TEXTE DE L'EPREUVE
DE PROBABILITES ET STATISTIQUES

Conventions et notations.

a. Dans ce problème, on aura lieu de considérer, sur l'ensemble $\mathbf{R}^n$,

— le *produit scalaire* usuel (noté .) et la *norme* qui lui est associée (norme euclidienne, notée $|\quad|$);

— la *topologie usuelle,* définie par la norme $|\quad|$ (pour toute partie A de $\mathbf{R}^n$, on note $\mathring{A}$ son intérieur et $\bar{A}$ sa fermeture) et la *tribu borélienne* (notée $\mathcal{B}^n$) qui lui est associée;

— l'ensemble des parties *convexes* : étant donné une partie A de $\mathbf{R}^n$, on appelle enveloppe convexe (resp enveloppe convexe fermée) de A la plus petite (au sens de l'inclusion) partie convexe (resp convexe fermée) contenant A.

b. Tous les espaces mesurables envisagés dans ce problème sont *euclidiens :* $(\Omega, \mathcal{A})$ est dit euclidien si, et seulement si, il existe n, entier strictement positif, tel que $\Omega \in \mathcal{B}^n$, et que $\mathcal{A}$ soit la tribu induite par $\mathcal{B}^n$ sur Ω (c'est-à-dire $\mathcal{A} = \{ A \in \mathcal{B}^n \; ; \; A \subset \Omega \}$).

c. Pour toute probabilité P sur $(\mathbf{R}^n, \mathcal{B}^n)$, on appelle *support* de P, et on note Supp (P) la plus petite (au sens de l'inclusion) partie fermée F de $\mathbf{R}^n$ vérifiant $P(F) = 1$ (on admettra son existence).

Définitions.

Soient deux entiers strictement positifs, n et m; soit $\Omega \in \mathcal{B}^n$ et soit $\Theta \subset \mathbf{R}^m$; soit (P_θ) une famille de probabilités sur l'espace mesurable euclidien $(\Omega, \mathcal{A})$, indicée par Θ; Ω est appelé l'ensemble des *résultats,* et Θ l'ensemble des *paramètres de* la famille (P_θ).

On dit que la famille (P_θ) est *exponentielle* si et seulement si il existe :

une mesure σ-finie μ sur $(\Omega, \mathcal{A})$,
un entier strictement positif k,
une application mesurable T de $(\Omega, \mathcal{A})$ dans $(\mathbf{R}^k, \mathcal{B}^k)$,
une application U de Θ dans $\mathbf{R}^k$,
une application mesurable a de $(\Omega, \mathcal{A})$ dans $(\mathbf{R}, \mathcal{B})$,
une application b de Θ dans $\mathbf{R}$,

tels que, pour tout θ, P_θ soit absolument continue par rapport à μ et admette, pour densité par rapport à μ, la fonction

$$\omega \longmapsto \exp [T(\omega) \,.\, U(\theta) - a(\omega) - b(\theta)] \,.$$

Le quadruplet (T, U, a, b) est appelé une *représentation* de *référence* μ (ou μ-représentation), de *dimension* k, de la famille (P_θ); cette représentation est dite :

résultats-identique si $n = k$ et, pour tout $\omega \in \Omega$, $T(\omega) = \omega$,
paramètres-identique si $m = k$ et, pour tout $\theta \in \Theta$, $U(\theta) = \theta$,
de *type nul* si, pour tout $\omega \in \Omega$, $a(\omega) = 0$;

la fonction b est appelée *fonction de cumul.*

I. *Généralités sur les familles exponentielles.*

1° Démontrer que les familles suivantes sont exponentielles :

a. Étant donné un entier strictement positif h, $\Omega = \{0, 1, \ldots, h\}$; $\Theta =]0, 1[$; pour tout θ, P_θ est la loi binomiale d'ordre h et paramètre θ.

b. $\Omega = \mathbf{N}$; $\Theta = \mathbf{R}_+^*$ $(=]0, +\infty[)$; pour tout θ, P_θ est la loi de Poisson de paramètre θ.

c. $\Omega = \mathbf{R}$; $\Theta = \mathbf{R}$; pour tout θ , P_θ est la loi de Laplace-Gauss (dite aussi loi normale) réduite (c'est-à-dire de variance égale à 1), de moyenne θ.

d. $\Omega = \mathbf{R}$; $\Theta = \mathbf{R}_+^*$; pour tout θ , P_θ est la loi de Laplace-Gauss centrée (c'est-à-dire de moyenne égale à 0), de variance θ.

e. $\Omega = \mathbf{R}$; $\Theta = \mathbf{R} \times \mathbf{R}_+^*$; pour tout $\theta = (\theta_1 , \theta_2)$, P_θ est la loi de Laplace-Gauss de moyenne θ_1 et variance θ_2.

2° Soit (P_θ) une famille exponentielle, d'ensemble des résultats Ω et ensemble des paramètres Θ ; soit (T, U, a, b) une μ-représentation de dimension k de (P_θ).

a. Deux mesures σ-finies sur $(\Omega , \mathcal{A})$ sont dites équivalentes si chacune est absolument continue par rapport à l'autre.

Démontrer que, pour toute mesure σ-finie ν équivalente à μ, la famille (P_θ) admet une représentation de référence ν et de dimension k.

Démontrer que, pour tout $\theta^0 \in \Theta$, la famille (P_θ) admet une représentation de référence P_{θ^0} et de dimension k, qui est de type nul.

b. Une application surjective φ, de Θ sur un ensemble Θ', est appelée un codage compatible avec la famille (P_θ) si, et seulement si, pour tout couple $(\theta^1 , \theta^2) \in \Theta^2$, tel que $\varphi(\theta^1) = \varphi(\theta^2)$, on a $P_{\theta^1} = P_{\theta^2}$; l'unique famille, notée $(({}_\varphi P)_{\theta'})$, admettant Ω pour ensemble des résultats et Θ' pour ensemble des paramètres, définie par

$$(\forall \theta \in \Theta) \qquad P_\theta = ({}_\varphi P)_{\varphi(\theta)} ,$$

est dite *codée* de (P_θ) par φ.

Soit $\Theta' = U(\Theta)$.

Démontrer que U est un codage compatible avec la famille (P_θ).

Démontrer que la famille $(({}_U P)_{\theta'})$, codée de (P_θ) par U, est exponentielle et admet une μ-représentation de dimension k, paramètres-identique.

c. f étant une application mesurable de $(\Omega , \mathcal{A})$ dans un espace mesurable $(\Omega' , \mathcal{A}')$, on appelle *transformée* de (P_θ) par f la famille, notée $(({}^f P)_\theta)$, admettant Ω' pour ensemble des résultats et Θ pour ensemble des paramètres, et où, pour tout $\theta \in \Theta$, $({}^f P)_\theta$ est l'image de P_θ par f (c'est-à-dire que, pour tout $A' \in \mathcal{A}'$, on a

$$({}^f P)_\theta (A') = P_\theta [f^{-1}(A')]) \quad .$$

Soit $\Omega' \in \mathcal{B}^k$, tel que $T(\Omega) \subset \Omega'$; T peut être considéré comme une application mesurable de $(\Omega , \mathcal{A})$ dans l'espace mesurable euclidien $(\Omega' , \mathcal{A}')$.

Soit ${}^T\mu$ la mesure, sur $(\Omega' , \mathcal{A}')$, image de μ par T; démontrer que la famille $(({}^T P)_\theta)$, transformée de (P_θ) par T, est exponentielle et admet une ${}^T\mu$-représentation de dimension k, résultats-identique.

II. *Représentation canonique d'une famille exponentielle.*

A partir de toute famille exponentielle admettant une représentation de dimension k on peut obtenir, par les opérations détaillées en I 2° (changement de référence, codage, transformation), une famille exponentielle (P_θ), admettant une représentation de référence P_{θ^0} (où θ^0 appartient à l'ensemble des paramètres), de dimension k, qui est de type nul, paramètres-identique et résultats-identique; si de plus $\theta^0 = 0$ (ce qui est toujours réalisable par un codage défini par une application bijective de $\mathbf{R}^k$ sur lui-même), la P_{θ^0}-représentation est dite *canonique*.

1° Soit P_0 une probabilité sur $(\mathbf{R}^k , \mathcal{B}^k)$.

a. On note b l'application de $\mathbf{R}^k$ dans $\mathbf{R} \cup \{ + \infty \}$ définie par

$$(\forall y \in \mathbf{R}^k) \qquad b(y) = \log \int_{\mathbf{R}^k} \exp(x \, . \, y) \, P_0(dx) \quad ;$$

soit $\Theta = b^{-1}(\mathbf{R})$; démontrer que Θ est une partie convexe non vide de $\mathbf{R}^k$.

b. On note Ω l'enveloppe convexe fermée de Supp (P_0); démontrer qu'il existe une famille exponentielle, dont l'ensemble des résultats est Ω et l'ensemble des paramètres Θ, admettant une P_0-représentation canonique de fonction de cumul b (on s'autorise les « abus de langage » consistant à confondre :

1° P_0, probabilité sur $(\mathbf{R}^k, \mathcal{B}^k)$, avec la probabilité qu'elle définit, par restriction, sur $(\Omega, \mathcal{A})$ (car $P_0(\Omega) = 1$),

2° b, application de $\mathbf{R}^k$ dans $\mathbf{R} \cup \{+\infty\}$, avec l'application, à valeurs dans $\mathbf{R}$, qui s'en déduit par restriction à Θ).

La famille exponentielle ainsi définie est dite *engendrée* par P_0.

2° Dans chacun des cas ci-dessous (de *a.* à *e.*), caractériser (en donnant l'ensemble des résultats (Ω'), l'ensemble des paramètres (Θ') et la fonction de cumul de la représentation P_0'-canonique (b')) la famille exponentielle engendrée par P_0', et démontrer que cette famille peut être obtenue, par codage et transformation, à partir de la famille exponentielle étudiée dans le cas correspondant de la question I 1°.

a. Étant donné un entier strictement positif h, P_0' est l'unique probabilité sur $(\mathbf{R}, \mathcal{B})$ qui vérifie

$$(\forall i \in \{0, 1, \dots, h\}) \qquad P_0'(\{i\}) = C^i_h \left(\frac{1}{2}\right)^h .$$

b. P_0' est l'unique probabilité sur $(\mathbf{R}, \mathcal{B})$ qui vérifie

$$(\forall i \in \mathbf{N}) \qquad P_0'(\{i\}) = \frac{1}{e}\,\frac{1}{i\,!} .$$

c. P_0' est la loi de Laplace-Gauss de dimension 1, centrée et réduite.

d. P_0' est la probabilité sur $(\mathbf{R}, \mathcal{B})$ qui admet, pour densité par rapport à la mesure de Lebesgue, la fonction f_0 définie par :

$$\left|\begin{array}{lll} \text{si} \quad x \leqslant 0 \quad , & & f_0(x) = 0 \\ \text{si} \quad x > 0 \quad , & & f_0(x) = \dfrac{1}{\sqrt{\pi}}\, e^{-x} x^{-\frac{1}{2}} . \end{array}\right.$$

e. P_0' est la probabilité définie sur $(\mathbf{R}^2, \mathcal{B}^2)$ par les conditions suivantes :

$P_0'(\{(x_1, x_2) \in \mathbf{R}^2 ; \quad x_2 = \frac{1}{2}(x_1)^2\}) = 1$,

la première marge de P_0' (c'est-à-dire l'image de P_0' par la projection $\pi_1 : (x_1, x_2) \longmapsto x_1$) est la loi de Laplace-Gauss centrée et réduite.

3° Soit P_0 une probabilité sur $(\mathbf{R}^k, \mathcal{B}^k)$; soit (P_θ) la famille exponentielle (d'ensemble des résultats Ω et ensemble des paramètres Θ) engendrée par P_0, et soit b (application de $\mathbf{R}^k$ dans $\mathbf{R} \cup \{+\infty\}$) la « fonction de cumul » de sa P_0-représentation canonique.

a. Démontrer que, en tout élément θ de l'intérieur $\mathring{\Theta}$ de Θ, la fonction b est dérivable à tous les ordres.

b. Pour tout $i \in \{1, \dots, k\}$, on note π_i la projection de Ω $(\subset \mathbf{R}^k)$ dans $\mathbf{R}$ définie par :

$$(\forall\ (x_j)_{1 \leqslant j \leqslant k} \in \Omega) \qquad \pi_i((x_j)_{1 \leqslant j \leqslant k}) = x_i .$$

Démontrer que, pour tout $\theta \in \mathring{\Theta}$ $(\theta = (\theta_j)_{1 \leqslant j \leqslant k})$, on a les résultats suivants :

— pour tout $i \in \{1, \dots, k\}$, l'*espérance mathématique* de π_i par rapport à P_θ est égale à $\dfrac{\partial b}{\partial \theta_i}(\theta)$;

— pour tout $i \in \{1, \dots, k\}$, la *variance* de π_i par rapport à P_θ est égale à $\dfrac{\partial^2 b}{\partial \theta_i^2}(\theta)$;

— pour tout couple $(i, j) \in \{1, \ldots, k\}^2$, tel que $i \neq j$, la *covariance* de π_i et π_j, par rapport à P_θ, est égale à $\dfrac{\partial^2 b}{\partial\theta_i \partial\theta_j}(\theta)$.

III. *Estimation par maximum de vraisemblance.*

Soit P_0 une probabilité sur $(\mathbf{R}^k, \mathcal{B}^k)$; soit (P_θ) la famille exponentielle (d'ensemble des résultats Ω et ensemble des paramètres Θ) engendrée par P_0, et soit b (application de $\mathbf{R}^k$ dans $\mathbf{R} \cup \{+\infty\}$) la « fonction de cumul » de sa P_0-représentation canonique (voir II 1°). On suppose que Supp (P_0) n'est contenu dans aucun hyperplan.

On va étudier le problème de l'estimation, en un résultat ω, du paramètre θ par la méthode du maximum de vraisemblance.

1° On appelle fonction convexe à k dimensions toute application f de $\mathbf{R}^k$ dans $\mathbf{R} \cup \{+\infty\}$ vérifiant :

$$\left|\begin{array}{l} (\forall (x^1, x^2) \in (\mathbf{R}^k)^2) \quad (\forall \lambda \in]0,1[) \\ \qquad\qquad f(\lambda x^1 + (1-\lambda)x^2) \leqslant \lambda f(x^1) + (1-\lambda) f(x^2), \\ f^{-1}(\mathbf{R}) \neq \varnothing \quad (f^{-1}(\mathbf{R}) \text{ est appelé le } \textit{domaine} \text{ de } f, \text{ et noté } D_f). \end{array}\right.$$

Si de plus la fonction convexe f vérifie :

$$(\forall (x^1, x^2) \in (\mathring{D}_f)^2) \quad (\forall \lambda \in]0,1[) \quad :$$

$$x^1 \neq x^2 \quad \Rightarrow \quad f(\lambda x^1 + (1-\lambda)x^2) < \lambda f(x^1) + (1-\lambda) f(x^2) \ ,$$

elle est dite *stricte.*

Démontrer que b est une fonction convexe stricte, de domaine Θ.

2° On rappelle qu'une fonction convexe f est continue sur $\mathring{D}_f \cup \complement \overline{D}_f$; elle est dite *fermée* si elle est semi-continue inférieurement sur $\mathbf{R}^k$, c'est-à-dire vérifie : $\quad (\forall x \in \mathbf{R}^k) \quad f(x) = \liminf\limits_{x' \to x} f(x')$.

(On rappelle que, pour que $l = \liminf\limits_{x' \to x} f(x')$, il faut que :

$$(\forall l' < l) \quad (\exists \eta > 0) \quad (\forall x' \in \mathbf{R}^k) \quad [\,|x' - x| < \eta \ \Rightarrow \ f(x') \geqslant l'\,])$$

Cette condition équivaut aussi au fait, pour f, d'être enveloppe supérieure des fonctions affines qu'elle majore, c'est-à-dire que, pour tout $x \in \mathbf{R}^k$,

$$f(x) = \sup \{\alpha \in \mathbf{R} \ ; (\exists y \in \mathbf{R}^k)(\forall x' \in \mathbf{R}^k) \ y.(x' - x) + \alpha \leqslant f(x')\} \ .$$

a. Démontrer que, pour qu'une fonction convexe f soit fermée, il suffit qu'elle vérifie :

$$(\forall x \in \mathbf{R}^k) \qquad f(x) \leqslant \liminf_{x' \to x} f(x') \ .$$

b. Démontrer que b est une fonction convexe fermée.

3° Deux fonctions convexes fermées à k dimensions, f et g, sont dites *compatibles* si, et seulement si, elles vérifient :

$$(\forall (x, y) \in (\mathbf{R}^k)^2) \qquad f(x) + g(y) \geqslant x.y \ .$$

a. Démontrer que, pour toute fonction convexe fermée f, l'ensemble des fonctions convexes fermées compatibles avec f est non vide, et admet un plus petit élément, noté f^*.

b. Démontrer que, pour toute fonction convexe fermée f, on a $(f^*)^* \leqslant f$.

Soit l une fonction affine; démontrer, en calculant l^* et $(l^*)^*$, que $l = (l^*)^*$.

En déduire que, pour toute fonction convexe fermée f, on a $(f^*)^* = f$.

On dit que deux fonctions convexes fermées f et g sont *conjuguées* si, et seulement si, elles vérifient $f^* = g$ (ce qui équivaut à $f = g^*$).

c. Démontrer que l'estimation du paramètre par maximum de vraisemblance, en l'observation $\omega \in \Omega$, est, *s'il existe et est unique,* l'élément θ de Θ tel que $b(\theta) + b^*(\omega) = \theta \cdot \omega$.

4° On sait que l'ensemble des paramètres Θ est égal à D_b, domaine de la fonction convexe b; le but de cette question est de comparer Ω, ensemble des résultats, et D_{b^*}, domaine de la fonction convexe b^*.

On note S la boule unité fermée de $\mathbf{R}^k$; pour tout $x \in \mathbf{R}^k$ et tout $\delta > 0$, $x + \delta S$ désigne la boule fermée de centre x et rayon δ

$$x + \delta S = \{ x' \in \mathbf{R}^k \quad ; \quad |x' - x| \leqslant \delta \}$$

a. Nous allons démontrer que $D_{b^*} \subset \Omega$.

Soit $x_0 \in \complement_{\mathbf{R}^k} \Omega$.

Démontrer qu'il existe $y_0 \in S$, et $\alpha < 0$, tels que, pour tout $\omega \in \Omega$, on ait $y_0 \cdot (\omega - x_0) \leqslant \alpha$.

Démontrer qu'on a :

$$\lim_{r \to +\infty} [(ry_0) \cdot x_0 - b(ry_0)] = +\infty.$$

En déduire

$$x_0 \notin D_{b^*}.$$

b. Nous allons démontrer que $\overset{\circ}{\Omega} \subset D_{b^*}$.

Soit $\omega_0 \in \overset{\circ}{\Omega}$.

On rappelle que Ω est l'enveloppe convexe fermée de Supp (P_0); on admet qu'il en résulte qu'il existe Ω', partie *finie* de Supp (P_0), telle que ω_0 appartienne à l'intérieur de l'enveloppe convexe de Ω'.

Soit y un élément de $\mathbf{R}^k$ de norme 1 ; on note :

$$\left|\begin{array}{lll} H(y) & \text{l'hyperplan} & \{x \in \mathbf{R}^k \ ; \ y \cdot x = y \cdot \omega_0\}, \\ K^+(y) & \text{le demi-espace} & \{x \in \mathbf{R}^k \ ; \ y \cdot x \geqslant y \cdot \omega_0\}, \\ K^-(y) & \text{le demi-espace} & \{x \in \mathbf{R}^k \ ; \ y \cdot x \leqslant y \cdot \omega_0\}; \end{array}\right.$$

pour tout $\omega' \in \Omega'$, on note $d(\omega', y)$ la distance de ω' à $H(y)$

$$(d(\omega', y) = \inf_{x \in H(y)} |\omega' - x|)$$

et on pose

$$d(y) = \min [\max_{\omega' \in K^+(y)} d(\omega', y) \ , \ \max_{\omega' \in K^-(y)} d(\omega', y)] \ ;$$

Démontrer que dans ces conditions on a : $d(y) > 0$.

Soit $d_0 = \inf_{|y|=1} d(y)$; démontrer qu'on a : $d_0 > 0$.

Démontrer que, pour tout élément ω de Supp (P_0) et tout $\delta > 0$, on a : $P_0(\omega + \delta S) > 0$.

On pose : $\alpha = \inf_{\omega' \in \Omega'} P_0(\omega' + d_0 S)$.

Démontrer qu'on a : $\alpha > 0$.

Démontrer que, pour tout y de norme 1, on a : $P_0[K^+(y)] \geqslant \alpha$.

En déduire : $\omega_0 \in D_{b^*}$.

c. Démontrer l'égalité $\overset{\circ}{D}_{b^*} = \overset{\circ}{\Omega}$.

5° On appelle *sous-gradient*, en un point $x \in \mathbf{R}^k$, de la fonction convexe à k dimensions f, la partie de $\mathbf{R}^k$.

$$\partial f(x) = \{ y \in \mathbf{R}^k \quad ; \quad (\forall\, x' \in \mathbf{R}^k) \qquad f(x') \geqslant f(x) + y\,.\,(x'-x) \} \,.$$

a. Soit $(x, y) \in (\mathbf{R}^k)^2$, et soit (f, g) un couple de fonctions convexes fermées à k dimensions, conjuguées. Démontrer l'équivalence des 3 propositions suivantes :

$y \in \partial f(x)$,
$f(x) + g(y) = x\,.\,y$,
$x \in \partial g(y)$.

En déduire une condition nécessaire et suffisante d'existence, en $\omega \in \Omega$, d'une estimation du paramètre par maximum de vraisemblance.

b. Une fonction convexe fermée à k dimensions est dite *douce* si, et seulement si, en tout $x \in \mathbf{R}^k$, $\partial f(x)$ a au plus un élément; on admettra que cette condition équivaut au couple de conditions suivantes :

pour tout $x \notin \mathring{\mathrm{D}}_f$, $\partial f(x) = \varnothing$,
pour tout $x \in \mathring{\mathrm{D}}_f$, f est dérivable en x et , $df(x)$ dénotant le gradient de f en x, on a $\partial f(x) = \{df(x)\}$.

Démontrer que, si la fonction convexe fermée f est stricte et douce, il en est de même de f^* (on pourra démontrer, à chaque fois par l'absurde, que f^* est douce, puis que f^* est stricte).

c. On suppose que la fonction b est douce.

Démontrer qu'alors l'estimation par maximum de vraisemblance

n'est pas définie aux points ω appartenant à la frontière de Ω,
est définie en tout point ω appartenant à $\mathring{\Omega}$, et vaut $db^*(\omega)$ (dire quelle est alors l'espérance mathématique, par rapport à $\mathrm{P}_{db^*(\omega)}$, de chacune des projections π_i $(1 \leqslant i \leqslant k)$ (voir II 3°)) .

6° Pour chacun des exemples étudiés en II 2°, on demande de :

a. s'assurer que la fonction de cumul est douce,

b. calculer la fonction $(b')^*$,

c. préciser $\mathring{\Omega}'$ et donner la valeur, pour tout $\theta' \in \Theta'$, de $\mathrm{P}'_{\theta'}(\mathring{\Omega}')$,

d. calculer, en tout $\omega' \in \mathring{\Omega}'$, la valeur de l'estimation par maximum de vraisemblance.