Devoir 1 : Matrices aléatoires et probabilité libres

Patrick Fournier

9 septembre 2019

Les probabilités libres (free probabilities) sont définies comme étant l'étude des variables aléatoires non commutatives. Pour ce devoir, nous nous limiterons aux matrices aléatoires à valeur dans $\mathbb{R}^{n \times n}$. Répondez à chaque question par une fonction. Chaque fonction devrait être relativement courte (10 lignes maximum). Soyez rusé !

Question 1

Fabienne est une mathématicienne aguerrie ainsi qu'une adepte de la programmation fonctionnelle. Son collègue Igor, pour sa part, est plus expérimenté en matière de programmation impérative. Lors des longues soirées froides d'hiver passées ensemble, ils se laissent souvent porter par leurs instincts et se lancent des défis à la frontière des mathématiques et de l'informatique. Aujourd'hui, ils s'intéressent à la distribution des valeurs propres des matrices de la forme

$$M = \begin{pmatrix} x & y \\ y & x \end{pmatrix}$$

où $x, y \in 0, \ldots, 9$. Comme Igor a bu beaucoup de café depuis la fin de ses études, il a de la difficulté à se rappeler son cours d'algèbre. Fabienne lui rappelle que les valeurs propres d'une matrice M sont les racines du polynôme en λ

$$|M - \lambda Id|.$$

Igor est rassuré.

a)

Igor, d'un tempérament explosif, se lance sur son ordinateur. Il ouvre R et écrit une fonction calculant *exactement* la distribution conjointe des deux valeurs propres. Pour se faire, il fait appel à deux boucles `for` imbriquées et à la fonction `eigen`.

b)

Fabienne, de son côté, commence par calculer explicitement les valeurs propres de M comme fonction de x et y. Par la suite, apercevant le travail de Igor, elle lui suggère qu'il n'a pas réellement besoin de `eigen`.

c)

Un peu froissé, Igor lance un défi à Fabienne : écrire une fonction plus rapide que la sienne. Après un moment de réflexion, elle va lire la documentation de `outer`. Comme elle se souvient qu'elle peut passer un opérateur infixe en argument en l'encadrant de backticks (‘-‘, ‘+‘), elle procède sans crainte.

d)

Les trois fonctions s'exécutent tellement rapidement que Fabienne et Igor n'arrivent pas à déterminer laquelle est la plus rapide. Pouvez-vous les départager ?

e)

Afin de visualiser le fruit de leurs efforts, Fabienne suggère de représenter la distribution graphiquement. Igor, un peu vieux jeu, se met à les dessiner à la main. Fabienne, sourire en coin, lui suggère plutôt de lire la documentation des fonctions `pdf`, `plot` et `dev.off`.

Question 2

Satisfaits de leur travail sur la matrice précédente, Fabienne et Igor décident de pousser les choses plus loin. Pour passer le temps, ils décident de vérifier par simulation certains résultats portant sur l'ensemble gaussien orthogonal (EGO). Fabienne a un blanc de mémoire et demande à Igor de lui rappeler la définition de l'EGO. Il lui répond qu'il s'agit du sous-espace de $\mathbb{R}^{n \times n}$ contenant les matrices symétriques à entrées i.i.d. gaussiennes. Pour simplifier les choses, on ne considère que le cas centré réduit.

a)

Fabienne écrit tout d'abord une fonction lui permettant de simuler un élément de l'EGO de taille arbitraire. Pour se faire, elle fait un usage astucieux des fonctions `lower.tri` et `diag`.

b)

Igor, qui est d'un naturel curieux, se demande quelle partie de la fonction de sa collègue est la plus demandante d'un point de vue computationnel. Ne reculant devant aucun effort, il profile son code.

c)

Comme les matrices dans EGO possèdent certaines caractéristiques, Fabienne décide de spécialiser la fonction `eigen` afin d'en extraire efficacement les valeurs propres et, optionnellement, de les normaliser par $2\sqrt{n}$.

d)

Fabienne désire vérifier le théorème limite central libre selon lequel la distribution des valeurs propres des matrices appartenant à EGO converge vers la distribution du demi-cercle dont la densité est donnée par

$$f(x) = \frac{2}{\pi}\sqrt{1 - x^2}.$$

Pour ce faire, elle procède comme suit :

for $n \in \{10, 100, 1000, 10000\}$ **do**
    Simuler $M_{n \times n} \in$ EGO.
    Calculer le vecteur de valeurs propres Λ.
    Faire un histogramme de Λ normalisé par $2\sqrt{n}$.
    Surimposer la densité de la distribution du demi-cercle.
end for

e)

Igor désire procéder plus rigoureusement. Habile intégrateur, il calcule que la fonction de répartition de la distribution du demi-cercle est

$$F(x) = \frac{1}{2} + \frac{1}{\pi}(x\sqrt{1 - x^2} + \sin^{-1} x)$$

Par la suite, pour les mêmes valeurs de n, il génère 10 matrices de EGO. Pour chaque matrice, il compare la distribution des valeurs propres à celle de la distribution du demi-cercle à l'aide de la statistique de Kolmogorov-Smirnov. Il calcule ensuite la moyenne de ces 10 statistiques. Afin de faire cela efficacement, il fait appel à la fonction `ecdf`.