

LES CAHIERS DE L'ANALYSE DES DONNÉES

J. P. BENZÉCRI

Les scalogrammes pondérés

Les cahiers de l'analyse des données, tome 1, n° 1 (1976),
p. 33-46

http://www.numdam.org/item?id=CAD_1976__1_1_33_0

© Les cahiers de l'analyse des données, Dunod, 1976, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

LES SCALOGRAMMES PONDERES

par J.P. Benzecri ⁽¹⁾

Dans la présente note, on étudie un modèle de correspondance, dérivé de celui de Guttman, et suggéré par l'analyse des réponses à un questionnaire (cf § 4).

1. Pondérations sur le parallélogramme de Guttman :

Sous sa forme continue, le modèle de l'échelle de Guttman (cf e.g. [Corr. Esp.] TII B n°7 § 3) peut être décrit comme suit. Soit X un ensemble continu de sujets, paramétré par le segment $(0,1)$; Q un ensemble de questions, paramétré également par $(0,1)$: on suppose que le sujet d'abscisse x répond correctement aux questions dont l'abscisse (ou niveau) est inférieur à x et à ces questions-ci seulement. Dédoublons selon l'usage l'ensemble Q (cf TII A n° 2 § 1.4.) et identifions l'ensemble $Q^- \cup Q^+$, avec le segment $(0,2) = Y$: à une question d'abscisse q correspond le point q^- d'abscisse q , et le point q^+ , d'abscisse $q + 1$. Si le sujet d'abscisse x répond correctement à la question d'abscisse q on note :

$$g(x,q) = 0 ; g(x,q + 1) = 1$$

sinon on a $g(x,q) = 1$, $g(x,q + 1) = 0$. Avec ces conventions le tableau g des réponses est défini par la formule logique :

$$g(x,y) = \text{si } y \in (x,x+1) \text{ alors } 1, \text{ sinon } 0.$$

A elle seule cependant, la fonction $g(x,y)$ ne définit pas une correspondance sur $X \times Y$, car elle n'est pas une mesure. D'ordinaire on pose $p_{XY} = g(x,y) dx dy$, mais d'autres pondérations sont possibles, tant sur les sujets que sur les questions. En effet il se peut par exemple que la population étudiée comprenne relativement beaucoup de sujets de niveau élevé ; ou encore qu'on ait multiplié dans le questionnaire les questions dont la difficulté (abscisse) est faible. Dans la pratique, pour les sujets comme pour les questions, il s'agira nécessairement d'ensembles finis, d'échantillons issus du segment $(0,1)$. Mais le profil de densité de ces échantillons est bien décrit par le cas limite du modèle continu. On appellera donc échelle de Guttman pondérée une correspondance p_{XY} défini par :

$$p(x,y) dx dy = g(x,y) p(x) p(y) dx dy = p_{XY} ;$$

(1) Laboratoire de Statistique - Université Pierre et Marie Curie - Paris

Dans cette formule, $p(x) dx = p_X$ est une loi de probabilité sur $X = (0,1)$: la loi de distribution des sujets ; et $\rho(y) dy$ est, pour $y \in (0,1)$, une loi de probabilité : la loi des questions ; la fonction ρ étant étendue à $Y = (0,2)$ suivant la règle $\forall y \in (0,1) : \rho(y) = \rho(1+y)$. On notera que $\rho(y) dy$ n'est pas la loi marginale p_Y de la correspondance p_{XY} ; en effet si on note $p_Y = b(y) dy$, il vient :

$$\forall y \in (0,1) : b(y) = \rho(y) \left(\int_0^y p(x) dx \right)$$

$$\forall y \in (1,2) : b(y) = \rho(y) \left(\int_{y-1}^1 p(x) dx \right)$$

On a la relation :

$$\forall y \in (0,1) : b(y) + b(1+y) = \rho(y) \left(\int_0^y p(x) dx + \int_y^1 p(x) dx \right) = \rho(y)$$

cette relation signifie que la somme des poids $p(y)$, $p(1+y)$ des réponses respectivement fausse et vraie à la question y , n'est autre que le poids $\rho(y)$ de question.

Le paramétrage adopté pour les questions et les sujets n'a pas jusqu'ici de caractère intrinsèque : il peut être modifié sans que cesse d'être vérifiée aucune des conditions que nous avons imposées. Soit en effet θ une application biunivoque croissante du segment $(0,1)$ sur lui-même, application qu'on étendra à $(0,2)$ par la condition de périodicité : $\forall y \in (0,1) : \theta(1+y) = 1 + \theta(y)$. On a :

$$\forall (x,y) \in X.Y : g(\theta(x), \theta(y)) = g(x,y)$$

En d'autres termes, si l'on fait sur l'ensemble des sujets et sur l'ensemble des questions la même transformation monotone de paramètre, θ , le tableau g des réponses subsiste avec sa forme parallélogrammatique. Cependant, la loi des réponses et la loi des questions changent de forme : elles sont transportées par θ . On peut normaliser le paramétrage simultané de X et de Y en imposant la condition $p(x) dx = dx$ (ou encore la condition $\rho(y) dy = dy$). Ce choix n'est pas toutefois préférable à tout autre : nous verrons sur un exemple de calculs (cf §§ 2.4, 2.5, 3.4, 3.5), qu'avec $p(x) = (N+1)x^N$ (N entier positif) et $\rho(y)$ convenablement choisi les facteurs $F_\alpha(x)$ sont tous des polynômes en x : ce qui ne serait pas vrai si, pour la même correspondance p_{XY} , on choisissait le paramétrage de X par $(0,1)$ en sorte que p_X fût la mesure de densité uniforme. Nous conservons donc toute liberté de paramétrer simultanément X et Y mais nous construirons un diagramme carré, qui lui est intrinsèque (déterminé univoquement par p_{XY})

Soit la courbe plane, ensemble de points paramétré par $t \in (0,1)$:

$$C = \left\{ \left(\int_0^t p(x) dx, \int_0^t \rho(y) dy \right) \mid t \in (0,1) \right\}.$$

En tant qu'ensemble de points, C ne dépend pas d'un changement de paramètre θ effectué simultanément sur les x et les y . Soit (ξ, η) un point de C : convenons de noter $i(\xi)$ un sujet qui dépasse (quant à la compétence sur les questions de Q) un ensemble de sujets dont la masse est ξ , et qui est dépassé par un ensemble dont la masse est $1 - \xi$ (il revient au même de dire qu'avec le paramétrage adopté le sujet $i(\xi)$ a une abscisse t telle que $\int_0^t p(x) dx = \xi$) ; notons de même $q(\eta)$ une question où échoue une fraction des sujets dont le poids est η , (i.e. $q(\eta)$ a une abscisse t telle que $\int_0^t \rho(y) dy = \eta$) : alors le sujet $i(\xi)$ résout les questions plus faciles que $q(\eta)$ et elles seulement ; ou encore la question $q(\eta)$ est résolue par les sujets dont la compétence est au moins égale à celle de $i(\xi)$ et par eux seulement.

Dans le carré $(0,1) \times (0,1)$, la courbe C relie le sommet $(0,0)$ au sommet $(1,1)$: le long de C , abscisse et ordonnée varient dans le même sens ; à ceci près C est quelconque. Dans le cas de l'échelle de Guttman usuelle (ou dans tout cas qui s'y réduit par un changement de paramètre ; i.e. $\forall t \in (0,1) : p(t) = \rho(t)$) la courbe C est la diagonale du carré.

Voici en figure 1 deux autres exemples que nous rencontrerons dans la suite (cf. figure 2 ; § 3.6) : dans le cas (a) on a $\xi < \eta$; le questionnaire est facile relativement à l'ensemble des sujets : on pourra dire selon le paramétrage choisi qu'il y a trop de questions faibles, ou beaucoup de sujets forts, ou les deux ... Indépendamment de tout paramétrage on voit que la question médiane $q(\frac{1}{2})$ fait échouer moins de la moitié des sujets et encore que le sujet médian $i(\frac{1}{2})$ résout plus de la moitié des questions. Dans le cas de la figure (b), c'est le contraire. On peut concevoir des cas plus complexes, où l'on a sur C tantôt $\xi < \eta$, tantôt $\eta < \xi$...

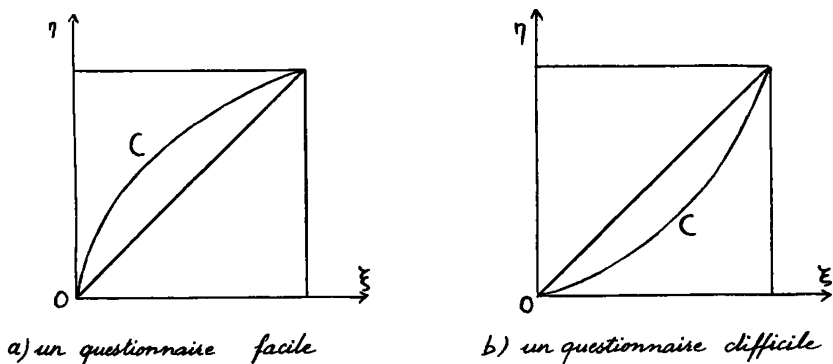


Figure 1 : deux exemples de courbe C

Quant à l'analyse de la correspondance $p(x,y)dx dy$ définie ci-dessus, on sait (cf TI B n°8 § 4) que quels que soient $p(x)$ et $\rho(y)$ elle fournit avec la plus grande valeur propre un facteur (couple φ^X, φ^Y) qui est croissant sur les deux ensembles X, Y . Mais il est également instructif de considérer des exemples numériques où l'analyse factorielle puisse être achevée par le calcul intégral : c'est l'objet du § qui suit ; qui est rédigé comme un problème et suivi de son corrigé.

2. Analyse d'une échelle de Guttman pondérée :

Problème

Sur le produit des deux ensembles $X = (0,1)$ et $Y = (0,2)$ on considère la correspondance p_{XY} définie par :

$$p_{XY} = g(x,y) p(x) \rho(y) dx dy ;$$

où $g(x,y)$ est donné par la formule usuelle de Guttman :

$$g(x,y) = \text{si } y \in (x, x+1) \text{ alors } 1, \text{ sinon } 0 ;$$

et où $p(x)$ et $\rho(y)$ sont des fonctions strictement positives assujetties aux conditions :

$$\int_0^1 p(x)dx = \int_0^1 \rho(y)dy = 1 ;$$

$$\forall y \in (0,1) : \rho(y) = \rho(1+y)$$

2.1. Exprimer par leur densité (i.e sous la forme $p_X = a(x)dx$, $p_Y = b(y)dy$) les lois marginales p_X et p_Y de la correspondance p_{XY} .

2.2. Soit α^X une fonction sur X : donner l'expression intégrale de la fonction β^Y image de α^X par la transition (*) π_{LY}^{LX} associée à la correspondance p_{XY} . (On exprimera $\beta(y)$ par le quotient de deux intégrales qu'on ne calculera pas).

2.3. Soit β^Y une fonction sur Y : donner l'expression intégrale de son image $\gamma^X = \pi_{LX}^{LY}(\beta^Y)$. (On exprimera $\gamma'(x)$ par une intégrale qu'on ne calculera pas).

2.4. Soit N un entier naturel (positif ou nul) fixé. Calculer explicitement $\beta_p^Y = \pi_{LY}^{LX}(\alpha_p^X)$ pour $\alpha_p^X(x) = x^p$ et $p(x) = (N+1)x^N = p_N(x)$. Donner alors l'expression intégrale de $\gamma_p^X = \pi_{LX}^{LY} \circ \pi_{LY}^{LX}(\alpha_p^X)$; (on décomposera $\gamma_p(x)$ en une somme de deux intégrales).

(*) Les transitions π_{LY}^{LX} et π_{LX}^{LY} sont définies comme d'usage ; cf e.g. [Corr. Esp.] TII B n°7 § 1.4.

2.5. L'expression donnée ci-dessus pour γ_p^X comporte la fonction $\rho(y)$ encore non fixée. Choisir une fonction $\rho_N(x)$ de telle sorte que, (N étant fixé), γ_p^X soit, pour tout entier $p \geq 0$, un polynôme en x de degré p .

2.6. Soit la courbe C_N , d'équation paramétrique :

$$C_N = \{(\xi(t), \eta(t)) \mid t \in (0,1)\} \text{ où on note :}$$

$$\xi(t) = \int_0^t p_N(t) dt ; \quad \eta(t) = \int_0^t \rho_N(t) dt.$$

Tracer C_N dans le plan des (ξ, η) ; préciser les tangentes aux extrémités (points de paramètres $t = 0$ et $t = 1$) ; dire quelle est la nature géométrique de C_1 .

2.7. Que peut-on dire des facteurs $\varphi_\alpha^{(N)}(x)$ sur X obtenus dans le cas où $p = p_N$, $\rho = \rho_N$? Dans le cas $N = 1$ calculer de façon précise les deux premiers facteurs non triviaux sur X : $\varphi_1^{(1)X}$, $\varphi_2^{(1)X}$.

2.8. Toujours sous l'hypothèse $p = p_N$, $\rho = \rho_N$, calculer les facteurs sur Y : $\varphi_1^{(1)Y}$ et $\varphi_2^{(1)Y}$; ainsi que les valeurs propres $\lambda_1^{(1)}$, $\lambda_2^{(1)}$ (on procédera comme pour l'échelle de Guttman). Tracer les ensembles en correspondance X et Y dans le plan rapporté aux deux premiers axes (i.e. x est représenté par $(F_1^{(1)}(x), F_2^{(2)}(x))$, et y par $(F_1^{(1)}(y), G_2^{(2)}(y))$).

2.9. Analyser la correspondance obtenue en prenant pour ρ et p les fonctions définies par :

$$\forall t \in (0,1) : \rho(t) = \rho_1(1-t) ; p(t) = p_1(1-t).$$

§ 3 : Solution sommaire :

$$3.1. \quad p_X = p(x)dx ; p_Y = b(y)dy ;$$

$$\forall y \in (0,1) : b(y) = \rho(y) \int_0^y p(x)dx ; \quad \forall y \in (1,2) : b(y) = \rho(y) \int_{y-1}^1 p(x)dx.$$

$$3.2. \quad \forall y \in (0,1) : \beta(y) = \left(\int_0^y \alpha(x) p(x)dx \right) / \left(\int_0^y p(x)dx \right) ;$$

$$\forall y \in (1,2) : \beta(y) = \left(\int_{y-1}^1 \alpha(x) p(x)dx \right) / \left(\int_{y-1}^1 p(x)dx \right) :$$

$$2.3. \quad \forall x \in (0,1) : \gamma'(x) = \int_x^{1+x} \beta'(y) \rho(y) dy.$$

$$3.4. \quad \forall y \in (0,1) : \beta_p(y) = \left(\int_0^x x^{N+p} dx \right) / \left(\int_0^y x^N dx \right) = (N+1)/(p+N+1) y^p ;$$

$$\forall y \in (1,2) : \beta_p(y) = \left(\int_{y-1}^1 x^{N+p} dx \right) / \left(\int_{y-1}^1 x^N dx \right) \\ = K_{Np} (1-(y-1)^{N=p+1}) / (1-(y-1)^{N+1}) ;$$

formule où on a noté K_{Np} le quotient $(N+1)/(p+N+1)$.

$$\forall x \in (0,1) : \gamma_p(x) = \int_x^1 \beta_p(y) \rho(y) dy + \int_1^{1-x} \beta_p(y) \rho(y) dy ;$$

$$\gamma_p(x) = K_{Np} \int_x^1 y^p \rho(y) dy + K_{Np} \int_0^x ((1-y)^{N+p+1}) / ((1-y)^{N+1}) \rho(y) dy$$

3.5. Pour que la première des intégrales dont la somme est $\gamma_p(x)$ soit un polynôme en x il suffit que sur $(0,1)$, $\rho(y)$ soit lui-même un polynôme ; tandis que dans la deuxième intégrale on a le quotient des polynômes :

$$(1-y^{N+p+1}) / (1-y^{N+1}) = (y^{N+p} + y^{N+p-1} + \dots + 1) / (y^N + y^{N-1} + \dots + 1).$$

Pour une valeur de p telle que $(N+p+1)$ et $(N+1)$ soient premiers entre eux le numérateur et le dénominateur de la fraction simplifiée de droite sont premiers entre eux (ils n'ont pas de racines communes, ces racines étant dans le plan complexe aux sommets des deux polygones réguliers inscrits dans le cercle de rayon 1). Pour que la deuxième intégrale fournisse un polynôme en x , il faut donc prendre pour $\rho(y)$ un polynôme qui soit un multiple du dénominateur $(y^N + y^{N-1} + \dots + 1)$.

Posons donc : $\forall y \in (0,1) : \rho(y) = r(y) (y^N + \dots + 1)$, où $r(y)$ est un polynôme encore indéterminé. Il vient :

$$K_{Np}^{-1} \gamma_p(x) = \int_x^1 (y^{N+p} + \dots + y^p) r(y) dy + \int_0^x (y^{N+p} + \dots + y^p + y^{p-1} + \dots + 1) r(y) dy ; \\ = \int_0^1 (y^{N+p} + \dots + y^p) r(y) dy + \int_0^x (y^{p-1} + \dots + 1) r(y) dy ;$$

La première de ces deux intégrales est une constante. La deuxième est un polynôme dont le degré surpasse de p celui du polynôme r : pour que $\gamma_p(x)$ soit de degré p , on prendra donc pour $r(y)$ une constante r_N choisie telle que $\int_0^1 \rho(y) dy = 1$:

$$r_N = 1 / \int_0^1 (y^N + \dots + 1) dy \quad ; \quad r_1 = 2/3 \quad ; \quad r_2 = 6/11 \quad ; \quad \text{etc...}$$

$r_N^{-1} = (N+1)^{-1} + N^{-1} + \dots + 1 \approx \text{Log} N$; r_N décroît vers zéro quand N croît. Notons, incidemment, que le cas $N = 0$ n'est autre que celui du modèle de Guttman usuel.

3.6. On a : $\xi_N(t) = t^{N+1}$; $\xi_1(t) = t^2$; etc ...

$$\eta_N(t) = r_N((t^{N+1}/(N+1)) + t^{N/N+\dots+t}) ;$$

$$\eta_1(t) = (2/3) ((t^2/2) + t) = (t^2/3) + (2t/3) ;$$

$$\eta_2(t) = (6/11) ((t^3/3) + (t^2/2) + t) ; \text{ etc ...}$$

La courbe C_N joint le point $(\xi(0), \eta(0)) = (0,0)$, au point $(\xi(1), \eta(1)) = (1,1)$; à l'origine on a, (si $N \neq 0$) : $\xi'_N(0) = 0$, $\eta'_N(0) = r_N$; la courbe est donc tangente à l'axe des ordonnées $O\eta$; au point $(1,1)$ on a : $\xi'_N(1) = N+1$;

$\eta'_N(1) = r_N(N+1)$; la tangente a donc une pente r_N d'autant plus faible que N est plus élevé. Dans le cas $N = 1$, C_1 est un segment de droite. Dans le cas

$N = 2$, C_2 est un arc de la parabole d'équation : $\xi = (3\eta - \xi)^2/4$. Sur la figure 2, on a tracé en tireté l'arc $t \in (0, -1)$: la corde joignant le point de paramètre t au point de paramètre $(-t)$ a son milieu sur la droite $\eta = \xi/3$; cette droite est le diamètre conjugué de la direction de l'axe des η ; elle est parallèle à l'axe de la parabole support de C_2 . Quand N tend vers l'infini, enfin, la courbe C_N tend vers la ligne brisée faite des deux segments $((0,0), (0,1))$ (porté par $O\eta$) et $((0,1), (1,1))$ (parallèle à $O\xi$).

3.7. Comme dans le cas particulier $N = 0$ (échelle de Guttman usuelle) il résulte de la stabilité des sous-espaces de polynômes de degré $\leq p$ par la transition $\pi \circ \pi = \pi_{LX}^{LY} \circ \pi_{LY}^{LX}$, que les facteurs sur X ne sont autres que les polynômes orthogonaux sur $(0,1)$ pour la loi p_X . Le facteur $\phi_1^{(1)X}$ est complètement défini (au signe près) par la condition d'être un polynôme du 1^o degré ayant moyenne nulle et variance 1 : $\phi_1^{(1)}(x) = ax + b$:

$$\int_0^1 (ax + b) x \, dx = (a/3) + (b/2) = 0$$

$$\int_0^1 (ax + b)^2 x \, dx = (2a^2/4) + (4ab/3) + (2b^2/2)$$

Le système se résout en posant $a = 3u$, $b = 2u$, (d'après la 1^o équation), d'où il vient dans la 2^o : $((9/2) - 8 + 4)u^2 = 1$; $u^2 = 2$. D'où $a = 3\sqrt{2}$; $b = -2\sqrt{2}$.

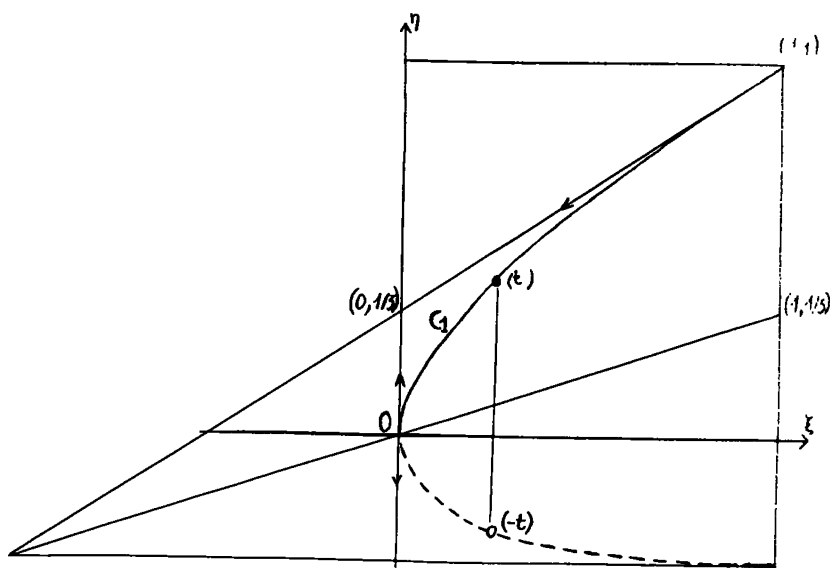
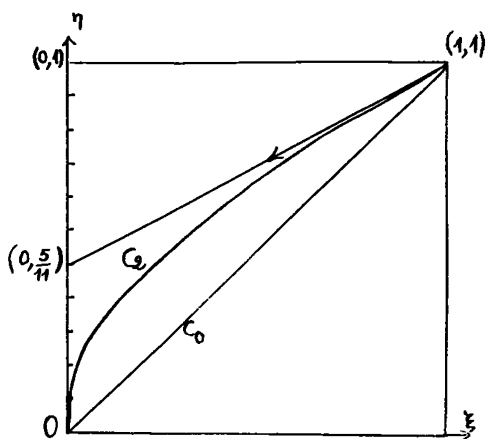


Figure 2 : Exemple de courbes C indicatrices du niveau relatif du questionnaire et de l'ensemble des sujets

$$\varphi_1^{(1)}(x) = 3\sqrt{2}x - 2\sqrt{2}.$$

Pour que $\varphi_2^{(1)}(x) = fx^2 + gx + h$, on a de même les conditions

$$\int_0^1 (fx^2 + gx + h)x \, dx = (f/4) + (g/3) + (h/2) = 0$$

$$\int_0^1 (fx^2 + gx + h)x^2 \, dx = (f/5) + (g/4) + (h/3) = 0$$

$$\int_0^1 (fx^2 + gx + h)^2 x \, dx = 1 = 2 \int_0^1 (fx^2 + gx + h) f x^3 \, dx$$

$$1 = (f^2/3) + (2fg/5) + (hf/2) = f((f/3) + (2g/5) + (h/2))$$

(dans la dernière équation on a tenu compte de ce que pour le poids $p(x)dx$, $\varphi_2^{(1)}$ est orthogonal à x et à la fonction constante). Par élimination entre les deux premières équations il vient : $f = 10v$; $g = 12v$; $h = 3v$; $v^2 = 3$. D'où :

$$\varphi_2^{(1)}(x) = 10\sqrt{3}x^2 - 12\sqrt{3}x + 3\sqrt{3}$$

3.8. On sait que l'on a en général : $\pi_{LY}^{LX}(\varphi_\alpha^X) = G_\alpha^Y$

Cette formule nous permet de calculer les G_α^Y par transition à partir des φ_α^X (cf 2°) :

$$\forall y \in (0,1) : G_\alpha(y) = \left(\int_0^y \varphi_\alpha(x) x \, dx \right) / \left(\int_0^y x \, dx \right)$$

$$\forall y \in (1,2) : G_\alpha(y) = \left(\int_{y-1}^1 \varphi_\alpha(x) x \, dx \right) / \left(\int_{y-1}^1 x \, dx \right)$$

Appliquons successivement cette formule à $\varphi_1^{(1)}$ et $\varphi_2^{(1)}$: il vient

$$\forall y \in (0,1) : G_1(y) = \sqrt{2} \left(\int_0^y (3x^2 - 2x) dx \right) / (y^2/2) = 2\sqrt{2}(y-1)$$

$$\begin{aligned} \forall y \in (1,2) : G_1(y) &= \sqrt{2} \left(\int_{y-1}^1 (3x^2 - 2x) dx \right) / ((1 - (y-1)^2)/2) \\ &= 2\sqrt{2}(-(y-1)^3 + (y-1)^2) / (1 - (y-1)^2) = 2\sqrt{2}(y-1)^2/y \end{aligned}$$

$$\forall y \in (0,1) : G_2(y) = \sqrt{3} \left(\int_0^y (10x^3 - 12x^2 + 3x) dx \right) / (y^2/2)$$

$$= \sqrt{3}(5y^2 - 8y + 3) = \sqrt{3}(y-1)(5y-3)$$

$$\begin{aligned}
 \forall y \in (1,2) : G_2(y) &= \sqrt{3} \left(\int_{y-1}^1 (10x^3 - 12x^2 + 3x) / ((1-(y-1)^2)/2) \right. \\
 &= \sqrt{3} [5x^4 - 8x^3 + 3x^2]_{y-1}^1 / [x^2]_{y-1}^1 \\
 &= \sqrt{3} (5(y-1)^3 - 3(y-1)^2)/y = \sqrt{3}(y-1)^2(5y-8)/y
 \end{aligned}$$

(dans le calcul on a noté selon l'usage $[P(x)]_a^b$ pour $P(b) - P(a)$).

Quant au calcul des valeurs propres, il se fait comme dans l'échelle de Guttman: λ_p est le coefficient de x^p dans le polynôme $\pi \circ \pi(x^p)$. Or on a (cf 5°)

$$\gamma_p(x) = K_{Np} r_N \int_0^x (y^{p-1} + \dots + 1) dy + Cte = (K_{Np} r_N / p) x^p + \dots$$

D'où : $\lambda_p^{(N)} = K_{Np} r_N / p = r_N(N+1)/(p(N+1))$

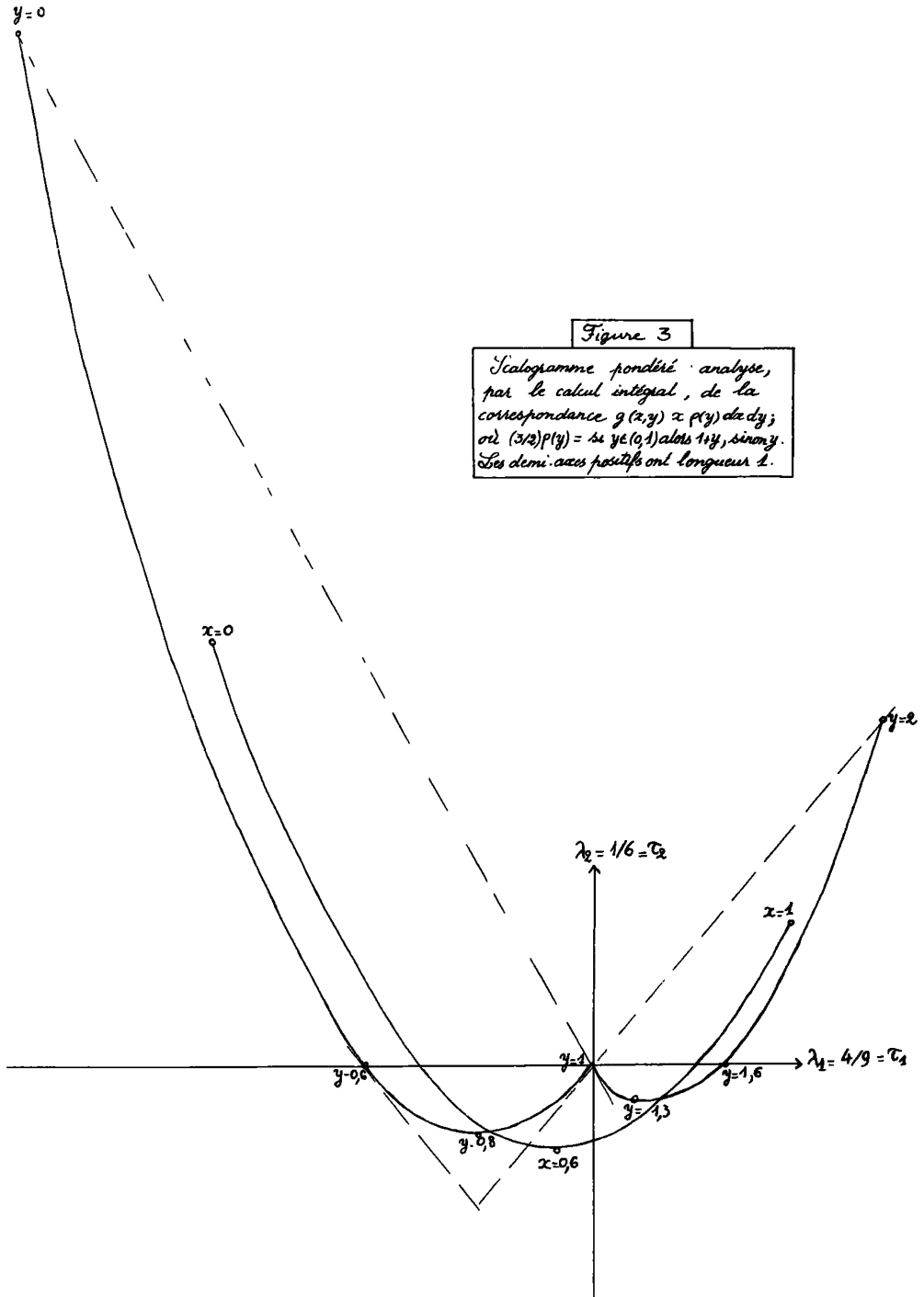
$$\lambda_p^{(1)} = (4/3)/(p(p+2)) ; \lambda_1^{(1)} = 4/9 ; \lambda_2^{(1)} = 1/6 ; \lambda_3^{(1)} = 4/45 ; \dots$$

On peut vérifier que, comme il est de règle pour un tableau logique dédoublé, la trace vaut 1 (cf [Corr. Esp.], § 6 Prob 71 9°). Le calcul des $\lambda_p^{(1)}$ fournit immédiatement les $F_p^{(1)}$ en fonction des $\varphi_p^{(1)X}$:

$$F_1^{(1)}(x) = (\lambda_1^{(1)})^{\frac{1}{2}} \varphi_1^{(1)}(x) = 2\sqrt{2} (3x - 2)/3$$

$$F_2^{(1)}(x) = (\lambda_2^{(1)})^{\frac{1}{2}} \varphi_2^{(1)}(x) = (10x^2 - 12x + 3)/\sqrt{2}$$

Sur la figure 3 on a tracé les ensembles X et Y dans le plan 1 x 2. L'ensemble X (ensemble des sujets) est un arc de parabole dont l'extrémité $x = 1$ (sujet omiscient) s'écarte moins de l'origine que l'extrémité $x = 0$ (sujet 0) : cette dissymétrie était prévisible d'après la densité des masses $p(x) = x$ qui s'élève d'une extrémité à l'autre. L'ensemble Y est formé de deux arcs. L'arc $(0,1)$ (ou ensemble Q^- des échecs) est un arc de parabole issu de l'origine sur laquelle la densité $p(y) = (2/3)y^2 (1+y)$ (cf 2°) croît du point $y = 0$ au point $y = 1$. L'arc $(1,2)$ (ou ensemble Q^+ des succès) est un arc de courbe du 3° degré, dont la portion utile diffère toutefois assez peu d'un arc de parabole. Sur l'arc Q^+ , la densité $p(y) = (2/3) y^2 (2 - y)$ croît à partir de l'origine jusqu'à un maximum atteint pour $y=4/3$, puis elle tend vers zéro quand $y \rightarrow 2$. On a tracé pour chacun des trois arcs X, Q^+ et Q^- les tangentes aux extrémités, ainsi que les tangentes parallèles au 1° axe (minima de F_2 et G_2). On notera que la tangente à l'origine à l'arc Q^+ (resp Q^-) n'est autre que la corde joignant les deux extrémités de l'arc Q^- (resp Q^+). Cette propriété s'explique par le principe du bras de levier (cf [Prat. Corr.] § 1.5.) : le segment joignant le point q^- (de paramètre $y \in (0,1)$) au point q^+ correspondant (de paramètre $(1+y) \in (1,2)$) passe par l'origine ; quand $y \rightarrow 0$ la corde Oq^+ a pour limite la tangente en 0 à Q^+ :



et cette limite est la même que celle de Q_1^- . Même raisonnement pour $y \rightarrow 1$. Il ne faut pas s'étonner de ce que les branches Q^+ et Q^- s'écartent le plus de l'origine dans la direction du 2° axe et non dans celle du 1° axe : comme nous l'avons déjà signalé à propos du tracé, ces branches sont très légères et n'apportent donc à λ_2 qu'une contribution modérée, faible vis-à-vis de celle que les parties moyennes de Q^+ et Q^- apportent à λ_1 qui est donc bien prépondérant.

3.9. Soit en général la correspondance $p_{XY} = g(x,y)p(x)\rho(y)dx dy$; et la correspondance $p_{XY}^S = g(x,y) p^S(x) \rho^S(y) dx dy$, où p^S et ρ^S sont définis par :

$$\forall t \in (0,1) ; \rho^S(t) = \rho(1-t) ; p^S(t) = p(1-t).$$

L'analyse de la correspondance p_{XY}^S se ramène à celle de p_{XY} par changement de variable. Soit $\{F_\alpha(x), G_\alpha(y), \lambda_\alpha\}$ issu de p_{XY} ; on aura pour p_{XY}^S les facteurs F_α^S et G_α^S relatifs à la même valeur propre λ_α et définis par :

$$F_\alpha^S(x) = F_\alpha(1-x) ; G_\alpha^S(y) = G_\alpha(2-y)$$

Les changements de fonctions ($p \rightarrow p^S, \rho \rightarrow \rho^S$) puis de variables ($x \rightarrow 1-x ; y \rightarrow 2-y$) appliqués successivement pour revenir au tableau initial peuvent être symbolisés par la figure 4 ci-dessous. En bref les rôles des succès et des échecs sont échangés.

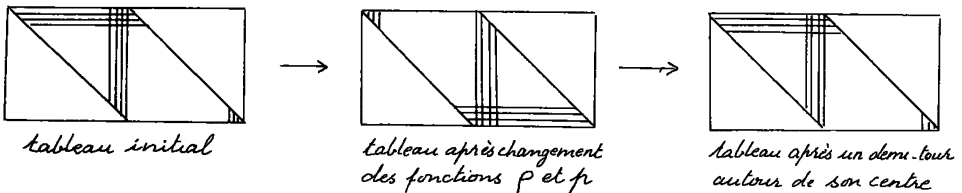


Figure 4 : Schéma des changements ; on a suggéré par des hachures les valeurs inégales des fonctions de densité $p(x)$ et $p(x)$.

On notera que par le changement de fonctions $(\rho, p) \rightarrow (\rho^S, p^S)$ la courbe C est changée en sa symétrique par rapport au centre du carré. Donc on passe du cas d'un questionnaire facile (fig. 1 - a ci-dessus) à celui d'un questionnaire difficile (fig. 1 - b).

4. Origine de la présente étude :

L'étude des scalogrammes pondérés nous a été suggérée par un rapport de stage dû à M. C. Gonthier et intitulé : "Etude d'une session expérimentale de l'épreuve de code à l'examen du permis de conduire".

On sait que l'épreuve de connaissance du code de la route, autrefois orale et individuelle, est maintenant remplacée par un questionnaire écrit commun et choisi parmi un certain nombre de questionnaires-types appuyés par des projections de diapositives. Une vue est projetée sur un écran, représentant une situation réelle de conduite dans toute sa complexité : multiplicité des panneaux, enseignes de magasins, visibilité réduite etc ... ; et le candidat doit choisir dans chaque cas, parmi les attitudes proposées A) B) C) ou D) une ou plusieurs qui conviennent à la situation. Soulignons qu'une question peut éventuellement donner lieu à une réponse multiple, double réponse par exemple ou même triple. Dès lors, le candidat ne sera considéré comme ayant donné la réponse exacte que s'il a fourni simultanément toutes ces réponses. Somme toute on conçoit que les questions, tant par leur format logique que par leur présentation visuelle, soient d'une grande diversité qu'il appartient à l'analyse factorielle de révéler.

Nous ne tenterons pas de donner ici du travail de M. Gonthier un résumé suffisant pour faire saisir les résultats spécifiques de son étude, mais seulement de montrer sur un graphique (cf figure 5) la particularité qui a attiré notre attention et provoqué l'étude des scalogrammes pondérés.

Dans le plan des axes 1 - 2 le nuage des questions dédoublées apparaît nettement divisé en deux branches : d'une part la branche Q^+ , très concentrée du côté du 1^o axe positif, d'autre part la branche Q^- très dispersée du côté du 1^o axe négatif. Seules deux questions, 7 et 16 font exception. La présentation de ces questions était mauvaise, le nombre des candidats qui les ont réussies est très faible : sur 1800 en effet, 30 seulement ont réussi la question 7, et 300 la question 16, alors qu'au contraire, la grande proximité des points q^+ opposée à l'éloignement des points q^- suffit à montrer que les points q^- sont très légers c'est-à-dire que les questions ont été en général réussies.

Quant à l'interprétation, bien que l'axe 2 permette d'isoler des groupes de question q^+ ou q^- interprétables, il semble bien qu'il ne s'agisse principalement que de l'étalement, dû à l'effet Guttman des deux ensembles Q^+ et Q^- que le 1^o axe a d'abord séparés. C'est afin de préciser mathématiquement cette hypothèse que le modèle des scalogrammes pondérés a été construit. De plus nous signalerons que les facteurs 3 et 4 obtenus par M. Gonthier et dont il a étudié l'interprétation, semblent indépendants du 1^o facteur, et liés seulement à la diversité des questions quant au type logique et à la présentation visuelle.

30-

Figure 5

L'épreuve du code à l'examen du permis de conduire. Analyse d'un tableau dédoublé 1800x80 des succès et échecs de 1800 sujets à 40 questions

