2 BCPST 1 & 2 Samedi 26 mars 2016

MATHÉMATIQUES
Devoir surveillé n°6
Durée : 3 heures 30

L'usage d'une calculatrice est indispensable pour cette épreuve. Si, au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il le signale sur sa copie et poursuit sa composition en expliquant les raisons des initiatives qu'il a été amené à prendre.

Cet épreuve est composée de deux problèmes indépendants.

PROBLEME N°1

On rappelle que si S et T sont deux variables aléatoires indépendantes de densités respectives f_S et f_T, alors $S+T$ est une variable à densité dont une densité est la fonction f_{S+T} définie sur $\mathbb{R}$ par :

$$\forall t \in \mathbb{R},\ f_{S+T}(x) = \int_{-\infty}^{+\infty} f_S(t)\ f_T(x-t)\ \mathrm{d}t$$

Préliminaires

1. Soit U une variable aléatoire distribuée selon la Loi Uniforme sur $[0,1[$ et soit λ un réel strictement positif.
On note $X = -\dfrac{1}{\lambda}\ln(1-U)$
Vérifiez que X est distribuée selon la Loi Exponentielle de paramètre λ.

2. On note $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes, de même Loi que X.
On note alors $S_0 = 0$ et $S_n = X_1 + X_2 + \cdots + X_n$ pour tout entier naturel non nul n.
Montrer que pour tout entier naturel $n \neq 0$, S_n admet pour densité la fonction f_n définie par :

$$\forall t \in \mathbb{R},\ f_n(t) = \begin{cases} 0 & \text{si} \quad t < 0 \\ \dfrac{\lambda e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!} & \text{si} \quad t \geq 0 \end{cases}$$

Dans la suite de ce problème, nous nommerons F_k la fonction de répartition de la variable S_k

Le paradoxe de l'autobus

On suppose qu'à un arrêt d'autobus A, les différences entre les temps de passage successifs d'un autobus sont des **variables aléatoires indépendantes** de même Loi Exponentielle de paramètre λ. Le service de bus commence à l'instant $S_0 = 0$. À cet instant, un premier bus quitte l'arrêt de bus .
On note $S_1,\ S_2,\ \cdots, S_n,\ \cdots$ les instants de passages successifs des autobus qui suivent :

• À l'instant	$S_0 = 0$	:	un bus	n°0	quitte l'arrêt de bus.
• À l'instant	$S_1 = X_1$	:	un bus	n°1	s'arrête et repart immédiatement.
• À l'instant	$S_2 = X_1 + X_2$	:	un bus	n°2	s'arrête et repart immédiatement.
......	....	...	......	...	....
• À l'instant	$S_n = X_1 + X_2 + \cdots + X_n$	:	un bus	n°n	s'arrête et repart immédiatement.
......	....	...	......	...	....

1. On note alors N_t la variable aléatoire égale au nombre d'autobus qui sont passés à l'arrêt de bus entre l'instant 0 et l'instant t, plus exactement dans l'intervalle de temps $]0, t]$.

(a) Justifier l'égalité $[N_t = n] = [S_n \leq t < S_{n+1}]$.

(b) En déduire $P(N_t = n) = P(S_n \leq t) - P(S_{n+1} \leq t)$

(c) En déduire que N_t suit la Loi de Poisson de paramètre λt

2. On suppose dans les deux questions suivantes que les temps de passages successifs d'un autobus ont pour moyenne <u>10 minutes</u>.
Un individu arrive à l'arrêt de Bus à l'instant t_0 <u>fixé</u> pour prendre le bus.
On pose alors : $W = t_0 - S_{N_{t_0}}$ et $Z = S_{N_{t_0}+1} - t_0$

(a) Que représente la variable aléatoire W ?

(b) Que représente la variable aléatoire Z ?

3. On vous propose le script PYTHON suivant :

```python
from math import log
from random import random
def BusStop(T):
    s=0;N=0
    while s<T:
        r=s
        s=s-10*log(random())
        print(s)
        N+=1
    u=s-T; v=s-r;
    return u,v,N
```

(a) Que représentent les variables aléatoires r, s, u, v, N dans ce script ?

(b) Lors de l'exécution des six instructions suivantes :

```python
[u,v,N]=BusStop(2);print('Resultat 1 : ',[u,v,N])
[u,v,N]=BusStop(10);print('Resultat 2 : ',[u,v,N])
[u,v,N]=BusStop(100);print('Resultat 3 : ',[u,v,N])
```

nous avons obtenu :

```
Resultat 1 : [2.1397149495555242, 3.7711362277272937, 2]
Resultat 2 : [9.099021055187752, 19.09902105518775, 1]
Resultat 3 : [0.2335693918375341, 1.6424737628208987, 14]
```

Commentez chaque résultat en deux lignes.

(c) Après exécution des instructions suivantes :

```python
T=100
a=0;b=0;c=0;N=10000;
for k in range (N):
    [u,v,Numero]=BusStop(T)
    a=a+u; b=b+v; c=c+Numero
print(a/N,b/N,c/N)
```

nous avons obtenu :

```
9.745710157030997 19.79428021053376 10.9577
```

En terme d'estimation, que représentent les valeurs affichées ?
En quoi deux d'entre elles pourraient sembler paradoxales ?
En quoi la dernière est-elle parfaitement explicable ?

4. Nous nous proposons dans cette question de débusquer le mystère soulevé par la simulation précédente. Soit t_0 un réel strictement positif fixé et t un réel quelconque de l'intervalle $]0, t_0[$.

(a) Justifier pour tout réel t de l'intervalle $]0, t_0[$:

$$\left[S_{N_{t_0}} \leq t\right] = \bigcup_{k=0}^{+\infty} \left[S_k \leq t < t_0 < S_{k+1}\right]$$

Nous poserons désormais : $\forall k \in \mathbb{N}^*, \forall t \in]0, t_0[\,,\ G_k(t) = P(S_k \leq t < t_0 < S_{k+1})$

(b) Soit (α, β) un couple de réels quelconques tels que : $0 < \alpha < \beta \leq t_0$.
Montrer pour tout entier naturel non nul k, les inclusions suivantes :

- $\left[\alpha < S_k \leq \beta < t_0 \leq S_{k+1}\right] \supset \left[\alpha < S_k \leq \beta\right] \cap \left[t_0 - \alpha \leq X_{k+1}\right]$
- $\left[\alpha < S_k \leq \beta < t_0 \leq S_{k+1}\right] \subset \left[\alpha < S_k \leq \beta\right] \cap \left[t_0 - \beta \leq X_{k+1}\right]$

Prouver alors l'inégalité suivante :

$$\frac{F_k(\beta) - F_k(\alpha)}{\beta - \alpha} \exp(-\lambda(t_0 - \alpha)) \leq \frac{G_k(\beta) - G_k(\alpha)}{\beta - \alpha} \leq \frac{F_k(\beta) - F_k(\alpha)}{\beta - \alpha} \exp(-\lambda(t_0 - \beta))$$

(c) En déduire : $\forall k \in \mathbb{N}^*,\ \forall t \in]0, t_0[\,,\ G_k'(t) = \dfrac{\lambda e^{-\lambda t_0} (\lambda t)^{k-1}}{(k-1)!}$
Donner l'expression de $G_k(t)$ pour tout $k \in \mathbb{N}^*$ et $t \in]0, t_0[$

(d) On note G la fonction définie sur $]0, t_0[$ par $G(t) = P\left(S_{N_{t_0}} \leq t\right)$.

- Nous rappelons $G_k(t) = P(S_k \leq t < t_0 < S_{k+1})$.
  En déduire : $\forall t \in]0, t_0[\,,\ G(t) = P\left(S_{N_{t_0}} \leq t\right) = e^{-\lambda(t_0 - t)}$
- Justifier que $S_{N_{t_0}}$ n'est pas une variable aléatoire à densité.
- Vérifier que $P\left(S_{N_{t_0}} = 0\right) + \int_0^{t_0} G'(t)\, dt = 1$

(e) On admettra que $S_{N_{t_0}}$ admet une espérance définie par :

$$E\left(S_{N_{t_0}}\right) = 0 \times P\left(S_{N_{t_0}} = 0\right) + \int_0^{t_0} t\, G'(t)\, dt$$

- Déterminer cette espérance.
- En déduire l'espérance de la variable aléatoire W.

5. Nous allons maintenant déterminer par un procédé analogue l'espérance de Z
Soit t_0 un réel strictement positif fixé et t strictement supérieur à t_0.

(a) Exprimer $\left[S_{N_{t_0}+1} > t\right]$ en fonction des événements $\left[S_k \leq t_0 < t < S_{k+1}\right]$.

(b) Soit (α, β) un couple de réels quelconques tels que : $0 \leq \alpha < \beta \leq t_0$
Nous poserons alors : $\forall k \in \mathbb{N}^*, \forall u \in [0, t_0]\,,\ H_k(u) = P(S_k \leq u \leq t_0 < t < S_{k+1})$
Comparer pour tout entier naturel non nul k, les trois événements suivants :
$\left[\alpha < S_k \leq \beta \leq t < S_{k+1}\right]$, $\left[\alpha < S_k \leq \beta\right] \cap \left[t - \alpha \leq X_{k+1}\right]$, $\left[\alpha < S_k \leq \beta\right] \cap \left[t - \beta \leq X_{k+1}\right]$
En déduire un encadrement de $\dfrac{H_k(\beta) - H_k(\alpha)}{\beta - \alpha}$

(c) En utilisant l'encadrement précédent, donner la valeur de $H_k'(u)$ pour tout $k \in \mathbb{N}^*$ et $u \in [0, t_0]$.

(d) En déduire l'expression de $H_k(u)$ pour tout $k \in \mathbb{N}^*$ et $u \in [0, t_0]$.
Déterminer pour tout entier naturel k l'expression de $H_k(t_0) = P(S_k \leq t_0 < t < S_{k+1})$.

(e) En déduire que : $\forall t > t_0,\ P\left(S_{N_{t_0}+1} > t\right) = e^{-\lambda(t - t_0)}$.
Vérifier que $Z = S_{N_{t_0}+1} - t_0$ est distribuée selon la Loi Exponentielle de paramètre λ

(f) Ce dernier résultat n'était-il pas prévisible ?

6. En utilisant les résultats obtenus aux questions 4 et 5, vérifier que l'espérance de $W+Z = S_{N_{t_0}+1} - S_{N_{t_0}}$ vaut :

$$E\left(S_{N_{t_0}+1} - S_{N_{t_0}}\right) = E\left(W+Z\right) = \frac{2-e^{-\lambda t_0}}{\lambda}$$

Ces résultats sont-ils en accord avec les estimations obtenues à la question (3c)

PROBLEME N˚2

Nous nous intéressons ici à l'incidence de la température diurne Θ sur l'allongement X d'une tige de tomate. Nous avons observé les résultats suivants sur un échantillon de 11 couples d'observations :

k	1	2	3	4	5	6	7	8	9	10	11
θ_k	5	7	10	13	15	18	20	22	25	28	30
x_k	1	2	3	6	8	10	11	15	18	20	23

Les températures sont mesurées en ˚C ; les allongements sont mesurés en mm/j

Nous allons tester deux modèles d'ajustements :

Le modèle I : ajustement affine $\left\{\begin{array}{l}\text{Pour une température } \theta \text{ donnée}\\ X = a + b\,\theta + \epsilon \quad \text{où } \epsilon \text{ suit la } \mathcal{N}\left(0,\sigma^2\right)\end{array}\right.$

Le modèle II : ajustement quadratique $\left\{\begin{array}{l}\text{Pour une température } \theta \text{ donnée}\\ X = a + b\,\theta + c\,\theta^2 + \epsilon \quad \text{où } \epsilon \text{ suit la } \mathcal{N}\left(0,\sigma^2\right)\end{array}\right.$

Dans l'un et l'autre cas, les variables aléatoires $X_1,\ X_2,\ X_3,\cdots$ relatives à une série de mesures sont indépen

Nous utiliserons dans l'un et l'autre cas la « **méthode des moindres carrés** »

Tentative d'ajustement affine

Nous désignerons par $\widehat{a}$ et $\widehat{b}$ les valeurs respectives de a et b qui rendent minimale la somme suivante :

$$S\left(a,b\right) = \sum_{k=1}^{11}\left(x_k - a - b\,\theta_k\right)^2$$

1. Rappeler les formules de cours permettant de déterminer $\widehat{a}$ et $\widehat{b}$.

2. Calculer $\sum\limits_{k=1}^{11} x_k$, $\sum\limits_{k=1}^{11}\theta_k$, $\sum\limits_{k=1}^{11}\theta_k\,x_k$, $\sum\limits_{k=1}^{11}\theta_k^2$ et $\sum\limits_{k=1}^{11} x_k^2$.

3. Déterminer le coefficient de corrélation R et le coefficient de détermination R^2 de cette série double.

4. Déterminer l'équation de la droite des moindres carrés : $x = \widehat{a} + \widehat{b}\,\theta$.

5. Tracez la droite de régression, les onze points et faites apparaître les 11 résidus : $\varepsilon_k = x_k - \widehat{a} - \widehat{b}\,\theta_k$.

6. Indiquez clairement pourquoi ce modèle n'est pas acceptable au regard des hypothèses.

Ajustement quadratique

Nous désignerons par $\alpha,\ \beta$ et γ les valeurs respectives de $a,\ b$ et c qui rendent minimale la somme suivante :

$$S\left(a,b,c\right) = \sum_{k=1}^{11}\left(x_k - a - b\,\theta_k - c\,\theta_k^2\right)^2$$

On note $\begin{array}{lcl}\overrightarrow{\mathrm{X}} & = & \left(x_1,\ x_2,\ \cdots,\ x_{11}\right)\\ \overrightarrow{\mathrm{U}} & = & \left(1,\ 1,\ \cdots,\ 1\right)\end{array}$ et $\begin{array}{lcl}\overrightarrow{\mathrm{B}} & = & \left(\theta_1,\ \theta_2,\ \cdots,\ \theta_{11}\right)\\ \overrightarrow{\mathrm{C}} & = & \left(\theta_1^2,\ \theta_2^2,\ \cdots,\ \theta_{11}^2\right)\end{array}$

1. Nous désignerons par $\overrightarrow{H}$ la projection orthogonale du vecteur $\overrightarrow{X}$ de $\mathbb{R}^{11}$ sur le sous espace $F = \operatorname{Vect}\left\langle\overrightarrow{\mathrm{U}},\ \overrightarrow{\mathrm{B}},\ \overrightarrow{\mathrm{C}}\right\rangle$

- Préciser le lien entre les vecteur $\overrightarrow{H}$ et les réels $\alpha,\ \beta$ et γ.
- Montrer que le triplet $\left(\alpha,\beta,\gamma\right)$ est l'unique solution d'un système linéaire de trois équations .

2. Après avoir calculé $\left\|\overrightarrow{\mathrm{U}}\right\|^2$, $\left\|\overrightarrow{\mathrm{B}}\right\|^2$, $\left\|\overrightarrow{\mathrm{C}}\right\|^2$, $\overrightarrow{\mathrm{B}} \bullet \overrightarrow{\mathrm{U}}$, $\overrightarrow{\mathrm{C}} \bullet \overrightarrow{\mathrm{U}}$, $\overrightarrow{\mathrm{B}} \bullet \overrightarrow{\mathrm{C}}$, $\overrightarrow{\mathrm{X}} \bullet \overrightarrow{\mathrm{U}}$, $\overrightarrow{\mathrm{X}} \bullet \overrightarrow{\mathrm{B}}$, $\overrightarrow{\mathrm{X}} \bullet \overrightarrow{\mathrm{C}}$, écrire le système sous forme matricielle.
3. Résoudre le système et donner les valeurs α, β et γ.(*L'usage de la calculatrice est fortement conseillé*)
4. Déterminer les relations entre $\left\|\overrightarrow{\mathrm{X}} - \alpha\,\overrightarrow{\mathrm{U}} - \beta\,\overrightarrow{\mathrm{B}} - \gamma\,\overrightarrow{\mathrm{C}}\right\|$, $\left\|\alpha\,\overrightarrow{\mathrm{U}} + \beta\,\overrightarrow{\mathrm{B}} + \gamma\,\overrightarrow{\mathrm{C}}\right\|$ et $\left\|\overrightarrow{\mathrm{X}}\right\|$
5. On pose $\mathrm{SCE}_{\text{Résiduelle}} = \left\|\overrightarrow{\mathrm{X}} - \alpha\,\overrightarrow{\mathrm{U}} - \beta\,\overrightarrow{\mathrm{B}} - \gamma\,\overrightarrow{\mathrm{C}}\right\|^2$, $\mathrm{SCE}_{\text{Expliquée}} = \left\|\alpha\,\overrightarrow{\mathrm{U}} + \beta\,\overrightarrow{\mathrm{B}} + \gamma\,\overrightarrow{\mathrm{C}} - \frac{1}{11}\sum_{k=1}^{11} x_k\,\overrightarrow{\mathrm{U}}\right\|^2$ et $\mathrm{SCE}_{\text{Totale}} = \left\|\overrightarrow{\mathrm{X}} - \frac{1}{11}\sum_{k=1}^{11} x_k\,\overrightarrow{\mathrm{U}}\right\|^2$.
   Justifier $\mathrm{SCE}_{\text{Résiduelle}} + \mathrm{SCE}_{\text{Expliquée}} = \mathrm{SCE}_{\text{Totale}}$.
6. Déterminer le rapport $R^2 = \frac{\mathrm{SCE}_{\text{Expliquée}}}{\mathrm{SCE}_{\text{Totale}}}$
7. Tracez la courbe d'équation $x = \alpha + \beta\,\theta + \gamma\,\theta^2$, les onze points et faites apparaître les 11 résidus : $\varepsilon_k = x_k - \alpha + \beta\,\theta_k + \gamma\,\theta_k^2$
8. Indiquez en quoi ce modèle est acceptable au regard des hypothèses.

Programmation - ajustement quadratique

Nous vous proposons de compléter le script suivant en vous conformant aux instructions données dans les questions 1 à 3 qui suivent :

```
1  from numpy import*
2  #****************************************
3  def ProdScal(X,Y):
4      R=________________________________
5      return R
6  #****************************************
7  def Matrice(B):
8      U=array([1]*len(B))
9      L1=_______________________________________
10     L2=_______________________________________
11     L3=_______________________________________
12     return matrix([L1,L2,L3])
13 #*****************************************
14 X=array([1,2,3,6,8,10,11,15,18,20,23])
15 B=array([5,7,10,13,15,18,20,22,25,28,30])
16 U=array([1]*len(B))
17 C=matrix([[ProdScal(X,U)],[ProdScal(X,B)],[ProdScal(X,B**2)]])
18 Estimation=Matrice(B)**(-1)*C
19 print("l'Estimation de a vaut ",Estimation[0,0])
20 print("l'Estimation de b vaut ",Estimation[1,0])
21 print("l'Estimation de c vaut ",Estimation[2,0])
22 Residus=______________________________________
23 SCR=ProdScal(Residus,Residus)
24 print("La somme des carrés des résidus vaut ",SCR)
```

1. Construire une fonction appelée ProdScal qui calcule le produit scalaire de deux vecteurs de type *array* (lignes 3 à 5).
2. Construire une fonction appelée Matrice qui contient dans l'ordre adéquat les produits scalaires des vecteurs $\overrightarrow{U}$, $\overrightarrow{B}$, $\overrightarrow{C}$ (lignes 7 à 12).
3. Construire le vecteur des résidus (ligne 22).