
> 2^2
> 3^2
> 4^2
> 5^2

ou encore

> for(i in c(1:5)) i^2
```

Si il faut revenir sur ce code plus tard, et transformer tous les 2 en 3 , la première solution implique de tout corriger manuellement. En utilisant une fonction, la logique est différente :

```
> f = function(x) x^2  
> f(2)
```

```
[1] 4
```

Peut importe le nombre de fois ou on devra effectuer l'opération contenue dans `f`, si on veut la modifier, elle sera toujours stockée au même endroit. Avec cet avantage en tête, quelques généralités sur les fonctions en R.

4.1 Généralités

4.1.1 Tout est fonction

4.1.2 Le scope

4.2 La commande *return*

5 Mise en application

5.1 Test par permutation

Lorsque les données devient de la normalité, on peut préférer réaliser un test paramétrique avec des permutations plutôt qu'un test non paramétrique. La plupart des programmes de statistique n'offrent pas cette possibilité qui demande pourtant très peu d'efforts pour être implémentée dans R. Dans cette mise en application, on veut effectuer un test `t`, pour comparer deux distributions, disponibles dans un fichier `s4-data.txt`.

Le principe d'un test par permutations est simple. La première étape est d'effectuer le test sur l'échantillon non permuté, pour obtenir la valeur de la statistique (T). Dans le cas du test `t`, R propose la fonction `t.test`, et un rapide survol de `?t.test` vous donnera les arguments nécessaires et la manière de récupérer la statistique. Commencent ensuite les permutations a proprement parler. Pour un nombre n d'itérations choisies (en général 9999), on mélange l'ensemble des valeurs des deux distributions. On reconstruit ensuite, en tirant au hasard dans le *pool* de valeurs ainsi formées, deux distributions de taille égale. Cette étape peut, par exemple, prendre la forme d'une fonction `resample`, qui prendrait une `data.frame` avec deux colonnes (la valeur, et le groupe d'origine) en argument. R propose la fonction `sample`, qui permettra de mélanger la colonne correspondant au groupe (ce qui recréera automatiquement les deux distributions – économisons nous!). Une fois les deux distributions reconstruites, on calcule la nouvelle statistique T' . Si la valeur de T' est

inférieure à la valeur de T , on incrémente une variable N de 1. Sinon, la valeur de N reste la même.

Le calcul de la p -value se fait de la manière suivante :

$$p = \frac{N}{n} \quad (1)$$

À partir de ces informations, et des informations données dans l'introduction de cette séance, vous devez être en mesure de programmer sans difficultés une fonction `t.test.permut`, qui réaliser un test `t` par permutations. En bonus, vous pouvez ajouter des arguments qui permettent de contrôler le nombre de permutations qui doivent être réalisées.

6 Solution des mises en application

6.1 Test par permutation