MATHEMATIQUES
Analyse-Statistiques-Dénombrements

Le sujet se compose d'un exercice et de deux problèmes auxquels on pourra consacrer respectivement une demi-heure, une heure et une heure et demi.
Il sera tenu compte de la présentation et en particulier de $\boxed{\textit{l'encadrement}}$ *des résultats.*
L'usage de la calculatrice ***est*** *autorisé au cours de l'épreuve.*

Exercice :

Pour tout entier n on définit f_n par $f_n(1) = \dfrac{1}{2}$ et $\forall x \in \mathbb{R}_+^* \setminus \{1\}$ $f_n(x) = \dfrac{x^n lnx}{x^2-1}$
On notera $(\mathcal{C}_n)$ la courbe représentative de f_n dans un repère orthonormé.

① a) Donner le développement limité à l'ordre 3 au voisinage de 0 de $u \longmapsto ln(1+u)$ et montrer que $\dfrac{1}{2+u} \underset{0}{=} \dfrac{1}{2} \cdot \left(1 - \dfrac{u}{2} + \dfrac{u^2}{4} + o(u^2)\right)$.

b) Donner le développement limité à l'ordre 2 au voisinage de 1 de f_0 puis de f_n pour $n \geq 1$.

c) Montrer que, pour $n \in \mathbb{N}$, f_n est dérivable en 1. Que vaut $f_n'(1)$?

d) Préciser la position, au voisinage de 1, de $(\mathcal{C}_n)$ par rapport à sa tangente au point d'abscisse 1 (discuter suivant les valeurs de n).

② a) Montrer que, pour un entier naturel n fixé, f_n est dérivable sur $\mathbb{R}_+^*$ et calculer $f_n'(x)$.

b) Montrer que, pour $n \geq 1$, f_n est prolongeable par continuité en une fonction g_n qu'on précisera. Pour quelles valeurs de n ce prolongement est-il dérivable à droite en 0 ?

③ a) Étudier φ_0 définie sur $\mathbb{R}_+^*$ par $\varphi_0(x) = 1 - \dfrac{1}{x^2} - 2lnx$. Quel est le signe de $f_0'(x)$?

b) En déduire les variations de f_0 et tracer $(\mathcal{C}_0)$ (on choisira en abscisse une unité de 5 cm)

Problème 1 :

1. Étude de la somme d'une série

Pour tout entier naturel $n \geq 1$, on pose :

$$S_n = \sum_{k=1}^{n} \frac{1}{k^2}$$

① Démontrer que $\displaystyle\sum_{n\geq 2} \frac{1}{n(n-1)}$ converge et en déduire que la série $\displaystyle\sum_{n\geq 1} \frac{1}{n^2}$ converge.

Dans la suite du problème, on notera $l = \displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2}$.

② Écrire une fonction Python `sommeP(n)` qui prend en argument n et qui renvoie la valeur de S_n.

③ a) Soit $f : t \longmapsto \dfrac{1}{t^2}$.
Pour tout entier $k \geq 2$, représenter graphiquement la courbe représentant f sur l'intervalle $[k, k+1]$ et montrer que $\forall x \in [k, k+1],\ f(t) \leq f(k)$. En déduire que $\displaystyle\int_k^{k+1} \frac{1}{t^2}dt \leq \frac{1}{k^2}$.
Montrer plus généralement que pour tout entier $k \geq 2$, on a :

$$\int_k^{k+1} \frac{1}{t^2}dt \leq \frac{1}{k^2} \leq \int_{k-1}^{k} \frac{1}{t^2}dt$$

b) Soit p et n deux entiers naturels non nuls.
Démontrer que :

$$\int_{n+1}^{n+p+1} \frac{1}{t^2}dt \leq S_{n+p} - S_n \leq \int_n^{n+p} \frac{1}{t^2}dt$$

c) Démontrer alors que :

$$\frac{1}{n+1} \leq l - S_n \leq \frac{1}{n}$$

puis l'équivalent :

$$l - S_n \underset{n\to\infty}{\sim} \frac{1}{n}$$

Que peut-on dire de $S_n + \dfrac{1}{n}$ pour n suffisamment grand ?

④ On considère la fonction suivante :

```
def estimation(n):
   s = sommeP(n)
   s = sqrt(6*(s+1/n))
   return (s)
```

L'instruction `estimation(1000)` renvoie la valeur : 3.141593130895433.
Que peut-on conjecturer quant à la probable valeur de l ?

2. Étude de la nature d'une série

Pour tout entier naturel $n \geq 1$, on pose :

$$u_n = \prod_{k=1}^{n} \left(1 + \frac{1}{k^2}\right) = (1+1)\left(1 + \frac{1}{4}\right)\cdots\left(1 + \frac{1}{n^2}\right)$$

① Donner, sous forme d'entiers ou de fractions simplifiées, les valeurs de u_1, u_2 et u_3.

② a) Montrer que, pour tout entier naturel $n \geq 1$, on a : $u_n \geq 2$.

b) Étudier la monotonie de la suite (u_n).

③ a) Montrer que, pour tout réel x strictement supérieur à -1, on a : $\ln(1+x) \leq x$.

b) Démontrer alors que, pour tout $n \in \mathbb{N}^*$, on a :

$$\ln(u_n) \leq l \qquad \text{(où } l = \sum_{k=1}^{\infty} \frac{1}{k^2} \text{ a été estimée dans la partie 1.)}$$

④ En utilisant les questions précédentes, montrer que la suite (u_n) converge vers un réel l', élément de $[2, e^l]$.

⑤ On se propose dans cette question de déterminer la nature de la série de terme général $l' - u_n$.

a) Justifier que la suite $(\ln(u_n))_{n\in\mathbb{N}^*}$ converge et que l'on a :

$$\ln(l') = \sum_{k=1}^{+\infty} \ln\left(1 + \frac{1}{k^2}\right)$$

b) Montrer que, pour tout $n \in \mathbb{N}^*$, on a :

$$\ln\left(\frac{l'}{u_n}\right) = \sum_{k=n+1}^{+\infty} \ln\left(1 + \frac{1}{k^2}\right)$$

c) Écrire en Python une fonction `reste` qui prend en argument n et qui renvoie une bonne valeur approchée de :

$$\sum_{k=n+1}^{\infty} \ln\left(1 + \frac{1}{k^2}\right)$$

d) Voici les résultats de quelques appels de cette fonction (supposée correctement écrite) :

```
> > > reste(9)
0.10497285546814518

> > > reste(49)
0.020198959583845037

> > > reste(99)
0.010048997486379833

> > > reste(199)
0.0050114998428795786

> > > reste(499)
0.0020009999955394315

> > > reste(999)
0.0009994999929165772
```

Que peut-on alors conjecturer comme équivalent en l'infini très simple de $\displaystyle\sum_{k=n+1}^{\infty} \ln\left(1 + \frac{1}{k^2}\right)$?

e) En admettant le résultat conjecturé dans la question précédente, justifier qu'il existe un entier naturel n_0 (qu'on ne cherchera pas à calculer) tel que, pour tout $n \geq n_0$, on a :

$$\sum_{k=n+1}^{\infty} \ln\left(1 + \frac{1}{k^2}\right) \geq \frac{0,9}{n}$$

f) En déduire que pour tout $n \geq n_0$, on a :

$$l' - u_n \geq u_n(e^{0,9/n} - 1)$$

g) On admet que pour tout réel x, on a : $e^x - 1 \geq x$.
Démontrer alors que :

$$\forall n \geq n_0,\ l' - u_n \geq \frac{1,8}{n}$$

Conclure quant à la nature de la série de terme général $l' - u_n$.

Problème 2 :

Les estimations et statistiques concernant la population des Etats-Unis de 1790 à 1950 donnent les chiffres suivants (exprimés en millions d'individus) :

dates t_i	1790	1800	1810	1820	1830	1840	1850	1860	1870
population y_i	3.929	5.308	7.240	9.638	12.866	17.069	23.192	31.443	38.558
dates t_i	1880	1890	1900	1910	1920	1930	1940	1950	
population y_i	50.156	62.498	75.995	91.972	105.711	122.775	131.669	150.697	

Nous noterons dans la suite $n = 17$.

Préambule : On supposera que la commande `import numpy as np` a été exécutée mais toutes les fonctions Python demandées dans ce problème devront être rédigées **sans utiliser** les fonctions suivantes : `np.sum(), np.mean(), np.std(), np.cov(), np.polyfit()`

① Initialiser en langage Python un tableau `P` à deux entrées, la première ligne contenant les dates, la seconde contenant le nombre d'habitants de telle façon que, par exemple, `P[0,4]=1830`, `P[1,9]=50.156` est la population des USA en millions d'habitants en 1880.

② Écrire une fonction `moyenne(P,i)` qui prend en argument d'entrée un tableau `P` de deux lignes tel que celui défini précédemment ainsi qu'un entier `i` à valeur dans $\{0, 1\}$ et qui retourne la moyenne des valeurs de la ligne i de P (à titre d'exemple `moyenne(P,1)` retourne la taille moyenne de la population entre 1790 et 1950).

Partie I :

Nous supposons que cette population a suivi sur ces décennies une loi logistique, c'est-à-dire que pour tout temps t compris entre 1790 et 1950, la valeur théorique $y(t)$ (exprimée elle aussi en millions d'individus) du nombre d'habitant des USA à l'instant t et donné par la formule :

$$y(t) = \frac{K}{1 + Cexp(-rt)}, \; \forall t \in [1790, 1950]$$

où K, C, et r sont des constantes réelles strictement positives. Le but de ce problème est de déterminer ces constantes à l'aide des seules données expérimentales ci-dessus.

① Montrer, en justifiant les différentes étapes, que :

$$\forall t \in [1790, 1950],\; -rt + \ln(C) = \ln\left(\frac{K - y(t)}{y(t)}\right)$$

Dans la suite, nous noterons $\begin{cases} \alpha & = -r \\ \beta & = \ln(C) \end{cases}$

② Dans cette partie, nous supposons connue la population limite K. Il s'agit donc de déterminer les constantes α et β et pour cela nous utilisons la méthode des moindres carrés dont nous rappelons le principe :
L'objectif est de déterminer les valeurs de α et β telles que les écarts entre les valeurs mesurées $z_i = \ln\left(\frac{K - y_i}{y_i}\right)$ et les valeurs théoriques $\ln\left(\frac{K - y(t)}{y(t)}\right)$ aux mêmes instants t_i ($i \in [\![0, n-1]\!]$) soient les plus petits possibles.

Il s'agit donc de déterminer les valeurs de α et β telles que la quantité :

$$E(\alpha, \beta) = \sum_{i=0}^{n-1} (\alpha t_i + \beta - z_i)^2$$

soit la plus petite possible.

a) En considérant que E est une fonction de deux variables en α et β, justifier qu'elle est de classe $\mathcal{C}^1$ sur $\mathbb{R}^2$ et rappeler une condition nécessaire sur $\dfrac{\partial E}{\partial \alpha}$ et $\dfrac{\partial E}{\partial \beta}$ pour que E admette un minimum en (α, β).

b) Montrer en résolvant le système obtenu qu'on obtient pour solution α^* et β^* qui sont les coefficients de la droite de régression du nuage de points $\{(t_i, z_i), i \in [\![0, n-1]\!]\}$.
On admettra par la suite que $E(\alpha^*, \beta^*)$ est bien un minimum.

c) Écrire une fonction `coeffRegression(P)` qui prend en entrée le tableau P, utilise la fonction `moyenne()` [exclusivement] du préambule et retourne les valeurs α^* et β^* qui minimisent E.

d) En supposant que $K = 195$, un appel à cette fonction : `alphaEtoile,betaEtoile= coeffRegression(P)` retourne `alphaEtoile`$= -0.031$ et `betaEtoile`$= 60.179$.
En déduire les valeurs de r et de C sous cette hypothèse.

Partie II :

α^* et β^* étant les valeurs qui minimisent la quantité $E(\alpha, \beta)$ définie en I/2., on note désormais $Err(L) = E(\alpha^*, \beta^*)$.
C'est une évaluation de l'erreur globale faite en fonction de la population limite L.

① Écrire une fonction Python `Err(P,L)` qui prend en argument d'entrée le tableau P et la population limite L et retourne $Err(L)$.

② Malheureusement nous ignorons la valeur de la population limite K. Nous savons seulement, sur la base des dernières valeurs, qu'elle est supérieure à 150.697 millions d'habitants. L'idée pour la déterminer consiste à minimiser l'erreur globale faite en fonction de K que nous avons notée $Err(K)$.
Pour cela nous partons d'une valeur arbitraire de K_0 ($K_0 = 160$ est une valeur plausible) et nous calculons successivement $K_j = K_0 + \dfrac{j}{10}$ et $Err(K_j)$, $j \in \mathbb{N}$.
Nous stopperons les itérations dès que :

$$Err(K_j) < Err(K_{j+1})$$

Nous estimons alors que la population limite K est comprise entre K_j et K_{j+1}.

a) Le critère d'arrêt suffit-il à justifier que le minimum de la fonction `Err` est compris entre K_j et K_{j+1} ?

b) Écrire un algorithme qui détermine la plus petite valeur de j telle que $Err(K_j) < Err(K_{j+1})$ et renvoie K_j.

c) On obtient le tracé suivant de $Err(L)$ en fonction de L. Estimez la valeur de K et l'effectif attendu pour la population des USA en 2016.
Que vous suggère comme commentaire le fait que la population américaine compte, lors du recensement de 2016, 324227000 habitants ?