

Initiation à R

Romain François¹ et Jean-Michel Marin²

¹ Mango Solutions, Chippenham, Grande-Bretagne

² Projet SELECTINRIA Futurs, Laboratoire de Mathématiques,
Université Paris-Sud, France

Auteur correspondant : Jean-Michel.Marin@inria.fr

Résumé Ce document a pour objectif de familiariser son lecteur avec le langage et l'environnement de programmation R dans sa version 2.4.1 (décembre 2006). Il ne constitue pas une référence complète mais plutôt un aperçu des capacités de R et un point d'entrée vers d'autres documents plus complets. À plusieurs reprises, le lecteur est invité à exécuter les commandes proposées dans une session R afin de s'habituer à la syntaxe.

1 Introduction

R est un langage de programmation interactif interprété et orienté objet contenant une très large collection de méthodes statistiques et des facilités graphiques importantes. R est une implémentation libre (au sens GNU/GPL) du langage S Carlier (1991); Fayet (1991); Venables and Ripley (2002), conçu par les laboratoires BELL puis racheté par la société INSIGHTFUL en vue d'une commercialisation sous le nom de S-PLUS. R est né en 1995 des travaux de Robert Gentleman et Ross Ihaka (Département de Statistique, Université d'Auckland, Nouvelle-Zélande) pour être ensuite adopté par une communauté d'utilisateurs et de développeurs de plus en plus importante. Aujourd'hui, la *R Development Core Team* est composée d'une vingtaine de personnes, principalement des chercheurs, qui contribuent à l'amélioration et la documentation des puissantes fonctionnalités de R. En quelques années, grâce notamment à son modèle de diffusion basé sur la liberté, R est devenu un laboratoire incontournable des nouvelles idées du paysage statistique mondial. L'une des forces de R est son extensibilité, sous forme de packages réalisés par les utilisateurs à des fins bien précises. Le CRAN (Comprehensive R Archive Network, <http://cran.r-project.org/>) recense aujourd'hui plus de 700 de ces packages, couvrant ainsi largement le spectre de la statistique actuelle.

La documentation sur R est très importante, parfois un peu éparpillée, et souvent difficile à appréhender au premier essai. Différents points d'entrée sont recensés sur site internet central de R : <http://www.r-project.org/>, on peut notamment citer R-Development-Core-Team (2006a,c,b). Nous conseillons la lecture de Broutaux (2002), Dalgaard (2002), Venables and Ripley (2002), Verzani (2005) qui constituent d'excellentes références sur R (ou S au sens large). Les ouvrages de Dalgaard (2002) et Verzani (2005) donnent de nombreux exemples de traitement de données à l'aide de R. Concernant l'impressionnante interface graphique qu'est R, le lecteur peut se référer à Maindonald and Braun (2003). Notons que R est depuis peu doté d'un système collaboratif de documentation, permettant à tout un chacun de partager ses ruses avec le reste de la communauté : <http://wiki.r-project.org>, et d'une galerie de graphiques : <http://addictedtor.free.fr/graphiques>.

Le but de cet article est de donner un bref aperçu des capacités de R en terme de