

Conception : HEC Paris

MATHÉMATIQUES

OPTION SCIENTIFIQUE

Mercredi 30 avril 2014, de 8 h. à 12 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.  
Ils ne doivent faire usage d'aucun document : **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

*Dans ce problème, on s'intéresse à des opérations de transport dans des situations déterministes ou aléatoires, modélisées de manière discrète ou continue, dans le but de trouver un programme de transport optimal dont le coût serait le plus faible possible.*

*Les parties I, II et III sont largement indépendantes.*

- Toutes les variables aléatoires considérées dans ce problème sont supposées définies sur le même espace probabilisé  $(\Omega, \mathcal{A}, P)$ .
- Sous réserve d'existence, on note  $E(Z)$  l'espérance d'une variable aléatoire  $Z$ .
- Pour tout entier  $N$  supérieur ou égal à 1, on note  $\mathcal{E}_N$  l'ensemble des applications de  $\llbracket 1, N \rrbracket$  dans  $\llbracket 1, N \rrbracket$ .

**Préliminaire**

1. Soit  $N$  un entier supérieur ou égal à 2.
  - a) Quel est le nombre d'éléments de l'ensemble  $\mathcal{E}_N$  ?
  - b) Parmi les éléments de  $\mathcal{E}_N$ , quel est le nombre d'applications injectives et parmi celles-ci, combien sont strictement monotones ?  
(les réponses aux questions 1.a) et 1.b) seront données sans démonstration)
2. Soit  $p$  un réel vérifiant  $0 < p < 1$ .

On considère une variable aléatoire  $X$  suivant la loi exponentielle de paramètre 1.  
Pour tout  $\omega \in \Omega$ , on pose :  $Y(\omega) = \lfloor pX(\omega) \rfloor$ , où  $\lfloor \cdot \rfloor$  désigne la fonction partie entière.

  - a) Vérifier que  $Y$  est une variable aléatoire discrète. Calculer pour tout  $n \in \mathbb{N}$ , la probabilité  $P([Y = n])$ .
  - b) Montrer que la variable aléatoire  $Y + 1$  suit une loi géométrique dont on précisera le paramètre.
  - c) Établir les inégalités strictes :  $0 < E(Y) < p$ .

3.a) Pour tout couple  $(r, s) \in \mathbb{N}^2$ , montrer que l'intégrale  $\int_0^1 x^r (\ln x)^s dx$  est convergente.

(on pourra utiliser le changement de variable  $u = -\ln x$  après avoir justifié précisément sa validité)

b) Établir pour tout couple  $(r, s) \in \mathbb{N}^2$ , l'égalité :  $\int_0^1 x^r (\ln x)^s dx = \frac{(-1)^s s!}{(r+1)^{s+1}}$ .

### Partie I. Transport dans une situation aléatoire

On dit que la loi d'une variable aléatoire  $Y$  est *accessible* depuis une variable aléatoire  $X$ , s'il existe une application  $T : X(\Omega) \rightarrow \mathbb{R}$  telle que la variable aléatoire  $T(X)$  suit la même loi que  $Y$ .

L'application  $T$  est alors appelée une *fonction de transport* de la variable aléatoire  $X$  vers la loi de  $Y$ .

On associe à  $T$  un *coût de transport*  $C(T)$  défini, sous réserve d'existence, par :  $C(T) = E\left((X - T(X))^2\right)$ .

Dans toute cette partie,  $X$  désigne une variable aléatoire vérifiant  $X(\Omega) = ]0, 1[$  et suivant la loi uniforme sur  $]0, 1[$ , c'est-à-dire admettant pour densité la fonction  $f_X$  définie par :

$$f_X(x) = \begin{cases} 1 & \text{si } x \in ]0, 1[ \\ 0 & \text{sinon} \end{cases} .$$

4. Soit  $p$  un réel vérifiant  $0 < p < 1$ . Pour tout réel  $a \in [0, 1 - p]$ , on note dans cette question,  $T_a$  la fonction définie sur  $]0, 1[$  par :

$$T_a(x) = \begin{cases} 1 & \text{si } x \in ]a, a + p[ \\ 0 & \text{sinon} \end{cases} .$$

a) Calculer la probabilité  $P([T_a(X) = 1])$  et en déduire que les fonctions  $T_a$  sont des fonctions de transport de  $X$  vers une même loi que l'on précisera.

b) Vérifier que le coût de transport  $C(T_a)$  est égal à  $\frac{1}{3} + p(1 - p) - 2ap$ .

c) En déduire la valeur de  $a$  qui minimise  $C(T_a)$  et exprimer le coût minimal correspondant en fonction de  $p$ .

5. Soit  $T_1$  et  $T_2$  les applications définies sur  $]0, 1[$  par  $T_1(x) = -\ln x$  et  $T_2(x) = -\ln(1 - x)$ .

a) Vérifier que  $T_1$  et  $T_2$  sont des fonctions de transport de  $X$  vers une loi que l'on précisera.

b) En utilisant les résultats de la question 3, comparer les coûts de transport  $C(T_1)$  et  $C(T_2)$ .

c) À l'aide de la question 2, montrer que toutes les lois géométriques sont accessibles depuis  $X$ .

6. Dans cette question,  $Y$  désigne une variable aléatoire admettant une densité  $f_Y$  continue et strictement positive sur  $\mathbb{R}$ .

a) Justifier que la fonction de répartition  $F_Y$  de  $Y$  réalise une bijection de  $\mathbb{R}$  sur l'intervalle ouvert  $]0, 1[$ .

b) On note  $F_Y^{-1}$  la bijection réciproque de  $F_Y$ .

Montrer que  $F_Y^{-1}$  est une fonction de transport de la variable aléatoire  $X$  vers la loi de  $Y$ .

7. *Cas particulier* : on suppose que  $Y$  suit la loi normale centrée réduite.

On note  $F_Y$  la fonction de répartition de  $Y$  et  $\varphi$  la densité continue sur  $\mathbb{R}$  de  $Y$ .

a) Établir la convergence de l'intégrale  $\int_{-\infty}^{+\infty} y F_Y(y) \varphi(y) dy$ .

À l'aide d'une intégration par parties, montrer que  $\int_{-\infty}^{+\infty} y F_Y(y) \varphi(y) dy = \frac{1}{2\sqrt{\pi}}$ .

b) Montrer que l'intégrale  $\int_{-\infty}^{+\infty} (y - F_Y(y))^2 \varphi(y) dy$  est convergente et la calculer.

c) En déduire que le coût de transport  $C(F_Y^{-1})$  est égal à  $\frac{4}{3} - \frac{1}{\sqrt{\pi}}$ .

## Partie II. Transport optimal dans une situation déterministe

Dans toute cette partie,  $N$  désigne un entier supérieur ou égal à 2.

On considère  $N$  réels  $d_1, d_2, \dots, d_N$  (appelés points de départ) et  $N$  réels  $a_1, a_2, \dots, a_N$  (appelés points d'arrivée) vérifiant  $d_1 < d_2 < \dots < d_N$  et  $a_1 < a_2 < \dots < a_N$ .

On pose :  $D = \{d_1, d_2, \dots, d_N\}$  et  $A = \{a_1, a_2, \dots, a_N\}$ .

8.a) Montrer que pour tout couple  $(k, \ell) \in \llbracket 1, N \rrbracket^2$ , on a :  $d_k a_k \geq d_k a_\ell + d_\ell a_k - d_\ell a_\ell$ .

b) En déduire à l'aide d'une double sommation que pour tout  $N$ -uplet  $(p_1, p_2, \dots, p_N) \in \mathbb{R}_+^N$  tel que  $\sum_{k=1}^N p_k = 1$ , on a :

$$\sum_{k=1}^N p_k d_k a_k \geq \left( \sum_{k=1}^N p_k d_k \right) \times \left( \sum_{k=1}^N p_k a_k \right) \quad (1)$$

9. Soit  $t \in \mathcal{E}_N$ . On réordonne la liste  $(t(1), t(2), \dots, t(N))$  selon les valeurs croissantes et on note alors  $(\hat{t}(1), \hat{t}(2), \dots, \hat{t}(N))$  la liste ordonnée obtenue. On a donc :  $\hat{t}(1) \leq \hat{t}(2) \leq \dots \leq \hat{t}(N)$ .

a) Justifier pour tout  $n \in \llbracket 1, N \rrbracket$ , l'inégalité :  $\sum_{k=n}^N a_{t(k)} \leq \sum_{k=n}^N a_{\hat{t}(k)}$ .

b) On pose  $d_0 = 0$ . Justifier l'égalité :  $\sum_{n=1}^N d_n a_{t(n)} = \sum_{n=1}^N \left( (d_n - d_{n-1}) \sum_{k=n}^N a_{t(k)} \right)$ .

c) Établir l'inégalité :  $\sum_{n=1}^N d_n a_{t(n)} \leq \sum_{n=1}^N d_n a_{\hat{t}(n)}$ . (2)

On appelle *programme de transport*, toute bijection  $T$  de  $D$  sur  $A$ , et *coût* d'un programme de transport  $T$ , la somme  $c(T)$  définie par :  $c(T) = \sum_{k=1}^N (d_k - T(d_k))^2$ .

10. Soit  $\hat{T}$  le programme de transport défini par : pour tout  $k \in \llbracket 1, N \rrbracket$ ,  $\hat{T}(d_k) = a_k$ .

Déduire des questions précédentes que le programme  $\hat{T}$  est optimal, c'est-à-dire que pour tout programme de transport  $T$ , on a :  $c(T) \geq c(\hat{T})$ .

11. *Interprétation probabiliste des inégalités (1) et (2).*

Soit  $h$  une application croissante de  $\mathbb{R}$  dans  $\mathbb{R}$ .

a) En utilisant l'inégalité (1), établir pour toute variable aléatoire discrète  $X$  ne prenant qu'un nombre fini de valeurs, l'inégalité :  $E(X h(X)) \geq E(X) E(h(X))$ .

b) Que peut-on en déduire pour le coefficient de corrélation linéaire de  $X$  et  $h(X)$  lorsque les variances de  $X$  et  $h(X)$  sont strictement positives ?

c) En utilisant l'inégalité (2), montrer que si  $X$  est une variable aléatoire discrète suivant la loi uniforme sur  $\llbracket 1, N \rrbracket$  et  $t$  un élément de  $\mathcal{E}_N$ , on a :  $E(h(X) t(X)) \leq E(h(X) \hat{t}(X))$ .

### Partie III. Transport optimal dans une situation aléatoire

Les définitions de fonction de transport et de coût de transport sont identiques à celles données dans le préambule de la partie I.

Dans toute cette partie,  $U$  désigne une variable aléatoire vérifiant  $U(\Omega) = [0, 1]$  et suivant la loi uniforme sur le segment  $[0, 1]$ .

Soit  $Y$  une variable aléatoire admettant une densité  $f_Y$  nulle hors d'un segment  $[\alpha, \beta]$  ( $\alpha < \beta$ ) et dont la restriction à ce segment est continue et strictement positive. On note  $F_Y$  la fonction de répartition de  $Y$ .

On suppose l'existence d'une fonction  $g$  de classe  $C^1$  sur  $[0, 1]$ , à valeurs dans  $[\alpha, \beta]$ , telle que la variable aléatoire  $Z = g(U)$  suit la même loi que  $Y$ .

12. Pour tout entier  $N \geq 1$ , on pose pour tout  $\omega \in \Omega$  :

$$X_N(\omega) = \begin{cases} [1 + NU(\omega)] & \text{si } 0 \leq U(\omega) < 1 \\ N & \text{si } U(\omega) = 1 \end{cases} \quad \text{et} \quad Y_N(\omega) = g\left(\frac{X_N(\omega)}{N}\right).$$

a) Trouver la loi de la variable aléatoire  $X_N$ .

b) Établir l'existence d'une constante  $\lambda > 0$ , indépendante de  $N$  telle que :  $\forall \omega \in \Omega, |Z(\omega) - Y_N(\omega)| \leq \frac{\lambda}{N}$ .

c) Montrer que pour tout réel  $y$ , on a :  $F_Y\left(y - \frac{\lambda}{N}\right) \leq P([Y_N < y])$ .

13. Pour tout  $k \in \llbracket 1, N \rrbracket$ , on pose :  $t_N(k) = g\left(\frac{k}{N}\right)$ . On définit alors  $\hat{t}_N$  à partir de  $t_N$ , comme  $\hat{t}$  à partir de  $t$  dans la question 9.

a) Établir pour tout  $k \in \llbracket 1, N \rrbracket$ , les inégalités :  $F_Y\left(\hat{t}_N(k) - \frac{\lambda}{N}\right) \leq P([Y_N < \hat{t}_N(k)]) < \frac{k}{N}$ .

b) On note  $F_Y^{-1}$  la fonction réciproque de la restriction à  $[\alpha, \beta]$  de la fonction  $F_Y$ .

Montrer que pour tout entier  $N \geq 1$ , on a :  $\frac{1}{N} \sum_{k=1}^N \frac{k}{N} g\left(\frac{k}{N}\right) \leq \frac{1}{N} \sum_{k=1}^N \frac{k}{N} \left(F_Y^{-1}\left(\frac{k}{N}\right) + \frac{\lambda}{N}\right)$ .

c) En déduire l'inégalité :  $E(U g(U)) \leq E(U F_Y^{-1}(U))$ .

14.a) Parmi les fonctions de transport de classe  $C^1$  de  $U$  vers la loi de  $Y$ , trouver une fonction de transport  $T^*$  de coût minimal.

b) On suppose que  $Y = |4U - 2|$ . Déterminer  $T^*$  et  $C(T^*)$ .