BANQUE COMMUNE D'ÉPREUVES

Code sujet : 288

Conception : HEC Paris – ESSEC BS

OPTION ÉCONOMIQUE

MATHÉMATIQUES

Mardi 28 avril 2020, de 14 h. à 18 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.
*Les candidats sont invités à **encadrer** dans la mesure du possible les résultats de leurs calculs.*
*Aucun document n'est autorisé. **L'utilisation de toute calculatrice et de tout matériel électronique est interdite.** Seule l'utilisation d'une règle graduée est autorisée.*
Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre.

On s'intéresse dans ce sujet au problème de la *double dépense* de *bitcoins* par un groupe d'individus mal intentionnés.

On rappelle que le bitcoin est une monnaie virtuelle dont l'utilisation pour des transactions est associée à une structure unique appelée *blockchain*, partagée sur le réseau des usagers de cette monnaie et ayant pour but de sécuriser ces transactions.

La modélisation étudiée ne nécessite pas de connaissances particulières sur le *bitcoin* et la *blockchain*.

Partie I - Deux résultats généraux

On démontre dans cette partie deux résultats préliminaires, aux questions 5 et 6. Ces résultats seront utilisés dans la suite du sujet et pourront être admis.

Calcul d'une probabilité

Soit X et Y deux variables aléatoires sur un espace probabilisé, à densité et indépendantes.

On note F_X et F_Y les fonctions de répartition de X et Y.

On suppose que Y est à valeurs positives et possède une densité f_Y dont la restriction à $[0, +\infty[$ est continue sur cet intervalle.

Pour tout $x \in \mathbb{R}^+$, on pose $H(x) = \mathbb{P}([X \leqslant Y] \cap [Y \leqslant x])$.

1. a) Montrer que H est une fonction croissante sur $\mathbb{R}^+$ qui admet une limite finie en $+\infty$.
   b) En utilisant la suite $(H(n))_{n\in\mathbb{N}}$, montrer que $\lim\limits_{x\to+\infty} H(x) = \mathbb{P}([X \leqslant Y])$.
   Que vaut $H(0)$?
2. Soit (u, v) un couple de réels positifs tels que $u < v$.
   a) Montrer que $H(v) - H(u) = \mathbb{P}([X \leqslant Y] \cap [u < Y \leqslant v])$ puis que :
   $$F_X(u)\,\frac{F_Y(v) - F_Y(u)}{v-u} \leqslant \frac{H(v) - H(u)}{v-u} \leqslant F_X(v)\,\frac{F_Y(v) - F_Y(u)}{v-u}$$
   b) En déduire que pour tout $x \in \mathbb{R}^+$, H est dérivable en x et $H'(x) = F_X(x) f_Y(x)$.
   c) En conclure que pour tout x réel positif, $H(x) = \displaystyle\int_0^x F_X(t) f_Y(t)\,\mathrm{d}t$.
3. Montrer que $\mathbb{P}([X \leqslant Y]) = \displaystyle\int_0^{+\infty} F_X(t) f_Y(t)\,\mathrm{d}t$.
4. En utilisant la fonction $K : x \mapsto \mathbb{P}([X < Y] \cap [Y \leqslant x])$, on montrerait de même et nous l'admettrons que :
   $$\mathbb{P}([X < Y]) = \int_0^{+\infty} F_X(t) f_Y(t)\,\mathrm{d}t = \mathbb{P}([X \leqslant Y])$$
   Que peut-on en déduire pour $\mathbb{P}([X = Y])$?
5. *Application aux lois exponentielles*
   On suppose que U et V sont deux variables aléatoires indépendantes suivant des lois exponentielles de paramètres respectifs λ et μ, réels strictement positifs.
   Soit θ un réel positif ou nul.
   a) Déterminer la fonction de répartition de la variable aléatoire $X = U - \theta$.
   b) En déduire que pour tout $\theta \geqslant 0$,
   $$\mathbb{P}([U - \theta \leqslant V]) = 1 - \frac{\mu}{\lambda + \mu}\,\mathrm{e}^{-\lambda\theta}$$

Inégalité de Boole

6. On considère $(B_k)_{k\in\mathbb{N}^*}$ une famille d'événements d'un espace probabilisé.

   a) Montrer par récurrence sur $n \in \mathbb{N}^*$ que $\mathbb{P}\left(\bigcup_{k=1}^{n} B_k\right) \leqslant \sum_{k=1}^{n} \mathbb{P}(B_k)$.

   b) On suppose que la série $\sum_{k\geqslant 1} \mathbb{P}(B_k)$ converge. Montrer que :

$$\mathbb{P}\left(\bigcup_{k\geqslant 1} B_k\right) \leqslant \sum_{k=1}^{+\infty} \mathbb{P}(B_k)$$

Partie II - Une compétition entre deux groupes

Dans toute la suite du sujet, on désigne par p un réel de l'intervalle $]0,1[$ et on pose $q = 1-p$.

On modélise une compétition entre deux groupes d'individus A et B avec les règles suivantes :

- Le groupe A doit résoudre une suite de problèmes $(P_k)_{k\geqslant 1}$ dans l'ordre des indices. Au temps $t=0$, le groupe commence la résolution du problème P_1, ce qui lui prend un temps représenté par la variable aléatoire X_1. Une fois P_1 résolu, le groupe aborde immédiatement le problème P_2, et on note X_2 le temps consacré à la résolution de P_2 par le groupe A, et ainsi de suite.

  Pour tout $k \in \mathbb{N}^*$, on note X_k la variable aléatoire donnant le temps consacré à la résolution du problème P_k par le groupe A.
- De même, le groupe B doit résoudre dans l'ordre une suite de problèmes $(Q_k)_{k\geqslant 1}$; la résolution du premier problème Q_1 commence au temps $t = 0$ et on note, pour tout $k \in \mathbb{N}^*$, Y_k la variable aléatoire donnant le temps consacré par le groupe B à la résolution du problème Q_k.
- À ce jeu est associé un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ sur lequel sont définies les suites de variables aléatoires $(X_k)_{k\geqslant 1}$ et $(Y_k)_{k\geqslant 1}$, et on fait les hypothèses suivantes :

  * pour tout $k \in \mathbb{N}^*$, X_k suit la loi exponentielle de paramètre p, notée $\mathcal{E}(p)$, et Y_k suit la loi exponentielle $\mathcal{E}(q)$;
  * pour tout $k \in \mathbb{N}^*$, les variables aléatoires $X_1, \dots, X_k, Y_1, \dots, Y_k$ sont indépendantes.
- On établit alors la liste de tous les problèmes résolus *dans l'ordre où ils le sont par les deux groupes.* En cas de simultanéité temporelle de la résolution par les deux groupes d'un de leurs problèmes, on placera d'abord le problème résolu par A dans la liste puis celui résolu par B.

  Pour tout $n \in \mathbb{N}^*$, on note U_n la variable aléatoire de Bernoulli associée à l'événement « le n-ème problème placé dans la liste est un problème résolu par le groupe A ».

  Par exemple, si la liste des cinq premiers problèmes résolus est $(P_1, P_2, Q_1, P_3, Q_2)$ alors $U_1 = 1$, $U_2 = 1$, $U_3 = 0$, $U_4 = 1$ et $U_5 = 0$.

- Pour tout $n \geqslant 0$, on note aussi S_n la variable aléatoire donnant le nombre de problèmes qui ont été résolus par A présents dans la liste des n premiers problèmes résolus. En particulier, S_0 vaut toujours 0.

7. a) Que représente la variable aléatoire $\sum_{k=1}^{n} X_k$?

   b) On suppose que $X_1 = 5$, $X_2 = 2$, $X_3 = 3$, $X_4 = 2$,$Y_1 = 2$, $Y_2 = 2$, $Y_3 = 4$, $Y_4 = 2$.
   Déterminer $U_1, ..., U_7$.
   Peut-on aussi en déduire la valeur de U_8 ?

   c) Compléter le script Scilab suivant pour qu'il simule le jeu et, pour n, p donnés, affiche la liste des valeurs U_1,U_2,...,U_n :

```
p = input('p=')
n = input('n=')
q = 1-p

U = zeros(1, n)

sommeX = grand(1, 1, "exp", 1/p)
sommeY = grand(1, 1, "exp", 1/q)

mini = min(sommeX, sommeY)

for k = 1:n
    if sommeX == ...
        U(k) = ...
        sommeX = sommeX + grand(1, 1, "exp", 1/p)
    else
        sommeY = ...
        end
     mini = min(sommeX, sommeY)
end
...
```

   d) Quelle(s) instruction(s) faut-il ajouter pour afficher la valeur de S_n ?

8. *Loi de U_n*
   Dans cette question, on démontre par récurrence sur $n \geqslant 1$ que $\mathbb{P}([U_n = 1]) = p$.

   a) Montrer que $\mathbb{P}([U_1 = 1]) = \mathbb{P}([X_1 \leqslant Y_1]) = p$.

   b) i. Montrer que pour tout réel $x < 0$, $\mathbb{P}_{[U_1=1]}([Y_1 - X_1 \leqslant x]) = 0$.

      ii. Soit x un réel positif ou nul.
      Établir : $\mathbb{P}_{[U_1=1]}([Y_1 - X_1 \leqslant x]) = \dfrac{1}{p}\mathbb{P}([X_1 \leqslant Y_1 \leqslant X_1 + x])$,
      puis calculer $\mathbb{P}_{[U_1=1]}([Y_1 - X_1 \leqslant x])$.

   c) On peut interpréter ce résultat en disant que la *loi conditionnelle de* $Y_1 - X_1$ *sachant* $[U_1 = 1]$ est une loi exponentielle. Quel est son paramètre ?

Par analogie, quelle est la loi conditionnelle de $X_1 - Y_1$ sachant $[U_1 = 0]$? (on n'attend pas une démonstration précise mais un argument de bon sens pour justifier le résultat proposé)

d) On suppose que $n \in \mathbb{N}^*$ et $\mathbb{P}([U_n = 1]) = p$.
Déduire de cette hypothèse et de la question précédente que $\mathbb{P}_{[U_1=1]}([U_{n+1} = 1]) = p$ et $\mathbb{P}_{[U_1=0]}([U_{n+1} = 1]) = p$.

e) Conclure.

9. On montrerait aussi par récurrence, et nous l'admettrons, que pour tout $n \in \mathbb{N}^*$, les variables aléatoires $U_1,\dots,U_n$ sont mutuellement indépendantes.
En déduire la loi de S_n.

Soit $r \in \mathbb{N}$, on s'intéresse, dans les questions qui suivent, à la probabilité a_r de l'événement,

A_r : « il existe un $n \geqslant r$ tel que, lorsque n problèmes en tout ont été résolus, le groupe A en a résolu r de plus que le groupe B ».

10. a) Justifier que $a_0 = 1$.

b) Montrer que pour tout $r \geqslant 1$, $\mathbb{P}_{[U_1=1]}(A_r) = \mathbb{P}(A_{r-1})$ et $\mathbb{P}_{[U_1=0]}(A_r) = \mathbb{P}(A_{r+1})$.

c) En déduire que pour tout $r \geqslant 1$, $a_{r+1} = \dfrac{1}{q}\, a_r - \dfrac{p}{q}\, a_{r-1}$.

d) En remarquant que $1 - 4pq = (1-2p)^2$, donner une expression de a_r en fonction de p, q, r et de deux constantes que l'on introduira.

11. Le cas $p \geqslant \dfrac{1}{2}$.

Montrer que, dans les cas $p = \dfrac{1}{2}$ et $p > \dfrac{1}{2}$, la suite $(a_r)_{r\in\mathbb{N}}$ est constante et égale à 1 .

12. Le cas $p < \dfrac{1}{2}$.

a) Soit k un entier naturel.

i. Établir : $A_{2k} = \displaystyle\bigcup_{i\geqslant k} [S_{2i} = i + k]$.

ii. Montrer que pour tout $i \geqslant k$, on a $\mathbb{P}([S_{2i} = i+k]) = \dbinom{2i}{i+k} p^{i+k} q^{i-k}$.

iii. Après avoir donné la valeur de la somme $\displaystyle\sum_{j=0}^{2i} \binom{2i}{j}$, montrer que :

pour tout entier $i \geqslant k$, $\dbinom{2i}{i+k} \leqslant 4^i$.

iv. En déduire l'inégalité :

$$\sum_{i=k}^{+\infty} \mathbb{P}([S_{2i} = k+i]) \leqslant \left(\frac{p}{q}\right)^k \frac{(4pq)^k}{1-4pq}$$

b) Montrer en utilisant l'inégalité de Boole (voir question 6) que si $p < \dfrac{1}{2}$, alors $\displaystyle\lim_{k\to+\infty} a_{2k} = 0$.

c) Conclure en utilisant la question 10.d. que si $p < \frac{1}{2}$, alors :

pour tout entier naturel r, $a_r = \left(\frac{p}{q}\right)^r$.

On a ainsi établi dans les questions 11 et 12 :

$$\forall r \in \mathbb{N}, \quad a_r = \begin{cases} \left(\frac{p}{q}\right)^r & \text{si } p < \frac{1}{2} \\ 1 & \text{si } p \geqslant \frac{1}{2}. \end{cases}$$

Ce résultat pourra être admis et utilisé dans le suite du sujet.

Partie III - La *blockchain* et la stratégie de la *double dépense*

On utilise, dans cette partie, les notations et résultats de la partie II.

Soit n un entier supérieur ou égal 1.

La *blockchain* est formée d'une suite de blocs, chacun associé à plusieurs transactions. Elle contient l'historique de toutes les transactions effectuées depuis la création du *bitcoin.*

Avant d'être placé dans la *blockchain* , un nouveau bloc doit être validé. Cette validation nécessite la mise en œuvre d'une grande puissance de calcul pour résoudre un problème dépendant fortement du contenu du bloc et des blocs qui le précèdent.

Les individus qui valident les blocs sont appelés mineurs.

Il est possible qu'à un instant donné, coexistent sur le réseau deux *blockchains*, valides et différentes. Dans ce cas, le réseau choisira celle qui comporte le plus de blocs et l'autre sera abandonnée.

Par prudence, lorsqu'un bloc est validé, il est recommandé d'attendre que $n-1$ blocs le suivant soient aussi validés pour considérer que les transactions incluses dans le bloc soient honnêtes.

Un groupe de mineurs mal intentionnés, noté A, peut essayer de dépenser deux fois les mêmes *bitcoins* en procédant ainsi :

- Le groupe A demande la validation de l'achat d'un bien d'un montant de s *bitcoins* qu'il a en sa possession.
- Lorsque le bloc K incluant cette transaction est proposé à la validation sur le réseau, A modifie ce bloc en K', qu'il ne diffuse pas, en remplaçant l'achat par une vente des s *bitcoins* en euros à son profit par exemple. Il se met alors à la validation de ce nouveau bloc et crée ainsi une deuxième instance de la *blockchain* qu'il continue à développer sans la diffuser.
- Lorsque le groupe B, représentant l'ensemble des autres mineurs du réseau, a validé K ainsi que les $n-1$ blocs suivants, le vendeur du bien considère que la transaction est valide et fournit le bien.

- Le groupe A attend alors d'avoir une *blockchain* plus longue que celle de B, qui est publique, pour la diffuser donc invalider la *blockchain* publique et l'achat du bien. Le crédit en *bitcoins* du vendeur du bien est alors annulé.

On reprend et on complète la modélisation de la partie précédente pour déterminer la probabilité que la stratégie de la *double dépense* réussisse et le choix de n pour que cette probabilité soit faible.

Une première phase du jeu, décrit dans la partie II, s'achève à l'instant aléatoire t où le problème Q_n est ajouté à la liste des problèmes résolus.

Le groupe de mineurs A est ensuite déclaré vainqueur s'il se trouve un instant $t' \geqslant t$ où le nombre de problèmes résolus par A dans la liste des problèmes résolus depuis le début du jeu, est strictement supérieur au nombre de ceux résolus par B dans cette même liste. On note G_n cet événement.

On détermine, dans cette partie, la probabilité de G_n en fonction de n et de p.

13. On s'intéresse tout d'abord à la loi de la variable aléatoire T_n égale au nombre de problèmes résolus par le groupe A lorsque l'on place Q_n dans la liste des problèmes résolus.
    a) Montrer que pour tout $k \in \mathbb{N}$, $[T_n = k] = [S_{n+k-1} = k] \cap [U_{n+k} = 0]$.
    b) En déduire que $\mathbb{P}([T_n = k]) = \binom{n+k-1}{k} p^k q^n$.
14. a) En utilisant la formule des probabilités totales, établir :

$$\mathbb{P}(G_n) = \mathbb{P}([T_n \geqslant n+1]) + \sum_{k=0}^{n} \mathbb{P}([T_n = k])\, a_{n+1-k}$$

    b) Dans le cas où $p \geqslant \frac{1}{2}$, en déduire que $\mathbb{P}(G_n) = 1$.
    c) De même lorsque $p < \frac{1}{2}$, montrer que :

$$\mathbb{P}(G_n) = 1 - \sum_{k=0}^{n} \binom{n+k-1}{k} \left(p^k q^n - p^{n+1} q^{k-1}\right)$$

15. *Une meilleure expression de $\mathbb{P}(G_n)$ lorsque $p < \frac{1}{2}$*
    Pour tout $x \in [0,1]$ et $n \in \mathbb{N}^*$, on pose :

$$u_n(x) = (1-x)^n \sum_{k=0}^{n} \binom{n+k-1}{k} x^k$$

    a) Vérifier que pour tout $n \in \mathbb{N}^*$: $\mathbb{P}(G_n) = 1 - u_n(p) + \frac{p}{q} u_n(q)$.
    b) Pour tout $x \in [0,1]$ et $n \in \mathbb{N}^*$, établir la relation :

$$u_{n+1}(x) = u_n(x) + (1-x)^n x^{n+1} \left(\binom{2n}{n+1} - \binom{2n+1}{n+1} x \right)$$

c) En déduire que pour tout $n \in \mathbb{N}^*$: $\mathbb{P}(G_{n+1}) = \mathbb{P}(G_n) - \left(1 - \frac{p}{q}\right)(pq)^{n+1}\binom{2n+1}{n+1}$.

d) Montrer par récurrence, que pour tout $n \in \mathbb{N}^*$:

$$\mathbb{P}(G_n) = \frac{p}{q} - \left(1 - \frac{p}{q}\right)\sum_{k=1}^{n}\binom{2k-1}{k}(pq)^k$$

16. *Application à la sécurisation des transactions*

Connaissant $p < \frac{1}{2}$, on cherche à limiter le risque que la stratégie mise en place par le groupe de mineurs A réussisse.

a) Après avoir établi la formule $\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}$ lorsque $k \in [\![1, n]\!]$, écrire une fonction Scilab qui calcule les coefficients binomiaux.

b) Écrire un script Scilab qui détermine n_p, le plus petit entier n tel que $\mathbb{P}(G_n) \leqslant \varepsilon$ pour $p < \frac{1}{2}$ et $\varepsilon > 0$ saisis au clavier par l'utilisateur.

NB : Pour $\varepsilon = 10^{-4} = 0,1‰$ et p variant entre 10% et 32%, on obtient pour la représentation de n_p en fonction de p :