

EXAMEN : INTRODUCTION AUX LOGICIELS MATHÉMATIQUES

Seuls les notes prises en TD et les documents distribués en TD sont autorisés.

Exercice 1 MATLAB :

Soit f une fonction continue sur un intervalle $[a, b]$. Pour un entier n non nul donné, soit $h = \frac{b-a}{n}$. On construit la subdivision uniforme $x_i = a + (i-1)h$ pour $i = 1, \dots, n+1$ et on calcule les ordonnées $y_i = f(x_i)$ correspondantes. L'objectif est d'approcher $I = \int_a^b f(t)dt$ par la formule des trapèzes

$$J_1(n) = \frac{h}{2} \left[y_1 + y_{n+1} + 2 \sum_{k=2}^n y_k \right].$$

1. Ecrire un fichier function, `f.m` qui calcule les valeurs de la fonction f . Exemple $f(x) = e^x$.

```
>> f([0.5 1])
ans =
    1.6487    2.7183
```

2. Ecrire un fichier function, `trap.m`, qui utilise la fonction f sous forme de fichier et permet le calcul approché de l'intégrale par `trap(namefun,a,b,n)`. Test avec $a = 0$, $b = 1$, $n = 10$. *boucle interdite!*

```
>> trap('f',0,1,10)
ans =
    1.7197
```

3. En faisant varier n , étudier $\log |J_1(n) - I|$ en fonction de $\log(n)$. Dans ce programme, on tracera en particulier la courbe correspondante. $a = 0$, $b = 1$, $f(x) = e^x$. Pour cette étude, on partira d'un tableau `arrn = [1, 2, 5, 7, 10, 15, 20, 35, 50, 75, 100]`, on construira le tableau correspondant `arrerr` puis on dessinera $\log(arrerr)$ en fonction de $\log(arrn)$. Sauvegarde dans `VI1.m` où `VI` sont vos initiales, en indiquant en commentaire l'ordre de la méthode, c'est à dire le coefficient α tel que $erreur \simeq C/n^\alpha$.

4. On utilise maintenant une approximation de l'intégrale par une quadrature de Gauss

$$J_2(n) = \frac{h}{2} \left[\sum_{i=1}^n (f(x_i + h_1) + f(x_i + h_2)) \right], \quad h_1 = \frac{h}{2} \left(1 - \frac{1}{\sqrt{3}} \right), \quad h_2 = \frac{h}{2} \left(1 + \frac{1}{\sqrt{3}} \right).$$

Ecrire un fichier function, `gauss.m`, qui utilise la fonction f et permet le calcul approché de l'intégrale par `gauss(namefun,a,b,n)`. Test avec $a = 0$, $b = 1$, $n = 10$. *boucle interdite!*

5. En faisant varier n , étudier $\log |J_2(n) - I|$ en fonction de $\log(n)$ en partant du tableau `arrn = [1, 2, 5, 7, 10, 15, 20, 35, 50, 75, 100]`. Sauvegarde du programme sous `VI2.m`

6. On souhaite accélérer la convergence en définissant la nouvelle approximation

$$J_3(n) = \frac{16J_2(2n) - J_2(n)}{15}.$$

Ecrire un programme `VI3.m` qui calcule ces nouvelles approximations pour différentes valeurs de n et étudier l'ordre de la méthode à partir du tableau `arrn = 10 : 20`. Autres données identiques à la question précédente. *Cette méthode est une extrapolation à la limite.*

7. **Bonus :** En reprenant les questions 2-3, à partir de la méthode des trapèzes, peut-on trouver une méthode d'approximation de l'intégrale, $J_4(n)$, accélérant la convergence ? Si oui, proposez un programme `VI4.m` avec étude d'erreur et ordre amélioré.

Exercice 2 Logiciel R

- Votre code R sera enregistré dans un seul script sous le nom `R_NOM_PRENOM.r`.
- Les questions 1, 2 et 3 sont indépendantes.

Pour commencer, copier-coller les fichiers `vols.txt`, `compagnies.csv` et `aeroports.txt` situés dans le répertoire public `X:\GM-345\GM 3A\IntroLogiciels\TPnote` dans votre répertoire de travail.

1. Trafic aérien en partance de New-York en 2013

Nous allons étudier un jeu de données décrivant les vols en partance des 3 aéroports de New York en 2013. Ces données proviennent du Bureau des Statistiques sur les transports des États-Unis. Le jeu de données `vols` comporte 336 776 observations et 15 variables. Le descriptif des variables est donné ci-dessous :

<code>month</code>	mois où le vol était programmé
<code>day</code>	jour du mois où le vol était programmé
<code>dep_time</code>	heure effective de départ du vol (format HHMM ou HMM)
<code>sched_dep_time</code>	heure programmée du départ du vol (format HHMM ou HMM)
<code>dep_delay</code>	différence entre l'heure effective et l'heure programmée du départ du vol en minutes
<code>arr_time</code>	heure effective d'arrivée du vol (format HHMM ou HMM)
<code>sched_arr_time</code>	heure programmée d'arrivée du vol (format HHMM ou HMM)
<code>arr_delay</code>	différence entre l'heure effective et l'heure programmée de l'arrivée du vol en minutes
<code>carrier</code>	code transporteur de la compagnie affrétant le vol
<code>flight</code>	numéro du vol
<code>tailnum</code>	numéro de l'avion ayant effectué le vol
<code>origin</code>	identifiant de l'aéroport de partance du vol
<code>dest</code>	identifiant de l'aéroport de destination du vol
<code>air_time</code>	durée du vol en minutes
<code>distance</code>	distance parcourue par l'avion en miles

Vous disposez également de deux autres fichiers de données.

- Le fichier `compagnies.csv` comprend 2 variables `carrier` et `name` qui donnent respectivement le code transporteur des compagnies et le nom des compagnies.
- Le fichier `aeroports.txt` comprend 4 variables `faa`, `name`, `lat` et `lon` qui donnent respectivement l'identifiant de l'aéroport, son nom, sa latitude et sa longitude. Dans ce fichier, les données manquantes sont codées par "?".

- (a) Lire les 3 fichiers `vols.txt`, `compagnies.csv` et `aeroports.txt` et stocker leur contenu dans des data-frames nommés respectivement `vols`, `compagnies` et `aeroports`.
- Attention**, pour la lecture du fichier `vols.txt`, vous utiliserez l'argument suivant dans votre fonction afin que les variables `month`, `day` et `flight` soient reconnues comme des facteurs : `colClasses=c(month="factor", day="factor", flight="factor")`
- (b) Déterminer le pourcentage de vols en partance de New York annulés en 2013.
Les vols annulés sont les vols dont l'heure effective de départ est manquante.
- (c) Supprimer du tableau `vols` les lignes contenant des données manquantes.
- (d) Construire un tableau `retardes` contenant les observations concernant les vols qui sont arrivés avec plus de deux heures de retard alors qu'ils étaient partis à l'heure.
- (e) Calculer la durée moyenne des vols en heures depuis New York pour chacune des destinations.
- (f) Ajouter au tableau `vols` une variable `vitesse` donnant la vitesse moyenne de chaque vol en km/h (1 mile = 1,60934 km), puis afficher dans la console les 5 vols ayant atteint les plus grandes vitesses moyennes.
- (g) Fusionner les tableaux `vols`, `compagnies` et `aeroports` afin d'obtenir un tableau plus complet contenant toutes les variables du tableau `vols`, le nom de la compagnie ayant affrété le vol, le nom de l'aéroport de départ et le nom de l'aéroport de destination.

2. Ultratrifoliophilie

- (a) Écrire une fonction `rotation` qui prend comme arguments l'angle de la rotation `angle` et une matrice `coord` contenant les coordonnées des points dont on veut calculer l'image par la rotation, et qui rend en sortie une matrice de même dimension que `coord` contenant les coordonnées des images des points par la rotation.

Rappel : La matrice de rotation d'angle α est définie par
$$\begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}.$$

- (b) On considère la courbe paramétrée Γ suivante :

$$\begin{cases} x(\theta) = \rho(\theta) \cos \theta \\ y(\theta) = \rho(\theta) \sin \theta \end{cases} \text{ où } \rho(\theta) = 1 + \cos(n\theta) + \sin^2(n\theta), \quad n \in \mathbb{N}^*, \quad \theta \in \mathbb{R}.$$

Écrire un fonction `courbe_par` qui prend comme arguments un vecteur `theta` qui contient les différentes valeurs de θ pour lesquelles l'on souhaite calculer les coordonnées des points $(x(\theta), y(\theta))$ et le paramètre `n`, et qui rend une matrice contenant les coordonnées des points $(x(\theta), y(\theta))$.

- (c) Vous allez devoir représenter un graphique pour lequel les ordonnées sont comprises entre -5 et 3, le ratio y/x est conservé (`asp=1`) et dont le titre est "Jour de chance".

Sur ce graphique,

- représenter en vert, avec un trait de largeur égale à 2, la courbe représentative de la fonction $f(x) = -1,41x^3 - 2,74x^2 + 2,37x$ pour $x \in [-1, 5; 0]$;
- représenter en marron la droite d'équation $y = -4,97$;
- représenter en vert la courbe paramétrée Γ avec θ variant entre $[0; 2\pi]$ et $n = 4$ après lui avoir fait subir une rotation d'angle $\pi/6$;
- colorier en vert l'aire de la surface délimitée par la courbe paramétrée Γ ainsi obtenue ;
- ajouter le texte suivant sur le graphique à l'emplacement de votre choix : "Je suis devenu un quadrifoliiste."

3. Test statistique du khi-deux

Soit $\mathbf{N} = (n_{ij})$ une matrice réelle non nulle de taille $p \times q$ ($p, q > 1$). On note $n_{i\bullet}$ la somme des termes de la $i^{\text{ème}}$ ligne, $n_{\bullet j}$ la somme des termes de la $j^{\text{ème}}$ colonne et n la somme des termes de la matrice.

Soit \mathbf{W} la matrice des contributions associée à \mathbf{N} , de taille $p \times q$, de terme général w_{ij} défini par :

$$w_{ij} = \frac{\left(n_{ij} - \frac{n_{i\bullet}n_{\bullet j}}{n}\right)^2}{\frac{n_{i\bullet}n_{\bullet j}}{n}}.$$

Soit s la statistique de test du khi-deux égale à $s = \sum_{i=1}^p \sum_{j=1}^q w_{ij}$.

On appelle probabilité critique du test la quantité $1 - F(s)$ où F désigne la fonction de répartition de la loi du khi-deux à $(p - 1) \times (q - 1)$ degrés de liberté $\chi_{(p-1)(q-1)}^2$.

Soit α un nombre réel compris entre 0 et 1 représentant le risque du test statistique.

- Écrire une fonction `contribution` qui prend comme argument une matrice `N` et qui retourne la matrice des contributions `W` associée à `N`.
- Écrire une fonction `test_khi2` qui prend comme arguments une matrice `N` et le risque `alpha` du test. Cette fonction doit rendre la valeur de la statistique de test du khi-deux `s` associée à la matrice `N`, la probabilité critique du test et une chaîne de caractères qui prend la valeur `H0` si la probabilité critique du test est supérieure ou égale au risque `alpha` et la valeur `H1` sinon (Vous utiliserez la fonction `pchisq`).
- Créer la matrice `Eff` ci-dessous et lui appliquer la fonction `test_khi2` en prenant un risque `alpha` égal à 0,05.

$$\text{Eff} = \begin{pmatrix} 400 & 1380 \\ 416 & 1823 \\ 188 & 1168 \end{pmatrix}$$